

VERITROOPER Scout — All Q&A Pairs



Every question in this audit (1000 total): the question, the ground truth, what the baseline (vanilla rag) answered (with its verdict), and what VERITROOPER answered. The heading verdict is the baseline (vanilla rag)'s — so misses on the deployed system show in red.

1. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_02`

HALLUCINATION

- **Question:** How many feet higher than flammable liquid storage should a bulk oxygen system be located?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "elevation", "parameters": {"elevation": "higher"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

2. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_07`

HALLUCINATION

- **Question:** Who within the Department of Labor is designated to conduct the investigations referenced in this part?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Compliance Safety and Health Officer", "parameters": {}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

3. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2542\_03`

- **Question:** Given the component location requirement in (iii) and the minimum-components principle in (iv), how should the system's components be both situated and selected?
- **Ground truth:** Components must be located to provide maximum practicable protection against damage from fire, explosion, corrosive atmosphere, or other environmental extremes consistent with adequate system performance, and the system must use the minimum number of components necessary to function, favoring components that do not require frequent servicing such as lubrication or cleaning.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "component_location_and_selection", "parameters": {"location": "maximum practicable protection against damage", "selection": "minimum number of components necessary"}}
- **✓ VERITROOPER (PASS):** {"name": "component_selection_and_situation", "parameters": {"situation": "maximum practicable protection against damage", "selection": "minimum number of components necessary"}}

4. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1227\_00`

- **Question:** Before performing welding on a generator, paragraphs (v) and (vi) together require what conditions concerning acetylene in both the generator being repaired and any other generators in the room?
- **Ground truth:** Per paragraph (v), all acetylene in the generator being repaired must be expelled by completely flooding the generator shell with water, with the carbide charge and feed mechanism completely removed and the generator disconnected from the piping system. Per paragraph (vi), if hot repairs are made in a room with other generators, all those generators and the piping must also be purged of acetylene.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "acetylene_purge", "parameters": {"generator_being_repaired": "True", "other_generators_in_room": "True"}}
- **✓ VERITROOPER (PASS):** {"name": "acetylene_conditions", "parameters": {"generator_being_repaired":

"completely removed", "other_generators_in_room": "purged of acetylene"}}

5. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_2159\_04`

- **Question:** Given a blank corrected total cadmium W_{bc} of 0.14 μg and an air volume sampled of 200 L, calculate the cadmium concentration in mg/m^3 using $\text{mg}/\text{m}^3 = W_{bc}/(\text{Air vol sampled, L})$.
- **Ground truth:** 0.0007 mg/m^3
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_cadmium_concentration", "parameters": {"W_bc": "0", "Air vol sampled": "200"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_cadmium_concentration", "parameters": {"W_bc": "0", "Air vol sampled": "200"}}

6. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_04`

- **Question:** The lower edge of the fixed guard shall extend at least how many inches outward from any handhold on the belt?
- **Ground truth:** 42 inches
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "fixed_guard_extension", "parameters": {"extension": "42"}}
- ✓ **VERITROOPER (PASS):** {"name": "fixed_guard_extension", "parameters": {"inches": "42"}}

7. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4247\_02`

- **Question:** Based on § 1955.13, who is the recipient of an administrative law judge's notice of self-withdrawal under paragraph (a), and who decides a party-initiated disqualification motion also under paragraph (a)?
- **Ground truth:** In both cases it is the Chief Administrative Law Judge - the judge's notice of self-withdrawal is directed to the Chief Administrative Law Judge, and the Chief Administrative Law Judge shall rule upon a party's motion to disqualify and remove.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "answer_question", "parameters": {"text": "Chief Administrative Law Judge", "question": "Based on \u00a7 1955.13, who is the recipient of an administrative law judge"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_recipient_and_decider", "parameters": {"recipient": "Chief Administrative Law Judge", "decider": "Chief Administrative Law Judge"}}

8. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0676\_03`

MODEL ERROR

- **Question:** When may a question of employee exposure to hazardous concentrations require monitoring under paragraph (h)?
- **Ground truth:** Where there may be a question of employee exposure to hazardous concentrations of hazardous substances, to assure proper selection of engineering controls, work practices, and personal protective equipment.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "monitoring", "parameters": {"h": "paragraph (h)"}}
- ✓ **VERITROOPER (PASS):** {"name": "Monitoring", "parameters": {"General": "where there may be a question of employee exposure to hazardous concentrations of hazardous substances"}}

9. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0075\_00`

- **Question:** If an employer's operations are physically dispersed such that posting at each place of alleged violation is not practicable, where must the citation be posted?
- **Ground truth:** It may be posted in a prominent place readily observable by all affected employees-for example, the location to which employees report each day.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "posting_citation", "parameters": {"location": "the location to"}}

which employees report each day"}}

- ✓ **VERITROOPER (PASS):** {"name": "post_citation", "parameters": {"location": "the location to which employees report each day"}}

10. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_02`

HALLUCINATION

- **Question:** What is the maximum allowable concentration variation between batches before a new classification under A.0.5.1.2 becomes mandatory?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "A.0.5.1.2", "parameters": {"concentration variation": "significant variation"}}

- ✓ **VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question.

11. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0250\_02`

- **Question:** Under which section of the Act are the shipyard employment standards adopted in § 1910.15?

- **Ground truth:** section 6(a)

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 6", "parameters": {"a": "§ 1910.15"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"number": "6"}}

12. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_02`

- **Question:** Which section of the Service Contract Act provides that the Act shall not apply to contracts for construction, alteration, and/or repair of public buildings or public works?

- **Ground truth:** Section 7

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 7 of the Service Contract Act", "parameters": {"section_number": "7"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"act_name": "Service Contract Act", "section_number": "7"}}

13. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2250\_03`

MODEL ERROR

- **Question:** Equipment may have become contaminated with blood, but the employer can demonstrate that decontamination of that equipment is not feasible. Must it still be decontaminated before servicing?

- **Ground truth:** No, decontamination is required unless the employer can demonstrate that decontamination of such equipment is not feasible.

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "decontaminate_equipment", "parameters": {"contaminated": "True", "feasible": "False"}}

- ✗ **VERITROOPER (FAIL):** {"name": "decontaminate_equipment", "parameters": {"contaminated": "True", "feasible": "False"}}

14. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_09`

HALLUCINATION

- **Question:** What is the required color of the warning signs marking abrasive-blasting enclosures?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_color", "parameters": {"tag_type": "Warning"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the

question.

15. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_01`

- **Question:** Under paragraph (l)(3)(i), if a previously exposed employee's initial results are all at or below the thresholds, what biological monitoring timing applies?
- **Ground truth:** The employer shall provide biological monitoring for CdU, β 2-M, and CdB one year after the initial biological monitoring, then comply with paragraph (l)(4)(v).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "biological_monitoring", "parameters": {"timing": "one year after the initial biological monitoring"}}
- ✓ **VERITROOPER (PASS):** {"name": "biological_monitoring", "parameters": {"timing": "one year after the initial biological monitoring"}}

16. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_01`

- **Question:** Compare the OMB control numbers for 1926.453(a)(2) and 1926.1427. Are they the same or different?
- **Ground truth:** They are different: 1926.453(a)(2) is OMB control No. 1218-0216, while 1926.1427 is OMB control No. 1218-0270.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_omb_control_numbers", "parameters": {"a": "1218-0216", "b": "1218-0270"}}
- ✓ **VERITROOPER (PASS):** {"name": "compare_omb_control_numbers", "parameters": {"x": "1218-0216", "y": "1218-0270"}}

17. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4138\_03`

- **Question:** Combining Option (4)(A) and Option (4)(C), what two distinct capabilities must the wire rope system satisfy with respect to load and crew control?
- **Ground truth:** Per (4)(A), the wire rope and attachments must be of sufficient size and strength to support the side load of the crane/derrick. Per (4)(C), the wire rope must be attached to the crane/derrick by appropriate methods (such as shackles or sheaves) on the undercarriage, and the method must allow the crew to secure the crane/derrick from movement during operation and to move it longitudinally along the vessel/flotation device for repositioning.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "centerline cable system", "parameters": {"Option (4)(A)": "support the side load of crane/derrick", "Option (4)(C)": "allow the crew to secure the crane/derrick from movement during operation and to move the crane/derrick longitudinally along the vessel/flotation device"}}
- ✓ **VERITROOPER (PASS):** {"name": "wire rope system", "parameters": {"capability1": "support the side load of crane/derrick", "capability2": "allow the crew to secure the crane/derrick from movement during operation and to move the crane/derrick longitudinally along the vessel/flotation device for repositioning"}}

18. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0022\_03`

- **Question:** If an interested person wishes to submit written data, views, and arguments on the proposal, within what period after publication of the notice in the Federal Register must they do so?
- **Ground truth:** Within 30 days after publication of the notice in the Federal Register.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "submit_written_comments", "parameters": {"days": "30"}}
- ✓ **VERITROOPER (PASS):** {"name": "submit_written_comments", "parameters": {"days_after_publication": "30"}}

19. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4228\_01`

- **Question:** If a developmental step's completion involves policies or procedures that differ from the Federal program, what must the State submit and within what time frame?
- **Ground truth:** The State must submit one copy of the required plan supplement within 60 days of adoption of the change.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "submit_required_plan_supplement", "parameters": {"documentation": "one copy of the required plan supplement", "time_frame": "60 days of adoption of the change"}}
- ✓ **VERITROOPER (PASS):** {"name": "submit_plan_supplement", "parameters": {"documentation": "one copy of the required plan supplement", "time_frame": "60 days"}}

20. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4032\_00`

- **Question:** Under what circumstance is the anti two-block device NOT required?
- **Ground truth:** The anti two-block device is not required when hoisting personnel in pile driving operations; instead, paragraph (p)(2) specifies how to prevent two-blocking during such operations.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "anti two-block device exception", "parameters": {"circumstance": "hoisting personnel in pile driving operations"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_anti_two_block_device", "parameters": {"circumstance": "hoisting personnel in pile driving operations"}}

21. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_01`

HALLUCINATION

- **Question:** What is the filing fee for submitting a discrimination complaint under this part?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_filing_fee", "parameters": {"complaint_type": "discrimination"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the filing fee for submitting a discrimination complaint under this part.

22. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3386\_01`

- **Question:** When mechanical equipment is being used, what is the minimum distance the warning line must be erected from the roof edge that is parallel to the direction of mechanical equipment operation?
- **Ground truth:** Not less than 6 feet (1.8 m).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "warning_line_distance_for_mechanical_equipment", "parameters": {"distance": "6 feet (1.8 m)"}}
- ✓ **VERITROOPER (PASS):** {"name": "warning_line_distance", "parameters": {"distance": "6"}}

23. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_06`

HALLUCINATION

- **Question:** How much open space, in inches, must be maintained between the duct and a combustible roof at the point of penetration?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "open_space", "parameters": {"inches": "18"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

24. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_1138\_00`

MODEL ERROR

- **Question:** If a critical component of the PSDI safety system fails, can the press remain in service after recertification by the employer alone, without revalidation?
- **Ground truth:** No; the failure of a critical component requires removal of the safety system from service until it is recertified AND revalidated.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "recertification_by_employer", "parameters": {"failure_type": "critical component"}}
- **✗ VERITROOPER (FAIL):** {"name": "recertification_by_employer", "parameters": {"component": "critical"}}

25. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0772\_00`

WRONG CALC/QUOTE

- **Question:** According to paragraph (a)(5), the threshold check solution can be prepared from the fit test solution referenced in which other paragraph, and how is it diluted?
- **Ground truth:** It is prepared by putting 1 ml of the fit test solution (referenced in (b)(5)) into 100 ml of distilled water.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "prepare_threshold_check_solution", "parameters": {"paragraph_number": "5", "dilution_factor": "100"}}
- **✓ VERITROOPER (PASS):** {"name": "prepare_threshold_check_solution", "parameters": {"paragraph_number": "5", "dilution_method": "1 ml of the fit test solution in 100 ml of distilled water"}}

26. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_04`

HALLUCINATION

- **Question:** What minimum ventilation rate is required for a noncombustible building housing a bulk oxygen system?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_ventilation_rate", "parameters": {"building_type": "noncombustible", "system_type": "bulk oxygen"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

27. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_06`

HALLUCINATION

- **Question:** What temperature must electrically powered spray liquid heaters be limited to during operation?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "temperature_limit", "parameters": {"temperature_limit": "199.4 \u00b0F (93 \u00b0C)"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

28. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_06`

- **Question:** If the administrative law judge receives a petition to participate, within how many days must he ascertain objections to it?
- **Ground truth:** Within 5 days of receipt of the petition.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "ascertain_objections", "parameters": {"days": "5"}}

- ✓ **VERITROOPER (PASS):** {"name": "ascertain_objections", "parameters": {"days": "5"}}

29. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_04`

- **Question:** For Class 1 equipment, what is the AC rms retest voltage?

- **Ground truth:** 10,000

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "1", "voltage_type": "AC rms"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "1", "voltage_type": "AC rms"}}

30. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4630\_01`

- **Question:** Under § 1991.109(e), how many days after the date of the decision do non-reinstatement portions of the ALJ's order become effective absent a timely petition for review?

- **Ground truth:** 30 days

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_days_after_decision", "parameters": {"days": "30"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"number": "30"}}

31. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_04`

HALLUCINATION

- **Question:** What is the maximum number of employees an employer can record before the COVID-19 log requirement no longer applies?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "COVID-19 log requirement exception", "parameters": {"number of employees": "10"}}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

32. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_03`

HALLUCINATION

- **Question:** What advance notice period is required for OSHA's 'announced' inspections under this section?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_advance_notice_period", "parameters": {"inspection_type": "announced"}}

- ✓ **VERITROOPER (PASS):** {"name": "no_function", "parameters": {}}

33. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3608\_02`

- **Question:** Based on subparts (ii) and (iii), what must the device allow the employee to do regarding hand use, and to what value must it limit the descending velocity of an employee?

- **Ground truth:** It must permit the employee to ascend or descend without continually having to hold, push or pull any part of the device (leaving both hands free for climbing), and it must limit the descending velocity to 7 feet/sec. (2.1 m/sec.) or less.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "device", "parameters": {"allow": "the employee to let go without obstructing the movement of the suspension ropes", "limit": "the descending velocity of an employee to 4.6 feet per second (1.4 meters per second)"}}

- ✓ **VERITROOPER (PASS):** {"name": "device_requirements", "parameters": {"hand_use": "leave both hands free for climbing", "velocity_limit": "7 feet/sec."}}

34. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_02`

HALLUCINATION

- **Question:** What grounding voltage threshold triggers the requirement to protect personnel from shock when floors are wet down?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "protect_personnel_from_shock", "parameters": {"voltage": "30"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

35. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0325\_02`

- **Question:** If no OSHA standard in this part requires a fire prevention plan, must an employer have one under this section?
- **Ground truth:** No, an employer must have a fire prevention plan only when an OSHA standard in this part requires one.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "application", "parameters": {"a": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If no OSHA standard in this part requires a fire prevention plan, must an employer have one under this section?", "answer": "No"}}

36. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_07`

HALLUCINATION

- **Question:** What is the maximum width permitted for aisles and walkways near abrasive-blasting operations?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_aisle_width", "parameters": {}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

37. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1593\_02`

- **Question:** Under what condition do entrances to buildings, rooms, or enclosures containing exposed live parts over 600 volts NOT have to be kept locked?
- **Ground truth:** They need not be kept locked when they are under the observation of a qualified person at all times.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "entrances_to_buildings_rooms_or_enclosures_containing_exposed_live_parts_over_600_volts_do_not_have_to_be_kept_locked", "parameters": {"condition": "when they are under the observation of a qualified person at all times"}}
- **✓ VERITROOPER (PASS):** {"name": "entrance_locking_condition", "parameters": {"condition": "under the observation of a qualified person at all times"}}

38. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4044\_00`

- **Question:** To what percent of the platform's rated capacity must the platform and rigging be proof tested before hoisting employees?
- **Ground truth:** 125 percent
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "proof_testing", "parameters": {"percent": "125"}}
- **✓ VERITROOPER (PASS):** {"name": "proof_testing", "parameters": {"percent_capacity": "125"}}

39. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_04`

HALLUCINATION

- **Question:** What temperature threshold qualifies as a 'fever' among the listed COVID-19 symptoms?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "COVID-19 symptoms", "parameters": {"Fever or chills": "100.4\u00b0F"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

40. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2920\_01`

- **Question:** Combining paragraph (vii) and paragraph (ix), what two separate facilities must be available near or for workers handling liquids that may burn or irritate the skin, including their key specifications?
- **Ground truth:** Per paragraph (vii), near each such tank there must be a supply of clean cold water with a water pipe (pressure not exceeding 25 pounds/11.325 kg) fitted with a quick opening valve and at least 48 inches (1.216 m) of hose not smaller than three-fourths inch. Per paragraph (ix), sufficient washing facilities including soap, individual towels, and hot water must be provided on the basis of at least one basin with a hot water faucet for every 10 employees.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_facilities", "parameters": {"facility1": "a supply of clean cold water", "spec1": "carrying a pressure not exceeding 25 pounds (11.325 kg), with a quick opening valve and at least 48 inches (1.22 m) of hose", "facility2": "basin with a hot-water faucet", "spec2": "one basin for every 10 employees"}}
- **✓ VERITROOPER (PASS):** {"name": "facility_availability", "parameters": {"facility1": "clean cold water", "specification1": "quick opening valve and at least 48 inches of hose", "facility2": "washing facilities", "specification2": "soap, individual towels, and hot water"}}

41. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0913\_02`

MODEL ERROR

- **Question:** Under what condition is the educational program of paragraph (g)(1) required, and how does its triggering condition differ from the training in paragraph (g)(3)?
- **Ground truth:** The educational program in (g)(1) is required where the employer has provided portable fire extinguishers for employee use in the workplace, while the training in (g)(3) is required for employees designated to use fire fighting equipment as part of an emergency action plan.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "educational_program", "parameters": {"condition": "when it can be shown that their work activities could result in exposure to blood or OPIM"}}
- **✓ VERITROOPER (PASS):** {"name": "compare_conditions", "parameters": {"condition1": "where the employer has provided portable fire extinguishers for employee use in the workplace", "condition2": "where employees have been designated to use fire fighting equipment as part of an emergency action plan"}}

42. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4448\_00`

- **Question:** Under paragraph (a), a complainant who received a written determination but for whom there has been no final decision of the Secretary brings a de novo action in district court. Under paragraph (b), within how many days of filing that complaint must they file a copy with the relevant DOL adjudicator, and within how many days after the written determination could they have filed the court action under (a)(1)?
- **Ground truth:** The complainant must file the file-stamped complaint within seven days after filing in federal court (paragraph (b)), and could have brought the action within 90 days after receiving the written determination under § 1983.105(a) (paragraph (a)(1)).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "district_court_jurisdiction_retaliation_complaints", "parameters": {"days_to_file_copy": "7", "days_to_file_court_action": "90"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"paragraph": "(a)", "filing_timeframe": "7", "filing_timeframe_2": "90"}}

43. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2632\_00`

- **Question:** If data on a skin sensitizer are not sufficient for sub-categorization, into which category must the substance be classified?
- **Ground truth:** It shall be classified in Category 1.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_skin_sensitizer", "parameters": {"substance": "Category 1"}}
- ✗ **VERITROOPER (FAIL):** {"name": "classify_skin_sensitizer", "parameters": {"sub_categorization": "insufficient"}}

44. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_04`

HALLUCINATION

- **Question:** Within what timeframe after exposure must asthma symptoms appear to count as evidence of respiratory hypersensitivity?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "A.4.2.1.2.3", "parameters": {"(iii)": "within hours"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

45. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4502\_02`

- **Question:** Based on paragraphs (e)(4) and (e)(5), what happens to the investigation if the respondent fails to make a timely response or fails to satisfy the clear and convincing evidence burden?
- **Ground truth:** If the respondent fails to make a timely response or fails to satisfy the burden set forth in paragraph (e)(4) (demonstrating by clear and convincing evidence it would have taken the same adverse action absent the protected activity), OSHA will proceed with the investigation, which will proceed whenever necessary or appropriate to confirm or verify the information provided by the respondent.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "OSHA_proceed_with_investigation", "parameters": {"respondent_fails_to_make_timely_response": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "OSHA proceeds with the investigation", "parameters": {"reason": "respondent fails to make a timely response or fails to satisfy the clear and convincing evidence burden"}}

46. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_04`

- **Question:** What diameter of particles must a HEPA filter remove with at least 99.97 percent efficiency?
- **Ground truth:** 0.3 micrometers in diameter
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "high-efficiency particulate air filter", "parameters": {"diameter": "0"}}
- ✓ **VERITROOPER (PASS):** {"name": "HEPA_filter_efficiency", "parameters": {"diameter": "0"}}

47. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_03`

HALLUCINATION

- **Question:** How many toxicologically active ingredients must mixtures A and B share for the interpolation principle in A.0.5.1.4 to apply?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "A.0.5.1.4", "parameters": {"n": "1"}}
- ✗ **VERITROOPER (FAIL):** {"name": "None", "parameters": {}}

48. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_1157\_01`

MODEL ERROR

- **Question:** When are sprocket wheels and chains exempt from the requirement that they be enclosed?
- **Ground truth:** When they are more than seven (7) feet above the floor or platform (the enclosure requirement does not apply to manually operated sprockets either).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "sprocket_wheels_and_chains_exempt", "parameters": {"height": "7 feet"}}
- **✓ VERITROOPER (PASS):** {"name": "exempt_sprocket_wheels_and_chains", "parameters": {"height": "7", "location": "above the floor or platform", "operation": "manually operated"}}

49. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3948\_04`

- **Question:** If a knuckle-boom truck crane transfers a precast concrete panel onto a structure using a fork/cradle while equipped with an overload prevention device, does the (c)(17) material-delivery exclusion apply?
- **Ground truth:** No; the exclusion does not apply when the material handled is a prefabricated component such as a precast concrete panel.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "apply_exclusion_17", "parameters": {"i": "False", "ii": "False", "iii_A": "False", "iii_B": "True", "iii_C": "False", "iii_D": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "exclusion_applies", "parameters": {"equipment": "knuckle-boom truck crane", "activity": "transferring a precast concrete panel onto a structure using a fork/cradle", "device": "overload prevention device"}}

50. PASS · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_01`

- **Question:** How wide must the hanging walkways in tunnels 16 feet or more in diameter be?
- **Ground truth:** UNANSWERABLE
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "tunnel_walkway_width", "parameters": {"diameter": "16 feet"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

51. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0772\_04`

MODEL ERROR

- **Question:** Using paragraphs (a)(3) and (a)(12), what must the test subject do during the screening test and what additional action is required once a taste response is elicited?
- **Ground truth:** During the test the subject must breathe through a slightly open mouth with tongue extended and report when a sweet taste is detected. Once a taste response is elicited, the subject must take note of the taste for reference in the fit test.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "taste_threshold_screening", "parameters": {"action": "don the test enclosure", "additional_action": "take note of the taste for reference in the fit test"}}
- **✓ VERITROOPER (PASS):** {"name": "test_subject_action", "parameters": {"action": "breathe through slightly open mouth with tongue extended", "additional_action": "take note of the taste for reference in the fit test"}}

52. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3092\_01`

MODEL ERROR

- **Question:** Combining the container requirements in paragraph (a)(1) and the storage restrictions in paragraph (a)(2), if a worker handles 4 gallons of flammable liquid, what type of container is required and in what areas may it not be stored?
- **Ground truth:** Since 4 gallons is 5 gallons or less, approved safety cans or DOT approved containers must

be used (per (a)(1)), and per (a)(2) it shall not be stored in areas used for exits, stairways, or normally used for the safe passage of people.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "container_requirements", "parameters": {"quantity": "4", "container_type": "closed containers", "storage_restrictions": "outside of an inside storage room or storage cabinet in a building or in any one fire area of a building"}}
- ✔ **VERITROOPER (PASS):** {"name": "container_requirements", "parameters": {"quantity": "4", "container_type": "approved safety cans or Department of Transportation approved containers", "storage_restrictions": "areas used for exits, stairways, or normally used for the safe passage of people"}}

53. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0250\_01`

- **Question:** As of what specific date were the part 1915 standards in effect when adopted as occupational safety or health standards for shipyard employment?
- **Ground truth:** April 28, 1971
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "part 1915 standards adoption date", "parameters": {"date": "April 28, 1971"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_date", "parameters": {"date": "April 28, 1971"}}

54. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3305\_02`

- **Question:** Using subpart (ii) and subpart (i), if a stairway-type ladder rises from a bottom step set at the maximum allowed 24 inches above the scaffold supporting level, what is the maximum vertical interval required between rest platforms?
- **Ground truth:** 12 feet (3.7 m), as specified in subpart (ii) for maximum vertical intervals between rest platforms, regardless of the 24-inch bottom step maximum in subpart (i).
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_max_vertical_interval", "parameters": {"subpart": "ii", "subpart_2": "i", "bottom_step_height": "24"}}
- ✖ **VERITROOPER (FAIL):** {"name": "max", "parameters": {"a": "12", "b": "3"}}

55. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_05`

HALLUCINATION

- **Question:** How far from bulk sulfur or baled paper storage may welding be conducted safely?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "welding_safety_distance", "parameters": {"x": "35"}}
- ✖ **VERITROOPER (FAIL):** {"name": "relocation of combustibles", "parameters": {"distance": "35"}}

56. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_00`

HALLUCINATION

- **Question:** What is the maximum cord- and plug-connected load permitted for a receptacle on a 40-ampere branch circuit supplying two or more receptacles?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_max_load", "parameters": {"circuit_rating": "40", "receptacle_rating": "40"}}
- ✖ **VERITROOPER (FAIL):** {"name": "Table S-4", "parameters": {"Circuit rating (amperes)": "40"}}

57. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3608\_00`

MODEL ERROR

- **Question:** Per the requirements in subparts (i) and (iii), what is the maximum drop weight a device must

withstand and within what distance must it activate after a fall?

- **Ground truth:** The device must withstand an 18-inch (41 cm) drop of a 500-pound (226 kg) weight, and it must be activated within 2 feet (.61 m) after a fall occurs.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "calculate_drop_test", "parameters": {"weight": "250", "distance": "4"}}
- **✓ VERITROOPER (PASS):** {"name": "get_device_requirements", "parameters": {"drop_weight": "500", "activation_distance": "2"}}

58. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_05`

HALLUCINATION

- **Question:** Within how many hours of a confirmed COVID-19 positive result must the employer notify the Assistant Secretary?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "reporting_covid_19_fatalities_and_hospitalizations_to_osh", "parameters": {"hours": "8"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

59. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0564\_01`

MODEL ERROR

- **Question:** When an LP-Gas-fueled vehicle is being repaired in a garage, may its container shutoff valve be open?
- **Ground truth:** Only when fuel is required for engine operation; otherwise the shutoff valve must be closed.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "LP-Gas-fueled vehicle repair in garage", "parameters": {"valve_status": "closed"}}
- **✗ VERITROOPER (FAIL):** {"name": "check_container_shutoff_valve", "parameters": {"status": "closed"}}

60. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_2809\_11`

MODEL ERROR

- **Question:** What is the street address (number and street) of the American National Standards Institute (ANSI) given under Availability of Referenced Material?
- **Ground truth:** 1430 Broadway, New York, NY 10018
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_address", "parameters": {"organization": "American National Standards Institute (ANSI)"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_street_address", "parameters": {"organization": "American National Standards Institute (ANSI)"}}

61. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_02`

HALLUCINATION

- **Question:** What form must the employee use to provide written consent for releasing prior medical examination records?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "appendix A to this section contains a sample form which may be used to establish specific written consent for access to employee medical records", "parameters": {}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

62. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_0598\_00`

MODEL ERROR

- **Question:** What is the minimum distance required between underground containers of over 2,000 gallons capacity?
- **Ground truth:** at least 5 feet
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_distance", "parameters": {"water_capacity": "2001"}}
- **✓ VERITROOPER (PASS):** {"name": "get_distance", "parameters": {"capacity": "2000", "distance": "5"}}

63. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_03`

- **Question:** In appendix A Table A, what is the threshold above which β 2-microglobulin falls into category B in the second listing?
- **Ground truth:** 750
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_threshold", "parameters": {"marker": "\u03b2 2 -microglobulin (\u03b2 2 -M) (\u03b2cb/g creatinine)", "category": "B", "listing": "2"}}
- **✓ VERITROOPER (PASS):** {"name": "get_category_b_threshold", "parameters": {"biological_marker": "\u03b22-microglobulin", "listing": "second"}}

64. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3020\_00`

- **Question:** Under subpart (vii) and subpart (viii), in an area where a major spill may occur and salvage equipment must be on hand, what two distinct requirements apply regarding salvage drums and spill containment?
- **Ground truth:** U.S. Department of Transportation specified salvage drums or containers and suitable quantities of proper absorbent must be kept available and used (vii), and a spill containment program-part of the employer's safety and health program required in paragraph (b)-must be implemented to contain and isolate the entire volume of the hazardous substance being transferred (viii).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "subpart_vii_and_viii_requirements", "parameters": {"salvage_drums": "shall be kept available and used in areas where spills, leaks, or ruptures may occur", "spill_containment": "shall be implemented to contain and isolate the entire volume of the hazardous substance being transferred"}}
- **✗ VERITROOPER (FAIL):** {"name": "salvage_drums_and_spill_containment_requirements", "parameters": {"subpart_vii": "True", "subpart_viii": "True"}}

65. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_04`

- **Question:** In appendix A Table A, what is the upper bound of category A for Cadmium in blood (CdB) in the second listing (>5 and ≤?)?
- **Ground truth:** 10
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_upper_bound_category_A", "parameters": {"marker": "Cadmium in blood (CdB)", "listing": "2"}}
- **✓ VERITROOPER (PASS):** {"name": "get_upper_bound", "parameters": {"category": "A", "marker": "Cadmium in blood (CdB)", "listing": "2"}}

66. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_03`

MODEL ERROR

- **Question:** If work near an insulated power line of 200 volts is being performed, what is the minimum distance required?

- **Ground truth:** 3 feet (0.9 m), since the line is less than 300 volts.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "minimum_distance", "parameters": {"voltage": "200"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_minimum_distance", "parameters": {"voltage": "200", "line_type": "insulated"}}

67. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_08`

- **Question:** If swinging loads being hoisted near a scaffold might contact it, what measure is required?
- **Ground truth:** Tag lines or equivalent measures to control the loads shall be used.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "tag_lines_or_equivalent_measures", "parameters": {"required": "True"}}
- ✔ **VERITROOPER (PASS):** {"name": "control_loads", "parameters": {"measure": "tag lines or equivalent measures"}}

68. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_4398\_00`

MODEL ERROR

- **Question:** What is the consequence when an ALJ approves a request to withdraw objections to the Assistant Secretary's findings and/or order, and there are no other pending objections?
- **Ground truth:** The Assistant Secretary's findings and/or order will become the final order of the Secretary.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "ALJ_approves_withdrawal_of_objections", "parameters": {"result": "The Assistant Secretary"}}
- ✖ **VERITROOPER (FAIL):** {"name": "consequence_of_withdrawal", "parameters": {"event": "ALJ approves withdrawal of objections and no other pending objections", "result": "Assistant Secretary"}}

69. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_2438\_00`

MODEL ERROR

- **Question:** If a removed employee files a claim for workers' compensation payments for a MDA-related disability, what must the employer do?
- **Ground truth:** The employer shall continue to provide medical removal protection benefits pending disposition of the claim.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "handle_workers_compensation_claim", "parameters": {"claim": "MDA-related disability"}}
- ✔ **VERITROOPER (PASS):** {"name": "continue_providing_benefits", "parameters": {"pending_disposition": "True"}}

70. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4269\_02`

- **Question:** Using the definition of 'issue' from § 1902.2(c) together with the statement about Federal standards in public employment, why does a State plan covering public employees meet the definition of section 18?
- **Ground truth:** Under § 1902.2(c), an issue is any industrial, occupational, or hazard grouping found to be administratively practicable and not in conflict with the purposes of the Act. Because Federal standards are in effect with regard to hazards found in public employment, a State plan covering this occupational category meets the definition of section 18 and the regulations.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "section_18_definition", "parameters": {"issue_definition": "any industrial, occupational, or hazard grouping that is found to be administratively practicable and not in conflict with the purposes of the Act", "federal_standards_in_public_employment": "Federal standards are in effect with regard to hazards found in public employment"}}
- ✔ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"prompt": "Using the definition of ", "section": "1902.2(c)", "federal_standards": "Federal standards are in effect with regard to hazards found in"}}

71. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_00`

HALLUCINATION

- **Question:** What penalty does an employer face for failing to record a positive case in the COVID-19 log within the 24-hour window?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "report_penalty", "parameters": {"amount": "10000"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

72. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_04`

- **Question:** If a person whose petition is denied wishes to appeal, how long do they have to appeal to the Secretary?
- **Ground truth:** Within 5 days of receipt of the notice of denial.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "appeal_denial", "parameters": {"days": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "appeal_to_secretary", "parameters": {"days_to_appeal": "5"}}

73. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2476\_03`

- **Question:** Based on the definitions of 'physician or other licensed health care professional' and 'symptom,' which paragraph governs the health care services they provide, and what cardiac effects are included among symptoms?
- **Ground truth:** The health care services are required by paragraph (j) of this section. The cardiac effects included among symptoms are chest pain or shortness of breath.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph", "parameters": {"j": "health care services", "cardiac effects": "chest pain or shortness of breath"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_health_care_services_and_cardiac_effects", "parameters": {"paragraph": "j", "cardiac_effects": "['chest pain', 'shortness of breath']"}}

74. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_03`

HALLUCINATION

- **Question:** What form number must employers use to record entries in the COVID-19 log?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Forms", "parameters": {"number": "300, 300A, and 301"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

75. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_03`

HALLUCINATION

- **Question:** Within how many minutes must dust spills outside an abrasive-blasting enclosure be cleaned up to comply with this section?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "clean_up_dust_spills", "parameters": {"minutes": "immediately"}}
- ✓ **VERITROOPER (PASS):** This information is not provided in the given evidence.

76. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_06`

- **Question:** Which ASTM standard for FRP rod and tube is referenced as deeming compliance with paragraph (j)(1)?
- **Ground truth:** ASTM F711-02 (2007)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ASTM F711-02", "parameters": {"year": "2007"}}
- ✓ **VERITROOPER (PASS):** {"name": "ASTM", "parameters": {"standard": "F711-02", "year": "2007"}}

77. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_00`

- **Question:** What is the publication date of ISO 10156:1996 (E) for the determination of fire potential and oxidizing ability?
- **Ground truth:** Feb. 15, 1996
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ISO 10156", "parameters": {"edition": "1996", "publication_date": "Feb. 15, 1996"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_publication_date", "parameters": {"standard": "ISO 10156:1996 (E)", "topic": "determination of fire potential and oxidizing ability"}}

78. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_03`

- **Question:** During the in-service test under paragraph (j)(2)(iii)(E), what voltage per meter applies for a wood tool and for how long?
- **Ground truth:** 164,000 volts per meter (50,000 volts per foot) for 1 minute
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "in-service test under paragraph (j)(2)(iii)(E)", "parameters": {"voltage_per_meter": "164000", "duration": "1 minute", "material": "wood"}}
- ✓ **VERITROOPER (PASS):** {"name": "in_service_test", "parameters": {"material": "wood", "voltage_per_meter": "164000", "duration": "1"}}

79. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0340\_00`

- **Question:** Each suspended unit component must support at least four times the maximum intended live load without failure - but which components are excepted from this requirement?
- **Ground truth:** Suspension ropes and guardrail systems are excepted from the requirement to support at least four times the maximum intended live load without failure.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "suspended_unit_component_exception", "parameters": {"components": "suspension ropes and guardrail systems"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_component_requirement", "parameters": {"component": "suspension ropes and guardrail systems"}}

80. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0012\_00`

- **Question:** If an employer wishes to contest both the abatement period and the proposed penalty for an alleged violation, does the State plan provide for review of these matters?
- **Ground truth:** Yes, the State must provide the employer the right of review of violations alleged, abatement periods, and proposed penalties.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "State plan review", "parameters": {"abatement_period": "True", "proposed_penalty": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "review_proceedings", "parameters": {"issues": "abatement period and proposed penalty"}}

81. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1114\_01`

- **Question:** What is the maximum nominal a.c. supply voltage permitted for a.c. control circuits and solenoid valve coils?
- **Ground truth:** 120-volt a.c.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_maximum_ac_supply_voltage", "parameters": {"voltage": "120"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_nominal_voltage", "parameters": {"voltage_type": "a.c.", "circuit_type": "control"}}

82. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3428\_03`

- **Question:** Under what circumstances may safety latches on hooks be deactivated or made inoperable?
- **Ground truth:** Only when (i) a qualified rigger has determined that hoisting and placing of purlins and single joists can be performed more safely by doing so, or (ii) when equivalent protection is provided in a site-specific erection plan.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "deactivate safety latches", "parameters": {"circumstances": "[\"When a qualified rigger has determined that the hoisting and placing of purlins and single joists can be performed more safely by doing so\", \"When equivalent protection is provided in a site-specific erection plan\"]"}}
- ✓ **VERITROOPER (PASS):** {"name": "deactivate safety latches", "parameters": {"when": "a qualified rigger has determined that the hoisting and placing of purlins and single joists can be performed more safely by doing so", "or": "when equivalent protection is provided in a site-specific erection plan"}}

83. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_1336\_00`

MODEL ERROR

- **Question:** Band wheels must be completely encased or guarded, but under what condition is part of the wheel allowed to be exposed?
- **Ground truth:** A portion of the upper wheel immediately around the point where the blade leaves the wheel may be exposed, to permit the operator to observe movement of equipment.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "band_wheels_encased_or_guarded", "parameters": {"condition": "to permit operator to observe movement of equipment"}}
- ✗ **VERITROOPER (FAIL):** {"name": "band_wheels_encased", "parameters": {"condition": "to permit operator to observe movement of equipment"}}

84. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1964\_00`

- **Question:** Under paragraph (l)(2)(i), what legend must the employer post at entrances to regulated areas, and what alternative legend could be used in its place prior to June 1, 2016 under paragraph (l)(2)(iii)?
- **Ground truth:** Under (l)(2)(i) the required legend is: 'DANGER VINYL CHLORIDE MAY CAUSE CANCER AUTHORIZED PERSONNEL ONLY'. Prior to June 1, 2016, under (l)(2)(iii), the employer could instead use: 'CANCER-SUSPECT AGENT AREA AUTHORIZED PERSONNEL ONLY'.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "post_signs", "parameters": {"entrance_legend": "DANGER VINYL CHLORIDE MAY CAUSE CANCER AUTHORIZED PERSONNEL ONLY", "alternative_legend": "CANCER-SUSPECT AGENT AREA AUTHORIZED PERSONNEL ONLY"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_legends", "parameters": {"paragraph": "(l)(2)(i)", "alternative_paragraph": "(l)(2)(iii)", "date": "June 1, 2016"}}

85. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_09`

- **Question:** What is the average extraction efficiency for the reanalyzed samples the next day?
- **Ground truth:** 98.7%
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "extraction_efficiency", "parameters": {"x": "98.7%"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_average_extraction_efficiency", "parameters": {"samples": "reanalyzed samples the next day"}}

86. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_03`

- **Question:** Where a final medical determination recommends special protective measures or exposure limitations for an employee, what must the employer do?
- **Ground truth:** The employer shall implement and act consistent with the recommendation.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "implement_special_protective_measures", "parameters": {"recommended_measures": "special protective measures or exposure limitations"}}
- ✓ **VERITROOPER (PASS):** {"name": "employer_action", "parameters": {"action": "implement and act consistent with the recommendation"}}

87. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1292\_02`

- **Question:** If a V-belt drive is located 9 feet from the floor and out of reach from any platform, and a set of gears is located in the same area, which must be enclosed according to the general machine guarding rules?
- **Ground truth:** Only the gears must be enclosed. Per paragraph (c)(2), all gears shall be completely enclosed regardless of location, but per paragraph (c)(3), V-belt drives need enclosure only if within reach from platforms/passageways or within 8 feet 6 inches from the floor - a V-belt drive 9 feet high and out of reach does not require enclosure.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "general_machine_guarding", "parameters": {"machine": "gears"}}
- ✓ **VERITROOPER (PASS):** {"name": "general_machine_guarding", "parameters": {"machine_part": "gears"}}

88. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2190\_03`

MODEL ERROR

- **Question:** If the employer monitored within one year prior to the standard's effective date and that monitoring satisfies all other requirements, must the employer still conduct new initial monitoring?
- **Ground truth:** No; the employer may rely on such earlier monitoring results to satisfy the initial monitoring requirement.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "initial_monitoring", "parameters": {"satisfies_requirements": "True", "monitored_within_one_year": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "monitoring_required", "parameters": {"prior_monitoring": "True", "satisfies_requirements": "True"}}

89. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4084\_01`

- **Question:** If a tower crane is NOT tied to or braced by the structure, from where must the tower (mast) bolts be inspected for loose or dislodged condition?
- **Ground truth:** From the base of the tower crane up.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "inspect_tower_bolts", "parameters": {"location": "from the base of the tower crane up"}}
- ✓ **VERITROOPER (PASS):** {"name": "inspect_tower_bolts", "parameters": {"location": "the base of the tower crane up"}}

90. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_07`

HALLUCINATION

- **Question:** Which OSHA office is responsible for approving the design specifications of the fan plenum?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Office of Variance Determination", "parameters": {"location": "Occupational Safety and Health Administration, U.S. Department of Labor, Room N3653, 200 Constitution Avenue, NW., Washington, DC 20210"}}
- **✓ VERITROOPER (PASS):** {"name": "None", "parameters": {}}

91. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1202\_00`

WRONG CALC/QUOTE

- **Question:** For transmission pipeline field shop operations described in paragraph (d)(1)(ii), which additional paragraph requirement applies that is NOT required under the general transmission pipeline requirements of paragraph (d)(1)(i)?
- **Ground truth:** Paragraph (d)(1)(i) general requirements observe paragraphs (b) and (c) and § 1910.254. Field shop operations in (d)(1)(ii) additionally require paragraph (a) and § 1910.253. So paragraph (a) (and § 1910.253) is the added requirement.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraph requirement", "parameters": {"(d)(1)(ii)": "1910.253 and 1910.254"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_additional_requirement", "parameters": {"paragraph": "(d)(1)(ii)", "general_requirement": "(d)(1)(i)"}}

92. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4450\_05`

- **Question:** Under which section of the Internal Revenue Code of 1986 may an employee receive a credit that is protected under section 18C of the FLSA?
- **Ground truth:** Section 36B (26 U.S.C. 36B)
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "section 36B of the Internal Revenue Code of 1986", "parameters": {}}
- **✓ VERITROOPER (PASS):** {"name": "get_credit_section", "parameters": {"section_number": "36B"}}

93. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_02`

HALLUCINATION

- **Question:** How long after an employee's imminent danger report must the agency or committee respond before OSHA may inspect under paragraph (a)(3)?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"a": "3"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

94. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0455\_00`

MODEL ERROR

- **Question:** If a metallic floorplate on which a container stands during filling is electrically connected to the fill stem, must the nozzle and container also be separately electrically interconnected to comply with the grounding section?
- **Ground truth:** No, that arrangement is deemed to have complied with the grounding provisions.

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "is_grounding_required", "parameters": {"is_metallic_floorplate_connected": "true"}}
- **✗ VERITROOPER (FAIL):** {"name": "answer_grounding_question", "parameters": {"electrically_connected": "True"}}

95. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0517\_03`

MODEL ERROR

- **Question:** Considering the internal-combustion engine requirement in subparagraph (g)(2)(ii) and the fuel oil flow requirement in subparagraph (g)(2)(iii), what two fire-related safety measures are required regarding engines and fuel oil flow?
- **Ground truth:** Internal-combustion engines used for electric power generation must be located outside the mixing plant building, or be properly ventilated and isolated by a firewall, with exhaust systems located so spark emission cannot be a hazard; and suitable means must be provided to prevent the flow of fuel oil to the mixer in case of fire, with an automatic spring-loaded shutoff valve with fusible link installed in gravity flow systems.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_fire_safety_measures", "parameters": {"engine_requirement": "internal-combustion engine requirement in subparagraph (g)(2)(ii)", "fuel_oil_flow_requirement": "fuel oil flow requirement in subparagraph (g)(2)(iii)"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_fire_safety_measures", "parameters": {"engine_requirement": "internal-combustion engine requirement in subparagraph (g)(2)(ii)", "fuel_oil_flow_requirement": "fuel oil flow requirement in subparagraph (g)(2)(iii)"}}

96. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_05`

HALLUCINATION

- **Question:** What form must a committee use to request inspection reports from an agency?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "request_inspection_report", "parameters": {"committee_name": "Establishment Committee", "agency_name": "Occupational Safety and Health Administration"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

97. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0455\_01`

- **Question:** If a pit has any part within a Division 1 or 2 classified area but is provided with mechanical ventilation, how must the pit be classified?
- **Ground truth:** It shall be classified Division 2 (rather than Division 1).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "classify_pit", "parameters": {"ventilation": "mechanical", "division": "2"}}
- **✗ VERITROOPER (FAIL):** {"name": "classify_pit", "parameters": {"has_mechanical_ventilation": "True", "is_within_division_1_or_2": "True"}}

98. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_00`

HALLUCINATION

- **Question:** What is the minimum thickness in millimeters required for the leather gloves worn by abrasive-blasting operators?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_thickness", "parameters": {"material": "leather", "gloves_type": "abrasive-blasting operators"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

99. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_05`

- **Question:** If a party wishes to obtain copies of the hearing transcript, what must they do?
- **Ground truth:** They must file a written application with the reporter and pay fees at the rate provided in the agreement with the reporter.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "obtain transcript copies", "parameters": {"method": "written application", "fee": "payment of fees at the rate provided in the agreement with the reporter"}}
- ✓ **VERITROOPER (PASS):** {"name": "obtain transcript copies", "parameters": {"action": "file written application", "entity": "reporter", "condition": "payment of fees"}}

100. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3162\_04`

- **Question:** If a proposed location for an air receiver would be underground or inaccessible, is that installation allowed?
- **Ground truth:** No, under no circumstances shall an air receiver be buried underground or located in an inaccessible place.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "installation_allowed", "parameters": {"location": "underground or inaccessible", "allowed": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_installation_allowed", "parameters": {"location": "underground or inaccessible"}}

101. PASS · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_01`

- **Question:** Which specific OSHA form number must be used to file a written retaliation complaint?
- **Ground truth:** UNANSWERABLE
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "No specific form number is required", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

102. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_08`

- **Question:** What is the average extraction efficiency for the six filters spiked at the target concentration (same-day analysis)?
- **Ground truth:** 99.6%
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "average_extraction_efficiency", "parameters": {"x": "99"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_extraction_efficiency", "parameters": {"num_filters": "6", "spiked_concentration": "target", "analysis_time": "same-day"}}

103. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4551\_03`

- **Question:** Under § 1990.106(b)(2), within how many days after receipt of a public petition must the Secretary respond?
- **Ground truth:** ninety (90) days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "response_to_petition", "parameters": {"days": "90"}}
- ✓ **VERITROOPER (PASS):** {"name": "respond_to_public_petition", "parameters": {"days": "90"}}

104. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3427\_00`

- **Question:** Under § 1926.752(c)(1), to which roads does the requirement for adequate access roads NOT apply?
- **Ground truth:** It does not apply to roads outside of the construction site.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_adequate_access_roads", "parameters": {"roads": "roads outside of the construction site"}}
- ✓ **VERITROOPER (PASS):** {"name": " roads outside of the construction site", "parameters": {}}

105. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1962\_00`

- **Question:** Under paragraph (k), which employees must a program of medical surveillance be instituted for, and per the same paragraph, who must perform all medical examinations and procedures and at whose cost to the employee?
- **Ground truth:** Medical surveillance must be instituted for each employee exposed, without regard to the use of respirators, to vinyl chloride in excess of the action level. All medical examinations and procedures must be performed by or under the supervision of a licensed physician, and must be provided without cost to the employee.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "medical_surveillance", "parameters": {"employees": "each employee with exposure to BD at concentrations at or above the action level on 30 or more days or for employees who have or may have exposure to BD at or above the PELs on 10 or more days a year", "physician": "a licensed physician", "cost": "at no cost to the employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "medical_surveillance_program", "parameters": {"employees": "each employee exposed, without regard to the use of respirators, to vinyl chloride in excess of the action level", "examinations_and_procedures_performed_by": "a licensed physician", "cost_to_employee": "without cost to the employee"}}

106. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3403\_02`

- **Question:** Based on the static charge protection in paragraph (j) and the personal protective equipment requirements in paragraph (e), what protective items might ground personnel need before touching a suspended load and while receiving a load?
- **Ground truth:** Per paragraph (j), before touching a suspended load ground personnel need the static charge dissipated by a grounding device, or they must wear protective rubber gloves. Per paragraph (e)(1), employees receiving the load must have complete eye protection and hard hats secured by chinstraps, and per (e)(2) they must not wear loose-fitting clothing likely to flap in the downwash.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "protective_items", "parameters": {"grounding_device": "True", "protective_rubber_gloves": "True", "complete_eye_protection": "True", "hardhats": "True", "chinstraps": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "protective_items", "parameters": {"static_charge_protection": "protective rubber gloves", "personal_protective_equipment": "complete eye protection and hard hats secured by chinstraps"}}

107. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_01`

- **Question:** In the worked example, what is the age-corrected threshold shift at 4000 Hz (after subtracting the aging contribution)?
- **Ground truth:** 17 dB
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_age_corrected_threshold_shift", "parameters": {"frequency": "4000", "hearing_threshold_baseline": "5", "hearing_threshold_recent": "25", "aging_contribution": "3"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_age_corrected_threshold_shift", "parameters": {"frequency": "4000", "hearing_threshold_most_recent_audiogram": "25", "hearing_threshold_baseline_audiogram": "5", "aging_contribution": "3"}}

108. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_02`

- **Question:** For a plan that was not developmental, the determination of eligibility shall not be sooner than how many years following the date of commencement of operations under the plan?
- **Ground truth:** 3 years
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "determination_of_eligibility", "parameters": {"years": "3"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_determination_of_eligibility", "parameters": {"years": "3"}}

109. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0227\_01`

- **Question:** Combining the UL listings, how many of the materials available from Underwriters Laboratories are IBR approved for § 1910.106(b)(1)(iii)(a)(1)?
- **Ground truth:** All three (UL 58-61, UL 80-63, and UL 142-68) are IBR approved for § 1910.106(b)(1)(iii)(a)(1).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "count", "parameters": {"es": ["UL 58-61 Steel Underground Tanks for Flammable and Combustible Liquids, 5th Ed.", "UL 80-63 Steel Inside Tanks for Oil-Burner Fuel", "UL 142-68 Steel Above Ground Tanks for Flammable and Combustible Liquids"]}}
- ✗ **VERITROOPER (FAIL):** {"name": "count", "parameters": {"iterable": "[58, 61, 80, 63, 142, 68]"}}

110. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0333\_03`

- **Question:** If a platform uses more than two suspension ropes in an intermittent stabilization system, which ropes require a stabilizer attachment?
- **Ground truth:** Only the two building-side suspension ropes at the platform ends require a stabilizer attachment.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "stabilizer_attachment_requirement", "parameters": {"number_of_suspension_ropes": "more than two"}}
- ✗ **VERITROOPER (FAIL):** {"name": "intermittent_stabilization_system", "parameters": {"number_of_suspension_ropes": "more than 2"}}

111. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_00`

HALLUCINATION

- **Question:** How many days does the Secretary have to respond to a certified committee's request before conducting an inspection under § 1960.31(a)(2)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_response_time", "parameters": {"days": "90"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

112. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3423\_03`

- **Question:** In the non-mandatory appendix, the column head area is defined as the distance between lines located at what multiple of the slab or drop panel thickness?
- **Ground truth:** one and one half times the thickness of the slab or drop panel
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "appendix_to_%C2%A7_1926.705", "parameters": {"x": "one and one half"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_column_head_area", "parameters": {"multiple": "1.5"}}

113. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_01`

HALLUCINATION

- **Question:** Which federal office is responsible for reviewing batching classification decisions made under A.O.5.1.2?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Assistant Secretary", "parameters": {}}
- **✓ VERITROOPER (PASS):** The evidence provided does not contain enough information to answer this question.

114. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0742\_02`

MODEL ERROR

- **Question:** If an alarm is being repaired, how do the personnel requirements differ between subpart (i) and subpart (iii)?
- **Ground truth:** Subpart (i) generally requires appropriately trained persons for repairs, but subpart (iii) imposes a stricter standard for alarms, requiring they be adjusted or repaired only by the manufacturer or a technician trained by the manufacturer.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "compare_personnel_requirements", "parameters": {"subpart_i": "alarm repair", "subpart_iii": "alarm repair"}}
- **✓ VERITROOPER (PASS):** {"name": "compare_personnel_requirements", "parameters": {"subpart_i": "Repairs or adjustments to respirators are to be made only by persons appropriately trained to perform such operations and shall use only the respirator manufacturer", "subpart_iii": "Reducing and admission valves, regulators, and alarms shall be adjusted or repaired only by the manufacturer or a technician trained by the manufacturer"}}

115. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_05`

- **Question:** Category 4 flammable liquids include liquids having flashpoints above what temperature?
- **Ground truth:** above 140 °F (60 °C)
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_flashpoint_category", "parameters": {"flashpoint": "140"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_flashpoint_category", "parameters": {"category": "4"}}

116. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_09`

HALLUCINATION

- **Question:** What penalty applies if ventilation ductwork is improperly connected to a chimney used for products of combustion?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "violation_penalty", "parameters": {"penalty_amount": "1000", "penalty_reason": "improperly connected ventilation ductwork"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

117. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_00`

HALLUCINATION

- **Question:** What is the minimum thickness required for the bulkheads used to contain compressed air?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "bulkhead_thickness", "parameters": {"wheel_diameter": "20", "wheel_hole_diameter": "12"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the minimum thickness required for the bulkheads used to contain compressed air.

118. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0517\_00`

- **Question:** Under the building requirements in subparagraph (g)(2)(ii), where must direct sources of heat be located, and under the composition requirements in subparagraph (g)(2)(iv), what is the minimum flashpoint requirement for hydrocarbon liquid fuel used?
- **Ground truth:** All direct sources of heat shall be provided exclusively from units located outside the mixing building, and no hydrocarbon liquid fuel with a flashpoint lower than that of No. 2 diesel fuel oil (125 °F minimum) shall be used.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_building_and_composition_requirements", "parameters": {"direct_sources_of_heat_location": "outside the mixing building", "minimum_flashpoint_requirement": "125"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_building_heat_source_location_and_flashpoint_requirement", "parameters": {"heat_source_location": "outside the mixing building", "flashpoint_requirement": "125"}}

119. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2565\_00`

MODEL ERROR

- **Question:** Under what statute is a 'consumer product' defined for purposes of subpart (v), and under what statute is a 'hazardous substance' defined for purposes of subpart (6)(ii)?
- **Ground truth:** A 'consumer product' is defined under the Consumer Product Safety Act (15 U.S.C. 2051 et seq.) in subpart (v), while a 'hazardous substance' in subpart (6)(ii) is defined under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) (42 U.S.C. 9601 et seq.).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_statute_definition", "parameters": {"term": "consumer product", "subpart": "v"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_statute_definition", "parameters": {"term": "consumer product", "subpart": "v"}}

120. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4450\_04`

- **Question:** Which section of the Affordable Care Act amended the Fair Labor Standards Act by adding new section 18C?
- **Ground truth:** Section 1558
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 1558 of the Patient Protection and Affordable Care Act", "parameters": {"Public Law": "111-148", "Stat": "124", "amended version": "124 Stat. 1029"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"section_number": "1558"}}

121. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4293\_00`

- **Question:** If a facility is classified in the interest of national security, may an employee representative who lacks access authorization accompany the inspector in those areas?
- **Ground truth:** No; only persons authorized to have access to such facilities may accompany a Safety and Health Inspector in those areas.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "1903.9", "parameters": {"d": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If a facility is classified in the interest of national security, may an employee representative who lacks access authorization accompany the inspector in those areas?", "answer": "No"}}

122. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2569\_01`

- **Question:** When a chemical manufacturer classifies a mixture it produces, may it rely on the safety data sheets of the individual ingredients if it knows the data sheet omits required information?
- **Ground truth:** No, it may not rely on the data sheet where it knows, or in the exercise of reasonable diligence should know, that the safety data sheet misstates or omits required information.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_mixture", "parameters": {"rely_on_sds": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "classify_mixture", "parameters": {"rely_on_sds": "False"}}

123. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_09`

HALLUCINATION

- **Question:** What is the penalty for an employer who fails to designate a person as isolated during the potential transmission period?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "notify_employer_of_failure_to_correct_violation_and_propose_additional_penalty", "parameters": {"penalty_amount": "1000", "violation_description": "Failure to designate a person as isolated during the potential transmission period"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

124. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_3327\_02`

WRONG CALC/QUOTE

- **Question:** What is the maximum clearance allowed above the walking/working surface for a securely fastened toeboard?
- **Ground truth:** 1/4 inch (0.7 cm)
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_clearance", "parameters": {"x": "0"}}
- ✗ **VERITROOPER (FAIL):** {"name": "max_clearance", "parameters": {"clearance": "0"}}

125. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3985\_02`

- **Question:** For which equipment are jib stops NOT required (when a jib is attached)?
- **Ground truth:** Derricks.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "jib_stops_are_not_required_for", "parameters": {"equipment": "derricks"}}
- ✓ **VERITROOPER (PASS):** {"name": "equipment_exception", "parameters": {"equipment": "derricks"}}

126. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3641\_00`

- **Question:** For an NPE system, given the air change specification and the pressure differential specification, what two numeric minimums must be maintained inside the enclosure?
- **Ground truth:** At least 4 air changes per hour and a minimum of -0.02 column inches of water pressure differential relative to outside pressure.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "NPE_system_specifications", "parameters": {"air_changes_per_hour": "4", "pressure_differential": "0"}}
- ✗ **VERITROOPER (FAIL):** {"name": "NPE_system_specifications", "parameters": {"air_changes": "4", "pressure_differential": "0"}}

127. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_01`

- **Question:** What is the upper flashpoint limit (in °F) for a liquid to be classified as a flammable liquid under this definition?
- **Ground truth:** 199.4 °F (93 °C)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "flashpoint", "parameters": {"°F": "199"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_flashpoint_limit", "parameters": {"flashpoint_limit": "199.4"}}

128. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4551\_02`

- **Question:** Under § 1990.106(b)(1), within how many days after receipt of a Director's recommendation must the Secretary publish a notice in the Federal Register?
- **Ground truth:** one hundred twenty (120) days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "response_to_recommendations_and_petitions", "parameters": {"days": "120"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_response_time", "parameters": {"days": "120"}}

129. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_04`

MODEL ERROR

- **Question:** Under paragraph (l)(3)(iii), if initial monitoring shows CdU of 16 µg/g Cr, but the second set of monitoring results during the medical exam does not show any mandatory removal trigger level exceeded, must the employee be removed by the mandatory provisions?
- **Ground truth:** No; if the second set does not show a mandatory removal trigger level exceeded, the employee is not required to be removed by the mandatory provisions of this paragraph.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"l": "3", "i": "iii", "CdU": "16", "trigger_level": "15"}}
- ✓ **VERITROOPER (PASS):** {"name": "paragraph_l_3_iii_mandatory_removal", "parameters": {"CdU": "16", "second_set_exceeds_trigger": "false"}}

130. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3285\_03`

- **Question:** If a worker is using a boatswains' chair rather than a pump jack scaffold, does the 12-inch minimum width requirement apply?
- **Ground truth:** No, there is no minimum width requirement for boatswains' chairs.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_minimum_width_requirement_applied", "parameters": {"scaffold_type": "boatswains"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_width_requirement", "parameters": {"scaffold_type": "boatswains"}}

131. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_05`

- **Question:** What edition of Bretherick's Hazards in the Chemical Laboratory is referenced?
- **Ground truth:** 3rd edition, 1986
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Bretherick", "parameters": {"edition": "3rd", "year": "1986"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_edition", "parameters": {"author": "Bretherick", "book_title": "Hazards in the Chemical Laboratory"}}

132. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1622\_03`

- **Question:** Cable tray systems may not be used where subjected to severe physical damage; under what comparable condition may conductors be separately enclosed in flexible nonmetallic tubing in open wiring systems?
- **Ground truth:** Cable tray systems may not be used where subjected to severe physical damage (paragraph (v)), while in open wiring systems conductors may be enclosed in flexible nonmetallic tubing only in dry locations where not exposed to severe physical damage (paragraph (iii)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "open wiring on insulators", "parameters": {"condition": "in dry locations, where not exposed to severe physical damage"}}
- ✓ **VERITROOPER (PASS):** {"name": "open_wiring", "parameters": {"condition": "not exposed to severe physical damage"}}

133. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3703\_01`

- **Question:** Based on the counting rules for the example structures, what is the combined fiber count for Structure No. 7 and Structure No. 12, and why are they counted that way?
- **Ground truth:** Both Structure No. 7 and Structure No. 12 are counted as 1/2 each (Structure 7 because the fiber crosses the circle once, Structure 12 because although split, the fiber only crosses once), giving a combined count of 1.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_fiber_count", "parameters": {"structure_no": "7", "count": "0", "explanation": "Fiber crosses circle once."}}
- ✓ **VERITROOPER (PASS):** {"name": "combined_fiber_count", "parameters": {"Structure No. 7": "1/2", "Structure No. 12": "1/2"}}

134. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2565\_03`

- **Question:** Subpart (vi) covers treated seed and subpart (6)(iv) covers wood products. Which exempted material's exception is administered by the Department of Agriculture, and what condition limits the wood products exemption?
- **Ground truth:** Agricultural or vegetable seed treated with pesticides (subpart (vi)) is labeled per the Federal Seed Act (7 U.S.C. 1551 et seq.) with regulations issued by the Department of Agriculture. The wood products exemption in subpart (6)(iv) is limited to cases where the only hazard is flammability or combustibility; treated wood, or wood that may be subsequently sawed or cut generating dust, is not exempted.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exempted_material_administration", "parameters": {"material": "treated seed", "administered_by": "Department of Agriculture", "condition": "the only hazard they pose to employees is the potential for flammability or combustibility"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_exemption_info", "parameters": {"material": "treated seed", "administered_by": "Department of Agriculture", "condition": "the only hazard they pose to employees is the potential for flammability or combustibility"}}

135. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_06`

- **Question:** The right of entry and inspection in a State plan must be at least as effective as that provided in which section of the Act, per § 1902.3(e)?
- **Ground truth:** section 8
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section", "parameters": {"x": "8"}}
- ✓ **VERITROOPER (PASS):** {"name": "section", "parameters": {"number": "8"}}

136. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_06`

HALLUCINATION

- **Question:** Who is responsible for supplying the heavy canvas or leather gloves to the operators?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "supply_gloves", "parameters": {"employer": "True"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question of who is responsible for supplying the heavy canvas or leather gloves to the operators.

137. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3059\_04`

- **Question:** If the same filters are being used for different coating materials, under what condition is alternate use prohibited?
- **Ground truth:** Where the combination of materials may be conducive to spontaneous ignition.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "alternate_use_prohibited", "parameters": {"condition": "where the combination of materials may be conducive to spontaneous ignition"}}
- **✓ VERITROOPER (PASS):** {"name": "alternate_filter_use_prohibited", "parameters": {"condition": "where the combination of materials may be conducive to spontaneous ignition"}}

138. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1206\_02`

- **Question:** Using the marking requirement in (b)(1)(ii) and the empty cylinder requirement in (b)(2)(iii), what two conditions must be met for an empty compressed gas cylinder?
- **Ground truth:** It must be legibly marked to identify the gas content by chemical or trade name (by stenciling, stamping, or labeling, not readily removable), and its valve must be closed.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "compressed_gas_cylinder_requirements", "parameters": {"marking_requirement": "legibly marked with either the chemical or the trade name of the gas", "empty_cylinder_requirement": "valves closed"}}
- **✓ VERITROOPER (PASS):** {"name": "check_compressed_gas_cylinder_conditions", "parameters": {"condition1": "The cylinder must be legibly marked with the chemical or trade name of the gas.", "condition2": "The valve of the empty cylinder must be closed."}}

139. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0325\_00`

- **Question:** If an employer has 8 employees and an OSHA standard in this part requires a fire prevention plan, must the plan be in writing?
- **Ground truth:** No, an employer with 10 or fewer employees may communicate the plan orally to employees.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "written_fire_prevention_plan_required", "parameters": {"employer_employees": "8", "OSHA_standard_requires_plan": "True"}}
- **✗ VERITROOPER (FAIL):** {"name": "is_fire_prevention_plan_required_to_be_in_writing", "parameters": {"number_of_employees": "8"}}

140. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2204\_01`

- **Question:** Using the normal total white blood count value and the adjustment noted for cigarette smokers, what could the upper white count value be for a cigarette smoker relative to the standard normal?
- **Ground truth:** The normal total white blood count is approximately 7,200/mm³ plus or minus 3,000 (so up to about 10,200/mm³). For cigarette smokers, the upper range may be 2,000 cells higher than normal for the laboratory, so the upper value could be about 12,200/mm³.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "calculate_upper_white_count_value", "parameters": {"normal_total_white_blood_count": "7200", "adjustment_for_smokers": "2000"}}

- ✓ **VERITROOPER (PASS):** {"name": "calculate_upper_white_count", "parameters": {"normal_total_white_blood_count": "7200", "adjustment_for_smokers": "2000"}}

141. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0455\_03`

- **Question:** If only Category 3 flammable liquids with a flashpoint at or above 100 °F or Category 4 flammable liquids are handled, is special Class I electrical equipment required?

- **Ground truth:** No, ordinary electrical equipment is satisfactory, though care must be used in locating it to prevent hot metal from falling into open equipment.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_special_class_i_electrical_equipment_required", "parameters": {"x": "False"}}

- ✗ **VERITROOPER (FAIL):** {"name": "electrical_equipment_required", "parameters": {"category": "3", "flashpoint": "100"}}

142. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_03`

- **Question:** What flashpoint threshold separates Category 1/2 from Category 3 flammable liquids?

- **Ground truth:** 73.4 °F (23 °C)

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_flashpoint_threshold", "parameters": {"category": "Category 1/2"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_flashpoint_threshold", "parameters": {"category": "Category 1/2", "threshold": "73.4"}}

143. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0075\_02`

- **Question:** If an employer files a notice of intention to contest under § 1903.17, is the employer still required to keep the citation posted?

- **Ground truth:** Yes, the posting responsibility continues unless and until the Review Commission issues a final order vacating the citation.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "posting_requirement", "parameters": {"still_required": "True"}}

- ✗ **VERITROOPER (FAIL):** {"name": "posting_requirement", "parameters": {"event": "filing a notice of intention to contest under \u00a7 1903.17"}}

144. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_00`

- **Question:** If an agency wants to realize exemption from unannounced OSHA inspections, what committees must it form?

- **Ground truth:** It must form a committee at the national level and at any establishment or grouping of establishments that is to be exempt, while keeping the Secretary advised of the locations and activities where such committees are functioning.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "committee_organization", "parameters": {"level": "national level and at any establishment or grouping of establishments that is to be exempt"}}

- ✓ **VERITROOPER (PASS):** {"name": "form_committees", "parameters": {"national_level": "True", "establishment_level": "True"}}

145. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4138\_01`

- **Question:** Comparing Option (1) physical attachment and Option (4)(B), how is the connection to the flotation device characterized in each?

- **Ground truth:** Option (1) physical attachment connects the crane/derrick physically to the barge, pontoons, vessel or other means of flotation using methods like crossed-cable systems, bolting, welding, or strapping with chains. Option (4)(B) requires that the wire rope is attached physically to the vessel/flotation device. Both require a physical attachment to the vessel/flotation device.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_options", "parameters": {"option1": "Option (1)", "characteristic1": "physically attached to the barge, pontoons, vessel or other means of flotation", "option2": "Option (4)(B)", "characteristic2": "attached physically to the vessel/flotation device"}}
- ✗ **VERITROOPER (FAIL):** {"name": "characterize_connection", "parameters": {"option": "Option (1) physical attachment"}}

146. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_03`

HALLUCINATION

- **Question:** How often must the medical examination referenced in paragraph (h)(2)(ii) be repeated for employees using respirators?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "medical examination", "parameters": {"frequency": "every 12 months"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

147. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_01`

- **Question:** Emergency landings are required where there is a travel of how many feet or more between floor landings?
- **Ground truth:** 50 feet
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "emergency_landings", "parameters": {"travel": "50"}}
- ✓ **VERITROOPER (PASS):** {"name": "emergency_landings_required", "parameters": {"travel": "50"}}

148. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_08`

HALLUCINATION

- **Question:** How frequently must containers and air-storage tanks under air pressure be retested for maintenance under the ASME Code?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "retest_for_maintenance", "parameters": {"frequency": "at regular intervals"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

149. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_01`

HALLUCINATION

- **Question:** What is the minimum thickness required for the noncombustible surfacing in liquid oxygen leakage areas?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_min_thickness", "parameters": {"material": "noncombustible surfacing", "location": "liquid oxygen leakage areas"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

150. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3285\_01`

MISINTERPRETED

- **Question:** If a platform's front edge is 16 inches from the face of the work and no guardrail or personal fall arrest system is used, does it comply with paragraph (b)(3)?
- **Ground truth:** No, because the front edge must not be more than 14 inches from the face of the work

unless guardrails and/or personal fall arrest systems are used.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "check_compliance", "parameters": {"distance": "16"}}
- ✔ **VERITROOPER (PASS):** {"name": "check_compliance", "parameters": {"front_edge_distance": "16", "guardrail_used": "False", "personal_fall_arrest_system_used": "False"}}

151. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2920\_00`

- **Question:** Under paragraph (i)(11)(vi), respirators must be used when employees are in areas where air contaminant concentrations exceed limits or where oxygen is below a certain percent. During which emergencies does this requirement apply, and what triggers the splashing protection referenced there?
- **Ground truth:** The requirement applies during the emergencies specified in paragraph (i)(11)(v), which concern situations where there is a danger of splashing - for example, when additions are made manually to the tanks, or when acids and chemicals are removed from the tanks. The oxygen threshold triggering respirator use is when oxygen concentrations are less than 19.5 percent.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph_i_11_vi", "parameters": {"emergency_situation": "when employees must be in areas where concentrations of air contaminants are greater than the limits set by paragraph (i)(2)(iii) of this section or oxygen concentrations are less than 19.5 percent", "splashing_protection_trigger": "when additions are made manually to the tanks, or when acids and chemicals are removed from the tanks"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_respirator_requirement", "parameters": {"emergency_type": "emergencies specified in paragraph (i)(11)(v)", "splashing_protection_trigger": "additions are made manually to the tanks, or when acids and chemicals are removed from the tanks"}}

152. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_01`

HALLUCINATION

- **Question:** What form must a certified safety and health committee use to submit its inspection request to the Secretary?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA-20", "parameters": {}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

153. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4456\_01`

- **Question:** Under what circumstance will the Office of Administrative Law Judges grant a respondent's motion for a stay of the Assistant Secretary's preliminary order of reinstatement?
- **Ground truth:** A stay shall be granted only based on exceptional circumstances.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "grant_motion_for_stay", "parameters": {"circumstance": "exceptional circumstances"}}
- ✔ **VERITROOPER (PASS):** {"name": "grant_motion", "parameters": {"circumstance": "exceptional circumstances"}}

154. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_04`

HALLUCINATION

- **Question:** What is the maximum allowable concentration of combustible dust before welding is prohibited under the explosive atmospheres provision?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_maximum_allowable_concentration", "parameters": {"class": "Class II, Division 1", "factor": "combustible dust"}}
- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

155. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_2159\_01`

- **Question:** Using $C_{ld} = k(sd)/m$ with $sd = 1.05$ and $m = 772.7$, what is the quantitative detection limit ($k = 10$) in $\mu\text{g/mL}$? Round to 3 decimal places.
- **Ground truth:** 0.014 $\mu\text{g/mL}$
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "C_ld", "parameters": {"k": "10", "sd": "1", "m": "772"}}
- ✓ **VERITROOPER (PASS):** {"name": "round", "parameters": {"number": "10 * 1.05 / 772.7", "ndigits": "3"}}

156. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_00`

HALLUCINATION

- **Question:** What minimum distance must a bulk oxygen system maintain from flammable gas lines?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "distance_from_exposures", "parameters": {"exposure_type": "flammable gas lines", "capacity": "less than 5000"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

157. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_00`

HALLUCINATION

- **Question:** What is the fee for filing a retaliation complaint under SPA?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_fee", "parameters": {"section": "1986.114", "complaint_type": "retaliation"}}
- ✓ **VERITROOPER (PASS):** There is no information in the provided evidence about a fee for filing a retaliation complaint under SPA.

158. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_04`

HALLUCINATION

- **Question:** What minimum temperature rating must the automatic fire dampers be certified to withstand?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "automatic fire damper", "parameters": {"temperature_rating": "250"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

159. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_01`

- **Question:** The operator must not leave the controls while the load is suspended. Under what conditions does this prohibition NOT apply (i.e., when may the operator leave)?
- **Ground truth:** Only when ALL of the following are met: (i) the operator remains adjacent to the equipment and is not engaged in any other duties; (ii) the load is to be held suspended for a period exceeding normal lifting operations; (iii) the competent person determines it is safe and implements measures to restrain the boom hoist and telescoping, load, swing, and outrigger/stabilizer functions; and (iv) barricades or caution lines and notices are erected to prevent all employees from entering the fall zone, with no employees permitted in the fall zone.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "operator_leave_controls", "parameters": {"condition": "The operator remains adjacent to the equipment and is not engaged in any other duties. The load is to be held suspended for a period of time exceeding normal lifting operations. The competent person determines that it is safe to do so and implements measures necessary to restrain the boom hoist and telescoping, load, swing,

and outrigger or stabilizer functions. Barricades or caution lines, and notices, are erected to prevent all employees from entering the fall zone."}}

- ✖ **VERITROOPER (FAIL):** {"name": "leaving_controls_while_load_suspended", "parameters": {"condition": "operator remains adjacent to the equipment and is not engaged in any other duties"}}

160. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_05`

- **Question:** Under § 1902.39(g), the Assistant Secretary shall publish a notice of informal hearing in the Federal Register not later than how many days after the last day for filing written views or comments?

- **Ground truth:** 30 days

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "publish_notice", "parameters": {"days": "30"}}

- ✓ **VERITROOPER (PASS):** {"name": "publish_notice", "parameters": {"days": "30"}}

161. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0370\_01`

- **Question:** For abrasive cutting-off wheels larger than 20 inches in diameter, the thickness must not be more than what fraction of their diameter?

- **Ground truth:** one-sixtieth

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_fraction", "parameters": {"diameter": "20", "thickness": "1/60"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_fraction", "parameters": {"diameter": "20", "fraction": "one-sixtieth"}}

162. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3208\_02`

UNIT/FORMAT MISMATCH

- **Question:** If live parts are elevated above the floor to exclude unqualified persons, what minimum elevation is required for this to serve as adequate guarding?

- **Ground truth:** 8 feet (2.44 m) or more above the floor or other working surface.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "elevation_for_guarding", "parameters": {"elevation": "8"}}

- ✓ **VERITROOPER (PASS):** {"name": "guarding_live_parts", "parameters": {"elevation": "8"}}

163. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_02`

- **Question:** For Class 2 equipment, what is the maximum proof-test current for a 280-mm (11-in) glove?

- **Ground truth:** 16

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_maximum_proof_test_current", "parameters": {"class_of_equipment": "2", "glove_size_mm": "280"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_max_proof_test_current", "parameters": {"class_of_equipment": "2", "glove_size_mm": "280"}}

164. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_00`

- **Question:** For manlift travel, an emergency landing must be provided so that there is a landing for every how many feet or less of travel?

- **Ground truth:** 25 feet

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "emergency_landing", "parameters": {"interval": "25"}}

- ✓ **VERITROOPER (PASS):** {"name": "emergency_landing", "parameters": {"travel": "25"}}

165. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1848\_00`

MODEL ERROR

- **Question:** Based on the permissible exposure limit in Section I.D and the respirator requirements in Section III.A, at what airborne fiber concentration must you begin wearing a respirator, and what type of respirator is required if concentrations reach 3 f/cc?
- **Ground truth:** You must wear a respirator when exposure exceeds the PEL of 0.1 f/cc. Since 3 f/cc exceeds 2 f/cc, an air-supplied, positive-pressure, full facepiece respirator must be used (HEPA air-purifying respirators are only allowed where concentrations do not exceed 2 f/cc).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "respirator_requirements", "parameters": {"PEL": "0", "airborne_fiber_concentration": "3"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_respirator_requirements", "parameters": {"airborne_fiber_concentration": "3"}}

166. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_3282\_00`

MODEL ERROR

- **Question:** What is the general capacity requirement for each scaffold and scaffold component, and under what conditions does it NOT apply?
- **Ground truth:** Each scaffold and scaffold component shall be capable of supporting, without failure, its own weight and at least 4 times the maximum intended load applied or transmitted to it - except as provided in paragraphs (a)(2), (a)(3), (a)(4), (a)(5) and (g) of this section.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "capacity_requirement", "parameters": {"scaffold_component": "scaffold", "capacity": "4 times the maximum intended load", "conditions": "aerial lifts"}}
- **✓ VERITROOPER (PASS):** {"name": "capacity_requirement", "parameters": {"capacity": "4 times the maximum intended load", "conditions": "Except as provided in paragraphs (a)(2), (a)(3), (a)(4), (a)(5) and (g) of this section"}}

167. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_00`

HALLUCINATION

- **Question:** What is the cadmium action level concentration referenced in work operations under subsection (v)?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_cadmium_action_level", "parameters": {"concentration": "2.5"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The action level is mentioned in subsection (v), but its concentration is not specified in the provided text.

168. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_4450\_03`

MODEL ERROR

- **Question:** What is the Statutes at Large citation (124 Stat.) for the Health Care and Education Reconciliation Act of 2010, Public Law 111-152?
- **Ground truth:** 124 Stat. 1029
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_statutes_at_large_citation", "parameters": {"public_law": "111-152", "act_name": "Health Care and Education Reconciliation Act of 2010"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_statutes_at_large_citation", "parameters": {"public_law": "111-152", "statutes_at_large": "124"}}

169. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2261\_01`

MODEL ERROR

- **Question:** When a healthcare professional is evaluating an employee after an exposure incident, is the employer required to provide the source individual's blood test results?
- **Ground truth:** Yes, but only if available.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "provide_source_individual_blood_test_results", "parameters": {"required": "true"}}
- **✗ VERITROOPER (FAIL):** {"name": "evaluate_exposure_incident", "parameters": {"provide_source_individual_blood_test_results": "True"}}

170. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1848\_01`

- **Question:** Combining the disposal requirement in Section IV.A.4 with the packaging requirement in Section IV.E, how must asbestos fireproofing material removed from a building be disposed of?
- **Ground truth:** Asbestos fireproofing or insulating material removed from buildings is a waste under IV.A.4, and per IV.E it must be disposed of in leak-tight 6-mil thick plastic bags, plastic-lined cardboard containers, or plastic-lined metal containers, and sealed while wet before drying out.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "dispose_of_asbestos_fireproofing_material", "parameters": {"disposal_method": "leak-tight 6-mil plastic bags, plastic-lined cardboard containers, or plastic-lined metal containers", "packaging_requirement": "Sealed in containers before they dry out to minimize the release of asbestos fibers during handling"}}
- **✗ VERITROOPER (FAIL):** {"name": "asbestos_disposal_requirements", "parameters": {"disposal_requirement": "Section IV.A.4", "packaging_requirement": "Section IV.E"}}

171. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0455\_04`

- **Question:** Where the provisions require Class I, Division 1 or Division 2 electrical equipment, may ordinary switchgear be used, and under what condition?
- **Ground truth:** Yes, ordinary electrical equipment including switchgear may be used if installed in a room or enclosure maintained under positive pressure with respect to the hazardous area.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "ordinary_switchgear_permitted", "parameters": {"Class": "I", "Division": "1 or 2", "condition": "installed in a room or enclosure which is maintained under positive pressure with respect to the hazardous area"}}
- **✓ VERITROOPER (PASS):** {"name": "electrical_equipment", "parameters": {"class": "I", "division": "1", "ordinary_switchgear": "True", "condition": "installed in a room or enclosure which is maintained under positive pressure with respect to the hazardous area"}}

172. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_12`

- **Question:** What is the injection volume specified in the GC conditions for MDA analysis?
- **Ground truth:** 5.0 uL
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "injection volume", "parameters": {"volume": "5"}}
- **✓ VERITROOPER (PASS):** {"name": "get_injection_volume", "parameters": {"volume": "5"}}

173. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2346\_00`

- **Question:** If an employee's designated representative wishes to observe AN exposure monitoring, but doing so requires entering an area where protective clothing is mandatory, what must the employer do before allowing entry?
- **Ground truth:** The employer must provide the observer with the required personal protective clothing and equipment, assure its use, and require the observer to comply with all other applicable safety and health

procedures.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_personal_protective_clothing_and_equipment", "parameters": {"observer": "employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_personal_protective_clothing_and_equipment", "parameters": {"require_use": "True", "comply_with_safety_procedures": "True"}}

174. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3809\_02`

- **Question:** Within how many days of the examination must the PLHCP explain the results to the employee from the CBD diagnostic center?
- **Ground truth:** 30 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "explain_results_from_cbd_diagnostic_center", "parameters": {"days": "30"}}
- ✓ **VERITROOPER (PASS):** {"name": "explain_results", "parameters": {"days": "30"}}

175. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_00`

- **Question:** For a flammable liquid, what is the maximum vapor pressure (in pounds per square inch absolute at 100 °F) that the definition specifies?
- **Ground truth:** 40 pounds per square inch (absolute)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "vapor_pressure", "parameters": {"pounds": "40"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_vapor_pressure", "parameters": {"vapor_pressure": "40"}}

176. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3386\_00`

- **Question:** When mechanical equipment is being used, what is the minimum distance the warning line must be erected from the roof edge that is perpendicular to the direction of mechanical equipment operation?
- **Ground truth:** Not less than 10 feet (3.1 m).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "warning_line_distance", "parameters": {"distance": "10 feet"}}
- ✓ **VERITROOPER (PASS):** {"name": "minimum_distance_warning_line", "parameters": {"distance": "10"}}

177. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_02`

HALLUCINATION

- **Question:** What air pressure value must compressed-air bulkheads be tested to in pounds per square inch?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "test_compressed_air_bulkheads", "parameters": {"pressure": "50"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

178. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2508\_03`

- **Question:** If a filter removes 99.95 percent of mono-dispersed particles of 0.3 micrometers in diameter, does it qualify as a HEPA filter?
- **Ground truth:** No, a HEPA filter must be at least 99.97 percent efficient in removing such particles.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "high-efficiency particulate air [HEPA] filter", "parameters": {"percent_efficient": "99", "particle_diameter": "0"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_HEPA_filter", "parameters": {"efficiency": "99", "particle_size": "0"}}

179. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_00`

- **Question:** If a party wishes to rely on an objection later in a proceeding, but that objection was never made before the presiding hearing examiner, can it be relied upon?
- **Ground truth:** No, only objections made before the presiding hearing examiner may be relied upon subsequently in a proceeding.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "objections_made_before_presiding_hearing_examiner", "parameters": {"relied_upon": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "objection_reliance", "parameters": {"made_before_examiner": "False"}}

180. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2565\_02`

MODEL ERROR

- **Question:** Combining subparts (6)(i) and (6)(ii), which federal agency administers the regulations referenced in both exclusions, and which two statutes are involved?
- **Ground truth:** The Environmental Protection Agency administers the regulations referenced in both subparts. Subpart (6)(i) involves the Solid Waste Disposal Act as amended by the Resource Conservation and Recovery Act of 1976 (42 U.S.C. 6901 et seq.), and subpart (6)(ii) involves CERCLA (42 U.S.C. 9601 et seq.).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "Combining_subparts_6i_and_6ii", "parameters": {"agency": "Federal Occupational Safety and Health Administration", "statutes": "Williams-Steiger Occupational Safety and Health Act of 1970, Contract Work Hours and Safety Standards Act"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_agency_and_statutes", "parameters": {"agency": "Environmental Protection Agency", "statutes": "['Solid Waste Disposal Act', 'Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)']"}}

181. PASS · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0478\_03`

- **Question:** What temperature must the water be maintained at when discarded filter rolls are kept completely in water?
- **Ground truth:** UNANSWERABLE
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "maintain_water_temperature", "parameters": {"temperature": "not specified"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

182. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_05`

- **Question:** Pending a final medical determination, may the employer return an employee who has been on removal status for the preceding six months due to MDA exposure?
- **Ground truth:** No; the employer shall await a final medical determination.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"option": "no"}}
- ✗ **VERITROOPER (FAIL):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"option": "return", "removal_status_duration": "6 months"}}

183. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_05`

HALLUCINATION

- **Question:** What is the minimum height required for the standard railings installed along the walkways?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "standard_railing_height", "parameters": {"height": "42 inches"}}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

184. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_02`

HALLUCINATION

- **Question:** Which specific federal office or agency receives the complaints filed under the Pipeline Safety Act?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Area Director", "parameters": {"address": "http://www.osha.gov"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

185. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_04`

MODEL ERROR

- **Question:** May an employer use a method of calculating incident heat energy not included in appendix E?
- **Ground truth:** Yes, if the chosen method reasonably predicts the incident energy to which the employee would be exposed.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "incident_heat_energy_calculation", "parameters": {"method_included_in_appendix_E": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "Note 1 to paragraph", "parameters": {"I": "8", "method": "calculating incident heat energy not included in appendix E"}}

186. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4060\_01`

- **Question:** If an employee is hoisted in a storage-tank operation using a personnel platform, which provisions apply?
- **Ground truth:** Paragraphs (a) through (n) of this section apply.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "storage-tank operation using a personnel platform", "parameters": {"provisions": "(a) through (n)"}}
- ✓ **VERITROOPER (PASS):** {"name": "apply_provisions", "parameters": {"provisions": "paragraphs (a) through (n) of this section"}}

187. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_07`

- **Question:** What was the average percentage of the spiked amount found on the top filters in the retention efficiency study?
- **Ground truth:** 91.8%
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "average", "parameters": {"x": "91"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_average", "parameters": {"values": "[91.8]"}}

188. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4553\_00`

- **Question:** Under what circumstance will the criteria for identification, classification, and regulation of potential occupational carcinogens NOT be applied?
- **Ground truth:** When the Secretary considers evidence under the provisions of §§ 1990.143, 1990.144 and 1990.145 and determines that such evidence warrants an exception to these criteria.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "identification, classification, and regulation of potential occupational carcinogens", "parameters": {"circumstance": "unless the Secretary considers evidence under the provisions of \u00a71990.143, 1990.144 and 1990.145 and determines that such evidence warrants an exception to these criteria"}}
- ✓ **VERITROOPER (PASS):** {"name": "circumstance_for_exception", "parameters": {"circumstance": "the

Secretary considers evidence under the provisions of 1990.143, 1990.144 and 1990.145 and determines that such evidence warrants an exception to these criteria"}}

189. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3038\_00`

MODEL ERROR

- **Question:** If an employer evacuates all employees from the worksite when an emergency occurs and does not permit any employees to assist in handling the emergency, are they still required to comply with the emergency response program requirements of paragraph (p)(8)?
- **Ground truth:** No, they are exempt from paragraph (p)(8) provided they have an emergency action plan complying with § 1926.35.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "is_exempt_from_emergency_response_program", "parameters": {"evacuate_employees": "True", "permit_employee_assistance": "False"}}
- **✗ VERITROOPER (FAIL):** {"name": "emergency_response_program", "parameters": {"employer_evacuates_employees": "True", "employees_assist_in_emergency": "False"}}

190. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_00`

- **Question:** If employees are on a scaffold, can it be moved horizontally provided it is a mobile scaffold following the provisions of § 1926.452(w)?
- **Ground truth:** Yes, mobile scaffolds may be moved horizontally with employees on them where the provisions of § 1926.452(w) are followed.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Scaffolds shall not be moved horizontally while employees are on them, unless they have been designed by a registered professional engineer specifically for such movement or, for mobile scaffolds, where the provisions of § 1926.452(w) are followed.", "parameters": {}}
- **✓ VERITROOPER (PASS):** {"name": "check_scaffold_movement", "parameters": {"scaffold_type": "mobile", "employees_on_scaffold": "True", "provisions_followed": "True"}}

191. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0643\_01`

- **Question:** Combining the requirement to develop a report of audit findings with the retention requirement, how many compliance audit reports must an employer keep on file?
- **Ground truth:** A report of the audit findings must be developed, and the employer must retain the two (2) most recent compliance audit reports.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "retain_compliance_audit_reports", "parameters": {"n": "2"}}
- **✓ VERITROOPER (PASS):** {"name": "retain_reports", "parameters": {"n": "2"}}

192. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1292\_00`

- **Question:** Under the general machine guarding provisions, which requires enclosure regardless of location and which only requires enclosure when located within reach from platforms or within 8 feet 6 inches from the floor?
- **Ground truth:** Per paragraph (c)(2), all gears shall be completely enclosed regardless of location. Per paragraph (c)(3), sprockets and V-belt drives must be completely enclosed only when located within reach from platforms or passageways or located within 8 feet 6 inches from the floor.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "General machine guarding", "parameters": {"(1)": "Gears", "(2)": "Sprockets and V-belt drives"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_machine_guarding_requirements", "parameters": {"location": "regardless of location", "component": "gears"}}

193. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0899\_00`

- **Question:** Under paragraph (c)(2), a fire brigade member who performs interior structural fire fighting must receive training how often, and per paragraph (c)(1), when must this training first be provided relative to performing fire brigade emergency activities?
- **Ground truth:** The interior structural fire fighting member must be provided with an education session or training at least quarterly, and per (c)(1) the training and education must be provided before they perform fire brigade emergency activities.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_training_frequency_and_initial_provision", "parameters": {"frequency": "at least quarterly", "initial_provision": "before they perform fire brigade emergency activities"}}
- ✓ **VERITROOPER (PASS):** {"name": "training_frequency", "parameters": {"frequency": "at least quarterly", "relative_time": "before performing fire brigade emergency activities"}}

194. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0325\_01`

MODEL ERROR

- **Question:** If an employer has 15 employees, does the option to communicate the fire prevention plan orally apply?
- **Ground truth:** No, the oral option only applies to employers with 10 or fewer employees, so this plan must be in writing, kept in the workplace, and made available to employees.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "check_oral_communication_eligibility", "parameters": {"number_of_employees": "15"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_oral_communication_applicable", "parameters": {"number_of_employees": "15"}}

195. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_08`

MODEL ERROR

- **Question:** When must the employer return an employee to former job status following temporary medical removal?
- **Ground truth:** When the employee no longer shows signs or symptoms of exposure to MDA, upon the advice of the physician, or when a subsequent final medical determination finds the employee no longer has a detected condition placing them at increased risk.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "return_to_former_job_status", "parameters": {"employee_removed_due_to_blood_lead_level": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "employer_return_employee", "parameters": {"reason": "When the employee no longer shows signs or symptoms of exposure to MDA, or upon the advice of the physician.", "alternative_reason": "When a subsequent final medical determination results in a medical finding, determination, or opinion that the employee no longer has a detected medical condition which places the employee at increased risk of material impairment to health from exposure to MDA."}}

196. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0882\_04`

MODEL ERROR

- **Question:** The energy control program required by paragraph (c)(1) consists of three components. Which paragraphs of subsection (c) address the second and third of those components?
- **Ground truth:** Paragraph (c)(1) lists energy control procedures, employee training, and periodic inspections. Employee training is addressed in paragraph (c)(7) (Training and communication) and periodic inspections are addressed in paragraph (c)(6) (Periodic inspection).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "energy_control_program_components", "parameters": {"c": "2-3"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_paragraphs", "parameters": {"component": "employee training"}}

197. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1622\_01`

WRONG CALC/QUOTE

- **Question:** Combining the support interval requirements for conductors smaller than No. 8 with the strap interval requirement for flexible nonmetallic tubing, what is the common maximum interval value that both share?
- **Ground truth:** Both share a maximum interval of 1.37 m (4.5 ft) - conductor supports at intervals not exceeding 1.37 m (4.5 ft) per paragraph (ii)(C), and tubing straps at intervals not exceeding 1.37 m (4.5 ft) per paragraph (iii).
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "max", "parameters": {"x": "305", "y": "152", "z": "137"}}
- ✔ **VERITROOPER (PASS):** {"name": "max", "parameters": {"a": "1.37", "b": "1.37"}}

198. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2239\_01`

- **Question:** Combining the retention requirement and the availability requirement, for how long must employee medical records be maintained, and to whom must they be provided upon request?
- **Ground truth:** The medical records must be maintained for at least 40 years, or for the duration of employment plus 20 years, whichever is longer (paragraph (iii)). These records must be provided upon request to employees, designated representatives, and the Assistant Secretary in accordance with 29 CFR 1910.1020(a)-(e) and (g)-(i) (paragraph (4)(ii) / (ii) of Availability).
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "medical_record_retention_and_availability", "parameters": {"retention_period": "duration of employment plus 30 years", "availability": "to the subject employee or former employee or to anyone having the specific written consent of the subject employee or former employee"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_retention_and_availability_requirements", "parameters": {"record_type": "employee medical records"}}

199. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_10`

- **Question:** What is the normal level for beta-2-microglobulin in urine listed in Attachment 3?
- **Ground truth:** ≤300 µg/g Cr
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "beta-2-microglobulin", "parameters": {"level": "300 \u00b5g/g Cr"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_normal_level", "parameters": {"marker": "Beta-2-microglobulin in Urine"}}

200. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0643\_00`

- **Question:** Under the compliance audit provisions, who must conduct the audit that employers are required to certify they have evaluated at least every three years, and what qualification must that person have?
- **Ground truth:** The compliance audit, which employers must certify they have evaluated at least every three years, shall be conducted by at least one person knowledgeable in the process.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "compliance_audit", "parameters": {"conducted_by": "at least one person knowledgeable in the process"}}
- ✔ **VERITROOPER (PASS):** {"name": "compliance_audit_conductor", "parameters": {"conductor": "at least one person", "qualification": "knowledgeable in the process"}}

201. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_06`

- **Question:** Within what percentage must the personal sampling pump be calibrated to the recommended flow rate?
- **Ground truth:** $\pm 5\%$
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "calibration_of_personal_pumps", "parameters": {"percentage": "5"}}
- **✓ VERITROOPER (PASS):** {"name": "calibration_percentage", "parameters": {"percentage": "5"}}

202. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_07`

HALLUCINATION

- **Question:** Which OSHA office reviews the examining physician's determination that an employee cannot function normally with a respirator?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "OSHA_review", "parameters": {"office": "Assistant Secretary"}}
- **✓ VERITROOPER (PASS):** The text does not mention which OSHA office reviews the examining physician's determination.

203. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_09`

- **Question:** For Class 3 equipment, what is the maximum use voltage AC rms?
- **Ground truth:** 26,500
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_max_use_voltage", "parameters": {"class_equipment": "3"}}
- **✓ VERITROOPER (PASS):** {"name": "get_maximum_use_voltage", "parameters": {"class_of_equipment": "3"}}

204. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3059\_00`

MISINTERPRETED

- **Question:** If a spray material is known to be highly susceptible to spontaneous heating and ignition, may filter rolls be used when applying it?
- **Ground truth:** No, filters or filter rolls shall not be used when applying such a material.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "filter_roll_usage", "parameters": {"susceptible": "True"}}
- **✓ VERITROOPER (PASS):** {"name": "print_answer", "parameters": {"answer": "No"}}

205. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4294\_01`

MODEL ERROR

- **Question:** Combining paragraph (c) and paragraph (d)(2), what minimum information about the reporting individual's identity may or may not be reflected in the establishment's log if confidentiality is requested?
- **Ground truth:** The log (per (d)(2)) requires date, time, code/reference/file number, location, description, classification, and action taken, but does NOT list the employee's name; under (c), upon request no person shall disclose the name of the individual making the report or names of employees referred to, except to authorized representatives of the Secretary. So the name is kept confidential and is not among the required log items.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "minimum_information_about_reporting_individual", "parameters": {"paragraph_c": "paragraph (c)", "paragraph_d2": "paragraph (d)(2)"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_log_info", "parameters": {"requested_confidentiality": "True"}}

206. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_02`

- **Question:** What is the detector zone temperature in the GC analysis conditions for MDA?
- **Ground truth:** 335 degrees C
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "GC conditions", "parameters": {"Zone temperatures": "Detector\u2014335 \u00b0C"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_detector_zone_temperature", "parameters": {"zone": "detector"}}

207. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_08`

HALLUCINATION

- **Question:** What exposure level triggers respirator use during the Class IV asbestos work described in subparagraph (vi)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "respirator_use", "parameters": {"exposure_level": "above the TWA or excursion limit"}}
- ✓ **VERITROOPER (PASS):** {"name": "None", "parameters": {}}

208. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4366\_00`

- **Question:** When a case is accepted for review by the ARB, the ALJ's decision becomes inoperative - but what type of order remains effective during the ARB's review despite this?
- **Ground truth:** Any order of reinstatement will be effective while review is conducted by the ARB, unless the ARB grants a motion by the respondent to stay that order based on exceptional circumstances.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_order_status", "parameters": {"order_type": "reinstatement"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer", "parameters": {"question": "When a case is accepted for review by the ARB, the ALJ", "answer": "An order of reinstatement remains effective during the ARB"}}

209. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3641\_03`

MODEL ERROR

- **Question:** Considering both the glove bag work practice for adjacent material and the negative pressure glove box work practice for adjacent ACM, how must loose or damaged material near the enclosure be handled in each case?
- **Ground truth:** In both cases, the loose/friable or damaged material adjacent to the operation shall be wrapped and sealed in two layers of six mil plastic (or otherwise rendered intact) before beginning the job.
- ✗ **Baseline (vanilla RAG) (FAIL):** [{"name": "wrap_and_seal_loose_ACM", "parameters": {"method": "glove bag work practice for adjacent material"}}; {"name": "wrap_and_seal_loose_ACM", "parameters": {"method": "negative pressure glove box work practice for adjacent ACM"}}]
- ✗ **VERITROOPER (FAIL):** {"name": "handle_adjacent_material", "parameters": {"material_condition": "loose", "enclosure_type": "glove bag"}}

210. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3948\_00`

- **Question:** An articulating/knuckle-boom truck crane delivers building supply plywood sheets and transfers them onto a structure using a fork/cradle, but the crane is NOT equipped with an automatic overload prevention device. Does the material-delivery exclusion in (c)(17)(ii) apply?
- **Ground truth:** No; the exclusion under (c)(17)(ii) applies only when the truck crane is equipped with a properly functioning automatic overload prevention device.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exclusion_in_1910_268_c_17_ii", "parameters": {"x": "False"}}

- ✓ **VERITROOPER (PASS):** {"name": "exclusion_applies", "parameters": {"equipment": "articulating/knuckle-boom truck crane", "material": "building supply plywood sheets", "transfer_method": "fork/cradle", "overload_prevention_device": "false"}}

211. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_01`

- **Question:** What is the publication date of ISO 10156-2:2005 (E) regarding oxidizing ability of toxic and corrosive gases?
- **Ground truth:** Aug. 1, 2005
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ISO 10156-2:2005 (E)", "parameters": {"publication_date": "Aug. 1, 2005"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_publication_date", "parameters": {"standard_number": "ISO 10156-2:2005 (E)", "topic": "oxidizing ability of toxic and corrosive gases"}}

212. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_02`

- **Question:** In the worked example, what is the difference at 4000 Hz attributed to aging between age 27 and age 32?
- **Ground truth:** 3 dB
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_difference", "parameters": {"x": "3", "y": "11"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_difference", "parameters": {"age1": "27", "age2": "32", "frequency": "4000"}}

213. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_04`

HALLUCINATION

- **Question:** How frequently must intermediate bulkheads with locks be re-inspected after installation?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "inspect_intermediate_bulkheads", "parameters": {"frequency": "at least once a week"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

214. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0742\_01`

MODEL ERROR

- **Question:** Combining the requirements for who may make repairs and how repairs must be conducted, what two standards govern a general repair to a respirator?
- **Ground truth:** Per subpart (i), repairs may be made only by persons appropriately trained to perform such operations using only the manufacturer's NIOSH-approved parts; and per subpart (ii), repairs shall be made according to the manufacturer's recommendations and specifications for the type and extent of repairs to be performed.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "combines_requirements_for_repairs", "parameters": {"employee": "authorized personnel", "location": "designated for such repairs"}}
- ✗ **VERITROOPER (FAIL):** {"name": "perform_repairs", "parameters": {"standard_1": "Repairs or adjustments to respirators are to be made only by persons appropriately trained to perform such operations and shall use only the respirator manufacturer", "standard_2": "Repairs shall be made according to the manufacturer"}}

215. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3162\_03`

WRONG CALC/QUOTE

- **Question:** If the maximum allowable working pressure of a receiver is 100 psi, what is the highest pressure the safety valves may allow before they must relieve?
- **Ground truth:** 110 psi, since pressure may not exceed the maximum allowable working pressure by more than 10 percent.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_highest_pressure", "parameters": {"x": "100"}}
- ✔ **VERITROOPER (PASS):** {"name": "calculate_highest_pressure", "parameters": {"maximum_allowable_working_pressure": "100", "relieving_capacity": "10"}}

216. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_01`

HALLUCINATION

- **Question:** How frequently must the inspection or clean-out doors themselves be inspected for proper functioning?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "inspect_inspection_or_clean_out_doors", "parameters": {"frequency": "every 30 days"}}
- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

217. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2016\_02`

- **Question:** Combining the note in Appendix A regarding expiration dates with the typical authorization period it implies, what is the default duration of the sample authorization letter if no expiration date is specified?
- **Ground truth:** One year. The note in Appendix A allows the worker to specify a particular expiration date 'if less than one year,' which implies the default authorization period is one year.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_default_duration", "parameters": {"authorization_period": "1 year"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_default_duration", "parameters": {"typical_authorization_period": "1 year"}}

218. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1249\_02`

- **Question:** At least how many exits, remote from each other, shall be provided for all rooms in which chlorine is stored?
- **Ground truth:** At least two exits
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "provide_exits_for_chlorine_storage", "parameters": {"number_of_exits": "2"}}
- ✔ **VERITROOPER (PASS):** {"name": "provide_exits", "parameters": {"number_of_exits": "2"}}

219. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4446\_01`

MODEL ERROR

- **Question:** If a timely petition for review is filed and the case was heard before an ALJ, who transmits the record to the appropriate court?
- **Ground truth:** The ARB or the ALJ, as the case may be, transmits the record (including the record of proceedings before the ALJ).
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "transmit_record", "parameters": {"person": "ALJ"}}
- ✖ **VERITROOPER (FAIL):** {"name": "transmit_record", "parameters": {"entity": "ALJ"}}

220. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_04`

BAD QUESTION

- **Question:** What is the maximum number of shifts' worth of flammable liquid supply permitted to be kept near spraying operations before bulk storage rules apply?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "bulk_storage", "parameters": {"quantity": "1"}}
- ✗ **VERITROOPER (FAIL):** {"name": "None", "parameters": {}}

221. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3420\_00`

- **Question:** Under § 1926.702, concrete mixers with loading skips of what minimum capacity must be equipped with a mechanical device to clear the skip and guardrails on each side?
- **Ground truth:** one cubic yard (.8 m3) or larger
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1926.702", "parameters": {"(b)": "Concrete mixers with one cubic yard (.8 m 3) or larger loading skips shall be equipped with the following: (1) A mechanical device to clear the skip of materials; and (2) Guardrails installed on each side of the skip."}}
- ✓ **VERITROOPER (PASS):** {"name": "get_equipment_requirements", "parameters": {"equipment_type": "concrete mixers", "skip_capacity": "1 cubic yard (.8 m^3)"}}

222. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_00`

MODEL ERROR

- **Question:** If equipment operations would breach the Table A minimum approach distance, but the employer determines it IS feasible to do the work without breaching that distance, is the operation permitted under the exception in this section?
- **Ground truth:** No; the exception requires the employer demonstrate it is infeasible to do the work without breaching the minimum approach distance (paragraph (a)), so if it is feasible, the operation closer than Table A is prohibited.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "is_operation_permitted", "parameters": {"determines_infeasible": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_operation_permitted", "parameters": {"feasible_to_do_work_without_breaching_distance": "True"}}

223. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4294\_00`

- **Question:** Under paragraph (d)(3), agency inspections for imminent danger conditions must be conducted within 24 hours; if instead the Secretary determines an inspection is warranted for such a report under paragraph (e)(3), what response time applies?
- **Ground truth:** The same response time of within 24 hours, because under (e)(3) the Secretary shall observe the same response times as required of the agencies under the Executive Order and § 1960.28(d)(3), which sets 24 hours for imminent danger conditions.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "response_time_for_secretary_inspection", "parameters": {"response_time": "24 hours"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_response_time", "parameters": {"paragraph": "e", "subparagraph": "3", "condition": "imminent danger"}}

224. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0250\_00`

- **Question:** What part of Title 29 CFR (formerly parts 1501-1503) provides the standards adopted for shipyard employment under § 1910.15?
- **Ground truth:** part 1915
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "part", "parameters": {"x": "1915"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_part", "parameters": {"part_number": "1915"}}

225. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4630\_04`

- **Question:** What standard of proof must the complainant meet to demonstrate that protected activity was a contributing factor in the adverse action?
- **Ground truth:** Preponderance of the evidence
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "prima facie showing", "parameters": {"burden": "the complainant must allege the existence of facts and evidence to make a prima facie showing that protected activity was a contributing factor in the adverse action"}}
- ✓ **VERITROOPER (PASS):** {"name": "preponderance of the evidence", "parameters": {"standard": "preponderance of the evidence"}}

226. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_04`

HALLUCINATION

- **Question:** What is the maximum load in amperes for a receptacle on a 50-ampere branch circuit?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_maximum_load", "parameters": {"circuit_rating": "50"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_maximum_load", "parameters": {"circuit_rating": "50"}}

227. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_03`

- **Question:** On a hand bailing machine, if the pawl has been released from the takeup gear teeth and the handle slips from the operator's hand, what must the stop guard prevent?
- **Ground truth:** It must prevent the handle from traveling beyond the vertical position.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "stop guard", "parameters": {"action": "prevent the handle from traveling beyond the vertical position"}}
- ✓ **VERITROOPER (PASS):** {"name": "handle_stop_guard", "parameters": {"action": "prevent the handle from traveling beyond the vertical position"}}

228. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_04`

MODEL ERROR

- **Question:** If a scaffold must be closer to power lines than the specified minimum distance because the clearance is necessary for performance of work, when is that permitted?
- **Ground truth:** Only after the utility company or electrical system operator has been notified and has deenergized the lines, relocated them, or installed protective coverings to prevent accidental contact.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "exception_to_paragraph", "parameters": {"f": "6", "clearance": "necessary for performance of work"}}
- ✗ **VERITROOPER (FAIL):** {"name": "check_clearance_exception", "parameters": {"distance": "necessary for performance of work"}}

229. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1593\_00`

- **Question:** For doors and covers of enclosures used solely as pull boxes, splice boxes, or junction boxes, what is the exception to the requirement that they be locked, bolted, or screwed on?
- **Ground truth:** Underground box covers that weigh over 45.4 kg (100 lb) are excepted from the requirement to be locked, bolted, or screwed on.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_locking_requirement", "parameters": {"doors_and_covers": "underground box covers that weigh over 45.4 kg (100 lb)"}}
- ✗ **VERITROOPER (FAIL):** {"name": "exception_to_locking_requirement", "parameters": {"weight": "100"}}

230. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_03`

- **Question:** In the worked example, what was the employee's hearing threshold at 4000 Hz in the baseline audiogram (age 27)?
- **Ground truth:** 5
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_baseline_hearing_threshold", "parameters": {"age": "27", "frequency": "4000"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_hearing_threshold", "parameters": {"age": "27", "frequency": "4000"}}

231. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_4456\_00`

MODEL ERROR

- **Question:** When a timely objection is filed, are all provisions of the preliminary order stayed?
- **Ground truth:** No. All provisions are stayed except for the portion requiring preliminary reinstatement, which is not automatically stayed and remains effective immediately upon the respondent's receipt of the findings and preliminary order.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "objections_to_findings_and_preliminary_order", "parameters": {"timely_objection": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_stayed", "parameters": {"provision": "preliminary order", "objection": "timely objection"}}

232. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0494\_01`

MODEL ERROR

- **Question:** If undercoating spray operations are conducted in a garage with adequate natural ventilation using solvents listed as having a flash point in excess of 100 °F, are they exempt from spray finishing requirements?
- **Ground truth:** Yes, they are exempt from the requirements pertaining to spray finishing operations.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "k", "parameters": {"Underwriters": "Laboratories in respect to fire hazard rating 30-40", "flash_point": "100"}}
- ✓ **VERITROOPER (PASS):** {"name": "is_exempt_from_spray_finishing_requirements", "parameters": {"has_adequate_ventilation": "true", "flash_point": "100"}}

233. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2920\_03`

MODEL ERROR

- **Question:** According to paragraphs (viii) and (ix), what must be provided to a worker who reports a small skin abrasion, and who treats it?
- **Ground truth:** Per paragraph (viii), any small skin abrasions, cuts, rash, or open sores which are found or reported shall be treated by a properly designated person so that chances of exposures to the chemicals are

removed. Per paragraph (ix), the worker must also have access to sufficient washing facilities, including soap, individual towels, and hot water, provided on the basis of at least one basin with a hot water faucet for every 10 employees.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "provide_treatment_and_first_aid", "parameters": {"treatment": "treatment by a properly designated person", "first_aid_supplies": "appropriate first-aid supplies"}}
- ✖ **VERITROOPER (FAIL):** {"name": "treatment_for_skin_abrasion", "parameters": {"treatment": "treated by a properly designated person"}}

234. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_02`

HALLUCINATION

- **Question:** What is the minimum negative pressure differential (in inches of water) that an AIIR must maintain?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "pressure_differential", "parameters": {"pressure": "0"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The evidence defines what an Airborne infection isolation room (AIIR) is, but it does not specify the minimum negative pressure differential that an AIIR must maintain.

235. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_00`

HALLUCINATION

- **Question:** What minimum air velocity (in feet per minute) must be maintained through the spray booth ventilation system?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "minimum_air_velocity", "parameters": {"velocity": "100"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

236. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3468\_01`

WRONG CALC/QUOTE

- **Question:** If an air quality test record relates to exposure to a toxic substance, how long must it be retained compared to a record that does not involve a toxic substance?
- **Ground truth:** Toxic substance exposure records are retained in accordance with § 1910.33, while all other air quality test records are retained only until completion of the project.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "recordkeeping", "parameters": {"toxic_substance_record": "30 years", "non_toxic_substance_record": "completion of the project"}}
- ✓ **VERITROOPER (PASS):** {"name": "compare_retention_periods", "parameters": {"toxic_substance_record": "in accordance with § 1910.33 of this chapter", "non_toxic_substance_record": "until completion of the project"}}

237. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_03`

HALLUCINATION

- **Question:** How many days does the employer have to forward the information to the licensed physician before the examination?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "information_to_PLHCP", "parameters": {"days": "15"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

238. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_08`

HALLUCINATION

- **Question:** How many square feet must an ambulatory surgical center provide per patient to qualify as an ambulatory care setting?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "ambulatory care setting space requirements", "parameters": {"square feet per patient": "100"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The evidence defines ambulatory care and lists examples of settings where it is provided, including ambulatory surgical centers, but it does not specify the required square footage per patient.

239. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2784\_01`

WRONG CALC/QUOTE

- **Question:** If an OSHA regulated substance has no action level, what triggers application of paragraphs (d) and (g)(1)(ii)?
- **Ground truth:** Routinely exceeding the permissible exposure limit (in the absence of an action level), for a substance with exposure monitoring and medical surveillance requirements.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraphs (d) and (g)(1)(ii) trigger", "parameters": {"action_level": "null", "PEL": "permissible exposure limit"}}
- **✗ VERITROOPER (FAIL):** {"name": "apply_paragraphs", "parameters": {"action_level": "None", "exposure_limit": "permissible exposure limit"}}

240. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_1317\_01`

MODEL ERROR

- **Question:** How does the definition of 'running line' distinguish it from other ropes?
- **Ground truth:** A running line means any moving rope as distinguished from a stationary rope such as a guyline.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "definition_of_running_line", "parameters": {"rope_type": "running line"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_definition", "parameters": {"term": "running line"}}

241. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_07`

- **Question:** If a scaffold is covered with snow or ice, are employees ever permitted to work on it?
- **Ground truth:** Only as necessary for removal of such materials.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "employees_prohibited_from_working_on_scaffolds_covered_with_snow_or_ice", "parameters": {"exception": "removal of such materials"}}
- **✓ VERITROOPER (PASS):** {"name": "work_on_scaffold", "parameters": {"snow_or_ice": "true", "removal_necessary": "false"}}

242. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4010\_00`

MODEL ERROR

- **Question:** If an auditor determines that the written and practical tests are valid and reliable but the auditor is an employee of the employer, can the auditor's approval satisfy the test development requirements?
- **Ground truth:** No, because paragraph (e)(1)(ii)(B) requires that the auditor is not an employee of the

employer.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "auditor_approval", "parameters": {"valid": "True", "reliable": "True", "auditor_employer": "False"}}
- ✖ **VERITROOPER (FAIL):** {"name": "auditor_approval", "parameters": {"valid_and_reliable": "True", "auditor_employed_by_employer": "True"}}

243. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2569\_00`

- **Question:** If an employer chooses to rely on the classification performed by the chemical manufacturer for a chemical, are they required to classify that chemical themselves?
- **Ground truth:** No, employers are not required to classify chemicals unless they choose not to rely on the classification performed by the chemical manufacturer or importer.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_employer_required_to_classify", "parameters": {"choose_to_rely_on_manufacturer": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If an employer chooses to rely on the classification performed by the chemical manufacturer for a chemical, are they required to classify that chemical themselves?", "answer": "No"}}

244. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0644\_03`

- **Question:** Using paragraph (p)(1), which paragraph governs emergency planning and response, and which governs the development of the process hazard analysis?
- **Ground truth:** Emergency planning and response is governed by paragraph (n), and the development of the process hazard analysis is governed by paragraph (e).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (p)(1)", "parameters": {"p": "n", "q": "e"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_paragraphs", "parameters": {"paragraph_number": "p", "subparagraph_number": "1"}}

245. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_07`

HALLUCINATION

- **Question:** What form must be used to document a close contact exposure under this section?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA 301", "parameters": {"form type": "Injury and Illness Incident Report"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

246. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2476\_02`

- **Question:** Using the definitions of 'action level' and 'methylene chloride (MC),' what is the concentration constituting the action level and the chemical formula of the substance to which it applies?
- **Ground truth:** The action level is a concentration of airborne MC of 12.5 parts per million (ppm) calculated as an eight (8)-hour time-weighted average (TWA), and the substance MC has the chemical formula CH₂Cl₂.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_action_level_and_mc_definition", "parameters": {"action_level": "12.5 ppm", "mc_definition": "CH₂Cl₂"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_action_level_and_mc", "parameters": {"concentration": "12", "chemical_formula": "CH₂Cl₂"}}

247. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3801\_02`

- **Question:** Per paragraph (h)(2)(i), when must employees remove all required PPE, and per paragraph (h)(3)(i), what must the employer do with reusable PPE so removed?
- **Ground truth:** Employees must remove all required PPE at the end of the work shift or at the completion of all tasks involving beryllium, whichever comes first (h)(2)(i), and the employer must ensure all reusable PPE is cleaned, laundered, repaired, and replaced as needed to maintain its effectiveness (h)(3)(i).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "remove_and_handle_PPE", "parameters": {"paragraph_h_2_i": "at the end of the activity, process, or work shift and prior to eating, drinking, smoking, chewing tobacco or gum, applying cosmetics, or using the toilet", "paragraph_h_3_i": "ensure that all reusable personal protective clothing and equipment required by this standard is cleaned, laundered, repaired, and replaced as needed to maintain its effectiveness"}}
- ✓ **VERITROOPER (PASS):** {"name": "remove_PPE", "parameters": {"time": "at the end of the work shift or at the completion of all tasks involving beryllium, whichever comes first"}}

248. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0644\_00`

- **Question:** Under paragraph (p)(1), trade secret information must be made available to those involved in incident investigations and those involved in compliance audits. Which lettered paragraphs of the section establish the requirements for these two activities respectively?
- **Ground truth:** Incident investigations are required by paragraph (m), and compliance audits are required by paragraph (o).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph_requirements", "parameters": {"incident_investigations": "m", "compliance_audits": "o"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_lettered_paragraphs", "parameters": {"activity1": "incident investigations", "activity2": "compliance audits"}}

249. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_08`

- **Question:** For Class 4 equipment, what is the DC average retest voltage?
- **Ground truth:** 70,000
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "4", "voltage_type": "DC avg"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "4", "voltage_type": "DC"}}

250. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_10`

HALLUCINATION

- **Question:** Which OSHA form must the employer use to record a worker as 'designated'?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA 301", "parameters": {"form": "Injury and Illness Incident Report"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about which OSHA form must be used to record a worker as 'designated'.

251. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_09`

HALLUCINATION

- **Question:** Which OSHA office is designated to handle complaints from seamen who reside outside the United States?

- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "OSHA office responsible for enforcement activities", "parameters": {"geographical_area": "outside the United States"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer this question.

252. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_03`

- **Question:** Which edition is ISO 10156:1996 (E)?
- **Ground truth:** Second Edition
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_edition", "parameters": {"std_id": "ISO 10156:1996 (E)"}}
- **✓ VERITROOPER (PASS):** {"name": "get_iso_edition", "parameters": {"iso_number": "ISO 10156:1996 (E)", "edition": "Second"}}

253. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_03`

HALLUCINATION

- **Question:** What is the minimum exposure concentration that triggers respiratory sensitizer classification under these criteria?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "classification_criteria_for_mixtures", "parameters": {"concentration_limit": "0"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

254. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_07`

HALLUCINATION

- **Question:** What fire-resistance rating must the shields used near combustible walls meet?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_fire_resistance_rating", "parameters": {"material": "shields used near combustible walls"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

255. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_03`

HALLUCINATION

- **Question:** For a 60-ampere branch circuit supplying two or more receptacles, what receptacle ratings are listed in Table S-5?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Table S-5", "parameters": {"circuit_rating_amperes": "60"}}
- **✓ VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question. The prompt asks for receptacle ratings for a 60-ampere branch circuit, but the evidence only lists ratings up to 50 amperes.

256. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3110\_03`

- **Question:** Combining the definition of a marine service station with the dispensing nozzle requirement, into what must flammable liquids be dispensed and what type of nozzle must be used?
- **Ground truth:** Per paragraph (k)(4), flammable liquids are dispensed into the fuel tanks of self-propelled craft. Per paragraph (k)(1)(iii), the dispensing nozzles must be automatic-closing without a hold-open latch.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "dispensing_nozzle_requirement", "parameters": {"liquid_type": "flammable liquids used as fuels", "nozzle_type": "automatic-closing without a hold-open latch"}}
- ✓ **VERITROOPER (PASS):** {"name": "dispensing_nozzle_requirement", "parameters": {"flammable_liquids": "fuel tanks of self-propelled craft", "nozzle_type": "automatic-closing without a hold-open latch"}}

257. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_05`

- **Question:** Sections 1926.800 and 1926.803 are listed separately. Do they have the same OMB control number?
- **Ground truth:** Yes, both 1926.800 and 1926.803 are assigned OMB control No. 1218-0067.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "check_omb_control_number", "parameters": {"a": "1218-0067", "b": "1218-0067"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_omb_control_numbers", "parameters": {"section1": "1926.800", "section2": "1926.803"}}

258. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1138\_04`

- **Question:** If modifications are made that may affect the safety of the system, when must the employer notify the third-party validation organization?
- **Ground truth:** Within five days.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "notify_third_party_validation_organization", "parameters": {"days": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "notify_third_party_validation_organization", "parameters": {"days": "5"}}

259. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_04`

HALLUCINATION

- **Question:** What penalty is imposed on an operator who fails to wear the required safety shoes?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "impose_penalty", "parameters": {"amount": "1000", "imprisonment": "6 months"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

260. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_01`

- **Question:** Under the Walsh-Healey Public Contracts Act, what contract amount must be exceeded for the safety and health provisions to apply to manufacturing or furnishing of materials?
- **Ground truth:** \$10,000
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Walsh-Healey Public Contracts Act", "parameters": {"amount": "10000"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_contract_amount", "parameters": {"contract_type": "Walsh-Healey Public Contracts Act", "provision_type": "safety and health"}}

261. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1441\_02`

- **Question:** When removing a ground from a line or equipment operating at 600 volts or less, what insulating equipment may be used other than a live-line tool, and under what conditions?
- **Ground truth:** The employer may permit the employee to use insulating equipment other than a live-line tool, if the employer ensures the line or equipment is not energized at the time the ground is disconnected, or if the employer can demonstrate that each employee is protected from hazards that may develop if the line or

equipment is energized.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "removing_ground", "parameters": {"voltage": "600", "equipment": "insulating equipment other than a live-line tool", "conditions": "the line or equipment is not energized at the time the ground is disconnected or the employer can demonstrate that each employee is protected from hazards that may develop if the line or equipment is energized"}}
- ✓ **VERITROOPER (PASS):** {"name": "remove_ground", "parameters": {"voltage": "600", "equipment": "insulating equipment other than a live-line tool", "conditions": "if the employer ensures that the line or equipment is not energized at the time the ground is disconnected or if the employer can demonstrate that each employee is protected from hazards that may develop if the line or equipment is energized"}}

262. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_00`

HALLUCINATION

- **Question:** What is the maximum capacity allowed for containers supplying spray nozzles by means other than gravity flow or air pressure?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_max_capacity", "parameters": {"container_type": "closed type or provided with metal covers kept closed"}}
- ✓ **VERITROOPER (PASS):** This information is not contained in the provided evidence.

263. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_10`

- **Question:** In the Table F-1 (male) example correction row, what is the difference value at 3000 Hz between age 32 and age 27?
- **Ground truth:** 3
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_difference", "parameters": {"x": "10", "y": "7"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_difference", "parameters": {"age1": "32", "age2": "27", "frequency": "3000"}}

264. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_05`

- **Question:** Wherever acids or caustics are used, what provision must be made?
- **Ground truth:** Provision must be made for a copious and flowing supply of fresh, clean water.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Provision for a copious and flowing supply of fresh, clean water", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "first_aid_provision", "parameters": {"provision": "a copious and flowing supply of fresh, clean water"}}

265. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_01`

HALLUCINATION

- **Question:** Which OSHA office must employers submit their COVID-19 log to for review?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Forms 300, 300A, and 301", "parameters": {"employer": "Assistant Secretary"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

266. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4456\_02`

- **Question:** Besides any party seeking review, who else may file objections and a request for a hearing under § 1984.106(a)?
- **Ground truth:** A respondent alleging that the complaint was frivolous or brought in bad faith who seeks an

award of attorney fees under section 18C of the FLSA.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "file objections and request for a hearing", "parameters": {"party": "respondent alleging that the complaint was frivolous or brought in bad faith who seeks an award of attorney fees under section 18C of the FLSA"}}
- ✓ **VERITROOPER (PASS):** {"name": "respond", "parameters": {"answer": "A respondent alleging that the complaint was frivolous or brought in bad faith who seeks an award of attorney fees under section 18C of the FLSA."}}

267. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0751\_02`

- **Question:** If neither the workplace nor the respirator type has changed and the employee's knowledge appears adequate, can retraining still be required?
- **Ground truth:** Yes, if any other situation arises in which retraining appears necessary to ensure safe respirator use.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "retraining", "parameters": {"situation": "other situation arises in which retraining appears necessary to ensure safe respirator use"}}
- ✓ **VERITROOPER (PASS):** {"name": "retraining_appears_necessary", "parameters": {"situation": "other situation"}}

268. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_05`

HALLUCINATION

- **Question:** What load multiple is applied as the proof load relative to the rated capacity?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "proof_test", "parameters": {"factor": "1"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

269. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4344\_00`

- **Question:** In which year were U.S. v. Silk, 331 U.S. 704, and Rutherford Food Corporation v. McComb, 331 U.S. 722, decided?
- **Ground truth:** 1947
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_year", "parameters": {"x": "U.S. v. Silk, 331 U.S. 704"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_year", "parameters": {"case_name": "U.S. v. Silk, 331 U.S. 704"}}

270. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1408\_00`

- **Question:** Under § 1910.269(a)(2)(iv), the employer must determine through regular supervision and inspections conducted on at least what frequency that each employee is complying with the safety-related work practices?
- **Ground truth:** an annual basis
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "employer_determination", "parameters": {"frequency": "annual"}}
- ✓ **VERITROOPER (PASS):** {"name": "frequency_of_inspection", "parameters": {"frequency": "annual"}}

271. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0022\_00`

MODEL ERROR

- **Question:** If the Assistant Secretary publishes a notice in the Federal Register on a given date, by what deadline must the applying State agency publish reasonable notice containing the same information within the State?
- **Ground truth:** No later than 5 days following the publication of the notice in the Federal Register.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "publish_notice", "parameters": {"days": "10"}}
- ✔ **VERITROOPER (PASS):** {"name": "calculate_deadline", "parameters": {"days": "5"}}

272. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_05`

- **Question:** What is the required accuracy to a 95 percent confidence level for the MDA monitoring method at the select PEL?
- **Ground truth:** plus or minus 25 percent
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "required_accuracy", "parameters": {"confidence_level": "95", "percent": "25"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_required_accuracy", "parameters": {"confidence_level": "95", "pel": "select PEL"}}

273. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_05`

- **Question:** On wheel-mounted equipment, lifting loads over the front area is prohibited except in what case?
- **Ground truth:** Except as permitted by the manufacturer.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "check_load_over_front_area", "parameters": {"allowed": "as permitted by the manufacturer"}}
- ✔ **VERITROOPER (PASS):** {"name": "check_load_lift", "parameters": {"case": "as permitted by the manufacturer"}}

274. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3938\_01`

- **Question:** Per § 1926.1204(e)(1), testing must occur before changes to natural ventilation and before entry is authorized. Under what exception does this not apply, and what must the employer do instead?
- **Ground truth:** The exception applies if the employer demonstrates that isolation of the space is infeasible because the space is large or is part of a continuous system (such as a sewer). In that case the employer must perform pre-entry testing to the extent feasible before entry is authorized, and if entry is authorized, continuously monitor entry conditions where entrants are working and provide an early-warning system for non-isolated engulfment hazards.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_testing_before_entry", "parameters": {"paragraph": "(e)(1)(i)", "condition": "isolation of the space is infeasible because the space is large or is part of a continuous system (such as a sewer)", "alternative": "perform pre-entry testing to the extent feasible before entry is authorized"}}
- ✔ **VERITROOPER (PASS):** {"name": "exception_to_testing_requirement", "parameters": {"exception": "isolation of the space is infeasible because the space is large or is part of a continuous system", "alternative_action": "perform pre-entry testing to the extent feasible before entry is authorized, and continuously monitor entry conditions in the areas where authorized entrants are working"}}

275. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1597\_04`

- **Question:** When a nongrounding-type receptacle is replaced with a grounding-type receptacle supplied through a GFCI (marked 'GFCI Protected' and 'No Equipment Ground'), what connection is prohibited?
- **Ground truth:** An equipment grounding conductor may not be connected to such grounding-type receptacles.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "connection_prohibited", "parameters": {"connection": "equipment grounding conductor"}}
- ✔ **VERITROOPER (PASS):** {"name": "connection_prohibited", "parameters": {"receptacle_type": "grounding-type", "GFCI": "True", "equipment_grounding_conductor": "False"}}

276. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_03`

- **Question:** Under § 1902.39(c), no later than how many days following publication of the notice in the Federal Register must the affected State agency publish notice within the State?
- **Ground truth:** 10 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "days_after_publication", "parameters": {"days": "10"}}
- ✓ **VERITROOPER (PASS):** {"name": "publish_notice", "parameters": {"days": "10"}}

277. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3038\_01`

- **Question:** If an employer's permit-required contingency plan fully addresses certain subjects, must the emergency response plan duplicate those subjects?
- **Ground truth:** No, the plan need not duplicate them, provided the contingency plan is made part of the emergency response plan.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "emergency_response_plan", "parameters": {"duplicate": "false"}}
- ✗ **VERITROOPER (FAIL):** {"name": "answer", "parameters": {"question": "If an employer"}}

278. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3428\_01`

- **Question:** The headache ball, hook, or load shall not be used to transport personnel - what is the stated exception?
- **Ground truth:** Except as provided in paragraph (c)(4) of this section (using cranes or derricks to hoist employees on a personnel platform).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_transporting_personnel", "parameters": {"exception": "as provided in paragraph (c)(4) of this section"}}
- ✗ **VERITROOPER (FAIL):** {"name": "exception_to_headache_ball_hook_or_load_usage", "parameters": {"exception": "when work under this subpart is being conducted, provided that all provisions of \u00a71926.1431 (except for \u00a71926.1431(a)) are met"}}

279. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_07`

- **Question:** In Table F-1 (males), what is the age correction value at 4000 Hz for age 27?
- **Ground truth:** 11
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1 (males)", "frequency": "4000 Hz", "age": "27"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "F-1", "gender": "male", "age": "27", "frequency": "4000"}}

280. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0035\_01`

- **Question:** At what Federal Register citation is Secretary's Order 12-71 published according to § 1902.30?
- **Ground truth:** 36 FR 8754
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_federal_register_citation", "parameters": {"section": "1902.30", "order": "12-71"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_published_federal_register_citation", "parameters": {"order_number": "12-71", "section_number": "1902.30"}}

281. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4448\_02`

- **Question:** Per paragraph (a)(3) and paragraph (b), who may request a jury trial, and on whom must the complaint be served after filing in federal court?
- **Ground truth:** At the request of either party, the action shall be tried by the court with a jury (a)(3); and a

copy of the complaint must be served on the OSHA official who issued the findings and/or preliminary order, the Assistant Secretary, and the Associate Solicitor, Division of Fair Labor Standards, U.S. Department of Labor (b).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "request_jury_trial_and_serve_complaint", "parameters": {"requester": "either party", "server": "OSHA official who issued the findings and/or preliminary order, the Assistant Secretary, and the Associate Solicitor, Division of Occupational Safety and Health, U.S. Department of Labor"}}
- ✓ **VERITROOPER (PASS):** {"name": "request_jury_trial_and_serve_complaint", "parameters": {"requester": "either party", "complaint_recipient": "OSHA official who issued the findings and/or preliminary order, the Assistant Secretary, and the Associate Solicitor, Division of Fair Labor Standards, U.S. Department of Labor"}}

282. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4330\_00`

- **Question:** According to the basic objective stated in the introductory text and function (a), what is the relationship between the field councils' overall purpose and their role as a clearinghouse?
- **Ground truth:** The basic objective is to facilitate the exchange of ideas and information to assist agencies to reduce the incidence, severity and cost of occupational accidents, injuries, and illnesses. Function (a) supports this by having councils act as a clearinghouse on information and data on occupational accidents, injuries, and illnesses and their prevention.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "field_councils_basic_objective", "parameters": {"field_councils_objective": "facilitate the exchange of ideas and information to assist agencies to reduce the incidence, severity and cost of occupational accidents, injuries, and illnesses"}}
- ✓ **VERITROOPER (PASS):** {"name": "facilitate_exchange_of_ideas", "parameters": {"field_councils": "clearinghouse", "ideas_and_information": "occupational accidents, injuries, and illnesses and their prevention"}}

283. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_10`

HALLUCINATION

- **Question:** What is the minimum ventilation airflow rate required in the mixing room when transferring these liquids?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_ventilation_airflow_rate", "parameters": {"room": "mixing room", "liquids": "flammable liquids or liquids with a flashpoint greater than 199.4 \u00b0F (93 \u00b0C)"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the minimum ventilation airflow rate required in the mixing room when transferring these liquids.

284. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0014\_00`

- **Question:** Which section of the Act requires State public employee programs to be as effective as standards contained in the State plan?
- **Ground truth:** section 18(c)(6)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 18", "parameters": {"c": "6"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"section_number": "18", "subsection": "c", "paragraph": "6"}}

285. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3468\_00`

MODEL ERROR

- **Question:** In an underground drill-and-blast operation, if employees are about to re-enter the area after blasting, must the air be tested for flammable gas, and is any further testing required while they work?
- **Ground truth:** Yes, the air must be tested for flammable gas prior to re-entry after blasting, and tested continuously while employees are working underground.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "test_for_flammable_gas", "parameters": {"re_entry": "True"}}
- **✓ VERITROOPER (PASS):** {"name": "air_quality_testing", "parameters": {"test_before_reentry": "True", "continuous_testing_while_working": "True"}}

286. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4084\_02`

- **Question:** When is the upper-most tie-in, braces, floor supports, and floor wedges required to be inspected for loose or dislodged components?
- **Ground truth:** Where the tower crane is supported by the structure.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "inspect_tower_crane", "parameters": {"component": "upper-most tie-in, braces, floor supports, and floor wedges", "frequency": "monthly"}}
- **✓ VERITROOPER (PASS):** {"name": "inspect_tie_in_braces", "parameters": {"frequency": "During the tower crane erection"}}

287. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4032\_01`

- **Question:** For controlled load lowering, what system other than the load line hoist brake must regulate the lowering rate of speed?
- **Ground truth:** The load line hoist drum must have a system, other than the load line hoist brake, which regulates the lowering rate of speed of the hoist mechanism.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "controlled_load_lowering_system", "parameters": {"system": "a system, other than the load line hoist brake, which regulates the lowering rate of speed of the hoist mechanism"}}
- **✓ VERITROOPER (PASS):** {"name": "controlled_load_lowering", "parameters": {"system": "a system, other than the load line hoist brake"}}

288. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_02`

HALLUCINATION

- **Question:** How many days does the employer have to provide a medical examination after an employee reports breathing difficulty during respirator use?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraph (l)(6)(iii)", "parameters": {"days": "as soon as possible"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

289. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2250\_02`

- **Question:** If outside contamination of the primary specimen container occurs AND the specimen could also puncture that primary container, what must be done?
- **Ground truth:** The primary container must be placed within a secondary container that prevents leakage, is labeled/color-coded, and is also puncture-resistant.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "container_placement", "parameters": {"primary_container": "placed within a second container", "secondary_container": "puncture-resistant and prevents leakage during"}}

handling, processing, storage, transport, or shipping", "labeling": "labeled or color-coded according to the requirements of this standard"]}

- ✓ **VERITROOPER (PASS):** {"name": "containerize_specimen", "parameters": {"primary_container": "True", "secondary_container": "True", "puncture_resistant": "True"}}

290. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2476\_01`

- **Question:** Combining the definitions of 'emergency' and 'employee exposure,' if maintenance personnel control an incidental MC release per paragraph (f) while not wearing respiratory protection, is this an emergency, and would the airborne MC they would be exposed to count as employee exposure?

- **Ground truth:** It is not considered an emergency, because an incidental release controlled by employees such as maintenance personnel in accordance with the leak/spill provisions of paragraph (f) is excluded from the emergency definition. However, the airborne MC they would be exposed to (or would be exposed to absent respiratory protection) still constitutes employee exposure.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "emergency_and_employee_exposure", "parameters": {"emergency": "false", "employee_exposure": "true"}}

- ✗ **VERITROOPER (FAIL):** {"name": "is_emergency", "parameters": {"release_controlled": "True", "respiratory_protection_used": "False"}}

291. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4520\_01`

- **Question:** Under what circumstance will the Office of Administrative Law Judges grant a respondent's motion for a stay of the Assistant Secretary's preliminary order of reinstatement?

- **Ground truth:** Only based on exceptional circumstances.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "grant_motion_for_stay", "parameters": {"circumstance": "exceptional circumstances"}}

- ✓ **VERITROOPER (PASS):** {"name": "grant_motion", "parameters": {"circumstance": "exceptional circumstances"}}

292. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_05`

- **Question:** Where petitions to participate as parties are made by groups with common interests, what may the administrative law judge require?

- **Ground truth:** He may require all such petitioners to designate a single representative, or recognize one or more of them to represent all.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "require_designated_representative", "parameters": {"petitioners": "groups with common interests"}}

- ✓ **VERITROOPER (PASS):** {"name": "require_designation_of_representative", "parameters": {"petitioners": "groups with common interests"}}

293. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_06`

HALLUCINATION

- **Question:** By what date must existing receptacle installations be brought into compliance with these rating provisions?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_installation_compliance_date", "parameters": {"year": "1990", "month": "7", "day": "23"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain information about the date by which existing receptacle installations must be brought into compliance with the rating provisions.

294. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3172\_02`

MODEL ERROR

- **Question:** For a horizontal belt where both runs are 36 inches from the floor, what guarding is required, and how does this differ from the general rule for runs that are 7 feet or less from the floor?
- **Ground truth:** Under paragraph (e)(1)(i), where both runs are 42 inches or less from the floor, the belt shall be fully enclosed. Since 36 inches is 42 inches or less, the belt must be fully enclosed. This differs from the general rule for runs 7 feet or less from the floor, where the guard need only extend to at least 15 inches above the belt or to a standard height.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "belt_guarding", "parameters": {"distance_from_floor": "36", "belt_width": "8"}}
- **✗ VERITROOPER (FAIL):** {"name": "guarding_required", "parameters": {"height": "36", "general_rule_height": "7"}}

295. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1695\_04`

- **Question:** Combining paragraph (a)(5)(i) and (a)(5)(ii), what handling restrictions apply when plugging in an energized cord connector that is wet from being immersed in water?
- **Ground truth:** Employees' hands may not be wet when plugging and unplugging energized equipment (per (a)(5)(i)), and the energized plug and receptacle connection may be handled only with insulating protective equipment because a wet connection could provide a conducting path to the employee's hand (per (a)(5)(ii)).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Connecting attachment plugs", "parameters": {"Employees": "hands may not be wet when plugging and unplugging flexible cords and cord- and plug-connected equipment, if energized equipment is involved", "Energized plug and receptacle connections": "may be handled only with insulating protective equipment if the condition of the connection could provide a conducting path to the employee"}}
- **✗ VERITROOPER (FAIL):** {"name": "restriction", "parameters": {"condition": "wet from being immersed in water", "handling": "only with insulating protective equipment"}}

296. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_00`

- **Question:** In the worked example, what is the threshold shift at 4000 Hz between the audiograms taken at ages 27 and 32 BEFORE any age correction is applied?
- **Ground truth:** 20 dB
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "calculate_threshold_shift", "parameters": {"freq": "4000", "age1": "27", "age2": "32", "threshold1": "5", "threshold2": "25"}}
- **✓ VERITROOPER (PASS):** {"name": "calculate_threshold_shift", "parameters": {"frequency": "4000", "age1": "27", "age2": "32"}}

297. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_05`

HALLUCINATION

- **Question:** What is the minimum steel pipe wall thickness required for piping systems conveying these flammable liquids?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_wall_thickness", "parameters": {"material": "steel"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the minimum steel pipe wall thickness required for piping systems conveying flammable liquids.

298. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4618\_00`

- **Question:** If an observer wishes to watch monitoring that requires entry into an area where protective clothing is required, what three obligations fall on the employer?
- **Ground truth:** The employer must provide the observer with the required personal protective clothing or equipment, assure its use, and require the observer to comply with all other applicable safety and health procedures.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "observation_procedures", "parameters": {"(i)": "provide the observer with personal protective clothing and equipment required to be worn by employees working in the area", "(ii)": "assure the use of such clothing and equipment", "(iii)": "require the observer to comply with all other applicable safety and health procedures"}}
- ✗ **VERITROOPER (FAIL):** {"name": "observation_procedures", "parameters": {"i": "3"}}

299. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3390\_00`

- **Question:** Under paragraph (i), the general maximum distance for a control line from the leading edge is 25 feet, but paragraph (ii) carves out an exception for precast concrete members. For a precast concrete member that is 90 feet long, what is the maximum distance the control line may be erected from the leading edge?
- **Ground truth:** 45 feet. Under paragraph (ii), the control line must be no more than 60 feet or half the length of the member, whichever is less. Half of 90 feet is 45 feet, which is less than 60 feet, so the maximum is 45 feet.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_control_line_distance", "parameters": {"member_length": "90"}}
- ✓ **VERITROOPER (PASS):** {"name": "min", "parameters": {"a": "60", "b": "45"}}

300. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2569\_02`

- **Question:** When classifying physical hazards, which appendix must be consulted?
- **Ground truth:** Appendix B to § 1910.1200 must be consulted for classification of physical hazards.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_physical_hazards", "parameters": {"appendix": "Appendix B to \u00a7 1910.1200"}}
- ✓ **VERITROOPER (PASS):** {"name": "consult_appendix", "parameters": {"appendix": "B"}}

301. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4486\_00`

MODEL ERROR

- **Question:** Under paragraph (a), the respondent is notified of its rights under paragraphs (b) and (f). Combining these, what are the two separate time limits the respondent has for responding under each of those paragraphs?
- **Ground truth:** Under paragraph (b), the respondent has 20 days of receipt of the notice of the filing of the complaint to submit a written statement, affidavits, or documents and to request a meeting. Under paragraph (f), the respondent must present its evidence within 10 business days of the Assistant Secretary's notification (or as soon thereafter as agreed if the interests of justice require).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "respondent_rights", "parameters": {"b": "20 days", "f": "20 days"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_time_limits", "parameters": {"paragraph_b": "20 days", "paragraph_f": "10 business days"}}

302. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_3130\_08`

MODEL ERROR

- **Question:** What is the maximum capacity (in gallons) of a safety can?
- **Ground truth:** not more than 5 gallons
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_max_capacity", "parameters": {"container_type": "safety can"}}
- **✓ VERITROOPER (PASS):** {"name": "get_safety_can_capacity", "parameters": {"capacity": "5"}}

303. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_08`

- **Question:** In Table F-1 (males), what is the age correction value at 3000 Hz for age 27?
- **Ground truth:** 7
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1 (males)", "frequency": "3000", "age": "27"}}
- **✓ VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "F-1", "gender": "male", "age": "27", "frequency": "3000"}}

304. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3990\_00`

- **Question:** Under § 1926.1420, when must the devices used to transmit signals be tested to ensure the signal transmission is effective, clear, and reliable?
- **Ground truth:** On site before beginning operations
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "test_signal_transmission_devices", "parameters": {"site": "on site", "time": "before beginning operations"}}
- **✓ VERITROOPER (PASS):** {"name": "test_signal_transmission", "parameters": {"test_time": "before beginning operations"}}

305. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3390\_03`

MODEL ERROR

- **Question:** For a precast concrete member that is 150 feet long, paragraph (ii) governs the maximum distance. How does this maximum compare to the general maximum given in paragraph (i)?
- **Ground truth:** For a 150-foot member, half the length is 75 feet, but the cap is 60 feet (whichever is less), so the maximum is 60 feet under paragraph (ii). This is greater than the 25-foot general maximum given in paragraph (i).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "compare_maxima", "parameters": {"x": "150"}}
- **✗ VERITROOPER (FAIL):** {"name": "compare_distances", "parameters": {"length": "150"}}

306. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4138\_00`

TABLE MISREAD

- **Question:** Both Option (2) corraling and Option (4) centerline cable system aim to prevent shifting, but they impose different standards on whether any movement is tolerated. What does each require regarding movement during operation?
- **Ground truth:** Option (2) corraling requires that the corraling system does not allow the equipment to shift by any amount of shifting in any direction. Option (4) centerline cable system requires that the crane/derrick is secured from movement during operation (per subpart (E)), though its attachment method must also allow the crew to move the crane/derrick longitudinally for repositioning.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "corraling_system_requirements", "parameters": {"system_type": "Option (2)"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_movement_requirement", "parameters": {"option": "2"}}

307. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_01`

- **Question:** For a wood live-line tool's initial test under paragraph (j)(1)(ii), what voltage per meter must it withstand and for how long?
- **Ground truth:** 246,100 volts per meter (75,000 volts per foot) for 3 minutes
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "live_line_tool_test", "parameters": {"voltage_per_meter": "246100", "duration_minutes": "3"}}
- ✓ **VERITROOPER (PASS):** {"name": "live_line_tool_test", "parameters": {"material": "wood", "voltage_per_meter": "246100", "duration": "3"}}

308. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_06`

HALLUCINATION

- **Question:** Within how many days must a sling manufacturer document a completed proof test?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "document_proof_test", "parameters": {"days": "7"}}
- ✓ **VERITROOPER (PASS):** No information is provided in the text regarding the time frame within which a sling manufacturer must document a completed proof test.

309. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3948\_03`

MODEL ERROR

- **Question:** A powered industrial truck (forklift) is configured to hoist and lower by means of a winch and horizontally move a suspended load. Is it excluded under (c)(8)?
- **Ground truth:** No; powered industrial trucks are excluded except when configured to hoist and lower (by winch or hook) and horizontally move a suspended load.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "check_exclusion", "parameters": {"i": "8", "condition": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "is_excluded", "parameters": {"equipment": "powered industrial truck (forklift)", "configuration": "configured to hoist and lower by means of a winch and horizontally move a suspended load", "exclusion": "8"}}

310. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1593\_03`

- **Question:** What is the exception that allows equipment operating at 600 volts or less to be installed within a high-voltage enclosure, room, or vault without separation?
- **Ground truth:** Switches or other equipment operating at 600 volts or less, and serving only equipment within the high-voltage vault, room, or enclosure, may be installed in the high-voltage enclosure if accessible to qualified persons only.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "install_equipment_in_high_voltage_enclosure", "parameters": {"voltage": "600", "equipment_type": "switches or other equipment", "condition": "serving only equipment within the high-voltage vault, room, or enclosure"}}
- ✖ **VERITROOPER (FAIL):** {"name": "allow_installation", "parameters": {"voltage": "600", "location": "high-voltage vault", "access": "qualified persons only"}}

311. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0478\_00`

- **Question:** For spraying operations other than electrostatic spraying, what is the minimum average air velocity required over the open face of the booth?
- **Ground truth:** 100 linear feet per minute
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "minimum_average_air_velocity", "parameters": {"velocity": "100 linear feet per minute"}}

"100"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_minimum_average_air_velocity", "parameters": {"operation_type": "non-electrostatic spraying"}}

312. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1114\_00`

- **Question:** Per the standard, the air-clutch machine control requirement applies only to clutch/brake air-valve controls manufactured and installed on or after what date?

- **Ground truth:** August 31, 1971

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "date_requirement", "parameters": {"date": "August 31, 1971"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_date", "parameters": {"year": "1971", "month": "8", "day": "31"}}

313. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_03`

HALLUCINATION

- **Question:** What monetary penalty may be imposed on a pipeline operator found to have discriminated against an employee?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "adjusted civil monetary penalties", "parameters": {"penalty per willful violation under section 17": "161323"}}

- ✓ **VERITROOPER (PASS):** {"name": "no_function", "parameters": {}}

314. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4618\_01`

- **Question:** Provided that observers do not interfere with the monitoring, are they entitled to record the results obtained?

- **Ground truth:** Yes, without interfering with the monitoring, observers are entitled to record the results obtained and receive results supplied by the laboratory.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "observation_procedures", "parameters": {"entitled_to_record_results": "True"}}

- ✓ **VERITROOPER (PASS):** {"name": "observer_entitlement", "parameters": {"entitlement": "True", "interference": "False", "entitlement_type": "record_results"}}

315. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_00`

- **Question:** Sections 1926.50 and 1926.250 each contain a collection of information requirement. Do they share the same OMB control number, and if so, what is it?

- **Ground truth:** Yes, both 1926.50 and 1926.250 are assigned OMB control No. 1218-0093.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "check_omb_control_number", "parameters": {"citation1": "1926.50", "citation2": "1926.250"}}

- ✓ **VERITROOPER (PASS):** {"name": "check_omb_control_number", "parameters": {"section1": "1926.50", "section2": "1926.250"}}

316. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2920\_02`

- **Question:** Based on paragraph (v) and paragraph (vii), what protections are required for an employee who manually adds chemicals to a tank containing a liquid that may irritate the skin if splashed?

- **Ground truth:** Per paragraph (v), the employee must wear either tight-fitting chemical goggles or an effective face shield because of the danger of splashing during manual additions. Per paragraph (vii), there must also be a nearby supply of clean cold water (with a quick opening valve and at least 48 inches of hose not smaller than three-fourths inch), or alternatively deluge showers and eye flushes, so liquids can be quickly washed off the skin or clothing.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "protect_employee_from_chemical_splash", "parameters": {"protective_gear": "tight-fitting chemical goggles or an effective face shield", "skin_protection": "rubber or other impervious boots or shoes, rubbers, or wooden-soled shoes sufficient to keep feet dry", "clothing_protection": "aprons, coats, jackets, sleeves, or other garments made of rubber, or of other materials impervious to liquids other than water, as are required to keep their clothing dry", "skin_washing_facilities": "a supply of clean cold water"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_protections", "parameters": {"action": "manually adds chemicals to a tank", "hazard": "liquid that may irritate the skin if splashed"}}

317. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3059\_02`

- **Question:** When spraying areas are illuminated through transparent panels, what type of lighting unit must be used?
- **Ground truth:** Only fixed lighting units shall be used as a source of illumination.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "fixed_lighting_unit", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "fixed_lighting_unit", "parameters": {"type": "fixed"}}

318. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4269\_00`

MODEL ERROR

- **Question:** According to the purpose and scope, public employees were not assured protections under the Act prior to a State plan because of a definition in one section, but a different section provides the route for ensuring their protection. Which sections are these, and what does each establish?
- **Ground truth:** Section 3(5) excludes State and local government agencies from the definition of 'employer,' which is why such public employees had not heretofore been assured any protections under the Act. Section 18 (and specifically section 18(b)) is the only provision under which such public employees are ensured protection through an approved State plan, permitting States to submit plans for any occupational safety and health issue for which a Federal standard has been promulgated under section 6.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_sections", "parameters": {"section1": "section 3(5)", "section2": "section 18(b)"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_sections", "parameters": {"section1": "3(5)", "section2": "18"}}

319. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_02`

MODEL ERROR

- **Question:** If an employee performs work within reaching distance of exposed energized parts while wearing a wristwatch that does not increase the hazards associated with contact, must the employer ensure it is removed or rendered nonconductive?
- **Ground truth:** No; the requirement to remove or render nonconductive does not apply if such articles do not increase the hazards associated with contact with the energized parts.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "remove_conductive_articles", "parameters": {"article": "wristwatch", "hazard_increase": "false"}}
- ✗ **VERITROOPER (FAIL):** {"name": "ensure_conductive_articles_removal", "parameters": {"increase_hazards": "False"}}

320. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2239\_03`

- **Question:** Regarding x-ray records, which specific x-rays must be kept according to the medical records provisions, combining the relevant subparts?
- **Ground truth:** The employer must keep the initial x-ray (ii)(c), the x-rays for the most recent five (5) years (ii)(d), and any x-ray with a demonstrated abnormality and all subsequent x-rays (ii)(e).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "keep_xray_records", "parameters": {"initial_xray": "True",

"recent_5_years": "True", "abnormality": "True", "subsequent_xrays": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "keep_xray_records", "parameters": {"initial_xray": "True",
"recent_five_years": "True", "abnormality": "True"}}

321. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0644\_01`

- **Question:** Combining the references in paragraph (p)(1), which two paragraphs cover the persons responsible for compiling process safety information and those responsible for developing operating procedures?
- **Ground truth:** Process safety information is required by paragraph (d), and operating procedures are required by paragraph (f).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "combine_references", "parameters": {"paragraph": "(p)(1)", "references": "['(d)', '(f)']"}}
- ✓ **VERITROOPER (PASS):** {"name": "reference_combiner", "parameters": {"ref1": "paragraph (d)", "ref2": "paragraph (f)"}}

322. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2542\_02`

- **Question:** Per the signal duration rule in (v) and the automatic response rule in (vi), what two operational requirements must the signal satisfy when an initiating event occurs?
- **Ground truth:** The signal must last a minimum duration sufficient to insure that all affected persons hear it, and it must sound automatically in response to the initiating event without requiring any human action.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "signal_duration_and_automatic_response_requirements", "parameters": {"signal_duration": "sufficient to insure that all affected persons hear the signal", "automatic_response": "respond automatically to an initiating event without requiring any human action to sound the signal"}}
- ✓ **VERITROOPER (PASS):** {"name": "signal_requirements", "parameters": {"duration": "sufficient to insure that all affected persons hear the signal", "response": "automatically to an initiating event without requiring any human action to sound the signal"}}

323. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_00`

- **Question:** For an FRP live-line tool's initial test under paragraph (j)(1)(i), what voltage per meter must it withstand and for how long?
- **Ground truth:** 328,100 volts per meter (100,000 volts per foot) for 5 minutes
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "live_line_tool_test", "parameters": {"voltage_per_meter": "328100", "duration_minutes": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "tool_test", "parameters": {"voltage_per_meter": "328100", "duration_minutes": "5"}}

324. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_05`

- **Question:** What is the Techstreet telephone toll-free number, as distinct from its fax number 734-780-2046?
- **Ground truth:** 800-699-9277
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_telephone_number", "parameters": {"telephone": "800-699-9277", "fax": "734-780-2046"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_telephone_number", "parameters": {"telephone": "800-699-9277", "fax": "734-780-2046"}}

325. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_06`

BAD QUESTION

- **Question:** What is the citation reference for the additivity formula mentioned as A.1.3.6 in another regulatory code?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_citation_reference", "parameters": {"code": "A.1.3.6", "reference": "additivity formula"}}
- **✓ VERITROOPER (PASS):** {"name": "citation_reference", "parameters": {"code": "A.1.3.6"}}

326. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1264\_01`

- **Question:** Combining the beater guard requirement in paragraph (d)(1) and the feed roll requirement in paragraph (d)(3), what is the protective purpose served by guards on opening and picking machinery in each case?
- **Ground truth:** The beater covers (d)(1) must prevent contact with the beater, while the feed roll guard (d)(3) must be designed to prevent the operator from reaching the nip while the machinery is in operation. Both serve to prevent operator contact with hazardous moving parts.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "guards_on_opening_and_picking_machinery", "parameters": {"beater_guard": "prevent contact with the beater", "feed_roll_guard": "prevent the operator from reaching the nip while the machinery is in operation"}}
- **✗ VERITROOPER (FAIL):** {"name": "protective_purpose", "parameters": {"guard_type": "beater guard", "protective_purpose": "prevent contact with the beater"}}

327. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_3423\_00`

MODEL ERROR

- **Question:** What is the minimum factor by which jacking equipment must be capable of supporting the load being lifted during jacking operations?
- **Ground truth:** two and one-half times
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_support_factor", "parameters": {"factor": "2"}}
- **✗ VERITROOPER (FAIL):** {"name": "jacking_equipment_support_factor", "parameters": {"factor": "2"}}

328. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2204\_00`

- **Question:** Based on the General surveillance section's timing requirements and the Hematology Guidelines' minimum battery of tests, what tests must an accredited laboratory perform at the initial examination provided within 60 days of the standard's effective date?
- **Ground truth:** The initial examination (provided within 60 days of the effective date or at time of initial assignment) must include the minimum battery of tests performed by an accredited laboratory: red cell, white cell, and platelet counts, white blood cell differential, hematocrit, and red cell indices.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "minimum_battery_of_tests", "parameters": {"laboratory": "accredited laboratory", "tests": "['Red cell count', 'White cell count', 'Platelet count', 'White blood cell differential', 'Hematacrit', 'Red cell indices']"}}
- **✓ VERITROOPER (PASS):** {"name": "perform_tests", "parameters": {"test_type": "minimum battery of tests", "laboratory_type": "accredited laboratory", "examination_type": "initial examination", "timeframe": "within 60 days of the standard"}}

329. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_01`

- **Question:** When a second physician is involved, under what condition must the employer provide that physician the same information given to the examining physician?
- **Ground truth:** Upon request either by the second physician or by the employee.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "information_provided_to_second_or_third_physician", "parameters": {"condition": "upon request either by the second or third physician, or by the employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "employer_provide_information", "parameters": {"condition": "upon request either by the second physician, or by the employee"}}

330. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_03`

- **Question:** A HEPA filter must be at least what percent efficient in removing particles?
- **Ground truth:** 99.97 percent
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "High-efficiency particulate air [HEPA] filter", "parameters": {"percent_efficient": "99.97", "particles_of": "0.3 micrometers in diameter"}}
- ✓ **VERITROOPER (PASS):** {"name": "HEPA filter efficiency", "parameters": {"efficiency": "99.97"}}

331. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3038\_03`

MODEL ERROR

- **Question:** If an employer arranges in advance for an outside fully-trained emergency response team to respond in a reasonable period, must all employees still be trained to the full degree specified?
- **Ground truth:** No, provided all employees who may come to the incident first have sufficient awareness training and have been instructed to call the designated outside team.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "emergency_response_to_hazardous_substance_releases", "parameters": {"arrangements_made": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "train_all_employees", "parameters": {"arrangements_made": "True"}}

332. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_03`

- **Question:** Which section of the Service Contract Act of 1965 requires safety and health standards as a condition of every Federal contract exceeding \$2,500?
- **Ground truth:** Section 2(e)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 2", "parameters": {"e": "Service Contract Act of 1965"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"act_name": "Service Contract Act of 1965", "year": "1965", "contract_amount": "2500"}}

333. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0517\_02`

- **Question:** Based on the general provision in (g)(1) and the building requirement in (g)(2)(ii), how must blasting agents (excluding water gels) be transported and stored, and what type of construction is required for mixing buildings?
- **Ground truth:** Blasting agents, excluding water gels, shall be transported, stored, and used in the same manner as explosives, and buildings used for mixing shall be of noncombustible construction or sheet metal on wood studs.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "transport_and_store_blasting_agents", "parameters": {"transport_and_store": "in the same manner as explosives", "building_construction": "noncombustible construction or sheet metal on wood studs"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_building_requirements", "parameters": {"transportation_storage": "same manner as explosives", "construction_type": "noncombustible construction or sheet metal on wood studs"}}

334. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4044\_01`

- **Question:** For at least how many minutes must the platform be held in a suspended position with the test load evenly distributed during proof testing?
- **Ground truth:** five minutes
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "proof testing", "parameters": {"minutes": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "print", "parameters": {"n": "5"}}

335. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_00`

- **Question:** In appendix A Table A, what is the upper bound of category A for Cadmium in urine (CdU) in the first table listing (>3 and ≤?)?
- **Ground truth:** 15
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_upper_bound_category_A_CdU", "parameters": {"marker": "Cadmium in urine (CdU)", "table_number": "1"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_upper_bound", "parameters": {"category": "A", "biological_marker": "Cadmium in urine (CdU)"}}

336. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0694\_02`

- **Question:** Considering both the nonpotable water rules in (n)(2) and the potable water rules in (n)(1), what must be done to distinguish unsafe water outlets and how must the two water systems relate?
- **Ground truth:** Outlets for nonpotable water, such as firefighting water, must be identified to clearly indicate the water is unsafe and not to be used for drinking, washing, or cooking; and there must be no cross-connection, open or potential, between a system furnishing potable water and a system furnishing nonpotable water.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "identify_and_separate_water_systems", "parameters": {"nonpotable_water_outlets": "be posted or otherwise marked to indicate clearly that the water is unsafe and is not to be used for drinking, washing, or cooking purposes", "relationship_between_water_systems": "there shall be no cross-connection, open or potential, between a system furnishing potable water and a system furnishing nonpotable water"}}
- ✗ **VERITROOPER (FAIL):** {"name": "identify_water_outlets", "parameters": {"potable_water": "false", "nonpotable_water": "true"}}

337. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_05`

HALLUCINATION

- **Question:** What civil penalty applies to an employer who selects filtering facepiece respirators against asbestos fibers in violation of this section?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "adjusted_civil_monetary_penalties", "parameters": {"penalty_per_willful_violation": "161323"}}
- ✓ **VERITROOPER (PASS):** This information is not provided in the given evidence.

338. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3863\_01`

- **Question:** Within how many days after an employee's request must the employer provide the employee with the information required to be provided to the examining physician under paragraph (l)(9)?
- **Ground truth:** 30 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_information_for_examining_physician", "parameters": {"days": "30"}}
- ✓ **VERITROOPER (PASS):** {"name": "time_to_provide_information", "parameters": {"days": "30"}}

339. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1947\_01`

- **Question:** This section applies to the transportation of vinyl chloride or polyvinyl chloride except under what circumstance?
- **Ground truth:** Except to the extent that the Department of Transportation may regulate the hazards covered by this section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "transportation exception", "parameters": {"circumstance": "to the extent that the Department of Transportation may regulate the hazards covered by this section"}}
- ✓ **VERITROOPER (PASS):** {"name": "applies_to_transportation", "parameters": {"circumstance": "to the extent that the Department of Transportation may regulate the hazards covered by this section"}}

340. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1202\_01`

- **Question:** Combining the general transmission pipeline provision (d)(1)(i) and the field shop operations provision (d)(1)(ii), which two § sections of part 1910 must be observed for field shop operations but not all are listed under the general provision?
- **Ground truth:** The general provision (d)(1)(i) requires § 1910.254 only. Field shop operations (d)(1)(ii) require both §§ 1910.253 and 1910.254, so § 1910.253 is the additional section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Combining the general transmission pipeline provision (d)(1)(i) and the field shop operations provision (d)(1)(ii), which two § sections of part 1910 must be observed for field shop operations but not all are listed under the general provision", "parameters": {"a": "1910.253", "b": "1910.254"}}
- ✓ **VERITROOPER (PASS):** {"name": "combine_sections", "parameters": {"general_provision": "['(b)', '(c)', '1910.254']", "field_shop_operations": "['(a)', '(b)', '(c)', '1910.253', '1910.254']"}}

341. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_05`

- **Question:** Under paragraph (l)(3)(iii), if both the initial and the medical-exam monitoring show CdU exceeding 15 µg/g Cr and CdU also exceeding 3 µg/g Cr, what must the physician do?
- **Ground truth:** The physician shall medically remove the employee from exposure to cadmium at or above the action level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "medical_removal", "parameters": {"CdU": "15", "CdB": "None", "B2_M": "None", "CdU_trigger": "3"}}
- ✗ **VERITROOPER (FAIL):** {"name": "physician_determination", "parameters": {"CdU": "15", "CdB": "5"}}

342. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0882\_00`

- **Question:** Under paragraph (c)(2)(ii), an energy isolating device that is capable of being locked out must use lockout unless the employer can demonstrate that a tagout system provides full employee protection. According to paragraph (c)(3)(i), where must the tagout device be attached in that situation?
- **Ground truth:** The tagout device shall be attached at the same location that the lockout device would have been attached, and the employer shall demonstrate that the tagout program will provide a level of safety equivalent to that obtained by using a lockout program.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "attach_tagout_device", "parameters": {"location": "the same location that the lockout device would have been attached"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_location", "parameters": {"text": "the same location that the lockout device would have been attached"}}

343. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2016\_00`

MODEL ERROR

- **Question:** Appendix B states the regulation applies to records of employees exposed to toxic substances or

harmful physical agents, and references the non-mandatory nature established in paragraph (i). Given paragraph (i)'s statement about appendices and Appendix B's description of RTECS, does an employer have a legal obligation to purchase RTECS to determine if the rule applies?

- **Ground truth:** No. Paragraph (i) states the appendices do not, by themselves, create any additional obligations not otherwise imposed by the section, and Appendix B explicitly notes that the final regulation does not require that employers purchase a copy of RTECS, and many employers need not consult RTECS to ascertain whether their records are subject to the rule.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "does_employer_have_legal_obligation_to_purchase_rtecs", "parameters": {"appendix_b": "Appendix B to \u00a7 1910.1020\u2014Availability of NIOSH Registry of Toxic Effects of Chemical Substances (RTECS) (Non-Mandatory)", "paragraph_i": "(i)"}}
- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Does an employer have a legal obligation to purchase RTECS to determine if the rule applies?", "answer": "No"}}

344. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_08`

- **Question:** What is the 'high level' threshold for beta-2-microglobulin associated with greater chance of developing kidney diseases?
- **Ground truth:** greater than 1000 micrograms per gram of creatinine (µg/g Cr)
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "high_level_beta_2_microglobulin", "parameters": {"level": "1000"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_high_level_threshold", "parameters": {"marker": "beta-2-microglobulin"}}

345. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_05`

HALLUCINATION

- **Question:** What is the mailing address of the Chief Administrative Law Judge that must appear in the findings?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_address_of_chief_administrative_law_judge", "parameters": {}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

346. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_0362\_01`

MODEL ERROR

- **Question:** Under what condition is wire rope permitted for use as the emergency stop rope?
- **Ground truth:** Wire rope may be used only if it is marlin-covered; unless marlin-covered, it shall not be used.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "is_wire_rope_permitted", "parameters": {"condition": "as the emergency stop rope"}}
- **✓ VERITROOPER (PASS):** {"name": "emergency_stop_rope_permitted", "parameters": {"condition": "marlin-covered"}}

347. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0458\_01`

- **Question:** Combining the fire protection requirements regarding portable fire extinguishers and the placement of materials on wharves, what is the minimum extinguisher rating required near hose connections and how must materials be arranged relative to firefighting equipment?
- **Ground truth:** Portable fire extinguishers must have a rating of not less than 12-BC and be located within 75 feet of areas like hose connections, pumps, and separator tanks. Additionally, material shall not be placed on wharves in such a manner as to obstruct access to firefighting equipment or important pipeline control

valves.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "fire_protection_requirements", "parameters": {"extinguisher_rating": "12-BC", "material_arrangement": "not obstruct access to firefighting equipment"}}
- ✓ **VERITROOPER (PASS):** {"name": "fire_protection_requirements", "parameters": {"extinguisher_rating": "12-BC", "material_arrangement": "not obstructing access to firefighting equipment"}}

348. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0767\_01`

- **Question:** According to paragraph (b), under what circumstances must the fit test be repeated, and what should be done if the respirant's comfort is found unacceptable upon questioning?
- **Ground truth:** The fit test must be repeated if any adjustment is made to the respirator once the fit test exercises begin, since any adjustment voids the test. If the respirator's comfort has become unacceptable when the test subject is questioned upon completion of the protocol, another model of respirator shall be tried.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "repeat_fit_test", "parameters": {"reason": "The respirator shall not be adjusted once the fit test exercises begin. Any adjustment voids the test, and the fit test must be repeated.", "action": "If the test subject states that the respirator is unacceptable, the employer must ensure that the test administrator repeats the protocol using another respirator model."}}
- ✓ **VERITROOPER (PASS):** {"name": "qlft_protocol", "parameters": {"circumstances": "if the respirator has become unacceptable", "action": "try another model of respirator"}}

349. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_06`

HALLUCINATION

- **Question:** Who is responsible for certifying that the compressed-air bulkhead test was successfully completed?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "compressed-air_bulkhead_test_certification", "parameters": {"requester": "appointed physician"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

350. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0916\_00`

- **Question:** Under § 1910.160(a)(2), what specific characteristic of certain fixed extinguishing systems causes them to be subject to the requirements of paragraphs (b)(4) through (b)(7) and (c)?
- **Ground truth:** Systems not installed to meet a particular OSHA standard but which, by means of their operation, may expose employees to possible injury, death, or adverse health consequences caused by the extinguishing agent.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "characteristic of fixed extinguishing systems", "parameters": {"characteristic": "may expose employees to possible injury, death, or adverse health consequences caused by the extinguishing agent"}}
- ✓ **VERITROOPER (PASS):** {"name": "fixed_extinguishing_systems", "parameters": {"characteristic": "may expose employees to possible injury, death, or adverse health consequences caused by the extinguishing agent"}}

351. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_07`

HALLUCINATION

- **Question:** What certification must an equivalent entity hold to be recognized for performing repairs?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_certification", "parameters": {"entity_type": "equivalent"}}

entity", "certification_type": "written certification"}}

- ✓ **VERITROOPER (PASS)**: The provided evidence does not contain enough information to answer the question about the certification required for an equivalent entity to be recognized for performing repairs.

352. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_05`

HALLUCINATION

- **Question**: What penalty applies to operators who store oxygen in a building used for non-exclusive purposes?

- **Ground truth**: UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL)**: {"name": "penalty_for_non_compliant_storage", "parameters": {"amount": "10000", "hazard": "oxygen storage in non-exclusive building"}}

- ✓ **VERITROOPER (PASS)**: The provided evidence does not contain enough information to answer the question.

353. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_03`

- **Question**: Who is permitted to refill a cylinder, with no other person allowed to do so?

- **Ground truth**: No one except the owner of the cylinder or a person authorized by him may refill a cylinder.

- ✓ **Baseline (vanilla RAG) (PASS)**: {"name": "refill_cylinder", "parameters": {"person": "the owner of the cylinder or person authorized by him"}}

- ✓ **VERITROOPER (PASS)**: {"name": "refill_cylinder", "parameters": {"person": "the owner of the cylinder or person authorized by him"}}

354. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1695\_03`

MODEL ERROR

- **Question**: A circuit has been deenergized by a circuit protective device and a worker wants to reenergize it. Combining paragraph (b)(2) and the following Note, when may the circuit be reenergized without examining the circuit or connected equipment?

- **Ground truth**: Per (b)(2) the circuit may not be manually reenergized until it has been determined the equipment and circuit can be safely energized; however, per the Note, no examination is needed before reenergizing when it can be determined from the design of the circuit and the overcurrent devices that the automatic operation was caused by an overload rather than a fault condition.

- ✗ **Baseline (vanilla RAG) (FAIL)**: {"name": "Reenergizing circuit", "parameters": {"circuit": "deenergized circuit", "condition": "overload rather than a fault condition"}}

- ✓ **VERITROOPER (PASS)**: {"name": "reenergize_circuit", "parameters": {"circuit_design": "overload", "overcurrent_devices": "automatic operation"}}

355. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_02`

MODEL ERROR

- **Question**: If conductive material is being handled on a scaffold near an uninsulated power line of less than 50 kv, what is the minimum required clearance?

- **Ground truth**: 10 feet (3.1 m).

- ✗ **Baseline (vanilla RAG) (FAIL)**: {"name": "clearance_distance", "parameters": {"voltage": "less than 50 kv", "power_line": "uninsulated"}}

- ✗ **VERITROOPER (FAIL)**: {"name": "get_minimum_clearance", "parameters": {"voltage": "less than 50 kv", "line_type": "uninsulated"}}

356. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_02`

MODEL ERROR

- **Question:** If an examining physician determines that an employee's abnormal liver function tests are NOT associated with MDA exposure, can the employee still be temporarily removed?
- **Ground truth:** Yes, if the physician also determines the abnormalities may be exacerbated as a result of occupational exposure to MDA.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "temporary_removal_due_to_abnormal_liver_function_tests", "parameters": {"associated_with_mda_exposure": "False"}}
- **✗ VERITROOPER (FAIL):** {"name": "temporary_removal_due_to_final_medical_determination", "parameters": {"reason": "abnormal liver function tests not associated with MDA exposure"}}

357. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_14`

- **Question:** What molar volume value (at 25 degrees C and 760 mm Hg) is used in the ppb conversion formula?
- **Ground truth:** 24.46
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "ppb conversion formula", "parameters": {"24.46": "24.46"}}
- **✓ VERITROOPER (PASS):** {"name": "get_molar_volume", "parameters": {"temperature": "25", "pressure": "760"}}

358. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4517\_03`

- **Question:** Using paragraph (b) on the nature of filing and paragraph (d) on what counts as the date of filing, if a complaint is made orally by telephone, what determines the filing date and how is the complaint recorded?
- **Ground truth:** Per paragraph (b), oral complaints (including by telephone) are accepted and will be reduced to writing by OSHA. Per paragraph (d), the date of the telephone call will be considered the date of filing.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Nature of filing and date of filing", "parameters": {"complaint made orally by telephone": "date of telephone call is considered the date of filing", "complaint recorded": "OSHA will reduce the oral complaint to writing"}}
- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If a complaint is made orally by telephone, what determines the filing date and how is the complaint recorded?", "paragraphs": "[\n(b) Nature of filing. No particular form of complaint is required. A complaint may be filed orally or in writing. Oral complaints will be reduced to writing by OSHA. If the complainant is unable to file the complaint in English, OSHA will accept the complaint in any language.\n(d) Time for filing. Within 180 days after an alleged violation of MAP-21 occurs, any person who believes that he or she has been retaliated against in violation of the MAP-21 may file, or have filed by any person on his or her behalf, a complaint alleging such retaliation. The date of the postmark, facsimile transmittal, electronic communication transmittal, telephone call, hand-delivery, delivery to a third-party commercial carrier, or

359. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_09`

HALLUCINATION

- **Question:** Within how many days must an employer notify other employees of a potential workplace exposure identified through the log?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "notify_other_employees", "parameters": {"days": "1"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

360. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1096\_00`

- **Question:** What is the maximum opening that work rests shall be kept adjusted to from the wheel on offhand grinding machines?
- **Ground truth:** one-eighth inch
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "work_rests_adjustment", "parameters": {"maximum_opening": "1/8 inch"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_maximum_opening", "parameters": {"machine_type": "offhand grinding machines"}}

361. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0767\_00`

- **Question:** Based on the test duration rules in paragraph (b), if a fit test protocol consists of four standard exercises plus the grimace exercise, what is the total time spent performing exercises?
- **Ground truth:** Four standard exercises at one minute each equals 4 minutes (240 seconds), plus the grimace exercise at 15 seconds, for a total of 4 minutes and 15 seconds (255 seconds).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_total_time", "parameters": {"x": "4", "standard_exercise_time": "15", "grimace_exercise_time": "15"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_total_time", "parameters": {"num_standard_exercises": "4", "grimace_duration": "15", "standard_duration": "60"}}

362. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0913\_00`

- **Question:** According to the training requirements, what two separate obligations does an employer have upon a worker's 'initial' point - one for general employees provided extinguishers and one for designated fire fighting equipment users?
- **Ground truth:** Per paragraph (g)(2), the employer must provide the educational program required in (g)(1) upon initial employment; and per paragraph (g)(4), the employer must provide the training required in (g)(3) upon initial assignment to the designated group of employees.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "training_requirements", "parameters": {"initial_employee": "education program", "initial_designated_employee": "training in the use of the appropriate equipment"}}
- ✓ **VERITROOPER (PASS):** {"name": "training_obligations", "parameters": {"event": "initial employment", "obligation": "educational program"}}

363. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_08`

HALLUCINATION

- **Question:** What form must the employer use to document the physician's respirator-use determination?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "Appendix C", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

364. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3703\_03`

- **Question:** Under Appendix E, who is permitted to classify chest X-rays, and what reference materials must such a person have immediately available when classifying digitally-acquired chest X-rays?
- **Ground truth:** X-rays may be classified only by a B-Reader, a board eligible/certified radiologist, or an experienced physician with known expertise in pneumoconioses (paragraph (b)). When classifying digitally-acquired chest X-rays, that person must have immediately available a complete set of ILO standard digital chest radiographic images provided for use with the Guidelines for the Use of the ILO International Classification of Radiographs of Pneumoconioses (revised edition 2011) (paragraph (d)).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_chest_x-rays", "parameters": {"physician": "B-Reader, board eligible/certified radiologist, or experienced physician with known expertise in pneumoconioses", "reference_materials": "a complete set of ILO standard digital chest radiographic images provided for use with the Guidelines for the Use of the ILO International Classification of Radiographs of Pneumoconioses (revised edition 2011)"}}
- ✓ **VERITROOPER (PASS):** {"name": "classify_chest_x-rays", "parameters": {"permitted_classifiers": ["B-Reader", "board eligible/certified radiologist", "experienced physician with known expertise in pneumoconioses"], "reference_materials": "ILO standard digital chest radiographic images"}}

365. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0333\_01`

- **Question:** Under the system utilizing angulated roping and building face rollers, is it acceptable where the suspended equipment is 140 feet above ground level?
- **Ground truth:** No; the system is acceptable only where the suspended portion does not exceed 130 feet above a safe surface or ground level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "System utilizing angulated roping and building face rollers", "parameters": {"acceptable": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_acceptable", "parameters": {"system": "angulated roping and building face rollers", "height": "140"}}

366. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_07`

HALLUCINATION

- **Question:** What is the duration of air supply required for the auxiliary positive-pressure SCBA used during Class I asbestos work above 1 f/cc?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_air_supply_duration", "parameters": {"type": "auxiliary positive-pressure SCBA", "work_class": "Class I", "asbestos_concentration": "above 1 f/cc"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the duration of air supply required for the auxiliary positive-pressure SCBA used during Class I asbestos work above 1 f/cc.

367. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3020\_01`

MODEL ERROR

- **Question:** Combining subpart (vi) and subpart (ix), what must happen before a drum that cannot be moved without rupture is emptied?
- **Ground truth:** Before any movement of drums or containers, all employees exposed to the transfer operation must be warned of the potential hazards associated with the contents (vi); and a drum that cannot be moved without rupture, leakage, or spillage must be emptied into a sound container using a device classified for the material being transferred (ix).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "emptying", "parameters": {"drum": "sound container", "device": "classified for the material being transferred"}}
- ✗ **VERITROOPER (FAIL):** {"name": "warn_employees", "parameters": {"drum_contents": "hazardous substance"}}

368. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_07`

HALLUCINATION

- **Question:** What is the reference number or citation of the footnote marked '1' regarding appropriate animal tests?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_footnote_reference", "parameters": {"n": "1"}}

- **✗ VERITROOPER (FAIL):** {"name": "get_footnote_reference", "parameters": {"footnote_number": "1"}}

369. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_02`

- **Question:** If some employees of an establishment are covered under collective bargaining arrangements and others are not, who must the nonmanagement members represent?

- **Ground truth:** Members shall be representative of both groups.

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "nonmanagement members representation", "parameters": {"represented groups": "both employees covered under collective bargaining arrangements and those not covered"}}

- **✗ VERITROOPER (FAIL):** {"name": "get_nonmanagement_members_representation", "parameters": {"collective_bargaining": "some", "other_employees": "not"}}

370. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4293\_02`

- **Question:** If the inspector determines that additional representatives will further aid the inspection, may they accompany the inspectors?

- **Ground truth:** Yes; additional representatives of the official in charge and of employees may accompany the inspectors if the inspector determines such additional representatives will further aid the inspection.

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "accompany_inspectors", "parameters": {"additional_representatives": "True"}}

- **✓ VERITROOPER (PASS):** {"name": "accompany_inspectors", "parameters": {"determined_by_inspector": "True"}}

371. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2164\_01`

- **Question:** A laboratory is seeking reevaluation for CDU analysis. According to the reevaluation guidelines combined with the internal QA/QC requirements, what additional analyte must it both establish a QA/QC program for and obtain CAP accreditation in?

- **Ground truth:** CRTU - the reevaluation guidelines require CAP accreditation for CRTU analysis, and the internal QA/QC procedures require that a QA/QC program for CRTU also be established for laboratories analyzing CDU.

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "establish_QA_QC_and_CAP_accreditation", "parameters": {"laboratory": "CDU analysis", "analyte": "CRTU"}}

- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "A laboratory is seeking reevaluation for CDU analysis. According to the reevaluation guidelines combined with the internal QA/QC requirements, what additional analyte must it both establish a QA/QC program for and obtain CAP accreditation in?", "answer": "CRTU"}}

372. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3047\_01`

MODEL ERROR

- **Question:** Although the Note states the appendices serve as non-mandatory guidelines, which paragraph does it identify as making the use of Level A and Level B PPE protection mandatory in certain circumstances?

- **Ground truth:** Paragraph § 1926.65(g) makes mandatory in certain circumstances the use of Level A and Level B PPE protection.

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"number": "1910.120(g)"}}

- **✓ VERITROOPER (PASS):** {"name": "get_paragraph", "parameters": {"number": "1926.65(g)"}}

373. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_03`

HALLUCINATION

- **Question:** What penalty applies to a respondent who fails to comply with the immediately effective reinstatement requirement?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "find_penalty", "parameters": {"respondent_action": "fails to comply with the immediately effective reinstatement requirement"}}
- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

374. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1249\_00`

- **Question:** Which American National Standard governs the exhaust ventilation requirement for the room in which bleach powder is mixed?
- **Ground truth:** Z9.1-1951
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "American National Standard", "parameters": {"Z9": "1-1951"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_standard", "parameters": {"standard_number": "Z9.1-1951"}}

375. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_07`

- **Question:** What is the NFPA number for Fire Protection for Laboratories Using Chemicals?
- **Ground truth:** NFPA-45, 1982
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "get_nfpa_number", "parameters": {"topic": "Fire Protection for Laboratories Using Chemicals"}}
- ✔ **VERITROOPER (PASS):** {"name": "NFPA", "parameters": {"number": "45"}}

376. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1349\_02`

- **Question:** Based on the requirements about footing and the requirement about hand grip, what conditions must the operator avoid when positioning the chain saw to cut?
- **Ground truth:** The operator must be certain of footing before cutting, and the chain saw shall not be used in a position or distance that could cause the operator to become off-balance, have insecure footing, or relinquish a firm grip (per (ix)); this firm grip relates to the requirement in (viii) that both hands encircle the handles during operation.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "check_positioning_conditions", "parameters": {"condition1": "off-balance", "condition2": "insecure footing", "condition3": "relinquish a firm grip on the saw"}}
- ✔ **VERITROOPER (PASS):** {"name": "operator_positioning_requirements", "parameters": {"footing_conditions": "off-balance, insecure footing", "grip_conditions": "relinquish a firm grip on the saw"}}

377. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_05`

- **Question:** Which section of the Contract Work Hours and Safety Standards Act provides that the Act shall not apply to work required to be done under the Walsh-Healey Public Contracts Act?
- **Ground truth:** Section 103(b)
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "section 103", "parameters": {"b": "the Act shall not apply to any work required to be done in accordance with the provisions of the Walsh-Healey Public Contracts Act"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"act_name": "Contract Work Hours and Safety Standards Act", "target_act": "Walsh-Healey Public Contracts Act"}}

378. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0012\_02`

- **Question:** If employees or their representatives wish to participate in review proceedings concerning alleged violations, does the State plan allow it?
- **Ground truth:** Yes, the plan provides employees or their representatives an opportunity to participate in review proceedings, such as a full hearing on the issues.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Provides for an employer to have the right of review of violations alleged by the State or any agency to which it has duly delegated authority, abatement periods and proposed penalties, where appropriate, for employees or their representatives to challenge the reasonableness of the period of time fixed in the citation for the abatement of the hazard, and for employees or their representatives to have an opportunity to participate in review, proceedings, by such means as providing for administrative review, with an opportunity for a full hearing on the issues.", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "review_proceedings", "parameters": {"participants": "employees or their representatives", "proceeding_type": "alleged violations"}}

379. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1848\_02`

MODEL ERROR

- **Question:** Using the recordkeeping requirements in Section V.C and the latency information in Section II.A, how long must medical records be kept, and how does this relate to when disease symptoms typically appear?
- **Ground truth:** Medical records must be kept for the period of employment plus thirty (30) years (V.C), which accommodates the fact that symptoms of asbestos-related diseases generally do not appear for 20 or more years after initial exposure (II.A).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_medical_record_keeping_time", "parameters": {"medical_record_keeping_time": "30", "latency_period": "20"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_recordkeeping_requirements", "parameters": {"medical_records": "true", "latency": "20"}}

380. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0694\_01`

WRONG CALC/QUOTE

- **Question:** Combining the toilet facility table in (n)(3)(i), how many toilet seats and urinals are required for a worksite with 160 employees versus one with 250 employees?
- **Ground truth:** For 160 employees (more than 20, fewer than 200): one toilet seat and one urinal per 40 employees = 4 toilet seats and 4 urinals. For 250 employees (more than 200): one toilet seat and one urinal per 50 employees = 5 toilet seats and 5 urinals.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_toilet_facilities", "parameters": {"employees": "160"}}; {"name": "calculate_toilet_facilities", "parameters": {"employees": "250"}}
- ✗ **VERITROOPER (FAIL):** [{"name": "calculate_toilet_facilities", "parameters": {"employees": "160"}}; {"name": "calculate_toilet_facilities", "parameters": {"employees": "250"}}]

381. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3468\_02`

MODEL ERROR

- **Question:** If the Secretary of Labor requests an air quality test record, where must it have been maintained and what details must it contain?
- **Ground truth:** It must be maintained above ground at the worksite and include the location, date, time, substance and amount monitored.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_air_quality_test_record", "parameters": {"location": "on-site", "details": "date, time, and results of the test"}}

- ✓ **VERITROOPER (PASS):** {"name": "maintain_record", "parameters": {"location": "above ground at the worksite", "details": "location, date, time, substance and amount monitored"}}

382. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1441\_03`

- **Question:** When an employee works on a cable remote from the cable terminal, under what condition must the cable not be grounded at the cable terminal?
- **Ground truth:** When there is a possibility of hazardous transfer of potential should a fault occur.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ensure_cable_not_grounded", "parameters": {"condition": "there is a possibility of hazardous transfer of potential should a fault occur"}}
- ✓ **VERITROOPER (PASS):** {"name": "additional_precautions", "parameters": {"condition": "possibility of hazardous transfer of potential should a fault occur"}}

383. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_11`

- **Question:** What is the retention time of the MDA derivative in the GC analysis?
- **Ground truth:** 3.5 min
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "GC conditions", "parameters": {"Zone temperatures": "Column\u2014220 degrees C Injector\u2014235 degrees C Detector\u2014335 degrees C", "Gas flows, Ar/CH 4": "Column\u201428 mL/min (95/5) Purge\u2014340 mL/min", "Injection volume": "5.0 uL", "Column": "6 ft \u00d7 1/8 in ID glass, 3% OV-101 on 100/120 Gas Chrom Q", "Retention time of MDA derivative": "3.5 min"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retention_time", "parameters": {"compound": "MDA derivative"}}

384. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4502\_03`

MODEL ERROR

- **Question:** Under paragraph (a) materials are redacted consistent with the Privacy Act, yet paragraph (d) addresses confidentiality of certain persons. How do these provisions treat the complainant differently regarding confidentiality protection?
- **Ground truth:** Paragraph (a) redactions are made consistent with the Privacy Act of 1974 and other confidentiality laws, but the complainant (or counsel) receives an unredacted copy of those materials. Paragraph (d) requires investigations to protect the confidentiality of any person who provides information on a confidential basis other than the complainant-so the complainant is expressly excluded from that confidential-informant protection.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "redact_complainant_info", "parameters": {"complainant_info": "false"}}
- ✗ **VERITROOPER (FAIL):** {"name": "compare_provisions", "parameters": {"provision1": "paragraph (a)", "provision2": "paragraph (d)", "regarding": "confidentiality protection"}}

385. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3531\_00`

- **Question:** When firing from a power circuit, must the firing switch always be locked in the open or 'Off' position?
- **Ground truth:** Yes, it shall be locked in the open or 'Off' position at all times, except when firing.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "locked_in_open_position", "parameters": {"switch": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "When firing from a power circuit, must the firing switch always be locked in the open or ", "answer": "Yes"}}

386. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_03`

HALLUCINATION

- **Question:** What efficiency percentage must HEPA filters meet to qualify under paragraph (3)(i)(C)?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "HEPA filter efficiency", "parameters": {"paragraph": "(3)(i)(C)", "efficiency": "99.97%"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

387. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4215\_01`

- **Question:** How many years passed between the initial approval of the New Mexico State plan and the Assistant Secretary's approval of the revised staffing benchmarks?
- **Ground truth:** The initial approval occurred December 10, 1975, and the revised benchmarks were approved August 11, 1994, which is 19 years later.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "calculate_years_between_dates", "parameters": {"date1": "December 10, 1975", "date2": "August 11, 1994"}}
- **✓ VERITROOPER (PASS):** {"name": "subtract", "parameters": {"a": "1994", "b": "1975"}}

388. FAIL · Calculation · `OSHA\_29CFR\_full\_2024\_4031\_00`

MODEL ERROR

- **Question:** A crane has a rated capacity of 60,000 pounds for the current radius and configuration. When using a suspended personnel platform, what is the maximum total load (platform loaded, including hook, load line and rigging) permitted, except during proof testing?
- **Ground truth:** 30,000 pounds
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "calculate_total_load", "parameters": {"rated_capacity": "60000"}}
- **✗ VERITROOPER (FAIL):** {"name": "calculate_max_total_load", "parameters": {"rated_capacity": "60000"}}

389. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_3938\_02`

MODEL ERROR

- **Question:** Under § 1926.1204(e)(1)(ii), continuous monitoring of entry conditions is required where entrants are working. Under what exception may employers instead use periodic monitoring?
- **Ground truth:** Employers may use periodic monitoring in accordance with paragraph (e)(2) for monitoring an atmospheric hazard if they can demonstrate that equipment for continuously monitoring that hazard is not commercially available.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "periodic monitoring", "parameters": {"paragraph": "(e)(2)"}}
- **✓ VERITROOPER (PASS):** {"name": "periodic_monitoring_exception", "parameters": {"condition": "employers can demonstrate that equipment for continuously monitoring the hazard is not commercially available"}}

390. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_06`

HALLUCINATION

- **Question:** How often must safety devices in a bulk oxygen system be inspected?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "inspect_bulk_oxygen_system", "parameters": {"frequency": "daily"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

391. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2164\_04`

- **Question:** Per the record keeping procedures, for how many years must QC sample analysis results be retained, and for how many years must duplicate copies of monitoring results sent to clients be retained?
- **Ground truth:** QC sample analysis results must be retained for the previous 2 years, while duplicate copies of monitoring results sent to clients must be retained for the previous 5 years.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "recordkeeping_procedures", "parameters": {"years_retained_QC": "2", "years_retained_monitoring": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retention_periods", "parameters": {"qc_sample_analysis_results": "2", "duplicate_copies_of_monitoring_results": "5"}}

392. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3703\_02`

- **Question:** Comparing Structure No. 10 and Structure No. 11, both have a count of 0 - what distinct reasons are given for each?
- **Ground truth:** Structure No. 10 has a count of 0 because the fiber is outside the graticule, while Structure No. 11 has a count of 0 because the fiber crosses the graticule twice.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "structure_count", "parameters": {"SN": "10", "Count": "0", "Explanation": "Fiber outside graticule"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_structure_explanation", "parameters": {"structure_no": "10"}}

393. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_06`

- **Question:** According to appendix A Table B, within how many days must excess cadmium exposure assessed conditions be corrected?
- **Ground truth:** 30 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "assess_and_correct", "parameters": {"issue": "excess cadmium exposure", "days": "30"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_correct_within_days", "parameters": {"category": "B 1"}}

394. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0242\_00`

- **Question:** How frequently does OSHA consider public requests to change the fees in the current fee schedule?
- **Ground truth:** annual
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "OSHA considers public requests to change the fees in the current fee schedule", "parameters": {"frequency": "annually"}}
- ✓ **VERITROOPER (PASS):** {"name": "OSHA considers public requests to change the fees", "parameters": {"frequency": "during its annual review of the fee schedule"}}

395. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_03`

- **Question:** Under paragraph (l)(3)(ii)(C), if the physician determines that medical removal is not necessary after a full medical examination, what must the employer provide until levels fall to or below the thresholds?
- **Ground truth:** The employer must provide biological monitoring semiannually and annual medical examinations until CdU falls to $\leq 3 \mu\text{g/g Cr}$, $\beta 2\text{-M}$ to $\leq 300 \mu\text{g/g Cr}$, and CdB to $\leq 5 \mu\text{g/lwb}$.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_biological_monitoring_and_medical_examinations", "parameters": {"CdU": "3 \u00b5g/g Cr", "CdB": "5 \u00b5g/lwb", "\u03b22-M": "300 \u00b5g/g Cr"}}
- ✗ **VERITROOPER (FAIL):** {"name": "provide_until_levels_fall", "parameters": {"CdU": "3", "\u03b22-M": "300", "CdB": "5"}}

396. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0333\_02`

- **Question:** For an intermittent stabilization system anchor, what is the minimum design wind load per anchor if two anchors share the wind load?
- **Ground truth:** 300 pounds (1334 n).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "minimum_design_wind_load", "parameters": {"load": "300"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_design_wind_load", "parameters": {"anchors": "2"}}

397. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2164\_03`

- **Question:** Comparing the control-limit criterion for CDB/CDU calibration standards with the control-limit criterion for B2MU control samples, what is the measurement error threshold in each case that signals the run has exceeded control limits?
- **Ground truth:** For CDB/CDU calibration, a standard more than $\pm 15\%$ of its theoretical value means the run has exceeded control limits; for B2MU, if the measurement error of one or more control samples is more than 15%, the run exceeds control limits. Both use a 15% threshold.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_control_limit_criteria", "parameters": {"cdb_cdu_threshold": "15", "b2mu_threshold": "15"}}
- ✗ **VERITROOPER (FAIL):** {"name": "control_limit_criterion", "parameters": {"criterion": "CDB/CDU calibration standards"}}

398. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3208\_03`

MODEL ERROR

- **Question:** If electric equipment is located where it would be exposed to physical damage, what requirement applies to the enclosures or guards?
- **Ground truth:** They must be arranged and of such strength as to prevent such damage.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "requirement_for_enclosures_or_guards", "parameters": {"location": "where electric equipment is likely to be exposed to physical damage"}}
- ✗ **VERITROOPER (FAIL):** {"name": "enclosures_or_guards_requirement", "parameters": {"location": "exposed to physical damage"}}

399. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4366\_02`

- **Question:** The conclusion of the hearing is deemed to be 14 days after the date of the ALJ's decision, but under what circumstance does this not apply?
- **Ground truth:** Unless a motion for reconsideration has been filed with the ALJ in the interim; in that case, the conclusion of the hearing is the date the motion for reconsideration is ruled upon or 14 days after a new decision is issued.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "conclusion_of_hearing", "parameters": {"deemed_date": "14 days after the date of the ALJ", "exception": "a motion for reconsideration has been filed with the ALJ in the interim"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_conclusion_of_hearing", "parameters": {"circumstance": "a motion for reconsideration has been filed with the ALJ in the interim"}}

400. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_05`

- **Question:** Standards developed or adopted under a State plan must be at least as effective as those promulgated under which section of the Act, per § 1902.3(c)(1)?
- **Ground truth:** section 6
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 6 of the Act", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "get_section_number", "parameters": {"section": "6"}}

401. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4551\_00`

- **Question:** Under § 1990.106(a)(1), how often must the Secretary request a review of this policy from the Directors of NCI, NIEHS and/or NIOSH?
- **Ground truth:** No later than every three (3) years from the effective date of this part, or from the last general review
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "request_review", "parameters": {"frequency": "every three (3) years"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_frequency_of_review_request", "parameters": {"frequency": "every three (3) years"}}

402. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1084\_01`

- **Question:** Combining the guard sizing rule and the prohibition in paragraph (m)(1), what minimum guard diameter applies to a circular shaper, and what type of device is never acceptable as a substitute?
- **Ground truth:** The diameter of circular shaper guards shall be not less than the greatest diameter of the cutter, and in no case is a warning device of leather or other material attached to the spindle acceptable.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "circular shaper guard size and substitute prohibition", "parameters": {"diameter": "the greatest diameter of the cutter", "substitute": "a warning device of leather or other material attached to the spindle"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_minimum_guard_diameter", "parameters": {"diameter": "greatest diameter of the cutter"}}

403. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3809\_00`

- **Question:** Within how many days of any medical examination must the employer ensure each employee receives a copy of the written medical opinion described in paragraph (k)(6)?
- **Ground truth:** 45 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "receive_written_medical_opinion", "parameters": {"days": "45"}}
- ✓ **VERITROOPER (PASS):** {"name": "ensure_employee_receive_written_medical_opinion", "parameters": {"days": "45"}}

404. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_01`

HALLUCINATION

- **Question:** How frequently must combustible floors be re-wetted to maintain protection during welding operations?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_frequency_of_rewetting_combustible_floors", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

405. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0517\_01`

- **Question:** Combining the equipment requirements in subparagraph (g)(2)(iii) and the composition provisions in subparagraph (g)(2)(iv), what must mixing and packaging equipment be constructed of, and which substances are prohibited from use?
- **Ground truth:** Mixing and packaging equipment shall be constructed of materials compatible with the fuel-ammonium nitrate composition. Prohibited substances include hydrocarbon liquid fuel with flashpoint lower than No. 2 diesel fuel oil (125 °F minimum), crude oil and crankcase oil, and peroxides and chlorates.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "mixing_and_packaging_equipment_requirements", "parameters": {"materials": "compatible with the fuel-ammonium nitrate composition", "prohibited_substances": "["crude oil", "crankcase oil", "peroxides", "chlorates"]"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_equipment_requirements", "parameters": {"subparagraph": "(g)(2)(iii)", "composition_provisions": "(g)(2)(iv)"}}

406. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_05`

HALLUCINATION

- **Question:** How long must the employer retain the written medical report after it is issued?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "retain_written_medical_report", "parameters": {"duration": "30 years"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

407. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0168\_01`

- **Question:** Given the requirements for affidavits in paragraph (d) and the response duty it imposes, plus the relief available in paragraph (e), what happens if a party opposing the motion cannot present essential facts by affidavit?
- **Ground truth:** Paragraph (d) requires the opposing party not to rest on mere allegations or denials but to set forth specific facts showing a genuine issue. However, under paragraph (e), if it appears from the opposing party's affidavits that he cannot for stated reasons present by affidavit facts essential to justify his opposition, the hearing examiner may deny the motion for summary decision, order a continuance to permit affidavits or discovery, or make such other order as is just.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "deny the motion for summary decision or may order a continuance", "parameters": {"reasons stated": "true", "affidavits": "true", "discovery": "true", "order": "just"}}
- ✓ **VERITROOPER (PASS):** {"name": "paragraph", "parameters": {"e": "If a party opposing the motion cannot present essential facts by affidavit, the hearing examiner may deny the motion for summary decision or may order a continuance to permit affidavits to be obtained or discovery to be had or may make such other order as is just."}}

408. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1332\_02`

MODEL ERROR

- **Question:** Combining the log hauls and slips control requirement with the log decks barricade requirement, what positioning protections are required for operators/sawyers relative to the logs?
- **Ground truth:** Per paragraph (vi)(e), log haul controls shall be located so the operator is at all times in the clear of logs, machinery, lines, and rigging (and where control is by lever exposed to incoming logs, the lever shall operate the haul only when moved toward the log slip or log pond); per paragraph (3)(iii), a barricade or other positive stop of sufficient strength to stop any log shall be erected between the sawyer's stand and the log deck.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "log_hauls_and_slips_control_requirement", "parameters": {}}
- ✗ **VERITROOPER (FAIL):** {"name": "log_hauls_and_slips_control", "parameters": {"requirement": "operators/sawyers must be positioned where they will, at all times, be in the clear of logs, machinery, lines, and rigging"}}

409. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2440\_00`

- **Question:** Under paragraph (o), when an affected employee exercises their right to observe MDA exposure monitoring conducted under paragraph (e), and that observation requires entry into an area where respirators are required, what three obligations does the employer have toward the observer?

- **Ground truth:** The employer must (1) provide the observer with the personal protective clothing and equipment or respirators required to be worn by employees working in the area, (2) assure the use of such clothing and equipment or respirators, and (3) require the observer to comply with all other applicable safety and health procedures.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "observation_procedures", "parameters": {"observer_provided_with_respirators": "True", "assure_use_of_respirators": "True", "comply_with_safety_procedures": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_personal_protective_equipment", "parameters": {"equipment": "respirators", "observer": "affected employee", "area": "area where respirators are required"}}

410. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0304\_00`

- **Question:** What is the consequence of supporting the line improperly such that pulling on one section takes up slack in adjacent sections?
- **Ground truth:** It causes the line to fall below the limits specified in paragraph (d)(2)(ii) of this section (i.e., below the required height of not less than 34 inches and not more than 39 inches above the walking-working surface).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "supporting_line_improperly", "parameters": {"consequence": "the line falls below the limits specified in paragraph (d)(2)(ii) of this section"}}
- ✓ **VERITROOPER (PASS):** {"name": "causes_line_to_fall_below_limits", "parameters": {}}

411. PASS · Exception · `OSHA\_29CFR\_full\_2024\_2929\_00`

- **Question:** For the purposes of MDA regulation, materials containing MDA in concentrations greater than 0.1% by weight or volume are covered, but what type of material is excluded from this category?
- **Ground truth:** "Finished articles" (materials other than "finished articles" containing MDA in concentrations greater than 0.1% are covered).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "finished_articles_containing_MDA", "parameters": {"concentration": "0.1%"}}
- ✓ **VERITROOPER (PASS):** {"name": "finished_article_containing_mda", "parameters": {"concentration": "0"}}

412. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_01`

- **Question:** If a rope washer is designed so that water cannot splash the operator, floor, or working surface, is a splash guard still required?
- **Ground truth:** No, splash guards are required on rope washers unless the machine is designed to prevent splashing.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "splash_guard_requirement", "parameters": {"required": "false"}}
- ✗ **VERITROOPER (FAIL):** {"name": "check_splash_guard_requirement", "parameters": {"machine_type": "rope washer", "splash_prevention_design": "True"}}

413. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_03`

HALLUCINATION

- **Question:** What is the minimum diameter required for the core in a cable laid rope?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_core_diameter", "parameters": {"rope_type": "cable_laid_rope"}}
- ✓ **VERITROOPER (PASS):** This information is not provided in the given text snippet.

414. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_06`

- **Question:** In which CFR part are the Secretary's rules concerning Walsh-Healey Public Contracts Act standards published?
- **Ground truth:** 41 CFR part 50-204
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "CFR part", "parameters": {"title": "50-204"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_cfr_part", "parameters": {"act": "Walsh-Healey Public Contracts Act"}}

415. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2261\_00`

MODEL ERROR

- **Question:** If a healthcare professional is responsible for an employee's Hepatitis B vaccination but is not evaluating an exposure incident, what is the employer required to provide them?
- **Ground truth:** A copy of this regulation.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "provide_information_to_healthcare_professional", "parameters": {"employee": "employee", "healthcare_professional": "healthcare professional responsible for the employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_information_to_healthcare_professional", "parameters": {"information": "a copy of this regulation"}}

416. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1593\_01`

- **Question:** On switchboard and control panels exceeding 1.83 m (6.0 ft) in width, an entrance is required at each end unless what conditions apply?
- **Ground truth:** Unless the location of the switchboards and control panels permits a continuous and unobstructed way of exit travel, or unless the work space required in paragraph (h)(5)(i) is doubled.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "entrance_and_access_to_work_space", "parameters": {"width": "1.83", "condition": "the location permits a continuous and unobstructed way of exit travel"}}
- ✓ **VERITROOPER (PASS):** {"name": "entrance_required", "parameters": {"width": "1.83", "condition": "the location of the switchboards and control panels permits a continuous and unobstructed way of exit travel, or the work space required is doubled"}}

417. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0913\_01`

MODEL ERROR

- **Question:** Combining the education requirement for general employees and the training requirement for designated equipment users, how often must each be repeated after the initial point?
- **Ground truth:** Both must be repeated at least annually thereafter - the educational program under (g)(2) at least annually after initial employment, and the equipment training under (g)(4) at least annually after initial assignment.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "training_and_education_frequency", "parameters": {"general_employee_training_frequency": "annually", "designated_equipment_user_training_frequency": "quarterly"}}
- ✓ **VERITROOPER (PASS):** {"name": "frequency_of_education_and_training", "parameters": {"education_frequency": "annually", "training_frequency": "annually"}}

418. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4551\_01`

- **Question:** Under § 1990.106(a)(1), within how many days must the institutes' answer to the Secretary's review request be provided?
- **Ground truth:** one hundred twenty (120) days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "answer_to_secretary", "parameters": {"days": "120"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_answer", "parameters": {"section": "1990.106(a)(1)", "question": "within how many days must the institutes"}}

419. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_07`

HALLUCINATION

- **Question:** What is the cost to the committee for obtaining the NIOSH hazard evaluation reports listed in paragraph (a)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "reimbursement", "parameters": {"c": "cost to the committee"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

420. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_06`

HALLUCINATION

- **Question:** What penalty does an employer face for failing to implement the respiratory protection program required in paragraph (2)(i)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"number": "2", "i": "penalty for failing to implement the respiratory protection program"}}
- ✓ **VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question.

421. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0485\_00`

- **Question:** In an unsprinklered building where sprinklers are installed only to protect spraying areas, must the installation conform to the requirements of § 1910.159?
- **Ground truth:** Yes, the installation shall conform to those standards insofar as they are applicable.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "installation_conformance", "parameters": {"conform": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "conformance_requirement", "parameters": {"building_type": "unsprinklered", "sprinkler_installation": "only to protect spraying areas"}}

422. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_00`

HALLUCINATION

- **Question:** What is the filing fee for the complainant to request a hearing after receiving the findings?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_filing_fee", "parameters": {"complainant_income": "50000"}}
- ✗ **VERITROOPER (FAIL):** {"name": "None", "parameters": {}}

423. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0035\_00`

- **Question:** Under which Secretary's Order was authority delegated to the Assistant Secretary, as cited in § 1902.30?
- **Ground truth:** Secretary's Order 12-71
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Secretary's Order", "parameters": {"number": "12-71", "volume": "36", "page": "8754"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_secretary_order", "parameters": {"order_number": "12-71"}}

424. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4486\_02`

- **Question:** Combining paragraphs (e)(4) and (e)(5), under what two conditions related to the respondent will the Assistant Secretary proceed with the investigation?
- **Ground truth:** The Assistant Secretary will proceed with the investigation if the respondent fails to make a timely response, or if the respondent fails to satisfy the burden set forth in paragraph (e)(4)-that is, fails to demonstrate by clear and convincing evidence that it would have taken the same adverse action in the absence of the protected activity.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "proceed_with_investigation", "parameters": {"condition1": "respondent fails to make a timely response", "condition2": "respondent fails to satisfy the burden set forth in the prior paragraph"}}
- ✓ **VERITROOPER (PASS):** {"name": "investigation_conditions", "parameters": {"condition1": "respondent fails to make a timely response", "condition2": "respondent fails to satisfy the burden set forth in paragraph (e)(4)"}}

425. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0884\_02`

MODEL ERROR

- **Question:** Under the general rule of paragraph (e)(3) and its exception, who normally removes a lockout device, and under what condition may someone else remove it?
- **Ground truth:** Under the general rule of (e)(3), each lockout or tagout device shall be removed by the employee who applied it. Under the exception, when that authorized employee is not available, the device may be removed under the direction of the employer, provided specific procedures and training have been developed, documented, and incorporated into the energy control program, and the employer demonstrates equivalent safety.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "remove_lockout_device", "parameters": {"employee_who_applied": "normally", "exception": "someone else", "condition": "when the authorized employee who applied the device is not available to remove it"}}
- ✓ **VERITROOPER (PASS):** {"name": "remove_lockout_device", "parameters": {"employee": "the employee who applied the device", "condition": "when the authorized employee who applied the lockout or tagout device is not available to remove it, under the direction of the employer"}}

426. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3020\_03`

MODEL ERROR

- **Question:** Per subpart (viii), where is the spill containment program required to originate, and what does it reference in another part of the section?
- **Ground truth:** The spill containment program is part of the employer's safety and health program required in paragraph (b) of this section, and it must be implemented to contain and isolate the entire volume of the hazardous substance being transferred where major spills may occur.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "answer_question", "parameters": {"question": "Per subpart (viii), where is the spill containment program required to originate, and what does it reference in another part"}}

of the section?"}}

- ✖ **VERITROOPER (FAIL):** {"name": "answer_origin_reference", "parameters": {"origin": "the employer", "reference": "paragraph (b) of this section"}}

427. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4215\_03`

- **Question:** Whose interests were served before the revised benchmarks were approved, and what was the same party's role in the original requirement for benchmarks?

- **Ground truth:** The AFL-CIO was served with notice before approval of the revised benchmarks, and the AFL-CIO was the plaintiff in the 1978 Court Order (AFL-CIO v. Marshall) that originally required the compliance staffing benchmarks.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_party_served_and_role", "parameters": {"party": "AFL-CIO", "role": "plaintiff"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_party_info", "parameters": {"party_name": "AFL-CIO", "role": "service"}}

428. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4188\_00`

- **Question:** Under the screw conveyor guarding rule, if a screw conveyor is located under a gin stand and the gin is about to be started, what guarding requirement applies and what additional action must occur per the warning device provision?

- **Ground truth:** A screw conveyor under a gin stand is considered guarded by location (so no additional cover or grating guard is required), but per the warning device provision, an audible signal of sufficient volume to be heard by employees must be sounded each time before starting the gin.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "screw conveyor guarding rule", "parameters": {"gin stand location": "under", "gin start status": "about to be started", "guarding requirement": "considered guarded by location", "additional action": "warning device must sound an audible signal to alert employees"}}

- ✖ **VERITROOPER (FAIL):** {"name": "guarding_requirement", "parameters": {"location": "under a gin stand", "action": "starting the gin"}}

429. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_04`

- **Question:** If a party objects to the admission of evidence, what must they do regarding the grounds of the objection?

- **Ground truth:** They must state briefly the grounds for such objection.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "objection_to_admission_of_evidence", "parameters": {"grounds": "state briefly the grounds for the objection"}}

- ✓ **VERITROOPER (PASS):** {"name": "objection", "parameters": {"grounds": "state briefly the grounds for the objection"}}

430. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_13`

- **Question:** How much MDA was used to spike each sample filter in the retention efficiency study?

- **Ground truth:** 0.814 µg

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_mda_used", "parameters": {"amount_mda": "0", "number_of_filters": "1"}}

- ✓ **VERITROOPER (PASS):** {"name": "calculate_mda_used", "parameters": {"amount": "0", "number_of_filters": "1"}}

431. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4240\_01`

- **Question:** Is the list of developmental steps limited to those items in the published developmental schedule?
- **Ground truth:** No. A developmental step 'includes, but is not limited to' those items listed in the published developmental schedule (or revisions), and also includes items in the plan as approved under section 18(c) and items in the approval decision subject to evaluations.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "developmental step", "parameters": {"includes": "those items listed in the published developmental schedule, or any revisions thereof, for each plan"}}
- ✓ **VERITROOPER (PASS):** {"name": "is_limited_to_published_schedule", "parameters": {"developmental_steps": "[items listed in the published developmental schedule, 'revisions thereto', 'items in the plan as approved under section 18(c) of the Act', 'items in the approval decision which are subject to evaluations']"}}

432. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_00`

- **Question:** What is the minimum undamaged area for a salvaged Class 1, 2, 3, or 4 rubber insulating blanket, in millimeters?
- **Ground truth:** 560 millimeters by 560 millimeters (22 inches by 22 inches)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_minimum_undamaged_area", "parameters": {"class_number": "1", "size_unit": "millimeters"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_minimum_undamaged_area", "parameters": {"class_number": "1", "unit": "millimeters"}}

433. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_09`

- **Question:** What is the SAMA standard number for Laboratory Fume Hoods?
- **Ground truth:** SAMA LF7-1980
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "SAMA", "parameters": {"standard_number": "LF7-1980"}}
- ✓ **VERITROOPER (PASS):** {"name": "SAMA", "parameters": {"standard_number": "LF7-1980"}}

434. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_02`

- **Question:** If the work is covered by subpart V of this part, does the minimum clearance distance determination under paragraph (c)(1) apply?
- **Ground truth:** No; for subpart V work, paragraph (c)(1) does not apply, and instead the minimum approach distances established under § 1926.960(c)(1)(i) apply.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "minimum clearance distance determination under paragraph (c)(1)", "parameters": {"apply": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "apply_minimum_clearance_distance_determination", "parameters": {"covered_by_subpart_V": "True"}}

435. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_08`

HALLUCINATION

- **Question:** What is the maximum period the time for filing may be tolled under applicable case law?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "time for filing a complaint equitably tolled", "parameters": {"days": "180"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

436. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_01`

- **Question:** Under § 1902.3(h), which section of the Intergovernmental Personnel Act of 1970 authorizes the U.S. Civil Service Commission to prescribe standards modifying or superseding the merit system standards?
- **Ground truth:** section 208
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 208 of the Intergovernmental Personnel Act of 1970", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"section_number": "208"}}

437. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3280\_03`

MODEL ERROR

- **Question:** Based on the definitions of 'Unstable objects' and 'Supported scaffold,' why would a barrel be inappropriate as part of a supported scaffold?
- **Ground truth:** A supported scaffold requires platforms supported by rigid support such as poles, legs, uprights, posts, or frames. Unstable objects, which include barrels, may become dislocated and shift and do not constitute a safe base support for scaffolds, so a barrel cannot serve as the rigid base support a supported scaffold needs.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_inappropriate_scaffold_part", "parameters": {"object_name": "barrel"}}
- ✓ **VERITROOPER (PASS):** {"name": "inappropriate_base_support", "parameters": {"object": "barrel", "reason": "Unstable objects do not constitute a safe base support for scaffolds, platforms, or employees."}}

438. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3305\_00`

- **Question:** According to subpart (i) and subpart (4), what is the maximum height in inches that the bottom step may be above the scaffold supporting level for both the stairway-type ladders and the staitowers?
- **Ground truth:** 24 inches (61 cm) for both, since subpart (i) and subpart (4) each specify the bottom step shall not be more than 24 inches above the scaffold supporting level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "stairway-type ladders and staitowers", "parameters": {"maximum height in inches": "24"}}
- ✓ **VERITROOPER (PASS):** {"name": "max_height", "parameters": {"height_in_inches": "24"}}

439. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_3999\_00`

MODEL ERROR

- **Question:** Under what circumstances may an operator-in-training operate the equipment in the listed prohibited circumstances despite not being certified under paragraph (c)?
- **Ground truth:** As provided in the exception under paragraph (b)(3)(v) of this section.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "operator-in-training operation in prohibited circumstances", "parameters": {"circumstance": "multiple-lift rigging operations, except where the operator"}}
- ✓ **VERITROOPER (PASS):** {"name": "operator_in_training_prohibited_circumstances_exception", "parameters": {"circumstance": "as provided in paragraph (b)(3)(v) of this section"}}

440. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4517\_01`

MODEL ERROR

- **Question:** Combining the time-for-filing rule in paragraph (d) with the tolling example it provides, if a complainant mistakenly files with an agency other than OSHA 150 days after the adverse action, is the complaint potentially still timely, and what is the base filing period?
- **Ground truth:** Yes, potentially. Paragraph (d) sets a 180-day filing period, and it states OSHA may consider

the time tolled if a complainant mistakenly files with an agency other than OSHA within 180 days after the alleged adverse action. Since 150 days is within 180 days, tolling may apply.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_filing_deadline", "parameters":

{"days_after_adverse_action": "150", "time_for_filing": "180"}}

- ✔ **VERITROOPER (PASS):** {"name": "calculate_filing_period", "parameters": {"days_after_adverse_action": "150", "tolling_example": "mistakenly files with an agency other than OSHA"}}

441. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_06`

- **Question:** How often must rubber insulating gloves be tested after the first issue?

- **Ground truth:** every 6 months thereafter

- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "test_frequency", "parameters": {"equipment_type": "Rubber insulating gloves", "frequency": "every 6 months"}}

- ✔ **VERITROOPER (PASS):** {"name": "test_frequency", "parameters": {"equipment_type": "Rubber insulating gloves", "frequency": "every 6 months"}}

442. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4502\_00`

MODEL ERROR

- **Question:** Under paragraph (a), what specific rights must OSHA notify the respondent of when notifying it of the filing of the complaint, and what are the two response timeframes associated with those rights set out in paragraphs (b) and (f)?

- **Ground truth:** OSHA must notify the respondent of its rights under paragraphs (b) and (f) of the section and § 1987.110(e). Under paragraph (b), the respondent may, within 20 days of receipt of the notice, submit a written statement and affidavits/documents and request a meeting. Under paragraph (f), the respondent must present its evidence within 10 business days of OSHA's notification (or as soon thereafter as OSHA and the respondent agree).

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "notify_respondent_of_filing", "parameters": {"rights": "rights under paragraphs (b) and (f) of this section and \u00a7 1982.110(e)", "response_timeframes": "[20 days, 20 days]"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_notification_rights", "parameters": {"rights": "paragraph (b) and (f)", "response_timeframes": "[20 days, 10 business days]"}}

443. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_02`

HALLUCINATION

- **Question:** What gauge of sheet steel is required for the exhaust ductwork construction?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "sheet_steel_gauge", "parameters": {"g": "18"}}

- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

444. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4630\_02`

- **Question:** Which U.S. Code provision's interest rate is used to calculate interest on back pay under § 1991.109(d)(1)?

- **Ground truth:** 26 U.S.C. 6621(a)(2)

- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_interest_on_back_pay", "parameters": {"code_section": "26 U.S.C. 6621(a)(2)"}}

- ✔ **VERITROOPER (PASS):** {"name": "get_interest_rate", "parameters": {"usc_provision": "26 U.S.C. 6621(a)(2)"}}

445. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1349\_01`

WRONG CALC/QUOTE

- **Question:** Combining the rules on how the chain saw must be started and how it must be held during operation, what is required regarding the operator's hands?
- **Ground truth:** When starting, drop starting is prohibited and the saw must be started on the ground or where otherwise firmly supported (per (vi)); during operation, the chain saw shall be held with the thumbs and fingers of both hands encircling the handles unless the employer demonstrates a greater hazard is posed by keeping both hands on the saw (per (viii)).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "chain_saw_operation", "parameters": {"hands_required": "both hands encircling the handles during operation"}}
- **✗ VERITROOPER (FAIL):** {"name": "operator_hands_requirement", "parameters": {"hands_position": "encircling the handles", "specific_situation_exception": "the employer demonstrates that a greater hazard is posed by keeping both hands on the chain saw"}}

446. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0494\_02`

MODEL ERROR

- **Question:** If an undercoating operation uses materials not more hazardous than kerosene but is conducted in a garage lacking adequate natural or mechanical ventilation, does the exemption apply?
- **Ground truth:** No, because adequate natural or mechanical ventilation is required for the exemption, so the operation is subject to all spray finishing requirements.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "k", "parameters": {"undercoating_materials": "not more hazardous than kerosene", "ventilation": "inadequate"}}
- **✗ VERITROOPER (FAIL):** {"name": "is_exempt", "parameters": {"is_hazardous": "False", "has_ventilation": "False"}}

447. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1336\_01`

- **Question:** A bandsaw with a crack greater than one-tenth the width of the saw shall not be placed in service unless what conditions are met?
- **Ground truth:** Until the width of the saw is reduced to eliminate the crack, until the cracked section is removed, or crack development is stopped.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "bandsaw_crack_conditions", "parameters": {"conditions": "[the width of the saw is reduced to eliminate the crack, the cracked section is removed, the crack development is stopped]"}}
- **✓ VERITROOPER (PASS):** {"name": "inspect_bandsaw", "parameters": {"crack_width": "one-tenth the width of the saw", "conditions": "[width of saw is reduced to eliminate crack, cracked section is removed, crack development is stopped]"}}

448. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4240\_00`

- **Question:** When the context clearly requires otherwise, do the definitions in § 1955.2 still apply?
- **Ground truth:** No. The definitions apply 'unless the context clearly requires otherwise.'
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Definitions", "parameters": {"a": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "apply_definitions", "parameters": {"context_requires_otherwise": "True"}}

449. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0485\_01`

- **Question:** If deposits build up on sprinklers protecting spraying areas, how often must they be cleaned to be kept free of deposits?
- **Ground truth:** By cleaning daily if necessary.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "cleaning_of_heads", "parameters": {"frequency": "daily if necessary"}}
- ✓ **VERITROOPER (PASS):** {"name": "clean_sprinklers", "parameters": {"frequency": "daily"}}

450. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3403\_01`

- **Question:** Combining the loose gear precaution in paragraph (f) and the approach distance rule in paragraph (o), what are the two specified distances and what does each govern?
- **Ground truth:** Paragraph (f) requires that all loose gear within 100 feet of the place of lifting/depositing the load be secured or removed, while paragraph (o) prohibits any unauthorized person from approaching within 50 feet of the helicopter when the rotor blades are turning.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "helicopter_operations", "parameters": {"distance1": "100", "description1": "loose gear within this distance of the place of lifting the load or depositing the load shall be secured or removed", "distance2": "50", "description2": "no employee shall approach within this distance of the helicopter when the rotor blades are turning"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_distance_governance", "parameters": {"distance1": "100", "governance1": "loose gear within this distance of the place of lifting the load, depositing the load, and all other areas susceptible to rotor downwash shall be secured or removed", "distance2": "50", "governance2": "unauthorized person shall be allowed to approach within this distance of the helicopter when the rotor blades are turning"}}

451. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_02`

- **Question:** What is the NIOSH Publication number for the Occupational Health Guidelines (NIOSH/OSHA)?
- **Ground truth:** NIOSH Pub. No. 81-123
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_niosh_publication_number", "parameters": {"guideline_title": "Occupational Health Guidelines (NIOSH/OSHA)"}}
- ✓ **VERITROOPER (PASS):** {"name": "NIOSH/OSHA", "parameters": {"Pub. No.": "81-123"}}

452. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4630\_03`

- **Question:** What standard of proof must the respondent meet to avoid relief by showing it would have taken the same adverse action absent protected activity?
- **Ground truth:** Clear and convincing evidence
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "clear_and_convincing_evidence", "parameters": {"standard_of_proof": "clear and convincing evidence"}}
- ✓ **VERITROOPER (PASS):** {"name": "show_standard_of_proof", "parameters": {"standard_of_proof": "clear and convincing evidence"}}

453. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0899\_02`

- **Question:** Based on paragraph (c)(1) and paragraph (c)(2), what are the differing training frequency and comprehensiveness requirements for fire brigade leaders compared to general members who do not perform interior structural fire fighting?
- **Ground truth:** Per (c)(1), fire brigade leaders and training instructors must be provided more comprehensive training and education than the general membership. Per (c)(2), all fire brigade members (including the general membership not performing interior structural fire fighting) must be provided training at least annually.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "training_frequency_and_comprehensiveness_requirements", "parameters": {"leaders_comprehensive_training": "more comprehensive than that provided to the general membership of the fire brigade", "leaders_training_frequency": "at least annually", "general_members_training_frequency": "at least quarterly"}}
- ✓ **VERITROOPER (PASS):** {"name": "training_frequency", "parameters": {"member_type": "general member", "interior_structural_fire_fighting": "False"}}

454. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0458\_03`

- **Question:** Considering the requirements for swivel joints in subpart (c) and the requirement in subpart (h) for hose or articulated swivel-joint connections, what dual design considerations must such a transfer system satisfy?
- **Ground truth:** The swivel joint design must ensure the mechanical strength of the joint will not be impaired if the packing material should fail, as by exposure to fire (subpart c), and the hose or articulated swivel-joint pipe connections must be capable of accommodating the combined effects of change in draft and maximum tidal range, with mooring lines kept adjusted to prevent surge stress (subpart h).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "swivel_joints_design", "parameters": {"design_consideration_1": "The design is such that the mechanical strength of the joint will not be impaired if the packing material should fail, as by exposure to fire.", "design_consideration_2": "The combined effects of change in draft and maximum tidal range"}}
- ✓ **VERITROOPER (PASS):** {"name": "design considerations for transfer system", "parameters": {"consideration1": "mechanical strength of the joint will not be impaired if the packing material should fail, as by exposure to fire", "consideration2": "capable of accommodating the combined effects of change in draft and maximum tidal range, and mooring lines shall be kept adjusted to prevent the surge of the vessel from placing stress on the cargo transfer system"}}

455. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0882\_01`

MODEL ERROR

- **Question:** Combining the durability requirement in paragraph (c)(5)(ii)(A)(3) and the training limitation in paragraph (c)(7)(ii)(D), what is required regarding tags used in corrosive environments such as areas where acid and alkali chemicals are handled?
- **Ground truth:** Tags shall not deteriorate when used in corrosive environments such as areas where acid and alkali chemicals are handled and stored, and employees shall be trained that tags and their means of attachment must be made of materials which will withstand the environmental conditions encountered in the workplace.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "requirement_combination", "parameters": {"durability_requirement": "paragraph (c)(5)(ii)(A)(3)", "training_limitation": "paragraph (c)(7)(ii)(D)"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_tag_requirements", "parameters": {"environment": "corrosive", "chemicals": "acid and alkali"}}

456. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3280\_02`

- **Question:** The evidence distinguishes a 'Walkway' from a work level. If a stair tower is used to reach a scaffold platform, and a portion of that platform is a walkway, what combined purpose do these two serve?
- **Ground truth:** The stair tower provides access to scaffold platforms and other elevated points such as floors and roofs, while a walkway is the portion of a scaffold platform used only for access and not as a work level. Combined, both serve as access rather than as work surfaces.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "combined_purpose", "parameters": {"stair_tower": "access", "walkway": "access"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide access", "parameters": {"to_scaffold_platforms_and_other_elevated_points_such_as_floors_and_roofs_as_well_as_a_work_level": "a portion of a scaffold platform used only"}}

for access"}}}

457. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2164\_02`

MODEL ERROR

- **Question:** Under the calibration provision and the internal QC analyses provision, what minimum frequency applies to running a standard versus running internal QC samples relative to compliance samples?
- **Ground truth:** A standard should be run about every fifth sample (to assure calibration), while internal QC samples should be run at a rate of 5% of the compliance samples or at least one set per analysis, whichever is greater.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_frequency", "parameters": {"provision": "calibration", "sample_type": "standard"}}
- **✓ VERITROOPER (PASS):** {"name": "get_minimum_frequency", "parameters": {"standard_frequency": "about every fifth sample", "internal_qc_frequency": "at a rate of 5% of the compliance samples or at least one set of QC samples per analysis of compliance samples, whichever is greater"}}

458. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_07`

- **Question:** Under what condition must the employer remove existing limitations or end special protective measures provided pursuant to a final medical determination?
- **Ground truth:** When a subsequent final medical determination indicates the limitations or special protective measures are no longer necessary.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "removal of other employee special protective measure or limitations", "parameters": {"condition": "when a subsequent final medical determination indicates that the limitations or special protective measures are no longer necessary"}}
- **✓ VERITROOPER (PASS):** {"name": "remove_limitations_or_special_protective_measures", "parameters": {"condition": "when a subsequent final medical determination indicates that the limitations or special protective measures are no longer necessary"}}

459. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_03`

- **Question:** At an establishment where employees are not represented under collective bargaining arrangements, how are nonmanagement members determined?
- **Ground truth:** They shall be determined through procedures devised by the agency which provide for effective representation of all employees.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "determine_nonmanagement_members", "parameters": {"method": "procedures devised by the agency which provide for effective representation of all employees"}}
- **✗ VERITROOPER (FAIL):** {"name": "determine_nonmanagement_members", "parameters": {"representation_type": "not represented under collective bargaining arrangements"}}

460. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1748\_01`

- **Question:** Combining the requirements of paragraph (g)(2) and paragraph (g)(3), what must the employer do regarding the room both during and after an aerosol-generating procedure on a person with suspected or confirmed COVID-19?
- **Ground truth:** During the procedure, per (g)(2), the employer must ensure it is performed in an existing AIIR if available; after the procedure is completed, per (g)(3), the employer must clean and disinfect the surfaces and equipment in the room or area where the procedure was performed.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "perform_aerosol_generating_procedure", "parameters": {"room_requirements": "ensure the procedure is performed in an existing AIIR, if available, and clean and disinfect the surfaces and equipment in the room or area where the procedure was performed"}}
- **✗ VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"question": "Combining the

requirements of paragraph (g)(2) and paragraph (g)(3), what must the employer do regarding the room both during and after an aerosol-generating procedure on a person with suspected or confirmed COVID-19?"}

461. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_04`

HALLUCINATION

- **Question:** What is the assigned protection factor for the tight-fitting PAPR that must be provided when an employee chooses to use one?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Assigned Protection Factor", "parameters": {"Type of respirator 1 2": "2. Powered Air-Purifying Respirator (PAPR); Half mask: 50; Full facepiece: 1,000; Helmet/hood: 4 25/1,000; Loose-fitting facepiece: 25"}}
- **✓ VERITROOPER (PASS):** The assigned protection factor for the tight-fitting PAPR is not explicitly stated in the provided evidence.

462. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3092\_02`

- **Question:** Per paragraph (a)(1), how is the container requirement different for highly viscid materials versus ordinary flammable liquids of 5 gallons or less, and per paragraph (a)(2), what storage areas are prohibited for both?
- **Ground truth:** Highly viscid (extremely hard to pour) materials may be used and handled in original shipping containers rather than approved safety cans or DOT approved containers required for ordinary flammable liquids of 5 gallons or less; under (a)(2) both are prohibited from storage in areas used for exits, stairways, or normally used for the safe passage of people.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "paragraph_a_1_difference_and_prohibited_areas", "parameters": {"highly_viscid_materials": "original shipping containers", "ordinary_flammable_liquids": "approved safety cans or Department of Transportation approved containers", "prohibited_areas": "areas used for exits, stairways, or normally used for the safe passage of people"}}
- **✗ VERITROOPER (FAIL):** {"name": "container_requirement_difference", "parameters": {"material_type": "highly viscid", "quantity": "5 gallons or less"}}

463. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3020\_02`

MODEL ERROR

- **Question:** According to subpart (v) and subpart (vi), what organizational and warning steps apply to drum movement operations?
- **Ground truth:** Site operations must be organized to minimize the amount of drum or container movement (v), and prior to any movement of drums or containers, all exposed employees must be warned of the potential hazards associated with the contents (vi).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "subpart", "parameters": {"v": "Organizational steps", "vi": "Warning steps"}}
- **✗ VERITROOPER (FAIL):** {"name": "organize_and_warn", "parameters": {"step": "drum movement operations"}}

464. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2204\_02`

- **Question:** Combining the referral provision in paragraph (i)(5) described in the General section with the listed examples, when may a physician decline to refer an employee despite persistent blood abnormalities, and what conditions could justify this?
- **Ground truth:** Persistent blood abnormalities must be referred unless the physician has good reason to believe such referral is unnecessary. Examples of conditions making referral unnecessary despite abnormal blood limits are iron or folate deficiency, menorrhagia, or blood loss due to some unrelated medical

abnormality.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "referral_provision", "parameters": {"paragraph": "(i)(5)", "examples": "[iron or folate deficiency, 'menorrhagia', 'blood loss due to some unrelated medical abnormality']"}}
- ✓ **VERITROOPER (PASS):** {"name": "referral_provision", "parameters": {"persistent_blood_abnormalities": "True", "justifying_conditions": "[iron or folate deficiency, 'menorrhagia', 'blood loss due to some unrelated medical abnormality']"}}

465. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1157\_02`

- **Question:** Which type of sprockets is excluded from the enclosure requirement of the sprockets and chains subparagraph?
- **Ground truth:** Manually operated sprockets - this subparagraph does not apply to them.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "sprockets_and_chains_subparagraph", "parameters": {"type": "manually operated sprockets"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_excluded_sprockets", "parameters": {"type": "manually operated"}}

466. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_06`

HALLUCINATION

- **Question:** What is the maximum dollar amount of compensatory damages that may be awarded in the preliminary order?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "issuance_of_findings_and_preliminary_orders", "parameters": {"a": "250,000"}}
- ✓ **VERITROOPER (PASS):** {"name": "None", "parameters": {}}

467. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0494\_00`

MODEL ERROR

- **Question:** If an automobile undercoating spray operation is conducted in a garage with adequate ventilation but uses materials more hazardous than kerosene with solvents having a flash point of 90 °F, is it exempt from spray finishing requirements?
- **Ground truth:** No, it does not conform to the exemption provisions and is therefore subject to all spray finishing requirements of this section.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "k", "parameters": {"Underwriters' Laboratories": "false", "flash point": "90"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_exempt_from_spray_finishing_requirements", "parameters": {"has_adequate_ventilation": "true", "is_kerosene_or_safer": "false", "flash_point": "90"}}

468. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_02`

- **Question:** Under paragraph (l)(3)(ii), if an employee's initial CdU is 4 µg/g Cr but β2-M and CdB are within limits, must the employer reassess occupational exposure, and within what time?
- **Ground truth:** Yes; because CdU exceeds 3 µg/g Cr, the employer must reassess the employee's occupational exposure within two weeks after receipt of biological monitoring results.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "reassess_occupational_exposure", "parameters": {"CdU": "4", "B2M": "within limits", "CdB": "within limits", "time": "two weeks"}}
- ✓ **VERITROOPER (PASS):** {"name": "reassess_occupational_exposure", "parameters": {"CdU": "4", "β2-M": "within limits", "CdB": "within limits", "time": "within two weeks"}}

469. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_07`

HALLUCINATION

- **Question:** How much does it cost to file objections electronically with the Office of Administrative Law Judges?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_filing_fees", "parameters": {"agency": "Office of Administrative Law Judges", "filing_type": "electronic"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

470. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_11`

- **Question:** In the female age correction table (F-2), what is the value at 4000 Hz for age 32?
- **Ground truth:** 10
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"gender": "Female", "age": "32", "frequency": "4000"}}
- **✓ VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "F-2", "frequency": "4000 Hz", "age": "32"}}

471. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_03`

- **Question:** If a petitioner's party participation is denied and they appeal to the Secretary, by when must the Secretary make the final decision?
- **Ground truth:** No later than 20 days following receipt of the appeal.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "appeal_decision", "parameters": {"days": "20"}}
- **✓ VERITROOPER (PASS):** {"name": "get_secretary_decision_deadline", "parameters": {"days_after_receipt_of_appeal": "20"}}

472. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0899\_03`

- **Question:** Per paragraph (c)(4), what must the employer develop regarding special hazards, and per paragraph (c)(1), to whom must the resulting training and education be provided?
- **Ground truth:** Per (c)(4), the employer must develop and make available for inspection written procedures describing actions for situations involving special hazards and include these in the training and education program. Per (c)(1), such training and education must be provided to all fire brigade members commensurate with the duties they are expected to perform.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "develop_and_provide_training", "parameters": {"employer_development": "written procedures that describe the actions to be taken in situations involving the special hazards", "training_recipients": "fire brigade members"}}
- **✓ VERITROOPER (PASS):** {"name": "develop_and_provide_training", "parameters": {"content": "written procedures that describe the actions to be taken in situations involving the special hazards", "target": "all fire brigade members"}}

473. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_07`

HALLUCINATION

- **Question:** How many days does a third-party commercial carrier have to deliver a complaint for it to count as timely?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "time_for_filing", "parameters": {"days": "180"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_days_for_timely_filing", "parameters": {"delivery_method":

"third-party commercial carrier"}}}

474. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_09`

HALLUCINATION

- **Question:** What is the maximum allowable resistance value for the bonding and grounding connections between containers during liquid transfer?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_max_resistance", "parameters": {"is_bonding": "True"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the maximum allowable resistance value for the bonding and grounding connections between containers during liquid transfer.

475. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_05`

MODEL ERROR

- **Question:** If there is a storm or high winds, when may employees work on a scaffold?
- **Ground truth:** Only when a competent person has determined it is safe and the employees are protected by a personal fall arrest system or wind screens.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "work_on_scaffold_during_storm", "parameters": {"wind_speed": "less than 30 mph"}}
- ✓ **VERITROOPER (PASS):** {"name": "work_on_scaffold", "parameters": {"storm_or_high_winds": "True", "competent_person_determination": "True", "personal_fall_arrest_system_or_wind_screens": "True"}}

476. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3863\_02`

- **Question:** On which OSHA form must medical events be reported under the reporting requirements?
- **Ground truth:** OSHA Form No. 200
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "OSHA Form No. 200", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "reporting_requirements", "parameters": {"form_number": "200"}}

477. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_06`

HALLUCINATION

- **Question:** Who is responsible for invoking the blue flag policy referenced in paragraph (c)(9)(i)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "invoke_blue_flag_policy", "parameters": {"policy": "blue flag policy", "paragraph": "(c)(9)(i)"}}
- ✓ **VERITROOPER (PASS):** {"name": "invoke_blue_flag_policy", "parameters": {"paragraph": "(c)(9)(i)", "responsible_party": "unspecified"}}

478. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4294\_02`

- **Question:** Per paragraph (d)(4), within how many days must an agency notify an employee that no hazard is found, and within how many days must a completed health-violation inspection report be made available to that employee?
- **Ground truth:** The agency must notify within 15 days if it finds no reasonable grounds and won't inspect; an inspection/investigation report must be made available within 15 days for safety violations or within 30 days for health violations.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "notify_employee_and_provide_inspection_report", "parameters": {"days_to_notify_employee": "15", "days_to_provide_inspection_report": "30"}}

- ✖ **VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"paragraph": "d", "subparagraph": "4"}}}

479. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4294\_03`

- **Question:** Under paragraph (e)(2) versus (e)(3), what determines whether the Secretary forwards a directly-received report to the agency or may initiate an inspection?
- **Ground truth:** Under (e)(2), where a certified safety and health committee exists, the Secretary shall forward the report to the agency for handling per § 1960.28(d). Under (e)(3), where there is no certified safety and health committee (or when requested by half the members of a committee), the Secretary may initiate an inspection or other appropriate action.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "determine_Secretary_action", "parameters": {"paragraph_e2": "existence of a certified safety and health committee", "paragraph_e3": "request by half the members of a committee"}}
- ✓ **VERITROOPER (PASS):** {"name": "determine_secretary_action", "parameters": {"existence_of_certified_committee": "True", "request_by_half_the_members": "False"}}

480. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3403\_04`

- **Question:** Combining the signal systems requirement in paragraph (n) and the communications requirement in paragraph (r), what must be established between the aircrew/pilot and ground personnel?
- **Ground truth:** Per paragraph (n), signal systems (radio or hand signals, with hand signals as shown in Figure N-1) must be understood and checked in advance of hoisting the load. Per paragraph (r), there must be constant reliable communication between the pilot and a designated ground crew employee acting as a distinctly recognizable signalman during loading and unloading.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "constant reliable communication", "parameters": {"signal systems": "understood and checked in advance of hoisting the load"}}
- ✓ **VERITROOPER (PASS):** {"name": "constant_reliable_communication", "parameters": {"signal_systems": "understood and checked in advance of hoisting the load"}}

481. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_04`

- **Question:** If the rigging may be within the Table A distance during the operation, what requirement applies to the rigging?
- **Ground truth:** Nonconductive rigging must be used.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Nonconductive rigging", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "Nonconductive rigging", "parameters": {"requirement": "Nonconductive rigging if the rigging may be within the Table A of \u00a71926.1408 distance during the operation."}}

482. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0881\_01`

MODEL ERROR

- **Question:** If an energy isolating device has neither a hasp nor a built-in locking mechanism, when is it still considered capable of being locked out?
- **Ground truth:** When lockout can be achieved without the need to dismantle, rebuild, or replace the device or permanently alter its energy control capability.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "is_capable_of_being_locked_out", "parameters": {"hasp": "False", "locking_mechanism": "False"}}
- ✖ **VERITROOPER (FAIL):** {"name": "is_capable_of_being_locked_out", "parameters": {"has_hasp": "false", "has_locking_mechanism": "false", "lockout_without_dismantling": "true"}}

483. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3423\_02`

- **Question:** What is the maximum number of manually controlled jacks/lifting units permitted on one slab?
- **Ground truth:** 14
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "maximum_number_of_manually_controlled_jacks", "parameters": {"n": "14"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_maximum_manually_controlled_jacks", "parameters": {"slab": "one slab"}}

484. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_03`

HALLUCINATION

- **Question:** How frequently must common areas like break rooms be cleaned under this section?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "cleaning_frequency", "parameters": {"frequency": "at least once a day"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

485. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_07`

HALLUCINATION

- **Question:** What qualifications must the licensed physician hold to conduct the examination under this standard?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "licensed_physician_qualifications", "parameters": {"standard": "this standard"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the qualifications of the licensed physician.

486. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2476\_00`

MODEL ERROR

- **Question:** Under the definitions, an 'authorized person' may enter a regulated area as a designated representative of employees to observe monitoring procedures under which paragraph, and a 'regulated area' is demarcated based on exposure exceeding which two limits?
- **Ground truth:** An authorized person may observe monitoring and measuring procedures under paragraph (d) of this section. A regulated area is demarcated where an employee's exposure to airborne MC exceeds or can reasonably be expected to exceed either the 8-hour TWA PEL or the STEL.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"r": "paragraph (r)", "PEL": "PEL", "action level": "action level"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_definition", "parameters": {"term": "authorized person", "paragraph": "d", "limits": "8-hour TWA PEL or the STEL"}}

487. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_07`

- **Question:** Floating type safety cones must be mounted on hinges at least how many inches below the underside of the floor?
- **Ground truth:** 6 inches
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "floating_type_safety_cones", "parameters": {"inches": "6"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Floating type safety cones must be mounted on hinges at least how many inches below the underside of the floor?", "answer": "6 inches"}}

"6"}}}

488. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_3017\_00`

MODEL ERROR

- **Question:** Which employees does the note to paragraph (h) state are NOT required to be monitored?
- **Ground truth:** Employees engaged in site characterization operations covered by paragraph (c) of this section.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_employees_not_required_to_be_monitored", "parameters": {}}
- **✓ VERITROOPER (PASS):** {"name": "note_to_paragraph_h", "parameters": {"employees": "employees engaged in site characterization operations covered by paragraph (c) of this section"}}

489. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4069\_00`

MODEL ERROR

- **Question:** If a tower crane manufacturer does not specify the maximum wind speed for operations, what wind speed limit applies?
- **Ground truth:** The wind must not exceed the speed determined by a qualified person.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_wind_speed_limit", "parameters": {"manufacturer_wind_speed": "null"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_wind_speed_limit", "parameters": {"manufacturer_specified": "false"}}

490. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_07`

- **Question:** Which CFR part contains the Department rules concerning the Service Contract Act safety and health standards?
- **Ground truth:** 29 CFR part 1925
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "CFR part", "parameters": {"title": "1925"}}
- **✓ VERITROOPER (PASS):** {"name": "get_CFR_part", "parameters": {"act": "Service Contract Act", "standards": "safety and health standards"}}

491. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_01`

HALLUCINATION

- **Question:** How many air changes per hour are required for an Airborne infection isolation room under this section?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Airborne infection isolation room", "parameters": {"number of air changes per hour": "12"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

492. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3208\_01`

WRONG CALC/QUOTE

- **Question:** If electric equipment operates at exactly 50 volts, must its live parts be guarded against accidental contact?
- **Ground truth:** Yes, live parts operating at 50 volts or more must be guarded against accidental contact.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "guarding_of_live_parts", "parameters": {"voltage": "50"}}

- ✖ **VERITROOPER (FAIL):** {"name": "guarding_of_live_parts", "parameters": {"voltage": "50"}}

493. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3641\_05`

- **Question:** Given the leak-testing requirement for NPE systems and the leak-testing requirement for glove boxes, what testing method is common to both before use?
- **Ground truth:** Both require smoke-testing for leaks and sealing any leaks prior to use - the NPE before beginning work and at the beginning of each shift, and the glove box prior to each use.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "smoke-testing", "parameters": {"system": "NPE systems and glove boxes"}}
- ✓ **VERITROOPER (PASS):** {"name": "smoke-testing", "parameters": {"system": "NPE systems and glove boxes"}}

494. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0075\_04`

- **Question:** If an employer fails to comply with the posting requirements of paragraphs (a) and (b), what consequence applies?
- **Ground truth:** The employer shall be subject to citation and penalty in accordance with § 1903.15(d).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraphs", "parameters": {"a": "The employer is subject to enforcement action by OSHA", "b": "The employer is required to provide interim protection for exposed employees"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_consequence", "parameters": {"action": "failing to comply with the posting requirements of paragraphs (a) and (b)"}}

495. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_06`

HALLUCINATION

- **Question:** What is the maximum cumulative time of aerosol-generating procedures allowed per patient before additional protections apply?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_aerosol_generating_procedure_time_limit", "parameters": {"procedure_type": "aerosol-generating", "patient_status": "suspected_or_confirmed_COVID-19"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

496. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1964\_01`

- **Question:** Comparing the label required for waste contaminated with vinyl chloride under paragraph (I)(3)(i) with the alternative permitted prior to June 1, 2015 under paragraph (I)(3)(ii), how does the cancer wording differ?
- **Ground truth:** Under (I)(3)(i) the label states 'CONTAMINATED WITH VINYL CHLORIDE MAY CAUSE CANCER'. The alternative permitted prior to June 1, 2015 under (I)(3)(ii) states 'CONTAMINATED WITH VINYL CHLORIDE CANCER-SUSPECT AGENT'. The difference is 'MAY CAUSE CANCER' versus 'CANCER-SUSPECT AGENT'.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_labels", "parameters": {"label1": "CONTAMINATED WITH VINYL CHLORIDE MAY CAUSE CANCER", "label2": "CONTAMINATED WITH VINYL CHLORIDE CANCER-SUSPECT AGENT"}}
- ✓ **VERITROOPER (PASS):** {"name": "compare_labels", "parameters": {"label1": "CONTAMINATED WITH VINYL CHLORIDE MAY CAUSE CANCER", "label2": "CONTAMINATED WITH VINYL CHLORIDE CANCER-SUSPECT AGENT"}}

497. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4188\_01`

MODEL ERROR

- **Question:** Combining the requirements for guard support arch spacing and the warning device, what are the two distinct numeric/operational conditions that must be satisfied when using horizontal bar guards on a screw conveyor in a gin that is about to start?
- **Ground truth:** First, the horizontal bars must be supported by arches that are not more than 8 feet apart; second, an audible warning signal must be sounded each time before starting the gin.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "screw conveyor guard support arch spacing and warning device", "parameters": {"spacing": "not more than 8 feet apart", "warning device": "sounds an alarm that will be recognized by each employee as a warning that the conveyor will start and that can be clearly heard at all points along the conveyor where personnel may be present"}}
- **✓ VERITROOPER (PASS):** {"name": "operational_conditions", "parameters": {"condition1": "arches are not more than 8 feet apart", "condition2": "an audible signal is sounded each time before starting the gin"}}

498. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_02`

HALLUCINATION

- **Question:** How many confirmed human cases are required before a substance can be classified in sub-category 1A?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "A.4.2.2.1", "parameters": {"human evidence": "positive responses at \u2264500 \u03bcg/cm² (Human Repeat Insult Patch Test (HRIPT), Human Maximization Test (HMT))\u2014induction threshold"}}
- **✓ VERITROOPER (PASS):** {}

499. PASS · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_03`

- **Question:** What minimum distance must the detached building for bulk storage of portable containers be from other important buildings?
- **Ground truth:** UNANSWERABLE
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "bulk storage", "parameters": {"distance": "not specified, but must be cut off in a standard manner"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

500. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_03`

- **Question:** If a party receives an adverse ruling, must they make a formal exception to preserve the issue?
- **Ground truth:** No, formal exception to an adverse ruling is not required.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "objection_to_ruling", "parameters": {"requires_formal_exception": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If a party receives an adverse ruling, must they make a formal exception to preserve the issue?", "answer": "No"}}

501. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_00`

- **Question:** What is the NIH Publication number for the NIH Guidelines for the Laboratory use of Chemical Carcinogens?
- **Ground truth:** NIH Pub. No. 81-2385
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_nih_publication_number", "parameters": {"title": "NIH"}}

Guidelines for the Laboratory use of Chemical Carcinogens"]}

- ✓ **VERITROOPER (PASS):** {"name": "get_NIH_Publication_number", "parameters": {"title": "NIH Guidelines for the Laboratory use of Chemical Carcinogens"]}}

502. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3428\_00`

- **Question:** Which provision of subpart CC does NOT apply to hoisting and rigging under § 1926.753(a)?

- **Ground truth:** All provisions of subpart CC apply except for § 1926.1431(a).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1926.1431", "parameters": {"a": "§ 1926.1431(a)"}}

- ✓ **VERITROOPER (PASS):** {"name": "print_provision_exception", "parameters": {"provision": "1926.1431(a)"}}

503. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_06`

- **Question:** The counterweight/ballast requirements in § 1926.1417(aa)(1) apply to what equipment, with what exception?

- **Ground truth:** They apply to equipment other than tower cranes (counterweight/ballast requirements for tower cranes are specified in § 1926.1435(b)(8)).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "counterweight/ballast requirements", "parameters": {"equipment": "equipment other than tower cranes", "exception": "Counterweight/ballast requirements for tower cranes are specified in § 1926.1435(b)(8)"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer", "parameters": {"section": "1926.1417(aa)(1)", "equipment": "equipment other than tower cranes"}}

504. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_05`

- **Question:** According to appendix A Table B, within how many days must a medical examination be administered for an employee in category B?

- **Ground truth:** 90 days

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Medical examination administration", "parameters": {"category": "B", "days": "90"}}

- ✗ **VERITROOPER (FAIL):** {"name": "get_required_actions", "parameters": {"category": "B"}}

505. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3801\_00`

- **Question:** Under paragraph (h)(2)(iii), an employee with reasonably expected exposure above the TWA PEL or STEL may not remove required personal protective clothing from the workplace unless it has been cleaned per paragraph (h)(3)(ii). What method of removing beryllium does (h)(3)(ii) prohibit?

- **Ground truth:** Removing beryllium by blowing, shaking, or any other means that disperses beryllium into the air is prohibited; the clothing cannot be removed from the workplace unless it has been cleaned without such air-dispersing methods.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "prohibited_beryllium_removal_methods", "parameters": {"method": "blowing, shaking, or any other means that disperses beryllium into the air"}}

- ✗ **VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"paragraph": "(h)(3)(ii)", "question": "What method of removing beryllium does (h)(3)(ii) prohibit?"}}

506. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_01`

- **Question:** If a scaffold is not a mobile scaffold and was not designed by a registered professional engineer for movement, may it be moved horizontally while employees are on it?

- **Ground truth:** No, it shall not be moved horizontally while employees are on it.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Scaffolds shall not be moved horizontally while employees are on them, unless they have been designed by a registered professional engineer specifically for such

movement or, for mobile scaffolds, where the provisions of § 1926.452(w) are followed.", "parameters": {}}

- ✖ **VERITROOPER (FAIL):** {"name": "is_scaffold_movement_allowed", "parameters": {"is_mobile_scaffold": "false", "is_designed_for_movement": "false"}}

507. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_06`

MODEL ERROR

- **Question:** How does the OMB control number assigned to the series 1926.1203 through 1926.1213 compare to the one assigned to the series 1926.1402 through 1926.1441?
- **Ground truth:** The 1926.1203-1926.1213 series all carry OMB control No. 1218-0258, while the 1926.1402-1926.1441 series (excluding 1926.1427) carry OMB control No. 1218-0261 - they are different.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "compare_omb_control_numbers", "parameters": {"a": "1218-0258", "b": "1218-0261"}}
- ✔ **VERITROOPER (PASS):** {"name": "compare_omb_control_numbers", "parameters": {"series1_start": "1926.1203", "series1_end": "1926.1213", "series2_start": "1926.1402", "series2_end": "1926.1441"}}

508. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_07`

- **Question:** At what Federal Register citation was § 1902.39(a) suspended indefinitely?
- **Ground truth:** 43 FR 11196
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "Federal Register citation", "parameters": {"citation": "43 FR 11196"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_federal_register_citation", "parameters": {"section": "1902.39(a)", "suspension_date": "January 20, 1978", "citation": "43 FR 11196"}}

509. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_07`

HALLUCINATION

- **Question:** What load capacity must the hanging walkways be designed to support?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_load_capacity", "parameters": {"material": "noncombustible material", "span": "full span of the arch between supports"}}
- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

510. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_01`

- **Question:** What is the NIOSH Publication number for the NIOSH/OSHA Pocket Guide to Chemical Hazards?
- **Ground truth:** NIOSH Pub. No. 85-114
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "NIOSH/OSHA Pocket Guide to Chemical Hazards", "parameters": {"NIOSH Pub": "No. 85-114"}}
- ✔ **VERITROOPER (PASS):** {"name": "NIOSH/OSHA Pocket Guide to Chemical Hazards", "parameters": {"NIOSH Pub. No.": "85-114"}}

511. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4205\_03`

- **Question:** According to paragraphs (c) and (d), what court order established the requirement for compliance staffing benchmarks, and what categories of employers does the plan cover?
- **Ground truth:** The 1978 Court Order in AFL-CIO v. Marshall established the requirement for compliance staffing benchmarks, and the plan covers all private-sector employers and employees (with several notable exceptions) as well as State and local government employers and employees within the State.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "get_court_order_and_covered_employers", "parameters": {}}

{"court_order": "AFL-CIO v. Marshall", "covered_employers": "private-sector employers and employees, with several notable exceptions, as well as State and local government employers and employees, within the State"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_court_order_and_covered_employers", "parameters":

{"court_order": "AFL-CIO v. Marshall", "covered_employers": "all private-sector employers and employees, with several notable exceptions, as well as State and local government employers and employees, within the State"}}

512. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_3130\_07`

MODEL ERROR

- **Question:** What is the minimum liquid capacity (in U.S. gallons) for a container to be considered a portable tank?

- **Ground truth:** more than 60 U.S. gallons

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "definition", "parameters": {"term": "portable tank", "value": "60"}}

- ✗ **VERITROOPER (FAIL):** {"name": "get_minimum_liquid_capacity", "parameters": {"capacity": "60"}}

513. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_12`

- **Question:** In the female age correction table (F-2), what is the value at 4000 Hz for age 27?

- **Ground truth:** 8

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"gender": "Female", "age": "27", "frequency": "4000"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "F-2", "frequency": "4000 Hz", "age": "27"}}

514. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_04`

- **Question:** If an employer wishes to hoist employees while rubber-tired equipment is traveling because there is no less hazardous way to perform the work, is this permitted?

- **Ground truth:** No; the no-less-hazardous-way exception does not apply to rubber-tired equipment.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "traveling", "parameters": {"equipment": "rubber-tired", "exception": "no less hazardous way to perform the work"}}

- ✓ **VERITROOPER (PASS):** {"name": "is_hoisting_permitted", "parameters": {"equipment_type": "rubber-tired", "hazardous_condition": "no less hazardous way to perform the work"}}

515. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1084\_00`

- **Question:** Under paragraph (m), for a double-spindle wood shaper, what guarding and control requirements must both be met regarding the cutting heads and the spindle controls?

- **Ground truth:** The cutting heads must be enclosed with a cage or adjustable guard designed to keep the operator's hand away from the cutting edge (with circular shaper guards no less than the greatest diameter of the cutter), as required by (m)(1); and, per (m)(3), the double-spindle shaper must be provided with a spindle starting and stopping device for each spindle.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "double-spindle_wood_shaper_guarding_and_control_requirements", "parameters": {"cutting_heads": "The cutting heads of each wood shaper, hand-fed panel raiser, or other similar machine not automatically fed, shall be enclosed with a cage or adjustable guard so designed as to keep the operator", "spindle_controls": "All double-spindle shapers shall be provided with a spindle starting and stopping device for each spindle."}}

- ✗ **VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"paragraph": "(m)", "machine_type": "double-spindle wood shaper"}}

516. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0014\_02`

- **Question:** Which section requires that the designated State agency retain overall responsibility for the entire plan?
- **Ground truth:** § 1902.3(b)(3)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section 1902.3", "parameters": {"b": "3"}}
- ✓ **VERITROOPER (PASS):** {"name": "section", "parameters": {"number": "1902.3", "subsection": "(b)(3)"}}

517. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1226\_03`

- **Question:** According to the general rule on the generator water chambers and the specific rule covering charging or carbide removal, when may the water chamber be allowed to not be at proper level?
- **Ground truth:** Generator water chambers must be kept filled to proper level at all times except while draining during the recharging operation; however, whenever repairs are made, the generator is charged, or carbide is removed, the water chamber shall be filled to the proper level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "allow_water_chamber_not_proper_level", "parameters": {"condition": "while draining during the recharging operation"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer", "parameters": {"question": "According to the general rule on the generator water chambers and the specific rule covering charging or carbide removal, when may the water chamber be allowed to not be at proper level?", "text": "while draining during the recharging operation"}}

518. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_08`

HALLUCINATION

- **Question:** How long must a duct or conveyor system remain shut down after welding is completed?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "shutdown_time", "parameters": {"duration": "15"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

519. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_01`

- **Question:** Under § 1902.39(a), no later than how many years following publication of the certification of completion of developmental steps shall the determination of eligibility occur?
- **Ground truth:** 2 years
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "determination_of_eligibility", "parameters": {"years": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "print_answer", "parameters": {"answer": "2"}}

520. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_12`

- **Question:** For a Class 3 glove, what is the DC proof test value?
- **Ground truth:** 4.0
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_dc_proof_test_value", "parameters": {"class_of_glove": "3"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_dc_proof_test_value", "parameters": {"class_of_glove": "3"}}

521. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0694\_00`

- **Question:** Per the potable water requirements in paragraph (n)(1), if drinking water is dispensed from a portable container, what two features must that container have, and additionally what marking requirement applies to it?
- **Ground truth:** The portable container must be capable of being tightly closed and equipped with a tap

(water shall not be dipped from containers), and any container used to distribute drinking water must be clearly marked as to the nature of its contents and not used for any other purpose.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "portable_container_requirements", "parameters": {"features": "capable of being tightly closed, equipped with a tap", "marking_requirement": "clearly marked as to the nature of its contents"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Per the potable water requirements in paragraph (n)(1), if drinking water is dispensed from a portable container, what two features must that container have, and additionally what marking requirement applies to it?", "answer": "The container must be capable of being tightly closed and equipped with a tap. Additionally, the container must be clearly marked as to the nature of its contents."}}

522. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2164\_00`

- **Question:** For a newly-formed laboratory seeking expedited reevaluation, how many proficiency test samples of each analyte must achieve 100% compliance, and over how many rounds of testing are these samples submitted to the CTQ program?
- **Ground truth:** 6 samples of each analyte must achieve 100% compliance, submitted during the first 2 rounds of proficiency testing.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "evaluate_proficiency_test_samples", "parameters": {"num_samples": "6", "num_rounds": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "evaluate_proficiency_test", "parameters": {"compliance_rate": "100", "number_of_samples": "6", "number_of_rounds": "2"}}

523. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3641\_01`

- **Question:** Combining the glove bag specifications and the negative pressure glove bag specifications, what plastic thickness and what additional attachment are required for a negative pressure glove bag system?
- **Ground truth:** The bag must be made of 6 mil thick plastic (from glove bag specifications) and, in addition, must have HEPA vacuum systems or other devices attached to the bag to prevent collapse during removal.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "negative_pressure_glove_bag_systems", "parameters": {"plastic_thickness": "6 mil", "additional_attachment": "HEPA vacuum systems"}}
- ✓ **VERITROOPER (PASS):** {"name": "glove_bag_system", "parameters": {"plastic_thickness": "6", "additional_attachment": "HEPA vacuum systems or other devices to prevent collapse during removal"}}

524. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1984\_03`

- **Question:** Lavatory facilities must comply with which CFR subsections?
- **Ground truth:** § 1910.141(d) (1) and (2)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "lavatory_facilities_compliance", "parameters": {"subsections": "[1910.141(d)(1), 1910.141(d)(2)]"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_CFR_subsections", "parameters": {"subsection": "1910.141(d) (1) and (2)"}}

525. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0040\_00`

- **Question:** Under § 1902.35, for at least how long after publication of the certification in the Federal Register will evaluation for the purposes of an 18(e) determination continue?
- **Ground truth:** at least one year
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "evaluation_for_18e_determination", "parameters": {"duration": "1 year"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_evaluation_period", "parameters": {"section": "1902.35", "certification_event": "publication of the certification in the Federal Register"}}

526. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_07`

- **Question:** What CFR section number is referenced as 'this standard' for the beryllium standard?
- **Ground truth:** 29 CFR 1910.1024
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_cfr_section_number", "parameters": {"text": "this standard", "section": "1926.1124"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_cfr_section_number", "parameters": {"standard_name": "beryllium standard"}}

527. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_02`

- **Question:** Who, other than the gas supplier, may attempt to mix gases in a cylinder?
- **Ground truth:** No person other than the gas supplier may attempt to mix gases in a cylinder.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "None", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "no one", "parameters": {}}

528. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_0598\_01`

MODEL ERROR

- **Question:** Above what container capacity does the minimum 5-foot spacing requirement for underground containers apply?
- **Ground truth:** over 2,000 gallons
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_minimum_spacing_requirement", "parameters": {"capacity": "1000"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"capacity": "2000"}}

529. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_01`

MODEL ERROR

- **Question:** If employees are occupying a personnel platform but the platform is being used over water, are they required to use a personal fall arrest system attached to a structural member within the platform?
- **Ground truth:** No; the personal fall arrest system requirement applies except over water, and when working over or near water the requirements of § 1926.106 apply instead.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "is_personal_fall_arrest_system_required", "parameters": {"over_water": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "fall_protection", "parameters": {"over_water": "true"}}

530. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0564\_02`

MODEL ERROR

- **Question:** May an LP-Gas-fueled vehicle be parked near an open pit?
- **Ground truth:** Not unless the pit is adequately ventilated.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "check_parking_proximity", "parameters": {"near_open_pit": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_parking_near_open_pit", "parameters": {"is_adequately_ventilated": "True"}}

531. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_04`

- **Question:** Which section of the Walsh-Healey Public Contracts Act establishes the general statutory duty regarding unsanitary or hazardous working conditions?
- **Ground truth:** Section 1(e)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Walsh-Healey Public Contracts Act", "parameters": {"section": "1(e)"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"act_name": "Walsh-Healey Public Contracts Act", "keyword": "unsanitary or hazardous working conditions"}}

532. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4344\_03`

- **Question:** What is the citation for NLRB v. Lamar Creamery?
- **Ground truth:** 246 F. 2d 8 (5th Cir., 1957)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "citation", "parameters": {"volume": "246", "reporter": "F. 2d", "page": "8"}}
- ✓ **VERITROOPER (PASS):** {"name": "citation", "parameters": {"case_name": "NLRB v. Lamar Creamery", "volume": "246", "reporter": "F. 2d", "page": "8", "court": "5th Cir.", "year": "1957"}}

533. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3948\_02`

- **Question:** If an articulating/knuckle-boom crane that otherwise delivers material is used to hold and stabilize material in place while it is attached to the structure, does the material-delivery exclusion apply?
- **Ground truth:** No; the exclusion does not apply when the crane is used to hold, support or stabilize material to facilitate a construction activity.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exclusion_applies", "parameters": {"exclusion_type": "material-delivery", "crane_use": "hold-and-stabilize"}}
- ✓ **VERITROOPER (PASS):** {"name": "exclusion_applies", "parameters": {"exclusion_type": "material-delivery", "crane_usage": "hold and stabilize material in place while it is attached to the structure"}}

534. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3110\_02`

- **Question:** Where dispensing is from a floating structure, what does the piping rule allow between shore piping and floating-structure piping, and what valve requirement applies at the point where that flexible hose is attached?
- **Ground truth:** Per paragraph (k)(3)(i), suitable lengths of oil-resistant flexible hose may be employed between the shore piping and the floating structure piping. Per paragraph (k)(3)(ii), a readily accessible shut-off valve must be provided at the shore end of each pipeline adjacent to the point where the flexible hose is attached.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "piping_rule", "parameters": {"shore_piping": "suitable lengths of oil-resistant flexible hose", "floating_structure_piping": "shore piping", "valve_requirement": "readily accessible valve to shut off the supply from shore"}}
- ✓ **VERITROOPER (PASS):** {"name": "piping_rule", "parameters": {"dispensing_location": "floating structure", "shore_piping": "True", "floating_structure_piping": "True", "flexible_hose": "True", "valve_requirement": "readily accessible valve to shut off the supply from shore"}}

535. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3948\_01`

- **Question:** If a digger derrick is used for augering holes for poles carrying electric lines, but the use does not comply with all provisions of subpart V, is it eligible for the (c)(4) exclusion?
- **Ground truth:** No; to be eligible for the exclusion, digger-derrick use in work subject to subpart V must comply with all of the provisions of that subpart.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_excluded", "parameters": {"x": "False"}}

- ✓ **VERITROOPER (PASS):** {"name": "is_eligible_for_exclusion", "parameters": {"equipment": "digger derrick", "use": "augering holes for poles carrying electric lines", "complies_with_subpart_v": "False"}}

536. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_01`

HALLUCINATION

- **Question:** How often must the air compressor used for abrasive-blasting respirators be inspected for harmful contaminants?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "check_air_compressor", "parameters": {"frequency": "periodically"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

537. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4269\_03`

- **Question:** Paragraph (a) states the purpose of adapting private-employee plan requirements to public employees only; paragraph (b) notes considerations in doing so. What two adjustments does paragraph (b) identify as flowing from differences between public and private employment?
- **Ground truth:** Paragraph (b) identifies that (1) a system of monetary penalties applicable to violations may not always be the most appropriate method of achieving compliance for public employers, and (2) the lack of Federal enforcement authority over public employers requires certain adjustments of private employer plan procedures when adapting them to plans covering only public employees.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "adjustments_identified", "parameters": {"a": "monetary penalties", "b": "Federal enforcement authority"}}
- ✗ **VERITROOPER (FAIL):** {"name": "identify_adjustments", "parameters": {"paragraph": "b", "differences": "public and private employment"}}

538. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2838\_02`

- **Question:** What is the date of IME Pub. No. 2, American Table of Distances for Storage of Explosives?
- **Ground truth:** Jun. 5, 1964
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IME Pub. No. 2", "parameters": {"date": "Jun. 5, 1964"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_date", "parameters": {"pub_no": "IME Pub. No. 2, American Table of Distances for Storage of Explosives"}}

539. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1597\_01`

- **Question:** Grounding-type receptacles must be installed only on circuits of the voltage class and current for which they are rated, except in what case?
- **Ground truth:** Except as provided in Table S-4 and Table S-5.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_grounding_type_receptacles", "parameters": {"case": "as provided in Table S-4 and Table S-5"}}
- ✓ **VERITROOPER (PASS):** {"name": "Table S-4 and Table S-5", "parameters": {}}

540. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0881\_00`

- **Question:** If an affected employee's duties are expanded to include performing servicing or maintenance covered under this section, what is that employee's status?
- **Ground truth:** That affected employee becomes an authorized employee.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Authorized employee", "parameters": {"An affected employee becomes an authorized employee when that employee": "performing servicing or maintenance covered under"}}

this section"]}

- ✖ **VERITROOPER (FAIL):** {"name": "employee_status", "parameters": {"duties": "performing servicing or maintenance covered under this section"}}

541. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0014\_01`

- **Question:** Which subsections containing criteria and indices of standards effectiveness apply to the public employee program, alongside § 1902.3(c)?

- **Ground truth:** §§ 1902.4(a) and (b)

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "apply_criteria_and_indices", "parameters": {"subsection_1": "1902.3(c)", "subsection_2": "1902.4(a)", "subsection_3": "1902.4(b)"}}

- ✓ **VERITROOPER (PASS):** {"name": "subsection", "parameters": {"number": "1902.4", "letters": "a and b"}}

542. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3985\_01`

- **Question:** For which equipment are boom stops NOT required?

- **Ground truth:** Derricks and hydraulic booms.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "boom stops", "parameters": {"equipment": "derricks and hydraulic booms"}}

- ✓ **VERITROOPER (PASS):** {"name": "equipment_exception", "parameters": {"equipment": "derricks and hydraulic booms", "device": "boom stops"}}

543. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_01`

- **Question:** For a Class 2 equipment, what is the AC rms proof-test voltage (in V)?

- **Ground truth:** 20,000

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_proof_test_voltage", "parameters": {"class_of_equipment": "2", "voltage_type": "AC rms"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_proof_test_voltage", "parameters": {"class_of_equipment": "2", "voltage_type": "AC rms"}}

544. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0694\_03`

MODEL ERROR

- **Question:** Given the general toilet requirement in (n)(3)(i) and the exception in (n)(3)(iv), would a mobile crew of 30 employees be required to provide toilet facilities on site if they have transportation readily available to nearby toilets?

- **Ground truth:** No. Although 30 employees would normally require one toilet seat and one urinal per 40 employees under the table, the requirements for sanitation facilities do not apply to mobile crews having transportation readily available to nearby toilet facilities.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "is_toilet_facility_required", "parameters": {"number_of_employees": "30"}}

- ✖ **VERITROOPER (FAIL):** {"name": "toilet_facilities_required", "parameters": {"number_of_employees": "30", "transportation_available": "True"}}

545. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3208\_00`

- **Question:** If live parts of electric equipment operate at 50 volts or more and are guarded by partitions with openings, but those openings are sized such that persons could bring conducting objects into contact with the live parts, does this satisfy the guarding requirement?

- **Ground truth:** No, openings in partitions or screens must be sized and located so persons are not likely to bring conducting objects into contact with the live parts.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "check_guarding_requirement", "parameters": {"voltage": "50", "guarding_method": "partitions with openings", "opening_size": "large enough to bring conducting objects into contact"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_guarding_requirement", "parameters": {"voltage": "50", "guarding_method": "partitions with openings", "opening_size": "large enough to bring conducting objects into contact"}}

546. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_02`

- **Question:** If a boiling tank is located so the operator can clearly see its contents from the floor, is an emergency shutoff valve controlled from a splash-safe point required?
- **Ground truth:** No, the emergency shutoff valve requirement applies only when the operator cannot see the contents from the floor or working area.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "open_tanks_and_vats_for_mixing_and_storage_of_hot_or_corrosive_liquids", "parameters": {"require_shutoff_valve": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "is_emergency_shutoff_valve_required", "parameters": {"operator_can_see_contents": "true"}}

547. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_05`

HALLUCINATION

- **Question:** What is the required tolerance range for the limit switch beyond the stated 30 degrees each way?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "limit_switch_tolerance", "parameters": {"angle": "45"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

548. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4205\_00`

WRONG CALC/QUOTE

- **Question:** Given the dates of initial and final approval in paragraphs (a) and (b), how many years elapsed between the Tennessee State plan's initial approval and its final approval?
- **Ground truth:** Initial approval was July 5, 1973, and final approval was July 22, 1985, so approximately 12 years elapsed.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_duration", "parameters": {"start_date": "July 5, 1973", "end_date": "July 22, 1985"}}
- ✗ **VERITROOPER (FAIL):** {"name": "years_between", "parameters": {"start_date": "July 5, 1973", "end_date": "July 22, 1985"}}

549. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4313\_01`

- **Question:** Combining paragraphs (a) and (b) of § 1960.59, what type of training does paragraph (a) provide to employees generally that paragraph (b) does not explicitly require for employee representatives?
- **Ground truth:** Paragraph (a) requires that general employee training inform employees of the agency occupational safety and health program, with emphasis on their rights and responsibilities. Paragraph (b), which addresses employee representatives, focuses instead on introductory and specialized courses to enable them to ensure safe working conditions and assist in conducting workplace safety and health inspections, and does not explicitly include the rights-and-responsibilities program emphasis stated in paragraph (a).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "training_of_employees_and_employee_representatives", "parameters": {"employee_type": "employees generally", "training_type": "specialized job safety and health"}}

training"]}

- ✖ **VERITROOPER (FAIL):** {"name": "training_type", "parameters": {"type": "general employee training", "paragraph": "paragraph (a)"}}

550. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3390\_01`

- **Question:** Combining paragraph (i) and paragraph (ii), what is the minimum distance a control line must be erected from the leading edge, and does it differ between general work and erecting precast concrete members?

- **Ground truth:** The minimum distance is 6 feet (1.8 m) in both cases. Paragraph (i) requires not less than 6 feet for general use, and paragraph (ii) also requires not less than 6 feet when erecting precast concrete members, so the minimum does not differ.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "control_line_distance", "parameters": {"distance": "6", "max_distance_general": "25", "max_distance_precast": "60"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_distance", "parameters": {"paragraph": "i"}}

551. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_0003\_03`

MODEL ERROR

- **Question:** What Statute at Large citation is given for the Intergovernmental Personnel Act of 1970 in § 1902.3(h)?

- **Ground truth:** 84 Stat. 1915

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_statute_at_large_citation", "parameters": {"section": "1902.3(h)", "act": "Intergovernmental Personnel Act of 1970"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_statute_at_large_citation", "parameters": {"section": "1902.3(h)", "act": "Intergovernmental Personnel Act of 1970"}}

552. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4520\_00`

- **Question:** When a timely objection is filed, all provisions of the preliminary order are stayed except for what?

- **Ground truth:** Except for the portion requiring preliminary reinstatement, which will not be automatically stayed (it is effective immediately upon the respondent's receipt of the findings and preliminary order, regardless of any objections).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "objections_to_findings_and_preliminary_order", "parameters": {"provisions_stayed": "all provisions of the preliminary order except for the portion requiring preliminary reinstatement"}}

- ✓ **VERITROOPER (PASS):** {"name": "identify_exception", "parameters": {"category": "provisions of the preliminary order that are not stayed when a timely objection is filed", "exceptions": "[portion requiring preliminary reinstatement]"}}

553. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0227\_00`

- **Question:** For a steel tank used both underground for flammable liquids and above ground for combustible liquids, which two UL standards apply and what single OSHA section are both approved for?

- **Ground truth:** UL 58-61 (Steel Underground Tanks) and UL 142-68 (Steel Above Ground Tanks); both are IBR approved for § 1910.106(b)(1)(iii)(a)(1).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "UL standards for steel tanks", "parameters": {"UL 58-61": "Steel Underground Tanks for Flammable and Combustible Liquids", "UL 142-68": "Steel Above Ground Tanks for Flammable and Combustible Liquids", "OSHA section": "1910.106(b)(1)(iii)(a)(1)"}}

- ✓ **VERITROOPER (PASS):** {"name": "UL standards", "parameters": {"UL 58-61": "Steel Underground Tanks for Flammable and Combustible Liquids", "UL 142-68": "Steel Above Ground Tanks for Flammable and Combustible Liquids"}}

554. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2190\_01`

- **Question:** If monitoring reveals employee exposure above the TWA, but the employer can otherwise alter the schedule, what condition must be met to switch from every six months to annual monitoring?
- **Ground truth:** Two consecutive measurements taken at least 7 days apart must indicate exposure has decreased to the TWA or below, but is at or above the action level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "alter_monitoring_schedule", "parameters": {"current_schedule": "every six months", "new_schedule": "annually", "condition": "two consecutive measurements taken at least 7 days apart indicate that employee exposure has decreased to or below the 8-hour TWA, but is at or above the action level"}}
- ✓ **VERITROOPER (PASS):** {"name": "alter_monitoring_schedule", "parameters": {"current_schedule": "every six months", "new_schedule": "annually", "condition": "two consecutive measurements taken at least 7 days apart indicate that the employee exposure has decreased to the TWA or below, but is at or above the action level"}}

555. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0734\_00`

- **Question:** According to the procedures for quantitative fit testing in subpart (ii) and the post-testing requirement in subpart (iii), after installing a permanent sampling probe onto a surrogate facepiece to allow sampling inside the breathing zone, what must be done before that facepiece can be used in the workplace?
- **Ground truth:** Any modifications to the respirator facepiece for fit testing must be completely removed, and the facepiece restored to its NIOSH-approved configuration, before it can be used in the workplace.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "restore_facepiece", "parameters": {"facepiece": "surrogate facepiece"}}
- ✓ **VERITROOPER (PASS):** {"name": "restore_facepiece", "parameters": {"modification": "permanent sampling probe"}}

556. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_4398\_03`

MODEL ERROR

- **Question:** What is the effect of OSHA's approval of a settlement reached by the respondent and the complainant?
- **Ground truth:** It demonstrates OSHA's consent and achieves the consent of all three parties.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA", "parameters": {"approval": "settlement reached by the respondent and the complainant"}}
- ✗ **VERITROOPER (FAIL):** {"name": "settlement_effect", "parameters": {"settlement_approval": "OSHA"}}

557. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_03`

- **Question:** If a factory-produced boom-mounted personnel platform has an original-equipment winch, under what conditions may a load be hoisted by that winch while employees occupy the platform?
- **Ground truth:** Only where the load on the winch line does not exceed 500 pounds and does not exceed the rated capacity of the winch and platform.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "load_hoisting_conditions", "parameters": {"load_weight": "less than or equal to 500 pounds", "winch_capacity": "within the rated capacity of the winch and platform"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_load_hoisting_conditions", "parameters": {"load_weight": "less than or equal to 500 pounds", "winch_capacity": "rated capacity of the winch and platform"}}

558. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4084\_00`

- **Question:** If a tower crane is tied to or braced by the structure, from where must the structural bolts be inspected for loose or dislodged condition?
- **Ground truth:** Those above the upper-most brace support.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "inspect_structural_bolts", "parameters": {"location": "above the upper-most brace support"}}
- ✓ **VERITROOPER (PASS):** {"name": "inspect_structural_bolts", "parameters": {"location": "above the upper-most brace support"}}

559. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_06`

- **Question:** What is the ANSI standard number for Fundamentals Governing the Design and Operation of Local Exhaust Systems?
- **Ground truth:** ANSI Z 9.2-1979
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ANSI standard number", "parameters": {"standard number": "Z9.2-1979"}}
- ✓ **VERITROOPER (PASS):** {"name": "ANSI", "parameters": {"Z": "9.2-1979"}}

560. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0742\_00`

- **Question:** Under the provisions governing respirator repairs, if a technician needs to adjust a regulator or alarm, what two conditions must be satisfied regarding who performs the work and what parts are used?
- **Ground truth:** Per subpart (iii), reducing and admission valves, regulators, and alarms shall be adjusted or repaired only by the manufacturer or a technician trained by the manufacturer; and per subpart (i), only the respirator manufacturer's NIOSH-approved parts designed for the respirator may be used.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "respirator repairs", "parameters": {"technician": "manufacturer or a technician trained by the manufacturer", "parts": "NIOSH-approved parts designed for the respirator"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_repair_conditions", "parameters": {"component": "regulator or alarm"}}

561. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0884\_00`

- **Question:** Under paragraph (e)(2)(ii) and paragraph (e)(3), after a lockout device has been removed by the employee who applied it, what must happen before the machine is started?
- **Ground truth:** Per (e)(3), the lockout or tagout device must be removed from each energy isolating device by the employee who applied it; then per (e)(2)(ii), after the device has been removed and before the machine is started, affected employees must be notified that the device(s) have been removed.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "notify_affected_employees", "parameters": {"message": "lockout device has been removed"}}
- ✓ **VERITROOPER (PASS):** {"name": "notify_affected_employees", "parameters": {"message": "lockout device has been removed"}}

562. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3162\_01`

- **Question:** If an air receiver is installed after the effective date of these regulations, what code must it be constructed in accordance with?
- **Ground truth:** The 1968 edition of the A.S.M.E. Boiler and Pressure Vessel Code Section VIII.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "new_and_existing_equipment", "parameters": {"i": "All new air receivers installed after the effective date of these regulations shall be constructed in accordance with the 1968 edition of the A.S.M.E. Boiler and Pressure Vessel Code Section VIII"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_construction_code", "parameters": {"installation_date": "after the effective date of these regulations"}}

563. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_01`

HALLUCINATION

- **Question:** What is the penalty for failing to agree on voice signals prior to beginning operations under § 1926.1421?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "penalty_for_failure_to_agree_on_voice_signals", "parameters": {}}
- **✓ VERITROOPER (PASS):** I don't have enough information to answer the question. The provided evidence does not mention penalties for failing to agree on voice signals prior to beginning operations under § 1926.1421.

564. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_04`

HALLUCINATION

- **Question:** How far apart must flag signals be spaced when protecting men during switching operations?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "flag signals spacing", "parameters": {"distance": "50 feet"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

565. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_04`

HALLUCINATION

- **Question:** Who pays for the powered air-purifying respirator when an employee chooses one over a negative-pressure respirator?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "provided by the employer", "parameters": {}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question of who pays for the powered air-purifying respirator.

566. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_10`

- **Question:** What is the recommended air volume for MDA sampling?
- **Ground truth:** 100 L
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "recommended air volume", "parameters": {"volume": "100"}}
- **✓ VERITROOPER (PASS):** {"name": "get_recommended_air_volume", "parameters": {"substance": "MDA"}}

567. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_04`

- **Question:** In the worked example, what was the employee's hearing threshold at 4000 Hz in the most recent audiogram (age 32)?
- **Ground truth:** 25
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_hearing_threshold", "parameters": {"age": "32", "frequency": "4000"}}
- **✓ VERITROOPER (PASS):** {"name": "get_hearing_threshold", "parameters": {"age": "32", "frequency": "4000"}}

568. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_09`

HALLUCINATION

- **Question:** How long must an employer retain records of respirator selection decisions made under this section?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "retain_records", "parameters": {"duration": "30 years"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

569. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_09`

HALLUCINATION

- **Question:** How many wraps are required for a strand laid grommet-hand tucked sling made from braided wire rope instead of strand?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_wraps", "parameters": {"sling_type": "strand laid grommet-hand tucked", "wire_rope_type": "braided"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

570. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_06`

HALLUCINATION

- **Question:** Which regulatory agency reviews the weight-of-evidence determination for respiratory sensitizer classification?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "A.4.2.1.1.1", "parameters": {"Effects seen in either humans or animals will normally justify classification in a weight of evidence approach for respiratory sensitizers.": "Substances may be allocated to one of the two sub-categories 1A or 1B using a weight of evidence approach in accordance with the criteria given in Table A.4.1 and on the basis of reliable and good quality evidence from human cases or epidemiological studies and/or observations from appropriate studies in experimental animals.": "US Environmental Protection Agency (EPA)"}}
- **✓ VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question.

571. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3172\_01`

MODEL ERROR

- **Question:** Under what specific conditions must a vertical belt running over a lower pulley more than 7 feet above the floor be guarded at the bottom, per paragraph (e)(4) read with the cross-referenced conditions?
- **Ground truth:** Per paragraph (e)(4), it must be guarded at the bottom like horizontal overhead belts if the conditions in paragraphs (e)(2)(ii)(a) and (c) apply - that is, if it is located over passageways or work places and traveling 1,800 feet or more per minute (condition (a)), and if the belt is 8 inches or more in width (condition (c)).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "guard_vertical_belt", "parameters": {"distance": "7", "conditions": "conditions are as stated in paragraphs (e)(2)(ii) (a) and (c) of this section"}}
- **✗ VERITROOPER (FAIL):** {"name": "guard_vertical_belt", "parameters": {"height": "more than 7 feet", "conditions": "conditions are as stated in paragraphs (e)(2)(ii)(a) and (c) of this section"}}

572. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3038\_02`

- **Question:** If a current employee's previous work experience shows they have had training equivalent to the initial 24-hour training, must they still complete the initial training?
- **Ground truth:** No, they are considered as meeting the initial training requirements, but they must still receive eight hours of refresher training annually.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "current_employee_training", "parameters": {"hours": "0"}}
- ✗ **VERITROOPER (FAIL):** {"name": "current_employee_training", "parameters": {"previous_experience": "equivalent to initial 24-hour training"}}

573. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_4398\_02`

- **Question:** If a complaint is withdrawn because of settlement, what must occur as a consequence?
- **Ground truth:** The settlement must be submitted for approval in accordance with paragraph (d) of this section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "submit_for_approval", "parameters": {"settlement": "settlement agreement"}}
- ✓ **VERITROOPER (PASS):** {"name": "settlement approval", "parameters": {"settlement": "submitted for approval in accordance with paragraph (d) of this section"}}

574. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2250\_01`

- **Question:** A facility uses Universal Precautions, but a specimen container is now leaving the facility for shipping. Is labeling or color-coding required?
- **Ground truth:** Yes, labeling or color-coding in accordance with paragraph (g)(1)(i) is required when such specimens/containers leave the facility.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_labeling_or_color_coding_required", "parameters": {"universal_precautions": "True", "leaving_facility": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "labeling_or_color_coding_required", "parameters": {"universal_precautions": "True", "leaving_facility": "True"}}

575. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4060\_00`

- **Question:** When hoisting an employee in a shaft operation, if the employer can demonstrate that use of a personnel platform is infeasible, what must be used?
- **Ground truth:** A boatswain's chair must be used.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "boatswain", "parameters": {"platform": "chair"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_hoisting_requirement", "parameters": {"operation": "shaft", "platform_feasibility": "False"}}

576. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0333\_04`

- **Question:** If building anchors have different spacing than the suspension wire rope, what must one building anchor and stabilizer tie be capable of sustaining?
- **Ground truth:** One building anchor and stabilizer tie shall be capable of sustaining the wind loads.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "sustaining_wind_loads", "parameters": {"anchor_spacing": "different", "wind_loads": "anticipated horizontal and vertical loads from winds specified for roof storage design"}}
- ✓ **VERITROOPER (PASS):** {"name": "building_anchor_capability", "parameters": {"spacing": "different", "capability": "wind loads"}}

577. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_04`

- **Question:** How often must live-line tools used for primary employee protection be removed from service for examination, cleaning, repair, and testing?
- **Ground truth:** Every 2 years
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "live-line tools used for primary employee protection", "parameters": {"frequency": "every 2 years"}}
- ✓ **VERITROOPER (PASS):** {"name": "remove_from_service", "parameters": {"frequency": "every 2 years"}}

578. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3092\_00`

- **Question:** Under paragraph (a)(1), for what maximum quantity of flammable liquids may approved safety cans or DOT approved containers be required, and under the same paragraph, for what maximum quantity may the original container be used for storage, use and handling?
- **Ground truth:** Approved safety cans or DOT approved containers are required for quantities of 5 gallons or less, while the original container may be used for quantities of one gallon or less.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph_a_1", "parameters": {"quantity_safety_cans": "5 gallons", "quantity_original_container": "1 gallon"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_quantity_requirements", "parameters": {"quantity": "5", "container_type": "approved safety cans or DOT approved containers"}}

579. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_05`

HALLUCINATION

- **Question:** Which OSHA office is designated to receive and process the inspection requests described in paragraph (a)(2)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Office", "parameters": {"office": "Area Office"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer this question.

580. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_08`

HALLUCINATION

- **Question:** What is the maximum permitted length of exhaust ductwork before a booster fan must be installed?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_ductwork_length", "parameters": {"ductwork_diameter": "12", "booster_fan_interval": "20"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

581. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_3034\_01`

- **Question:** A temporary workplace has 400 employees (more than 200). How many urinals are required at minimum per Table D-65.2?
- **Ground truth:** 8
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_urinals_required", "parameters": {"number_of_employees": "400"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_urinals_required", "parameters": {"number_of_employees": "400"}}

582. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3852\_00`

- **Question:** Under what condition may employees enter lunchroom facilities while wearing protective work clothing or equipment?
- **Ground truth:** Only if surface cadmium has been removed from the clothing and equipment by HEPA vacuuming or some other method that removes cadmium dust without dispersing it.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "enter_lunchroom_facilities", "parameters": {"condition": "surface lead dust has been removed by vacuuming, downdraft booth, or other cleaning method"}}
- ✓ **VERITROOPER (PASS):** {"name": "enter_lunchroom_facilities", "parameters": {"condition": "surface cadmium has been removed from the clothing and equipment by HEPA vacuuming or some other method that removes cadmium dust without dispersing it"}}

583. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3327\_00`

- **Question:** What is the minimum force a toeboard must withstand without failure when applied in any downward or horizontal direction?
- **Ground truth:** 50 pounds (222 n)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "toeboard_force", "parameters": {"force": "50"}}
- ✓ **VERITROOPER (PASS):** {"name": "min_force", "parameters": {"force": "50"}}

584. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3428\_04`

- **Question:** Pre-planned routes must ensure no employee works directly below a suspended load - which employees are excepted from this requirement?
- **Ground truth:** Employees engaged in the initial connection of the steel, and employees necessary for the hooking or unhooking of the load.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "pre-planned_routes", "parameters": {"employees": "employees engaged in the initial connection of the steel; or employees necessary for the hooking or unhooking of the load"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_pre_planned_routes", "parameters": {"employees_excepted": "[employees engaged in the initial connection of the steel, 'employees necessary for the hooking or unhooking of the load']"}}

585. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_02`

HALLUCINATION

- **Question:** What color must derrails be painted during switching operations?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "Safety color code for marking physical hazards", "parameters": {"(a)": "Color identification", "(1)": "Red", "(i)": "Danger"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

586. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1441\_01`

- **Question:** When attaching a ground to a line or equipment operating at 600 volts or less, what alternative to a live-line tool may the employer permit, and under what conditions?
- **Ground truth:** The employer may permit the employee to use insulating equipment other than a live-line tool, if the employer ensures the line or equipment is not energized at the time the ground is connected, or if the employer can demonstrate that each employee is protected from hazards that may develop if the line or equipment is energized.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "alternative_to_live-line_tool", "parameters": {"voltage": "600", "conditions": "the line or equipment is not energized at the time the ground is connected or the employer can"}}

demonstrate that each employee is protected from hazards that may develop if the line or equipment is energized"]}

- ✓ **VERITROOPER (PASS):** {"name": "insulating_equipment", "parameters": {"voltage": "600", "condition": "the line or equipment is not energized at the time the ground is connected or the employer can demonstrate that each employee is protected from hazards that may develop if the line or equipment is energized"]}}

587. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_01`

- **Question:** In appendix A Table A, what is the upper bound of category A for Cadmium in urine (CdU) in the second table listing (>3 and ≤?)?

- **Ground truth:** 7

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_upper_bound_category_A", "parameters": {"marker": "Cadmium in urine (CdU)", "table_number": "2"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_upper_bound", "parameters": {"category": "A", "biological_marker": "Cadmium in urine (CdU)", "table_number": "2"}}

588. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_05`

- **Question:** In the worked example, what is the adjusted hearing level at 4000 Hz after subtracting the aging contribution from the most recent audiogram value?

- **Ground truth:** 22

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_adjusted_hearing_level", "parameters": {"frequency": "4000", "most_recent_audiogram_value": "25", "aging_contribution": "3"}}

- ✓ **VERITROOPER (PASS):** {"name": "subtract", "parameters": {"a": "25", "b": "3"}}

589. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_2160\_02`

WRONG CALC/QUOTE

- **Question:** What were the consequences for analytical bias and OAE resulting from the average recovery of 94.2% of the spiked filter samples in the AAS-HGA analysis?

- **Ground truth:** The analytical bias was -5.8% and the OAE was ±14.2%.

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_consequences", "parameters": {"recovery": "94", "cv": "0"}}

- ✗ **VERITROOPER (FAIL):** {"name": "calculate_consequences", "parameters": {"recovery": "94", "analysis_type": "AAS-HGA"}}

590. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_05`

- **Question:** If a derrick is hoisting personnel, may it travel?

- **Ground truth:** No; derricks are prohibited from traveling while personnel are hoisted.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Traveling\u2014derricks", "parameters": {"derricks": "prohibited from traveling while personnel are hoisted"}}

- ✓ **VERITROOPER (PASS):** {"name": "derrick_travel", "parameters": {"traveling": "false"}}

591. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_00`

HALLUCINATION

- **Question:** What specific sensitization frequency percentage distinguishes sub-category 1A from sub-category 1B?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_sensitization_frequency", "parameters": {"sub_category": "1A"}}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

592. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3423\_01`

- **Question:** During lifting, within how many inches must all slab support points be kept to maintain the slab in a level position?
- **Ground truth:** 1/2 inch
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "all_points_at_which_the_slab_is_supported", "parameters": {"x": "1/2"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_support_point_tolerance", "parameters": {"tolerance": "0"}}

593. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3403\_00`

- **Question:** Per the cargo hooks requirement in paragraph (d) and the briefing requirement in paragraph (b), which two activities must each be performed 'prior to each day's operation'?
- **Ground truth:** Per paragraph (b), a briefing setting forth the plan of operation for the pilot and ground personnel must be conducted prior to each day's operation; and per paragraph (d), the cargo hooks must be tested prior to each day's operation to determine that the release functions properly, both electrically and mechanically.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "cargo_hooks_and_briefing", "parameters": {"activity1": "testing cargo hooks", "activity2": "conducting a briefing"}}
- ✓ **VERITROOPER (PASS):** {"name": "perform_activity", "parameters": {"activity": "testing cargo hooks to determine that the release functions properly, both electrically and mechanically", "time": "prior to each day"}}

594. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_04`

HALLUCINATION

- **Question:** What form must be submitted to document a new classification when batch toxicity changes under A.0.5.1.2?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "change_form", "parameters": {"description": "new classification due to batch toxicity change", "purpose": "documenting change in batch toxicity", "technical_basis": "A.0.5.1.2", "safety_and_health_considerations": "potential impact on human health and environment", "documentation_of_changes": "updated classification and associated documentation", "approvals_and_authorization": "authorized personnel"}}
- ✓ **VERITROOPER (PASS):** This information is not contained in the provided text.

595. PASS · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_00`

- **Question:** What percentage increase in Category 1 ingredient concentration triggers the requirement to reclassify an untested mixture under A.0.5.1.3?
- **Ground truth:** UNANSWERABLE
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "A.0.5.1.3", "parameters": {"concentration_increase": "any increase"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

596. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_07`

- **Question:** If the incident heat energy estimated under paragraph (I)(8)(ii) is exactly 2.0 cal/cm², must the outer layer of clothing be flame resistant under condition (D)?
- **Ground truth:** No; condition (D) requires flame resistance only when the estimated incident heat energy

exceeds 2.0 cal/cm2.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "must_outer_layer_be_flame_resistant", "parameters": {"incident_heat_energy": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "must_be_flame_resistant", "parameters": {"incident_heat_energy": "2"}}

597. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_03`

- **Question:** What angle or greater shall be used for the cone where ceiling heights permit?
- **Ground truth:** 60°
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "cone_angle", "parameters": {"angle": "60"}}
- ✓ **VERITROOPER (PASS):** {"name": "angle_greater_than", "parameters": {"angle": "60"}}

598. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1249\_01`

- **Question:** Which American National Standard applies to the chlorine gas exhaust system carrying gas away from the work place and breathing area?
- **Ground truth:** Z9.2-1960
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "American National Standard", "parameters": {"Z9": "2-1960"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_standard", "parameters": {"standard_number": "Z9.2-1960"}}

599. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2542\_00`

- **Question:** Under design objective (i), how quickly must the system produce the desired signal after activation, and per (vi), what human involvement is permitted to sound the signal upon an initiating event?
- **Ground truth:** The system must produce the desired signal within one-half second of activation, and no human action is permitted to sound the signal-the signal-generating system must respond automatically to an initiating event.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "design_objective_i_and_vi", "parameters": {"response_time": "one-half second", "human_involvement": "none"}}
- ✓ **VERITROOPER (PASS):** {"name": "system_performance", "parameters": {"time_to_produce_signal": "one-half second", "human_involvement": "none"}}

600. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1332\_01`

- **Question:** Under the log hauls and slips section, which two separate provisions address protecting employees from logs that are moving or traveling too far?
- **Ground truth:** Paragraph (vi)(f) requires a positive stop installed on all log hauls to prevent logs from traveling too far ahead in the mill, and paragraph (vi)(g) requires overhead protection for employees working below logs being moved to the log deck.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "log hauls and slips", "parameters": {"provisions": "['(f) A positive stop shall be installed on all log hauls to prevent logs from traveling too far ahead in the mill.', '(vi) Walkways along log hauls shall have a standard handrail on the outer edge, and cleats or other means to assure adequate footing and enable employees to walk clear of the log chute.]"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_provisions", "parameters": {"section": "log hauls and slips", "topic": "protecting employees from logs"}}

601. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4228\_03`

- **Question:** If the State fails to submit required developmental documentation when the step is scheduled for completion, what must OSHA do?
- **Ground truth:** OSHA shall notify the State that documentation or a supplement is required and set a timetable for submission, generally not to exceed 60 days.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "notify_state_of_required_documentation", "parameters": {"timetable_for_submission": "60 days"}}
- ✓ **VERITROOPER (PASS):** {"name": "OSHA_action", "parameters": {"action": "notify the State that documentation or a supplement is required and set a timetable for submission of any required documentation or supplement"}}

602. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_06`

- **Question:** How many borderline test results obtained within the period are needed for a 'confirmed positive' determination?
- **Ground truth:** three borderline test results
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "confirmed_positive", "parameters": {"test_results": "three"}}
- ✓ **VERITROOPER (PASS):** {"name": "confirmed_positive", "parameters": {"abnormal_test_results": "0", "borderline_test_results": "3"}}

603. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3428\_02`

- **Question:** When cranes or derricks may be used to hoist employees on a personnel platform, all provisions of § 1926.1431 must be met except for which one?
- **Ground truth:** Except for § 1926.1431(a).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1926.1431", "parameters": {"exception": "§ 1926.1431(a)"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_exception", "parameters": {"section": "1926.1431", "exception": "1926.1431(a)"}}

604. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_2959\_00`

UNIT/FORMAT MISMATCH

- **Question:** What is the airborne concentration of lead that defines the action level, calculated as an 8-hour time-weighted average?
- **Ground truth:** 30 micrograms per cubic meter of air (30 µg/m³)
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "action_level", "parameters": {"x": "30"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_action_level", "parameters": {"concentration": "30", "time_weighted_average": "8"}}

605. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_00`

- **Question:** If wind gusts reach 25 mph at the personnel platform, who must make the safety determination and what happens to an in-progress lift if it is deemed unsafe?
- **Ground truth:** A qualified person must determine if it is safe to lift personnel, and if it is not, the in-progress lifting operation must be terminated.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "determine_safety", "parameters": {"wind_speed": "25", "person": "qualified person", "action": "terminate"}}
- ✓ **VERITROOPER (PASS):** {"name": "safety_determination", "parameters": {"wind_speed": "25", "person_affected": "a qualified person", "action_taken": "the lifting operation must be terminated"}}

606. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4069\_01`

- **Question:** If the manufacturer does not specify a plumb tolerance for the crane tower, what tolerance must the tower be erected to?
- **Ground truth:** The crane tower must be plumb to a tolerance of at least 1:500 (approximately 1 inch in 40 feet).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "erection_tolerance", "parameters": {"tolerance": "1:500"}}

- ✓ **VERITROOPER (PASS):** {"name": "plumb_tolerance", "parameters": {"tolerance": "1:500"}}

607. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_04`

HALLUCINATION

- **Question:** What is the BeLPT and which laboratory must perform it under this standard?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "BeLPT", "parameters": {"test": "measurement of blood lymphocyte proliferation in a laboratory test when lymphocytes are challenged with a soluble beryllium salt", "laboratory": "a laboratory certified under the College of American Pathologists/Clinical Laboratory Improvement Amendments (CLIA) guidelines"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

608. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4065\_00`

- **Question:** Under § 1926.1434, modifications or additions affecting capacity or safe operation are generally prohibited - under what circumstance are they permitted?
- **Ground truth:** Where the requirements of paragraphs (a)(1), (a)(2), (a)(3), (a)(4), or (a)(5) are met (manufacturer review and approval, manufacturer refusal to review, unavailable manufacturer, manufacturer fails to complete review within 120 days, or multiple manufacturers for marine work sites).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Modifications or additions affecting capacity or safe operation are permitted", "parameters": {"circumstance": "where the requirements of paragraphs (a)(1), (a)(2), (a)(3), (a)(4), or (a)(5) of this section are met"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_permitted_circumstances", "parameters": {"section": "1926.1434", "modification_type": "capacity or safe operation"}}

609. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_06`

- **Question:** When the physician review mechanism has not yet produced a final medical determination, on whose findings may the employer base a removal or limitation?
- **Ground truth:** On the medical findings, determinations, or recommendations of any of the physicians who have reviewed the employee's health status.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"physician": "any of the physicians who have reviewed the employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"physician": "any of the physicians who have reviewed the employee"}}

610. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4069\_02`

- **Question:** On a jobsite where more than one fixed jib (hammerhead) tower crane is installed, how must the cranes be located, and may they pass over one another?
- **Ground truth:** The cranes must be located such that no crane can contact the structure of another, but cranes are permitted to pass over one another.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "tower crane location", "parameters": {"crane_type": "fixed jib (hammerhead)", "location_requirement": "such that no crane can come in contact with the structure of another crane", "pass_over_permission": "yes"}}
- ✓ **VERITROOPER (PASS):** {"name": "crane_location", "parameters": {"location": "such that no crane can come in contact with the structure of another crane", "pass_over": "True"}}

611. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_09`

- **Question:** What is the normal level for cadmium in blood listed in Attachment 3?
- **Ground truth:** ≤ 5 $\mu\text{g/lwb}$
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "normal_cadmium_in_blood", "parameters": {"level": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_normal_level", "parameters": {"marker": "Cadmium in Blood"}}

612. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1984\_02`

- **Question:** Change rooms must be equipped with storage facilities in accordance with which CFR section?
- **Ground truth:** 29 CFR 1910.141(e)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "change rooms", "parameters": {"paragraph": "29 CFR 1910.141(e)"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_cfr_section", "parameters": {"section_number": "29 CFR 1910.141(e)"}}

613. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4010\_01`

- **Question:** If a program is audited 3 months after it begins, when must the next audit occur?
- **Ground truth:** At least every 3 years thereafter.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "audit_timing", "parameters": {"audit_start_date": "3 months after program begins", "audit_frequency": "3 years"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_next_audit_date", "parameters": {"current_audit_date": "3 months after the program begins", "audit_frequency": "3 years"}}

614. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_02`

- **Question:** In appendix A Table A, what is the threshold above which β_2 -microglobulin falls into category B in the first listing?
- **Ground truth:** 1500
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_threshold", "parameters": {"marker": "\u03b22 2 -microglobulin (\u03b22 2 -M) (\u03b2cb/g creatinine)", "category": "B", "listing": "1"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_category_b_threshold", "parameters": {"biological_marker": "\u03b22-microglobulin", "listing": "first"}}

615. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_09`

- **Question:** Which ASTM standard is referenced for the Pensky Martens Closed Tester flashpoint method?
- **Ground truth:** ASTM D-93-69
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ASTM", "parameters": {"D": "93-71"}}
- ✓ **VERITROOPER (PASS):** {"name": "ASTM", "parameters": {"D": "93-69"}}

616. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_04`

- **Question:** What is the purge gas flow in the GC conditions for MDA analysis?
- **Ground truth:** 40 mL/min
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "GC Conditions", "parameters": {"Purge gas flow": "40"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_GC_condition", "parameters": {"condition_name": "Purge gas flow"}}

617. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_11`

- **Question:** Below what amount of beta-2-microglobulin in the urine is normally not reabsorbed into the blood and released in urine?
- **Ground truth:** less than 300 µg/g Cr
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "beta-2-microglobulin", "parameters": {"amount": "300 \u03bcg/g Cr"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_beta_2_microglobulin_normal_level", "parameters": {"level_type": "normal"}}

618. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0117\_00`

- **Question:** Why is a partial day of work recorded as a day of job transfer or restriction for recordkeeping purposes (with one exception)?
- **Ground truth:** Because the worker works only for a partial work shift due to a work-related injury or illness; the exception is the day on which the injury occurred or the illness began.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "record_partial_day", "parameters": {"reason": "A partial day of work is recorded as a day of job transfer or restriction for recordkeeping purposes, except for the day on which the injury occurred or the illness began."}}
- ✓ **VERITROOPER (PASS):** {"name": "record_partial_day", "parameters": {"reason": "work-related injury or illness"}}

619. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_0219\_02`

MODEL ERROR

- **Question:** What is the publication date of ISO 13943:2000 (E/F), Fire Safety-Vocabulary?
- **Ground truth:** April, 15, 2000
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_publication_date", "parameters": {"standard": "ISO 13943:2000 (E/F)"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_publication_date", "parameters": {"standard": "ISO 13943:2000 (E/F)"}}

620. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_04`

HALLUCINATION

- **Question:** How many cross rods are required per foot of metal mesh fabric?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "cross_rod_requirement", "parameters": {"number_of_cross_rods": "1"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

621. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_00`

- **Question:** Under § 1902.3, the conformity with the Standards for a Merit System of Personnel Administration references which CFR part issued by the Secretary of Labor?
- **Ground truth:** 45 CFR part 70
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "conformity_with_standards", "parameters": {"standard": "45 CFR part 70"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_reference", "parameters": {"title": "45 CFR part 70"}}

622. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_05`

BAD QUESTION

- **Question:** What is the required spacing between connecting flanges for ductwork larger than 12 inches in diameter?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_required_spacing", "parameters": {"duct_diameter": "larger than 12 inches"}}
- **✓ VERITROOPER (PASS):** {"name": "not_found", "parameters": {"item": "spacing between connecting flanges for ductwork larger than 12 inches in diameter"}}

623. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3059\_03`

- **Question:** Where conveyors carry work into or out of spray booths, how large may the openings be?
- **Ground truth:** The openings shall be as small as practical.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "conveyors_openings", "parameters": {"size": "as small as practical"}}
- **✓ VERITROOPER (PASS):** {"name": "conveyors_openings", "parameters": {"size": "as small as practical"}}

624. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0133\_00`

MODEL ERROR

- **Question:** Why must State-Plan States consult with and obtain approval for requirements that are more stringent than or supplemental to the Federal requirements for certain part 1904 provisions?
- **Ground truth:** Because of the unique nature of the national recordkeeping program.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "consult_and_obtain_approval", "parameters": {"requirement_type": "more stringent than or supplemental to the Federal requirements", "provisions": "industry exemptions, reporting of fatalities and hospitalizations, record retention, or employee involvement"}}
- **✓ VERITROOPER (PASS):** {"name": "get_approval", "parameters": {"reason": "because of the unique nature of the national recordkeeping program"}}

625. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2346\_01`

- **Question:** When an observer is entitled to record monitoring results, is there any condition that limits how they exercise this entitlement?
- **Ground truth:** Yes, the observer must exercise the entitlement without interfering with the monitoring.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "observation_procedures", "parameters": {"entitlement_limitation": "Without interfering with the monitoring"}}
- **✓ VERITROOPER (PASS):** {"name": "observer_recording_results", "parameters": {"limit": "Without interfering with the monitoring"}}

626. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_08`

HALLUCINATION

- **Question:** Which government office issues approvals for bulk oxygen system installations?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "local_fire_prevention_office", "parameters": {}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

627. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_00`

HALLUCINATION

- **Question:** What is the maximum weight capacity of a powered hand truck permitted under this section?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_maximum_weight_capacity", "parameters": {"truck_type": "powered hand truck"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

628. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0546\_01`

- **Question:** Open flames or other sources of ignition are not permitted in vaporizer rooms, with what exception?
- **Ground truth:** Except those vaporizer rooms housing direct-fired vaporizers.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "open_flames_exception", "parameters": {"exception": "direct-fired vaporizers"}}
- **✓ VERITROOPER (PASS):** {"name": "exception_to_open_flames_in_vaporizer_rooms", "parameters": {"exception": "direct-fired vaporizers"}}

629. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0040\_01`

- **Question:** In § 1902.35, which section of the Act provides the Federal standards authority that will not be relinquished during the evaluation period?
- **Ground truth:** section 6
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_section", "parameters": {"number": "6"}}
- **✓ VERITROOPER (PASS):** {"name": "get_section_of_Act", "parameters": {"section_number": "6"}}

630. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2632\_03`

- **Question:** If a substance shows only a low to moderate potency in animals and low to moderate frequency in humans, which sub-category applies?
- **Ground truth:** Sub-category 1B.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "sub_category", "parameters": {"x": "1B"}}
- **✓ VERITROOPER (PASS):** {"name": "get_sub_category", "parameters": {"potency_in_animals": "low to moderate", "frequency_in_humans": "low to moderate"}}

631. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0333\_00`

MODEL ERROR

- **Question:** If angulated roping is employed but the exterior building design makes tie-in guides feasible, may the tie-in guides be eliminated for the uppermost 75 feet?
- **Ground truth:** No; elimination is only permitted if it is infeasible due to exterior building design (and an angulation force of at least 10 pounds is maintained).
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "eliminate_tie_in_guides", "parameters": {"employed": "True", "design": "False", "feet": "75"}}
- **✓ VERITROOPER (PASS):** {"name": "check_tie_in_guides", "parameters": {"condition": "False"}}

632. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_02`

- **Question:** What is the Public Law number cited for the Intergovernmental Personnel Act of 1970 in § 1902.3(h)?
- **Ground truth:** Pub. L. 91-648
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_public_law_number", "parameters": {"citation": "section 208 of the Intergovernmental Personnel Act of 1970"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_public_law_number", "parameters": {"citation": "91-648"}}

633. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_2159\_00`

- **Question:** Using the IUPAC equation $C_{ld} = k(sd)/m$ with $sd = 1.05$ abs. units and $m = 772.7$ abs. units/($\mu\text{g/mL}$), what is the qualitative detection limit ($k = 3$) in $\mu\text{g/mL}$? Round to 4 decimal places.
- **Ground truth:** 0.0041 $\mu\text{g/mL}$
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "C_ld", "parameters": {"k": "3", "sd": "1", "m": "772"}}
- ✓ **VERITROOPER (PASS):** {"name": "round", "parameters": {"number": "3 * 1.05 / 772.7", "ndigits": "4"}}

634. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_10`

- **Question:** For Class 00 equipment, what is the AC rms retest voltage?
- **Ground truth:** 2,500
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "00", "voltage_type": "AC rms"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "00", "voltage_type": "AC rms"}}

635. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1264\_00`

- **Question:** According to the requirements for beater guards in paragraph (d)(1) and the rule for cleanout holes in paragraph (d)(2), what must be true about both the beater covers and the cleanout hole covers while the machine is in motion?
- **Ground truth:** The beater covers must be provided with an interlock that prevents the cover from being raised while the machine is in motion, and the cleanout hole covers must be securely fastened and not opened while the machine is in motion. In both cases, the covers cannot be opened/raised while the machine is moving.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "beater_covers_and_cleanout_hole_covers_requirements", "parameters": {"beater_covers": "prevents contact with the beater and is provided with an interlock that prevents the cover from being raised", "cleanout_hole_covers": "are securely fastened and not opened"}}
- ✓ **VERITROOPER (PASS):** {"name": "both_covers_must_be", "parameters": {"beater_covers": "closed", "cleanout_hole_covers": "closed"}}

636. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_02`

- **Question:** To what items do the leaving-equipment-unattended provisions of § 1926.1417(e)(1) NOT apply?
- **Ground truth:** They do not apply to working gear (such as slings, spreader bars, ladders, and welding machines) where the weight of the working gear is negligible relative to the lifting capacity of the equipment as positioned, and the working gear is suspended over an area other than an entrance or exit.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1926.1417", "parameters": {"e": "1", "ii": "working gear (such as slings, spreader bars, ladders, and welding machines) where the weight of the working gear is negligible relative to the lifting capacity of the equipment as positioned, and the working gear is suspended over an area other than an entrance or exit"}}
- ✓ **VERITROOPER (PASS):** {"name": "leaving-equipment-unattended provisions of § 1926.1417(e)(1) do not apply to", "parameters": {"items": "working gear (such as slings, spreader bars, ladders, and welding"}}

machines) where the weight of the working gear is negligible relative to the lifting capacity of the equipment as positioned, and the working gear is suspended over an area other than an entrance or exit"]}}

637. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1226\_02`

- **Question:** Based on the rule for the carbide charge in subpart (ii) and the rule for the water chamber in subpart (iv), what two conditions must be met when recharging a generator with carbide?
- **Ground truth:** The carbide added must be sufficient to refill the space provided for carbide without ramming the charge (and no steel or other ferrous tools may be used in distributing it), and the water chamber must be filled to the proper level when the generator is being charged.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "check_carbide_charge_and_water_chamber", "parameters": {"carbide_charge": "sufficient to refill the space provided for carbide without ramming the charge", "water_chamber": "filled to the proper level"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_conditions", "parameters": {"condition1": "The carbide added each time the generator is recharged shall be sufficient to refill the space provided for carbide without ramming the charge", "condition2": "The water chamber shall be filled to the proper level"}}

638. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_00`

- **Question:** If a laundry washing machine's outer door on the shell is open, can the inside cylinder be operated under power per the interlock requirement?
- **Ground truth:** No, the interlock device must prevent power operation of the inside cylinder when the outer door is open.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "interlock_device", "parameters": {"operates": "false"}}
- ✓ **VERITROOPER (PASS):** {"name": "interlock_device", "parameters": {"is_outer_door_open": "true"}}

639. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2239\_04`

- **Question:** Combining the two availability provisions, to which parties must the employer make all paragraph (m) records available, and which specific records get provided to employees and the Assistant Secretary?
- **Ground truth:** Under (3)(i), the employer shall make available all records required by paragraph (m) to the Secretary and the Director for examination and copying. Under (3)(ii), employee exposure measurement records and employee medical records shall be provided upon request to employees, designated representatives, and the Assistant Secretary.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "availability_provision", "parameters": {"paragraph": "m", "parties": "[Assistant Secretary, Director, affected employees, former employees, designated representatives]", "records_for_employees_and_Assistant_Secretary": "exposure records"}}
- ✓ **VERITROOPER (PASS):** {"name": "make_available", "parameters": {"parties": "[Secretary, Director, employees, designated representatives, Assistant Secretary]", "records": "[employee exposure measurement records, employee medical records]"}}

640. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_05`

- **Question:** Which IEEE standard provides guidelines for the examination, cleaning, repairing, and in-service testing of live-line tools?
- **Ground truth:** IEEE Std 516-2009
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IEEE Std 516-2009", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "get_ieee_standard", "parameters": {"standard_number": "516", "year": "2009"}}

641. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0546\_00`

- **Question:** Electrical equipment and wiring shall be of a type specified by and installed in accordance with subpart S for ordinary locations, except in what circumstance?
- **Ground truth:** Except that fixed electrical equipment in classified areas shall comply with subparagraph (18) of this paragraph.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "electrical equipment and wiring", "parameters": {"type": "specified by and installed in accordance with subpart S for ordinary locations", "circumstance": "fixed electrical equipment in classified areas"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_circumstance", "parameters": {"circumstance": "fixed electrical equipment in classified areas"}}

642. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_07`

HALLUCINATION

- **Question:** What is the maximum allowable storage temperature for liquid oxygen in these systems?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_max_storage_temperature", "parameters": {"material": "liquid oxygen"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

643. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3809\_03`

- **Question:** Within how many days of the medical examination must the employer obtain a written medical opinion from the CBD diagnostic center?
- **Ground truth:** 30 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "obtain_written_medical_opinion", "parameters": {"days": "30"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_days_for_written_medical_opinion", "parameters": {"days": "30"}}

644. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_03`

HALLUCINATION

- **Question:** What languages other than English does OSHA provide official complaint forms in?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "complaint forms in other languages", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

645. FAIL · Calculation · `OSHA\_29CFR\_full\_2024\_4031\_02`

MODEL ERROR

- **Question:** When hoisting personnel without a personnel platform, the equipment's rated capacity for the radius and configuration is 17,800 pounds. What is the maximum total load (including hook, load line, rigging and any other equipment that imposes a load) allowed, except during proof testing?
- **Ground truth:** 8,900 pounds
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_total_load", "parameters": {"rated_capacity": "17800"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_max_load", "parameters": {"rated_capacity": "17800"}}

646. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_01`

HALLUCINATION

- **Question:** What is the maximum penalty an employer faces for failing to provide the written medical report within the 45-day deadline?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "calculate_penalty", "parameters": {"penalty_type": "failure_to_correct_violation", "days_late": "45"}}
- **✓ VERITROOPER (PASS):** I don't have enough information to answer this question. The evidence provided does not mention any penalties for failing to provide the written medical report within the 45-day deadline.

647. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0478\_01`

- **Question:** For electrostatic spraying operations, what is the minimum air velocity allowed over the open face of the booth?
- **Ground truth:** 60 linear feet per minute
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "minimum_air_velocity", "parameters": {"velocity": "60"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_minimum_air_velocity", "parameters": {"operation_type": "electrostatic spraying"}}

648. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_01`

- **Question:** What is the injector zone temperature in the GC analysis conditions for MDA?
- **Ground truth:** 235 degrees C
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "GC conditions", "parameters": {"Injector": "235"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_GC_analysis_conditions", "parameters": {"zone": "injector"}}

649. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_00`

HALLUCINATION

- **Question:** What is the minimum thickness required for the metal or asbestos guards used to shield combustibles when relocation is impracticable?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_thickness", "parameters": {"material": "metal or asbestos"}}
- **✓ VERITROOPER (PASS):** I don't have enough information to answer the question.

650. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3386\_02`

- **Question:** If mechanical equipment is not being used, how far from the roof edge must the warning line be erected?
- **Ground truth:** Not less than 6 feet (1.8 m).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "warning_line_erection", "parameters": {"distance": "6 feet (1.8 m)"}}
- **✓ VERITROOPER (PASS):** {"name": "erect_warning_line", "parameters": {"distance": "6", "unit": "feet"}}

651. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_1593\_04`

MODEL ERROR

- **Question:** What is the stated exception to the minimum clear working space requirements specified in Table S-2?

- **Ground truth:** The Table S-2 minimum clear working space applies except as elsewhere required or permitted in this subpart.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "minimum_clear_working_space", "parameters": {"exception": "the location of the switchboards and control panels permits a continuous and unobstructed way of exit travel"}}
- **✓ VERITROOPER (PASS):** {"name": "exception_to_minimum_clear_working_space", "parameters": {"requirement": "elsewhere required or permitted in this subpart"}}

652. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_06`

HALLUCINATION

- **Question:** What fee may an employer charge for copying the COVID-19 records when requested?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "reasonable, non-discriminatory administrative costs", "parameters": {"search": "and copying expenses but not including overhead expenses"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

653. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1186\_02`

- **Question:** In what direction may deadman controls operate to disengage the drive, as an exception to the general control direction requirement?
- **Ground truth:** They may operate in any direction to disengage the drive.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "deadman controls direction", "parameters": {"direction": "any direction"}}
- **✓ VERITROOPER (PASS):** {"name": "operate_deadman_control", "parameters": {"direction": "any direction"}}

654. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_04`

- **Question:** After a fuel gas valve stem leak is found, the cylinder must normally be removed from the work area if tightening the gland nut fails. What is the exception that allows the cylinder to remain in the work area?
- **Ground truth:** If a regulator attached to a cylinder valve will effectively stop a leak through the valve seat, the cylinder need not be removed from the work area.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "answer_question", "parameters": {"question": "After a fuel gas valve stem leak is found, the cylinder must normally be removed from the work area if tightening the gland nut fails. What is the exception that allows the cylinder to remain in the work area?", "answer": "If a regulator attached to a cylinder valve will effectively stop a leak through the valve seat."}}
- **✓ VERITROOPER (PASS):** {"name": "exception_to_removal", "parameters": {"condition": "a regulator attached to a cylinder valve will effectively stop a leak through the valve seat"}}

655. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2016\_01`

MODEL ERROR

- **Question:** According to Appendix B, which two paragraphs of the regulation define the scope of coverage (what records the rule applies to) and the meaning of 'toxic substance or harmful physical agent'?
- **Ground truth:** Paragraph (b)(2) establishes that the regulation applies to all employee exposure and medical records of employees exposed to toxic substances or harmful physical agents, and paragraph (c)(13) defines the term 'toxic substance or harmful physical agent' to encompass chemical substances, biological agents, and physical stresses for which there is evidence of harmful health effects.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "specific_written_consent", "parameters": {"i": "1", "ii": "2"}}
- **✓ VERITROOPER (PASS):** {"name": "get_appendix_b_paragraphs", "parameters": {"paragraph_numbers": "1, 2"}}

"[2, 13]"}}

656. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2508\_02`

- **Question:** If an area's airborne respirable crystalline silica concentration can reasonably be expected to exceed the PEL but the employer has not demarcated it, does it qualify as a regulated area?
- **Ground truth:** No, a regulated area must be demarcated by the employer.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "regulated_area", "parameters": {"demarcated": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "is_regulated_area", "parameters": {"area_demarcated": "False"}}

657. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_03`

- **Question:** What is the column gas flow (Ar/CH₄) in the GC conditions for MDA analysis?
- **Ground truth:** 28 mL/min
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "GC conditions", "parameters": {"Zone temperatures": "Column\u201414220 degrees C Injector\u201414235 degrees C Detector\u201414335 degrees C", "Gas flows, Ar/CH": "Column\u20141428 mL/min (95/5) Purge\u20141440 mL/min", "Injection volume": "5.0 uL"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_gc_conditions", "parameters": {"parameter_name": "Column gas flow (Ar/CH₄)"}}

658. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1008\_00`

- **Question:** According to the requirements for ropes that have been idle for a month or more, who must perform the thorough inspection, and what three items must the resulting certification record include?
- **Ground truth:** The inspection must be performed by an appointed or authorized person (whose approval is required for further use of the rope). The certification record must include the date of inspection, the signature of the person who performed the inspection, and an identifier for the rope which was inspected.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "thorough_inspection_and_certification_record", "parameters": {"inspector": "a competent person", "items_included_in_record": "['date of inspection', 'signature of the person who performed the inspection', 'identifier for the ropes which were inspected']"}}
- ✓ **VERITROOPER (PASS):** {"name": "inspect_rope", "parameters": {"inspector": "appointed or authorized person", "items_to_include": "['date of inspection', 'signature of the person who performed the inspection', 'identifier for the rope which was inspected']"}}

659. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3172\_03`

- **Question:** When passage of persons between the upper and lower runs of a horizontal belt is possible and is regarded as necessary, what guarding combination is required, and how does this requirement change in powerplants?
- **Ground truth:** Under paragraph (e)(1)(iii)(b), there shall be a platform over the lower run guarded on either side by a railing completely filled in with wire mesh or other filler, or by a solid barrier, and the upper run shall be guarded to prevent contact by the worker or objects he carries. However, in powerplants only the lower run of the belt need be guarded.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "guarding_combination", "parameters": {"upper_run_guarding": "a railing completely filled in with wire mesh or other filler, or by a solid barrier", "lower_run_guarding": "a platform over the lower run guarded on either side by a railing completely filled in with wire mesh or other filler, or by a solid barrier", "powerplant_exception": "only the lower run of the belt need be guarded"}}
- ✗ **VERITROOPER (FAIL):** {"name": "guarding_combination", "parameters": {"passage_regarded_as_necessary": "True", "location": "powerplants"}}

660. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2838\_03`

- **Question:** What is the date of IME Pub. No. 20, Radio Frequency Energy?
- **Ground truth:** Mar. 1968
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IME Pub. No. 20", "parameters": {"date": "Mar. 1968"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_date", "parameters": {"pub_no": "IME Pub. No. 20, Radio Frequency Energy-A Potential Hazard in the Use of Electric Blasting Caps"}}

661. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3305\_03`

- **Question:** Based on subpart (iii) and subpart (iv), what is the difference between the minimum step width of a non-mobile stairway-type ladder and a mobile one, and what tread requirement applies to both?
- **Ground truth:** The non-mobile minimum step width is 16 inches (41 cm) and the mobile minimum is 11 1/2 inches (30 cm), a difference of 4 1/2 inches; both must have slip-resistant treads on all steps and landings per subpart (iv).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "stairway-type_ladder", "parameters": {"minimum_step_width_non_mobile": "16 inches", "minimum_step_width_mobile": "11 1/2 inches", "tread_requirement": "slip-resistant treads on all steps and landings"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_difference", "parameters": {"a": "16", "b": "11"}}

662. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_08`

HALLUCINATION

- **Question:** Who is designated as the custodian responsible for maintaining the confidentiality of the COVID-19 log?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_custodian", "parameters": {"log_type": "COVID-19 log"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

663. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_04`

- **Question:** Do 1926.502 and 1926.503 share an OMB control number, and how does it compare to that of 1926.550(b)(2)?
- **Ground truth:** 1926.502 and 1926.503 share OMB control No. 1218-0197, which differs from 1926.550(b)(2)'s OMB control No. 1218-0232.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_omb_control_numbers", "parameters": {"a": "1218-0197", "b": "1218-0232"}}
- ✓ **VERITROOPER (PASS):** {"name": "compare_omb_control_numbers", "parameters": {"x": "1218-0197", "y": "1218-0232"}}

664. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3403\_03`

MODEL ERROR

- **Question:** Using the weight limitation in paragraph (k) and the operator responsibility in paragraph (h), who is responsible for the load's weight and what is the maximum allowable external load weight?
- **Ground truth:** Per paragraph (h), the helicopter operator is responsible for the size, weight, and manner in which loads are connected to the helicopter; and per paragraph (k), the weight of an external load shall not exceed the manufacturer's rating.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "operator_responsibility_and_weight_limitation", "parameters": {"k": "1926", "h": "operator"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_load_weight_responsibility", "parameters": {"weight_limitation":

"The weight of an external load shall not exceed the manufacturer", "operator_responsibility": "The helicopter operator shall be responsible for size, weight, and manner in which loads are connected to the helicopter"]}]}

665. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1622\_02`

- **Question:** Open wiring on insulators is permitted only on systems of what voltage, and what is the maximum voltage condition combined with the requirement for supports near a dead-end connection to a lampholder?

- **Ground truth:** Open wiring on insulators is only permitted on systems of 600 volts, nominal, or less (paragraph (i)), and supports must be installed within 305 mm (12 in.) of a dead-end connection to a lampholder or receptacle (paragraph (ii)(B)).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "open_wiring_on_insulators", "parameters": {"voltage": "600", "maximum_voltage_condition": "305 mm (12 in.)"}}

- ✓ **VERITROOPER (PASS):** {"name": "open_wiring_on_insulators", "parameters": {"voltage": "600", "location": "near a dead-end connection to a lampholder"}}

666. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_4398\_04`

- **Question:** If objections or a petition for review are withdrawn because of settlement, what is required as a result?

- **Ground truth:** The settlement must be submitted for approval in accordance with paragraph (d) of this section.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "settlement", "parameters": {"approval": "required"}}

- ✓ **VERITROOPER (PASS):** {"name": "settlement_requirement", "parameters": {"event": "withdrawal of objections or petition for review", "requirement": "submission for approval"}}

667. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3420\_01`

- **Question:** Under § 1926.702(e), concrete pumping systems using discharge pipes must be provided with pipe supports designed for what overload percentage?

- **Ground truth:** 100 percent

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "pipe_supports", "parameters": {"overload_percentage": "100"}}

- ✗ **VERITROOPER (FAIL):** {"name": "answer_section_question", "parameters": {"section_number": "1926.702", "question": "What overload percentage must pipe supports be designed for in concrete pumping systems using discharge pipes under \u00a7 1926.702(e)?"}}

668. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0676\_00`

- **Question:** If there are no permissible exposure limits for a hazardous substance, what standard must monitoring ensure employees are not exposed beyond?

- **Ground truth:** Published exposure levels for that hazardous substance.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "monitoring", "parameters": {"paragraph": "published exposure levels"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If there are no permissible exposure limits for a hazardous substance, what standard must monitoring ensure employees are not exposed beyond?", "answer": "published exposure levels"}}

669. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_01`

- **Question:** If a power line is equipped with a device that automatically reenergizes the circuit upon contact, but the device's design does not permit disabling the automatic reclosing feature, must the feature still be made inoperative?

- **Ground truth:** No; the automatic reclosing feature must be made inoperative only if the design of the device

permits.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "disable automatic-reclosing feature", "parameters": {"made_inoperative": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "make_inoperative", "parameters": {"device_design_permits_disabling": "false"}}

670. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1226\_01`

- **Question:** Combining the electrical apparatus rule with the recharging carbide requirement, if a generator is being recharged, what condition allows electrical equipment in the generator room to merely conform to Class I, Division 2 provisions, and how much carbide should be added during that recharge?
- **Ground truth:** Electrical equipment may conform to subpart S Class I, Division 2 provisions only where the generator system is designed so that no carbide fill opening or other part of the generator is open to the generator house/room during operation and residue is carried in closed piping from the residue discharge valve to a point outside the house/room. When recharging, the carbide added must be sufficient to refill the space provided for carbide without ramming the charge.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "generator_recharge", "parameters": {"condition": "the carbide fill opening or other part of the generator is not open to the generator house or room during the operation of the generator, and residue is carried in closed piping from the residue discharge valve to a point outside the generator house or room", "carbide_amount": "sufficient to refill the space provided for carbide without ramming the charge"}}
- ✓ **VERITROOPER (PASS):** {"name": "check_generator_condition", "parameters": {"condition": "no carbide fill opening or other part of the generator is open to the generator house or room during the operation of the generator", "carbide_amount": "sufficient to refill the space provided for carbide without ramming the charge"}}

671. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_04`

- **Question:** What edition of Bretherick's Handbook of Reactive Chemical Hazards is referenced?
- **Ground truth:** 2nd edition, 1979
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Bretherick", "parameters": {"edition": "2nd", "year": "1979"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_edition", "parameters": {"author": "Bretherick", "book_title": "Handbook of Reactive Chemical Hazards"}}

672. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3285\_00`

- **Question:** If an employer uses a scaffold platform in an area that is too narrow to allow an 18-inch width, what must the employer do to comply?
- **Ground truth:** Make the platform as wide as feasible and protect employees from fall hazards using guardrails and/or personal fall arrest systems.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "narrow_scaffold_platform_protection", "parameters": {"width": "as wide as feasible", "guardrails": "true", "personal_fall_arrest_systems": "true"}}
- ✓ **VERITROOPER (PASS):** {"name": "protect_from_fall_hazards", "parameters": {"width": "as wide as feasible", "protection": "guardrails and/or personal fall arrest systems"}}

673. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3047\_02`

- **Question:** Combining the equipment requirement in paragraph (ii) with the training requirements it lists, what two conditions must be satisfied for clean-up done on plant property using plant or workplace employees?
- **Ground truth:** First, the employees must have completed the training requirements of 29 CFR 1926.35, 1926.59, and 1926.103 plus other appropriate safety and health training (e.g., personal protective equipment and decontamination procedures); second, all equipment used in the clean-up must be in serviceable

condition and have been inspected prior to use.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "clean_up_on_plant_property", "parameters": {"condition1": "All equipment to be used in the performance of the clean-up work shall be in serviceable condition and shall have been inspected prior to use.", "condition2": "Such employees shall have completed the training requirements of the following: 29 CFR 1926.35, 1926.59, and 1926.103, and other appropriate safety and health training made necessary by the tasks that they are expected to be performed such as personal protective equipment and decontamination procedures."}}
- ✓ **VERITROOPER (PASS):** {"name": "clean_up_conditions", "parameters": {"condition1": "All equipment to be used in the performance of the clean-up work shall be in serviceable condition and shall have been inspected prior to use.", "condition2": "Such employees shall have completed the training requirements of the following: 29 CFR 1926.35, 1926.59, and 1926.103, and other appropriate safety and health training made necessary by the tasks that they are expected to be performed such as personal protective equipment and decontamination procedures."}}

674. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1227\_01`

MODEL ERROR

- **Question:** Combining the requirements of paragraphs (v) and (vi), what must be done before making hot repairs to one generator located in a room containing additional generators?
- **Ground truth:** The carbide charge and feed mechanism of the generator being repaired must be completely removed, all its acetylene expelled by completely flooding the shell with water, and it must be disconnected from the piping system (paragraph (v)); additionally, all the other generators and piping in the room must be purged of acetylene (paragraph (vi)).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "remove_carbide_charge_and_feed_mechanism", "parameters": {}}
- ✗ **VERITROOPER (FAIL):** {"name": "completely flooding the generator shell with water", "parameters": {"generator": "the generator to be repaired"}}

675. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4486\_01`

OVERCONFIDENT

- **Question:** Per paragraph (e)(1) a complaint will be dismissed unless a prima facie showing is made, but per paragraph (e)(4), what additional circumstance can prevent an investigation even after that showing, and by what evidentiary standard?
- **Ground truth:** Even if the complainant has made a prima facie showing, an investigation will not be conducted or will be discontinued if the respondent demonstrates by clear and convincing evidence that it would have taken the same adverse action in the absence of the complainant's protected activity.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "investigation_prevention", "parameters": {"circumstance": "respondent demonstrates that it would have taken the same adverse action in the absence of the complainant", "evidentiary standard": "clear and convincing evidence"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Per paragraph (e)(1) a complaint will be dismissed unless a prima facie showing is made, but per paragraph (e)(4), what additional circumstance can prevent an investigation even after that showing, and by what evidentiary standard?"}}

676. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_3034\_02`

- **Question:** A site has 120 employees. Combining the required toilet seats and urinals from Table D-65.2, what is the total number of these facilities?
- **Ground truth:** 6
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_total_facilities", "parameters": {"number_of_employees": "120"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_toilet_facilities", "parameters": {"number_of_employees":

"120"}}

677. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0035\_03`

- **Question:** What U.S.C. citation is given for the Occupational Safety and Health Act of 1970 in § 1902.30?
- **Ground truth:** 29 U.S.C. 667
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_us_citation", "parameters": {"section": "29 U.S.C. 667"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_usc_citation", "parameters": {"usc_citation": "29 U.S.C. 667"}}

678. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_1288\_04`

MODEL ERROR

- **Question:** If dye kettles or vats may at any time contain hot or corrosive liquids, where may the installed pipes or drains empty?
- **Ground truth:** They must carry the contents safely away from the working area and shall not empty directly onto the floor.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "Pipes or drains for dye kettles and vats", "parameters": {"location": "a safe location away from the working area"}}
- ✓ **VERITROOPER (PASS):** {"name": "empty_pipes_or_drains", "parameters": {"location": "safely away from the working area"}}

679. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0914\_00`

- **Question:** In extremely cold climates, what is the permitted consequence for hose storage when standard installation could damage equipment?
- **Ground truth:** Where such installation may result in damaged equipment in extremely cold climates, the hose may be stored in another location provided it is readily available and can be connected when needed.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "hose_storage_in_cold_climates", "parameters": {"consequence": "the hose may be stored in another location provided it is readily available and can be connected when needed"}}
- ✗ **VERITROOPER (FAIL):** {"name": "hose_storage", "parameters": {"climate": "extremely cold", "installation": "standard", "equipment_damage": "true"}}

680. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1292\_01`

MODEL ERROR

- **Question:** Combining the lubrication requirement and the hot pipes requirement, what two protective measures are required to keep employees safe from machine hazards?
- **Ground truth:** Per paragraph (c)(5), stationary lubrication fittings inside a machine must be provided with extension piping to a point of safety so employees do not reach into dangerous parts when lubricating; and per paragraph (c)(8), exposed hot water and steam pipes must be covered with insulating material wherever necessary to protect employees from contact.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "protective_measures", "parameters": {"measure1": "automatic oiling device", "measure2": "insulating material or guarding"}}
- ✓ **VERITROOPER (PASS):** {"name": "protective_measures", "parameters": {"measure1": "stationary lubrication fittings inside a machine with extension piping to a point of safety", "measure2": "covering exposed hot water and steam pipes with insulating material"}}

681. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_02`

HALLUCINATION

- **Question:** Within how many days must the respondent submit documentation to the Social Security Administration after the preliminary order issues?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "submit_documentation_to_social_security_administration", "parameters": {"days": "10"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

682. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2784\_02`

MISINTERPRETED

- **Question:** For laboratories, does the full set of provisions of an OSHA health standard apply, or only part of it, and under what exceptions?
- **Ground truth:** Only the requirement to limit employee exposure to the specific permissible exposure limit applies, unless that standard states otherwise or unless the conditions of paragraph (a)(2)(iii) apply.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "apply_osh_a_health_standard", "parameters": {"provisions": "partially", "exceptions": "paragraphs (d) and (g)(1)(ii) of this section"}}
- **✓ VERITROOPER (PASS):** {"name": "apply_osh_a_health_standard", "parameters": {"full_set": "False", "exceptions": "['paragraph (a)(2)(iii) of this section applies', 'standard states otherwise']"}}

683. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_05`

- **Question:** If the equipment is equipped with a device that automatically limits range of movement, must it be used during the operation?
- **Ground truth:** Yes; it must be used and set to prevent any part of the equipment, load line, or load from breaching the minimum approach distance established under paragraph (c).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "device_that_automatically_limits_range_of_movement", "parameters": {"used_and_set": "True"}}
- **✓ VERITROOPER (PASS):** {"name": "equipment_operation", "parameters": {"equipment_type": "crane", "operation_type": "movement", "device_status": "automatic", "usage_requirement": "must be used"}}

684. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4486\_03`

- **Question:** Both paragraph (c) and paragraph (f) require providing materials to the complainant. What redaction standard applies in each case, and how does paragraph (f) additionally address witness statements?
- **Ground truth:** In both paragraphs (c) and (f), before providing materials to the complainant the Agency will redact them, if necessary, in accordance with the Privacy Act of 1974, 5 U.S.C. 552a, and other applicable confidentiality laws. Paragraph (f) additionally provides that witness statements will be redacted to protect the identity of confidential informants, and if statements cannot be redacted without revealing those identities, summaries of their contents will be provided.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "redaction_standard", "parameters": {"paragraph_c": "Privacy Act of 1974, 5 U.S.C. 552a, and other applicable confidentiality laws", "paragraph_f": "consistent with the Privacy Act of 1974, 5 U.S.C. 552a, and other applicable confidentiality laws", "witness_statements": "redacted to protect the identity of confidential informants where statements were given in confidence; if the statements cannot be redacted without revealing the identity of confidential informants, summaries of their contents will be provided"}}
- **✓ VERITROOPER (PASS):** {"name": "redaction_standard", "parameters": {"paragraph_c": "Privacy Act of 1974, 5 U.S.C. 552a, and other applicable confidentiality laws", "paragraph_f": "Privacy Act of 1974, 5 U.S.C.

552a, and other applicable confidentiality laws", "witness_statements": "redacted to protect the identity of confidential informants where statements were given in confidence; if the statements cannot be redacted without revealing the identity of confidential informants, summaries of their contents will be provided"}}}

685. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_3327\_04`

MODEL ERROR

- **Question:** Under the Effective Date Note, until what date were certain requirements of § 1926.451(b)(2)(i) stayed?
- **Ground truth:** Nov. 25, 1997, or until further rulemaking has been completed, whichever is later
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Effective Date Note", "parameters": {"date": "Nov. 25, 1997"}}
- **✗ VERITROOPER (FAIL):** {"name": "date_stayed_until", "parameters": {"date": "Nov. 25, 1997"}}

686. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4297\_00`

- **Question:** Under § 1960.31(a)(2), inspections may be conducted in response to a request from what portion of the membership of record of a certified safety and health committee?
- **Ground truth:** half the membership of record
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "inspection_conducted_in_response_to_request", "parameters": {"portion_of_membership": "half"}}
- **✓ VERITROOPER (PASS):** {"name": "inspection_request", "parameters": {"portion_of_membership": "half"}}

687. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4502\_01`

- **Question:** Combining the prima facie showing requirement of paragraph (e)(1) with the respondent's affirmative defense in paragraph (e)(4), under what circumstance will further investigation not be conducted even though the complainant has met the prima facie standard?
- **Ground truth:** Even though the complainant has made the prima facie showing (a non-frivolous allegation that protected activity was a contributing factor in the adverse action), further investigation will not be conducted if the respondent demonstrates by clear and convincing evidence that it would have taken the same adverse action in the absence of the complainant's protected activity.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "investigation_not_conducted", "parameters": {"circumstance": "respondent demonstrates by clear and convincing evidence that it would have taken the same adverse action in the absence of the complainant"}}
- **✓ VERITROOPER (PASS):** {"name": "investigation_circumstance", "parameters": {"complainant_met_prima_facie": "True", "respondent_affirmative_defense": "clear and convincing evidence that it would have taken the same adverse action in the absence of the complainant"}}

688. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_02`

- **Question:** If a petitioner's request for party participation is denied by the administrative law judge, what is the requirement for the written notice given to that petitioner?
- **Ground truth:** The notice must briefly state the grounds for denial.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "written notice", "parameters": {"content": "briefly state the grounds for denial"}}
- **✓ VERITROOPER (PASS):** {"name": "written notice", "parameters": {"grounds_for_denial": "briefly stated"}}

689. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4319\_00`

- **Question:** Under which CFR provision must agencies report fatalities and multiple hospitalization incidents to OSHA, as noted in § 1960.70?
- **Ground truth:** 29 CFR 1904.39

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "29 CFR 1904.39", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "get_cfr_provision", "parameters": {"number": "29", "section": "1904.39"}}

690. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0458\_02`

- **Question:** Using the requirements for pumps and for hoses and couplings, what testing/inspection conditions apply to relief devices versus pressure hoses?
- **Ground truth:** Relief devices on loading pumps shall be tested at not more than yearly intervals to determine that they function satisfactorily at the pressure at which they are set. Pressure hoses and couplings shall be inspected at intervals appropriate to the service and tested with the hose extended using the in-service maximum operating pressures.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Pumps and Hoses/Couplings Testing/Inspection", "parameters": {"Relief Devices": "tested at not more than yearly intervals to determine that they function satisfactorily at the pressure at which they are set", "Pressure Hoses": "inspected at intervals appropriate to the service and tested with the hose extended and using the in-service maximum operating pressures"}}
- ✓ **VERITROOPER (PASS):** {"name": "test_relief_devices_and_hoses", "parameters": {"test_interval": "yearly", "test_pressure": "the pressure at which they are set", "hose_test_pressure": "inservice maximum operating pressures"}}

691. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2838\_04`

- **Question:** What is the suite number of the Institute of Makers of Explosives at 1120 19th Street, NW., Washington, DC?
- **Ground truth:** Suite 310
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_suite_number", "parameters": {"address": "1120 19th Street, NW., Washington, DC"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_suite_number", "parameters": {"street_address": "1120 19th Street, NW.", "city": "Washington, DC"}}

692. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2261\_02`

- **Question:** Once an evaluation is completed, within what timeframe must the employer obtain and provide the employee with the healthcare professional's written opinion?
- **Ground truth:** Within 15 days of the completion of the evaluation.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_written_opinion", "parameters": {"timeframe": "15 days"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_timeframe_for_obtaining_written_opinion", "parameters": {"timeframe": "15 days"}}

693. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3959\_03`

- **Question:** If the work is covered by subpart V, is the employer still required to install an elevated warning line or barricade under paragraph (d)(3)?
- **Ground truth:** No; the elevated warning line/barricade provision does not apply to work covered by subpart V of this part.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "install_elevated_warning_line_or_barricade", "parameters": {"required": "false"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If the work is covered by subpart V, is the employer still required to install an elevated warning line or barricade under paragraph (d)(3)?", "evidence": "No, the provision does not apply to work covered by subpart V of this part."}}

694. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1597\_03`

- **Question:** When a nongrounding-type receptacle is replaced with a GFCI-type receptacle marked 'No Equipment Ground,' what connection is prohibited?
- **Ground truth:** An equipment grounding conductor may not be connected from the GFCI-type receptacle to any outlet supplied from the GFCI receptacle.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "connection_prohibited", "parameters": {"connection": "equipment grounding conductor"}}
- ✓ **VERITROOPER (PASS):** {"name": "connection_prohibited", "parameters": {"receptacle_type": "nongrounding-type", "replacement_receptacle": "GFCI-type", "marking": "No Equipment Ground"}}

695. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2838\_01`

- **Question:** Which IME publication number is IBR approved for § 1926.900(k)?
- **Ground truth:** IME Pub. No. 20
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IME Pub. No.", "parameters": {"20": "Radio Frequency Energy\u20142014A Potential Hazard in the Use of Electric Blasting Caps"}}
- ✓ **VERITROOPER (PASS):** {"name": "IME Pub", "parameters": {"No": "20", "IBR approved for": "§ 1926.900(k)"}}

696. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1264\_03`

MODEL ERROR

- **Question:** Based on the interlock requirement in paragraph (d)(1) and the feed roll guard requirement in paragraph (d)(3), which operating condition do both provisions specifically restrict?
- **Ground truth:** Both restrict access to hazardous parts while the machine is operating: the (d)(1) interlock prevents operating the machine while the cover is open (and prevents raising the cover while in motion), and the (d)(3) feed roll guard must prevent reaching the nip while the machinery is in operation.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "operating_condition", "parameters": {"condition": "operation of the machine with the cover open"}}
- ✓ **VERITROOPER (PASS):** {"name": "restriction", "parameters": {"condition": "while the machine is in motion"}}

697. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_05`

BAD QUESTION

- **Question:** By what time of day must the employer provide records following a request?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "provide_records", "parameters": {"time": "four business hours"}}
- ✗ **VERITROOPER (FAIL):** {"name": "provide_records", "parameters": {"time": "end of the next business day"}}

698. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_4240\_02`

MODEL ERROR

- **Question:** Is a separable portion of a plan always limited to an issue as defined in 29 CFR 1902.2(c)?
- **Ground truth:** No. There is an exception: wherever the Assistant Secretary has determined that other industrial, occupational or hazard groupings are administratively practicable, such groupings shall be considered separable portions of a plan.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "is_separable_portion_limited", "parameters": {"plan": "29 CFR 1902.2(c)"}}

- ✓ **VERITROOPER (PASS):** {"name": "is_separable_portion_of_plan", "parameters": {"issue_defined_in_29_cfr_1902_2c": "False"}}

699. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_00`

- **Question:** Under what circumstance is the operator permitted to use a cellular phone while actually engaged in operating the equipment, despite the prohibition on attention-diverting practices?
- **Ground truth:** When the cellular phone is used for signal communications.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "operator_use_cellular_phone", "parameters": {"circumstance": "when used for signal communications"}}
- ✓ **VERITROOPER (PASS):** {"name": "operator_permitted_to_use_cellular_phone", "parameters": {"circumstance": "when used for signal communications"}}

700. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_04`

- **Question:** Which section of the Act does the product standards provision in § 1902.3(c)(2) reflect?
- **Ground truth:** section 18(c)(2)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section_reflection", "parameters": {"section_number": "18", "subsection": "c", "paragraph": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_reflected_section", "parameters": {"section": "1902.3(c)(2)", "act_section": "18(c)(2)"}}

701. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2838\_00`

- **Question:** Which IME publication number is IBR approved for § 1926.914(a)?
- **Ground truth:** IME Pub. No. 2
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IME Pub. No.", "parameters": {"2": "American Table of Distances for Storage of Explosives"}}
- ✓ **VERITROOPER (PASS):** {"name": "IME Pub", "parameters": {"No": "2", "IBR approved for": "§ 1926.914(a)"}}

702. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1249\_03`

- **Question:** Under what CFR section must the employer provide self-contained breathing apparatuses or supplied-air respirators for emergency and rescue operations in bleach mixing rooms?
- **Ground truth:** 29 CFR 1910.134
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "29 CFR 1910.134", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "get_CFR_section", "parameters": {"section_number": "29", "subsection": "1910.134"}}

703. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3280\_00`

- **Question:** A tank builders' scaffold and a top plate bracket scaffold are both described in the evidence. Based on the definition of 'Supported scaffold,' which of these two relies on a platform held up by a rigid support like brackets, and what is the specific bracket arrangement for each?
- **Ground truth:** Both are supported scaffolds because they use brackets (a rigid support). The tank builders' scaffold has a platform resting on brackets directly attached to a cylindrical tank or to devices attached to such a tank, while the top plate bracket scaffold is supported by brackets that hook over or are attached to the top of a wall.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Supported scaffold", "parameters": {"type": "Tank builders", "bracket_arrangement": "brackets directly attached to a cylindrical tank"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_scaffold_info", "parameters": {"scaffold_type": "tank builders"}}

704. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2565\_01`

- **Question:** The term 'hazardous substance' appears in both subpart (v) and subpart (6)(ii). Which different statutes define the term in each, and which agency's regulations apply in each case?
- **Ground truth:** In subpart (v), 'hazardous substance' is defined by the Federal Hazardous Substances Act (15 U.S.C. 1261 et seq.) with regulations issued by the Consumer Product Safety Commission. In subpart (6)(ii), 'hazardous substance' is defined by CERCLA (42 U.S.C. 9601 et seq.) with regulations issued by the Environmental Protection Agency.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_hazardous_substance_definition", "parameters": {"subpart": "(v)", "statute": "Consumer Product Safety Act (15 U.S.C. 2051 et seq.) and Federal Hazardous Substances Act (15 U.S.C. 1261 et seq.)", "agency": "Consumer Product Safety Commission"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_statute_and_agency", "parameters": {"subpart": "(v)", "term": "hazardous substance"}}

705. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1654\_00`

- **Question:** If an irrigation machine has a stationary point, what must be connected to it and for what purpose?
- **Ground truth:** A grounding electrode system must be connected to the machine at the stationary point for lightning protection.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "connect_grounding_electrode_system", "parameters": {"machine": "irrigation machine", "purpose": "lightning protection"}}
- ✓ **VERITROOPER (PASS):** {"name": "connect_grounding_electrode_system", "parameters": {"machine": "irrigation machine", "point": "stationary point", "purpose": "lightning protection"}}

706. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_00`

- **Question:** When connecting deenergized lines to an energized circuit using a conducting wire, which end must an employee attach first?
- **Ground truth:** The employee shall first attach the wire to the deenergized part.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Connecting", "parameters": {"deenergized_part": "first"}}
- ✓ **VERITROOPER (PASS):** {"name": "attach_conducting_wire", "parameters": {"end": "deenergized"}}

707. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1332\_03`

- **Question:** Which two provisions-one from the log hauls section and one from the log decks section-address guarding or controlling chains so employees are not struck?
- **Ground truth:** Paragraph (vi)(c) requires that log haul gear and bull chain drive mechanism shall be guarded (and (vi)(d) requires substantial troughs for the return strand of log haul chains over passageways), while paragraph (3)(iv) prohibits loose chains from overhead canting devices or other equipment from hanging over the log deck in such manner as to strike employees.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Provisions", "parameters": {"log_hauls": "Substantial troughs for the return strand of log haul chains shall be provided over passageways.", "log_decks": "Loose chains from overhead canting devices or other equipment shall not be allowed to hang over the log deck in such manner as to strike employees."}}
- ✗ **VERITROOPER (FAIL):** {"name": "log_hauls", "parameters": {"provision": "vi", "subsection": "d"}}

708. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4448\_01`

MODEL ERROR

- **Question:** Combining paragraph (a)(2) and paragraph (b), if a complainant files a district court action because there was no final decision within 210 days of filing the complaint, what subsequent filing obligation

arises and how soon?

- **Ground truth:** Because there was no final decision of the Secretary within 210 days of the filing of the complaint (a)(2), the complainant may bring the action; then within seven days after filing the federal complaint (b), they must file a file-stamped copy with the Assistant Secretary, ALJ, or ARB, depending on where the proceeding is pending.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "file_stamped_complaint_filing_obligation", "parameters": {"days": "7"}}
- **✗ VERITROOPER (FAIL):** {"name": "filing_obligation", "parameters": {"days": "7"}}

709. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_07`

HALLUCINATION

- **Question:** How often must the limit switch on the powered hand truck operating handle be inspected?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "inspect_limit_switch", "parameters": {"frequency": "at the beginning of each operator"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

710. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_4450\_02`

MODEL ERROR

- **Question:** What is the Statutes at Large citation (124 Stat.) for the Patient Protection and Affordable Care Act, Public Law 111-148?
- **Ground truth:** 124 Stat. 119
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_statutes_at_large_citation", "parameters": {"public_law": "111-148", "statutes_at_large": "124"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_statutes_at_large_citation", "parameters": {"public_law": "111-148"}}

711. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0022\_02`

MODEL ERROR

- **Question:** If particularized written objections are filed and the Assistant Secretary finds that substantial objections have been filed, what action must he take?
- **Ground truth:** He shall afford a formal or informal hearing on the subjects and issues involved under § 1902.13 or § 1902.14, or shall commence a proceeding under § 1902.17.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "afford an informal hearing", "parameters": {}}
- **✓ VERITROOPER (PASS):** {"name": "afford_hearing", "parameters": {"type": "formal or informal", "subjects": "subjects and issues involved"}}

712. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0772\_01`

- **Question:** Combining paragraph (a)(2) and paragraph (a)(4), what feature of the test enclosure accommodates the nebulizer used by the test conductor, and what is its size?
- **Ground truth:** The test enclosure has a 3/4-inch (1.9 cm) hole in front of the test subject's nose and mouth area to accommodate the nebulizer nozzle of the DeVilbiss Model 40 Inhalation Medication Nebulizer (or equivalent) used by the test conductor to spray the threshold check solution.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "test_enclosure_feature", "parameters": {"feature": "hole", "size": "3/4 inch"}}
- **✓ VERITROOPER (PASS):** {"name": "get_test_enclosure_feature", "parameters": {"feature": "hole", "size": "3/4-inch (1.9 cm)"}}

713. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0882\_03`

- **Question:** Reading paragraph (c)(5)(ii)(C)(2) on tagout devices together with the training limitation in paragraph (c)(7)(ii)(F), what does the standard require regarding the attachment of tags?
- **Ground truth:** Tagout device attachment means shall be non-reusable, attachable by hand, self-locking, non-releasable with a minimum unlocking strength of no less than 50 pounds (at least equivalent to a one-piece, all-environment-tolerant nylon cable tie), and tags must be securely attached to energy isolating devices so that they cannot be inadvertently or accidentally detached during use.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "requirement_for_tag_attachment", "parameters": {"standard": "The standard requires that tags must be securely attached to energy isolating devices so that they cannot be inadvertently or accidentally detached during use."}}
- ✓ **VERITROOPER (PASS):** {"name": "standard_requirement", "parameters": {"requirement": "Tags must be securely attached to energy isolating devices so that they cannot be inadvertently or accidentally detached during use.", "standard": "OSHA"}}

714. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4344\_01`

- **Question:** What is the U.S. Reports citation number for U.S. v. Silk?
- **Ground truth:** 331 U.S. 704
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_us_reports_citation", "parameters": {"case_name": "U.S. v. Silk"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_us_reports_citation", "parameters": {"case_name": "U.S. v. Silk"}}

715. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4010\_02`

MODEL ERROR

- **Question:** If the auditor determines there is a significant deficiency in the program, what is triggered?
- **Ground truth:** The employer must ensure certain follow-up actions (as listed under the Deficiencies provision).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "audit_determination", "parameters": {"determination": "significant deficiency"}}
- ✗ **VERITROOPER (FAIL):** {"name": "determine_action", "parameters": {"deficiency": "significant deficiency"}}

716. PASS · Exception · `OSHA\_29CFR\_full\_2024\_2228\_01`

- **Question:** What beverage is excepted from the prohibition on consuming beverages in the regulated area?
- **Ground truth:** Drinking water may be consumed in the regulated area.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "prohibition_of_activities_in_regulated_areas", "parameters": {"food_or_beverage": "drinking water"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_excepted_beverage", "parameters": {"beverage": "drinking water"}}

717. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2190\_02`

MODEL ERROR

- **Question:** If initial monitoring reveals an employee's exposure is below the action level, may the employer discontinue monitoring for that employee unconditionally?
- **Ground truth:** No; the employer may discontinue except as otherwise required by paragraph (e)(5).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "termination_of_monitoring", "parameters": {"initial_monitoring_reveals_exposure_below_action_level": "True"}}

- ✓ **VERITROOPER (PASS):** {"name": "terminate_monitoring", "parameters": {"exposure_level": "below action level", "condition": "except as otherwise required by paragraph (e)(5)"}}

718. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_06`

- **Question:** In Table F-1 (males), what is the age correction value at 4000 Hz for age 32?
- **Ground truth:** 14
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1 (males)", "frequency": "4000 Hz", "age": "32"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1", "gender": "males", "frequency": "4000 Hz", "age": "32"}}

719. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_05`

- **Question:** For Class 1 equipment, what is the DC average retest voltage?
- **Ground truth:** 40,000
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_dc_average_retest_voltage", "parameters": {"class_of_equipment": "1"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retest_voltage", "parameters": {"class_of_equipment": "1", "voltage_type": "DC average"}}

720. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4228\_02`

MODEL ERROR

- **Question:** When a developmental step is missed and the State has an operational status agreement with OSHA under § 1954.3, what additional assurance must the State provide?
- **Ground truth:** The State must provide an assurance that the missed step will not affect the effectiveness of State enforcement in any issues for which the State program has been deemed to be operational.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "provide_assurance", "parameters": {"agreement": "operational status agreement", "section": "1954.3"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_additional_assurance", "parameters": {"agreement": "§ 1954.3", "status": "operational"}}

721. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1226\_00`

MODEL ERROR

- **Question:** Under the maintenance and operation provisions, when servicing a batch-type generator whose carbide charge is exhausted, what must be done to the generating chamber before adding carbide, and according to subpart (D), what restriction applies to the mixture drained out during this process?
- **Ground truth:** Before additional carbide is added, the generating chamber must always be flushed out with water, renewing the water supply in accordance with the manufacturer's instruction card. The water-carbide residue mixture drained from the generator shall not be discharged into sewer pipes or stored in areas near open flames (only clear water from residue settling pits may be discharged into sewer pipes).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "flush_out_generating_chamber", "parameters": {"water_supply": "renewed in accordance with the instruction card furnished by the manufacturer"}}
- ✓ **VERITROOPER (PASS):** {"name": "maintenance_operation", "parameters": {"action": "flush out with water", "restriction": "not be discharged into sewer pipes or stored in areas near open flames"}}

722. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3938\_00`

- **Question:** Under § 1926.1204(d)(8), the employer must provide rescue and emergency equipment needed to comply with paragraph (i). Under what circumstance is the employer NOT required to provide this equipment?
- **Ground truth:** To the extent that the equipment is provided by rescue services.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_rescue_and_emergency_equipment", "parameters": {"needed": "false"}}
- ✓ **VERITROOPER (PASS):** {"name": "NOT required to provide equipment", "parameters": {"circumstance": "the equipment is provided by rescue services"}}

723. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1848\_03`

- **Question:** Per Section III.A on respirators and Section III.B on protective clothing, what is the common triggering condition for both requirements?
- **Ground truth:** Both respirators (III.A) and protective clothing (III.B) are required when asbestos fiber concentrations exceed the permissible exposure limit of 0.1 f/cc.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "respirators_and_protective_clothing", "parameters": {"condition": "operations in which engineering controls or work-practice controls are not adequate or feasible to reduce exposure to the permissible limit"}}
- ✓ **VERITROOPER (PASS):** {"name": "common_triggering_condition", "parameters": {"respirator_condition": "asbestos exposure exceeds the permissible exposure limit (PEL) of 0.1 f/cc", "protective_clothing_condition": "asbestos fiber concentrations exceed the permissible exposure limit"}}

724. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2440\_01`

- **Question:** Combining the employee observation right in paragraph (o)(1) with the observation procedures in paragraph (o)(2), whose exposure measuring or monitoring is being observed and what triggers the requirement for protective equipment to be provided to the observer?
- **Ground truth:** The observation is of the measuring or monitoring of employee exposure to MDA conducted pursuant to paragraph (e). The requirement for the employer to provide protective clothing/equipment or respirators to the observer is triggered when the observation requires entry into areas where the use of protective clothing and equipment or respirators is required.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "observation_procedures", "parameters": {"employee": "affected employees or their designated representatives", "trigger": "entry into areas where the use of protective clothing or equipment is required"}}
- ✓ **VERITROOPER (PASS):** {"name": "employee_observation", "parameters": {"employee": "affected employees, or their designated representatives", "trigger": "entry into areas where the use of protective clothing and equipment or respirators is required"}}

725. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_2019\_01`

WRONG CALC/QUOTE

- **Question:** What minimum percentage of beryllium by weight must materials contain to define a beryllium work area when processed during listed operations?
- **Ground truth:** 0.1 percent by weight
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "define_beryllium_work_area", "parameters": {"percentage": "0"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_beryllium_work_area_percentage", "parameters": {"percentage": "0"}}

726. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_13`

- **Question:** For Class 4 equipment, what is the maximum proof-test current for a 410-mm (16-in) glove?
- **Ground truth:** 24
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_maximum_proof_test_current", "parameters": {"class_of_equipment": "4", "glove_size_mm": "410"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_max_proof_test_current", "parameters": {"class_of_equipment": "4", "glove_size_mm": "410"}}

727. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3110\_01`

- **Question:** Considering both the dispensing location requirement and the tank/pump location requirement, where must a dispensing unit be placed relative to ignition sources, and where must non-integral pumps be located?
- **Ground truth:** Per paragraph (k)(1)(i), dispensing units must be at least 20 feet (6.08 m) from any activity involving fixed sources of ignition, and per paragraph (k)(2)(i), pumps not integral with the dispensing unit must be on shore or on a pier of the solid fill type (except as provided in (k)(2)(ii) and (iii)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "dispensing_unit_placement", "parameters": {"dispensing_location_requirement": "at least 20 feet from any activity involving fixed sources of ignition", "tank_pump_location_requirement": "on shore or on a pier of the solid fill type"}}
- ✓ **VERITROOPER (PASS):** {"name": "location_requirements", "parameters": {"dispensing_location": "at least 20 feet away from ignition sources", "tank_pump_location": "on shore or on a pier of the solid fill type"}}

728. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1317\_00`

- **Question:** According to the definition of 'lumber hauling truck,' which two types of industrial trucks are explicitly excluded from being a lumber hauling truck?
- **Ground truth:** A lift truck or a carrier - a lumber hauling truck is an industrial truck other than a lift truck or a carrier, used for the transport of lumber.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "lumber hauling truck", "parameters": {"definition": "an industrial truck, other than a lift truck or a carrier, used for the transport of lumber"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_excluded_truck_types", "parameters": {"truck_type": "lumber hauling truck"}}

729. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1654\_01`

- **Question:** For a center pivot irrigation machine, if the main disconnecting means is not located at the point of connection of electrical power to the machine, what conditions must it satisfy?
- **Ground truth:** It must be visible and not more than 15.2 m (50 ft) from the machine.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "center_pivot_irrigation_machine_disconnect", "parameters": {"location": "visible and not more than 15.2 m (50 ft) from the machine"}}
- ✓ **VERITROOPER (PASS):** {"name": "center_pivot_irrigation_machine_disconnecting_means", "parameters": {"location": "visible and not more than 15.2 m (50 ft) from the machine"}}

730. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1984\_01`

- **Question:** Shower facilities must be provided in accordance with which specific CFR section?
- **Ground truth:** § 1910.141(d)(3)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Showers", "parameters": {"(i)": "29 CFR 1910.141(d)(3)"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_CFR_section", "parameters": {"section_number": "1910.141", "subsection": "d", "subsubsection": "3"}}

731. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_0075\_03`

MODEL ERROR

- **Question:** If the citation issued to an employer is for a de minimis violation, must it be posted?
- **Ground truth:** No, notices of de minimis violations need not be posted.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "posting_of_citations", "parameters": {"citation_type": "de minimis violation"}}
- ✓ **VERITROOPER (PASS):** {"name": "post_citation", "parameters": {"citation_type": "de minimis violation", "posting_requirement": "No"}}

732. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0168\_02`

MODEL ERROR

- **Question:** Combining the standard in paragraph (c) with the discovery-related provision in the same paragraph, under what two distinct circumstances does the hearing examiner have authority over the motion?
- **Ground truth:** Under paragraph (c), the hearing examiner may grant the motion if the pleadings, affidavits, material obtained by discovery or otherwise, or officially noticed matters show there is no genuine issue as to any material fact and a party is entitled to summary decision; alternatively, the hearing examiner may deny the motion whenever the moving party denies access to information by means of discovery to a party opposing the motion.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "motion_for_summary_decision", "parameters": {"paragraph": "c"}}
- ✓ **VERITROOPER (PASS):** {"name": "circumstances_for_hearing_examiner_authority", "parameters": {"standard": "no genuine issue as to any material fact and that a party is entitled to summary decision", "discovery_related_provision": "the moving party denies access to information by means of discovery to a party opposing the motion"}}

733. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_05`

HALLUCINATION

- **Question:** What is the telephone number for the OSHA office that handles SPA complaints?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_telephone_number", "parameters": {"agency": "OSHA", "complaint_type": "SPA"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

734. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3327\_03`

- **Question:** When using a toeboard with openings, what is the maximum size in the greatest dimension that openings may not exceed?
- **Ground truth:** one inch (2.5 cm)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "toeboard_openings", "parameters": {"size": "1 inch"}}
- ✓ **VERITROOPER (PASS):** {"name": "toeboard_openings", "parameters": {"size": "2.5"}}

735. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_05`

- **Question:** Over what period must the tests be conducted to determine a 'confirmed positive' result?
- **Ground truth:** a three-year period
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "confirmed_positive", "parameters": {"test_results": "3", "period": "3 years"}}
- ✓ **VERITROOPER (PASS):** {"name": "confirmed_positive", "parameters": {"test_results": "two abnormal"}}

BeLPT test results, an abnormal and a borderline test result, or three borderline test results", "period": "three-year"}}}

736. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3390\_02`

- **Question:** Using paragraph (ii) and paragraph (iv) together, for a precast concrete member that is 100 feet long, what is the maximum control line distance from the leading edge, and what must the control line be connected to on each side?
- **Ground truth:** The maximum distance is 50 feet (half the 100-foot length, since that is less than 60 feet, per paragraph (ii)), and under paragraph (iv) the control line must be connected on each side to a guardrail system or wall.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "control_line_distance_and_connection", "parameters": {"distance": "60", "connection": "guardrail system or wall"}}
- ✓ **VERITROOPER (PASS):** {"name": "control_line_distance_and_connection", "parameters": {"max_distance": "60", "connection": "guardrail system or wall"}}

737. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2508\_01`

- **Question:** If air monitoring data reflects workplace conditions with a lower exposure potential than the employer's current operations, does it qualify as objective data under this section?
- **Ground truth:** No, objective data must reflect workplace conditions closely resembling or with a higher exposure potential than the employer's current operations.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "objective_data", "parameters": {"data": "False"}}
- ✗ **VERITROOPER (FAIL):** {"name": "is_objective_data", "parameters": {"data": "air monitoring data", "exposure_potential": "lower"}}

738. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4047\_02`

MODEL ERROR

- **Question:** If personnel are being hoisted during a pile driving operation, may lifts be made on another of the equipment's load lines?
- **Ground truth:** Yes; the prohibition on lifts on other load lines while personnel are hoisted does not apply in pile driving operations.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "k", "parameters": {"11": "i"}}
- ✗ **VERITROOPER (FAIL):** {"name": "check_lifts_during_pile_driving", "parameters": {"operation": "pile driving"}}

739. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2138\_00`

- **Question:** Under paragraph (l)(3)(i), if a currently exposed employee's initial biological monitoring shows CdU at 3 µg/g Cr, β2-M at 300 µg/g Cr, and CdB at 5 µg/lwb, what must the employer provide?
- **Ground truth:** The employer shall provide the minimum level of periodic medical surveillance in accordance with paragraph (l)(4)(i), since all levels are at or below the thresholds.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide the minimum level of periodic medical surveillance", "parameters": {"CdU": "3 \u00b5g/g Cr", "\u03b22-M": "300 \u00b5g/g Cr", "CdB": "5 \u00b5g/lwb"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_minimum_level_of_periodic_medical_surveillance", "parameters": {"CdU": "3", "β2-M": "300", "CdB": "5"}}

740. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2602\_05`

HALLUCINATION

- **Question:** Within what timeframe must a new classification be completed after detecting significant batch variation under A.0.5.1.2?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "A.0.5.1.2", "parameters": {"duration": "before the next batch is produced"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

741. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3703\_00`

MODEL ERROR

- **Question:** Using the field diameter calibration requirement and the worked example for field area, what is the smallest field diameter D (in mm) still within the acceptable range, and does the worked example use a value within that range?
- **Ground truth:** The acceptable range is $100 \pm 2 \mu\text{m}$, so the smallest acceptable diameter is $98 \mu\text{m} = 0.098 \text{ mm}$. The worked example uses $D = 100 \mu\text{m} = 0.1 \text{ mm}$, which is within the acceptable range.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Field Area", "parameters": {"D": "0"}}
- **✗ VERITROOPER (FAIL):** {"name": "min", "parameters": {"a": "0", "b": "0"}}

742. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1138\_03`

- **Question:** When a point of operation injury occurs while a press is used in the PSDI mode, within what timeframe must the employer notify the OSHA-recognized third-party validation organization?
- **Ground truth:** Within five days of the occurrence.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "notify_osha_third_party_validation_organization", "parameters": {"timeframe": "five days"}}
- **✓ VERITROOPER (PASS):** {"name": "notify_osha", "parameters": {"timeframe": "five days"}}

743. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0564\_00`

- **Question:** If an LP-Gas-fueled vehicle has no leaks in its fuel system but its fuel tank is filled beyond the maximum filling capacity specified in paragraph (b)(12)(i), may it be stored inside a garage?
- **Ground truth:** No, both conditions must be met-no leaks AND tanks not filled beyond maximum capacity-so it may not be stored inside.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "LP-Gas-fueled vehicles garaged", "parameters": {"i": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "store_lp_gas_fueled_vehicle", "parameters": {"has_leaks": "false", "is_filled_beyond_capacity": "true"}}

744. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_03`

MODEL ERROR

- **Question:** Within 1926.550, the paragraphs (a)(1), (a)(6), and (a)(11) each carry an OMB control number. Are all three the same?
- **Ground truth:** No. 1926.550(a)(1) is 1218-0115, 1926.550(a)(6) is 1218-0113, and 1926.550(a)(11) is 1218-0054 - all three are different.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "check_omb_control_numbers", "parameters": {"citation": "1926.550", "paragraphs": ["(a)(1)", "(a)(6)", "(a)(11)"]}}
- **✗ VERITROOPER (FAIL):** {"name": "check_omb_control_numbers", "parameters": {"paragraphs": ["(a)(1)",

'(a)(6)', '(a)(11)]"]}]}

745. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1695\_00`

- **Question:** Under paragraph (a)(2), a flexible cord set is found to have a damaged outer jacket that might expose an employee to injury. Combining the visual inspection provisions, what must happen before any employee may use that cord set again?
- **Ground truth:** The defective or damaged item shall be removed from service, and no employee may use it until repairs and tests necessary to render the equipment safe have been made (per (a)(2)(ii)), even though such defects are exactly the kind identified during the required pre-shift visual inspection under (a)(2)(i).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (a)(2) visual inspection provisions", "parameters": {"action": "remove from service", "repair": "make repairs", "test": "perform tests necessary to render the equipment safe"}}
- ✗ **VERITROOPER (FAIL):** {"name": "remove_from_service", "parameters": {"item": "flexible cord set", "reason": "damaged outer jacket that might expose an employee to injury"}}

746. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4330\_02`

- **Question:** Based on functions (e) and (g), which groups outside the participating Federal agencies are field councils expected to engage with, and in what manner?
- **Ground truth:** Under (e), councils provide Federal Executive Boards, Federal Executive Associations, labor union organizations and other employee representatives with information on the administrative and technical aspects of safety and health programs. Under (g), councils develop a cooperative relationship with local community leaders by informing them of the council's functions and objectives and calling on them for support and participation in meetings and activities.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "field councils", "parameters": {"groups": "Federal Executive Boards, Federal Executive Associations, labor union organizations, and other employee representatives", "manner": "provide information on the administrative and technical aspects of safety and health programs"}}
- ✓ **VERITROOPER (PASS):** {"name": "engage_with_groups", "parameters": {"groups": "[Federal Executive Boards, Federal Executive Associations, labor union organizations, local community leaders]", "manner": "provide information on administrative and technical aspects of safety and health programs, and inform them of existing functions and objectives of the council"}}

747. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_4398\_01`

- **Question:** What results when the ARB approves a request to withdraw a petition for review of an ALJ decision, and there are no other pending petitions for review of that decision?
- **Ground truth:** The ALJ's decision will become the final order of the Secretary.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_final_order", "parameters": {"decision": "ALJ"}}
- ✗ **VERITROOPER (FAIL):** {"name": "determine_outcome_of_withdrawal", "parameters": {"entity": "ARB", "request_type": "petition for review", "decision_type": "ALJ decision", "pending_requests": "none"}}

748. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_00`

HALLUCINATION

- **Question:** How many feet of distance must be specified in a voice signal before the operator may begin a boom movement?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "voice signal", "parameters": {"distance": "10"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

749. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_05`

HALLUCINATION

- **Question:** What weighting score is assigned to epidemiological studies versus individual human case reports in the weight-of-evidence approach?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "A.0.3.3", "parameters": {"human data": "reliable, good quality human data shall generally have precedence over other data", "animal data": "positive results from well-conducted animal studies are not necessarily negated by the lack of positive human experience but require an assessment of the robustness, quality and statistical power of both the human and animal data"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

750. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0643\_03`

- **Question:** Given the certification frequency requirement and the report requirement, what must the employer certify and what document must result from the audit process?
- **Ground truth:** Employers must certify that they have evaluated compliance at least every three years, and a report of the findings of the audit must be developed.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "compliance_audit", "parameters": {"certification_frequency": "every three years", "report_requirement": "a report of the findings of the audit"}}
- **✓ VERITROOPER (PASS):** {"name": "certify_and_document", "parameters": {"certification_frequency": "every three years", "document_result": "a report of the findings of the audit"}}

751. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_02`

- **Question:** On the ascending side, the bevel guard or cone shall make an angle of not less than how many degrees with the horizontal?
- **Ground truth:** 45°
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_angle", "parameters": {"x": "45"}}
- **✓ VERITROOPER (PASS):** {"name": "math", "parameters": {"x": "45"}}

752. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0429\_03`

HALLUCINATION

- **Question:** What is the maximum permissible service pressure at the point where oxygen enters the supply line?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "pressure_regulating_devices", "parameters": {"maximum_pressure": "250"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

753. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4293\_03`

- **Question:** If an inspector deems it necessary, may the inspector interview an employee in private?
- **Ground truth:** Yes; the inspector may interview any employee in private if the inspector deems it necessary.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "inspectors", "parameters": {"may_interview": "yes", "in_private": "if deemed necessary"}}
- **✓ VERITROOPER (PASS):** {"name": "deems_necessary", "parameters": {"inspector": "True", "action": "interview an employee in private"}}

754. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1114\_02`

- **Question:** What is the maximum nominal d.c. supply voltage permitted for d.c. control circuits?
- **Ground truth:** 240-volt d.c.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_maximum_dc_supply_voltage", "parameters": {"voltage": "240"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_nominal_voltage", "parameters": {"voltage_type": "d.c.", "circuit_type": "control"}}

755. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_01`

- **Question:** If the administrative law judge determines that a final decision could substantially affect an interested person, may that person be granted party status?
- **Ground truth:** Yes, the administrative law judge may, at his discretion, grant the right to participate as a party.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "determination_of_parties", "parameters": {"interested_person": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "determine_party_status", "parameters": {"person_affected": "True"}}

756. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3972\_00`

- **Question:** Under the monthly inspection rule, what is the minimum retention period for the inspection documentation?
- **Ground truth:** three months
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "monthly inspection documentation retention period", "parameters": {"minimum retention period": "12 months"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_retention_period", "parameters": {"months": "3"}}

757. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4215\_02`

- **Question:** Combining the benchmark requirements with their origin, under what court order were compliance staffing benchmarks required, and how many safety officers did New Mexico propose in its 1992 reassessment?
- **Ground truth:** The benchmarks were required under the 1978 Court Order in AFL-CIO v. Marshall, and New Mexico proposed 7 safety compliance officers in its May 1992 reassessment.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "combining_benchmark_requirements_with_origin", "parameters": {"court_order": "AFL-CIO v. Marshall", "state": "New Mexico", "year": "1992", "s": "7"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_benchmark_requirements", "parameters": {"court_order": "AFL-CIO v. Marshall", "safety_officers": "7"}}

758. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3985\_04`

- **Question:** When equipment is on rails, what is the only exception to needing rail clamps and rail stops?
- **Ground truth:** Portal cranes are excepted from the rail clamps and rail stops requirement.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "rail_clamps_and_rail_stops_exception", "parameters": {"equipment_type": "portal cranes"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_rail_clamps_and_stops", "parameters": {"equipment_type": "portal cranes"}}

759. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_01`

HALLUCINATION

- **Question:** How many days does an employer have to offer a transfer position to an employee who cannot function while using a respirator?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "time_to_offer_transfer_position", "parameters": {"days": "10"}}
- ✔ **VERITROOPER (PASS):** The text does not contain information on the number of days an employer has to offer a transfer position to an employee who cannot function while using a respirator.

760. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_09`

- **Question:** The depth of the floating cone need not exceed how many inches?
- **Ground truth:** 12 inches
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "floating cone depth", "parameters": {"depth": "12"}}
- ✔ **VERITROOPER (PASS):** {"name": "floating_cone_depth", "parameters": {"depth": "12"}}

761. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1206\_03`

MODEL ERROR

- **Question:** Per paragraph (b)(1)(i), construction must follow which regulations, and per (b)(1)(iii), connections must comply with which standard?
- **Ground truth:** Construction and maintenance must follow the U.S. Department of Transportation regulations, 49 CFR parts 171-179, and connections must comply with ANSI B57.1-1965 (the American National Standard Compressed Gas Cylinder Valve Outlet and Inlet Connections).
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "construction_and_connections", "parameters": {"regulations": "29 CFR 1910.119", "standard": "Regulator Connection Standards, 1958, Compressed Gas Association"}}
- ✔ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"paragraph": "(b)(1)(i)", "regulations": "U.S. Department of Transportation, 49 CFR parts 171-179", "standard": "American National Standard Compressed Gas Cylinder Valve Outlet and Inlet Connections, ANSI B57.1-1965"}}

762. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_08`

HALLUCINATION

- **Question:** What form must be filed to certify that air compressor quality meets the 29 CFR 1910.134(i) requirements?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA_certification_form", "parameters": {"form_type": "29 CFR 1910.134(i) Air Compressor Quality Certification", "company_name": "ABC Corporation", "date": "2022-01-01"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

763. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_05`

- **Question:** Do committee terms have to be at least two years duration in every case?
- **Ground truth:** No; terms should be at least two years duration except when the committee is initially organized.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "committee_terms", "parameters": {"duration": "at least two years, except when the committee is initially organized"}}
- ✔ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Do committee terms"}}

have to be at least two years duration in every case?", "answer": "No"}}

764. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4344\_02`

- **Question:** What is the U.S. Reports citation number for Rutherford Food Corporation v. McComb?
- **Ground truth:** 331 U.S. 722
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_us_reports_citation", "parameters": {"case_name": "Rutherford Food Corporation v. McComb"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_us_reports_citation", "parameters": {"case_name": "Rutherford Food Corporation v. McComb"}}

765. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_09`

HALLUCINATION

- **Question:** What is the specific interest rate under 26 U.S.C. 6621 that applies to back pay calculations?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_interest_rate", "parameters": {"year": "2020"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

766. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2861\_00`

- **Question:** Under the Service Contract Act of 1965, what is the contract value threshold that a Federal contract must exceed to require safety and health standards be met?
- **Ground truth:** \$2,500
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Service Contract Act of 1965", "parameters": {"threshold": "2500"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_contract_value_threshold", "parameters": {"year": "1965"}}

767. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_04`

- **Question:** Which edition is ISO 10156-2:2005 (E)?
- **Ground truth:** First Edition
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_edition", "parameters": {"std_number": "ISO 10156-2:2005 (E)"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_edition", "parameters": {"standard": "ISO 10156-2:2005 (E)"}}

768. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4205\_01`

MODEL ERROR

- **Question:** Combining the revised compliance staffing benchmarks in paragraph (c), what is the total number of compliance officers (safety plus health) proposed for Tennessee?
- **Ground truth:** 22 safety plus 14 health compliance officers equals 36 total compliance officers.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "add", "parameters": {"a": "22", "b": "14"}}
- ✗ **VERITROOPER (FAIL):** {"name": "add", "parameters": {"a": "22", "b": "14"}}

769. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1206\_01`

- **Question:** Combining the general inside-building storage rule in (b)(2)(ii) and the fuel-gas cylinder storage limit in (b)(3), what is the minimum distance from highly combustible materials and the maximum total gas capacity allowed for stored fuel-gas cylinders inside a building?
- **Ground truth:** Cylinders must be stored at least 20 feet (6.1 m) from highly combustible materials, and fuel-gas cylinders (except those in actual use or attached ready for use) are limited to a total gas capacity of

2,000 cubic feet (56 m³) or 300 pounds (135.9 kg) of liquefied petroleum gas.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_storage_distance_and_capacity", "parameters": {"distance": "20", "capacity": "2000"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_storage_requirements", "parameters": {"distance_from_combustible_materials": "20", "total_gas_capacity": "2000"}}

770. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_01`

HALLUCINATION

- **Question:** How many inches must the carton-stitching machine guard extend beyond the stitching head?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "carton-stitching machine guard", "parameters": {"extension_length": "3"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

771. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3327\_01`

- **Question:** What is the minimum height of a toeboard from its top edge to the level of the walking/working surface?
- **Ground truth:** three and one-half inches (9 cm)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "minimum toeboard height", "parameters": {"height": "3.5 inches"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_toeboard_height", "parameters": {"height": "9"}}

772. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0772\_03`

- **Question:** Combining the threshold check solution preparation in paragraph (a)(5) with the nebulizer usage guidance in paragraph (a)(13), how much sodium saccharin and how much liquid are involved per nebulizer filling?
- **Ground truth:** The threshold check solution is made by dissolving 0.83 gram of sodium saccharin USP in 100 ml of warm water, and correct use of the nebulizer means approximately 1 ml of that liquid is used at a time in the nebulizer body.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "prepare_threshold_check_solution_and_nebulizer_usage", "parameters": {"sodium_saccharin": "0", "liquid": "1"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_nebulizer_filling", "parameters": {"sodium_saccharin": "0", "liquid": "1"}}

773. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0899\_01`

- **Question:** Combining paragraph (c)(3) and the example it provides, what training schools should the program be similar to for a general fire brigade versus one in the oil refinery industry?
- **Ground truth:** For a general fire brigade, the program should be similar to schools such as the Maryland Fire and Rescue Institute, Iowa Fire Service Extension, West Virginia Fire Service Extension, Georgia Fire Academy, New York State Department Fire Prevention and Control, Louisiana State University Firemen Training Program, or Washington State's Fire Service Training Commission for Vocational Education. For the oil refinery industry, the program should be similar to those conducted by Texas A & M University, Lamar University, Reno Fire School, or the Delaware State Fire School.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "training_schools", "parameters": {"general_fire_brigade": "['Maryland Fire and Rescue Institute', 'Iowa Fire Service Extension', 'West Virginia Fire Service Extension', 'Georgia Fire Academy', 'New York State Department, Fire Prevention and Control', 'Louisiana State University Firemen Training Program', 'Washington State', 'oil_refinery_industry': '['Texas A & M University', 'Lamar University', 'Reno Fire School', 'Delaware State Fire School']"]}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_training_schools", "parameters": {"industry": "general"}}; {"name": "get_training_schools", "parameters": {"industry": "oil refinery"}}

774. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_09`

HALLUCINATION

- **Question:** What temperature must damp sand be maintained at to qualify as adequate floor protection?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "temperature_maintenance", "parameters": {"temperature": "below 85 \u00b0F"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

775. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3863\_00`

- **Question:** Within how many working days after receipt must the employer provide a copy of the physician's written medical opinion to the examined employee?
- **Ground truth:** five working days
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "provide_copy_of_physician_written_medical_opinion", "parameters": {"days": "5"}}
- ✔ **VERITROOPER (PASS):** {"name": "provide_copy_of_physician_opinion", "parameters": {"days": "5"}}

776. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_08`

HALLUCINATION

- **Question:** What is the minimum reach measurement for an alloy steel chain sling used in a vertical hitch?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "Reach", "parameters": {"x": "vertical hitch", "y": "alloy steel chain sling", "z": "minimum"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

777. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_10`

- **Question:** What edition of Sax's Dangerous Properties of Industrial Materials is referenced?
- **Ground truth:** 5th edition, 1979
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "Sax", "parameters": {"edition": "4th"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_edition", "parameters": {"book_title": "Dangerous Properties of Industrial Materials", "author": "Sax"}}

778. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0599\_00`

- **Question:** Under what condition do the applicable rules of paragraph (b) NOT apply to refrigerated storage systems under paragraph (d)?
- **Ground truth:** All applicable rules of paragraph (b) apply unless otherwise noted.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "apply_to_refrigerated_storage_systems", "parameters": {"condition": "unless otherwise noted"}}
- ✔ **VERITROOPER (PASS):** {"name": "apply_rules", "parameters": {"paragraph": "d", "condition": "unless otherwise noted"}}

779. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_3059\_01`

MODEL ERROR

- **Question:** If a spray booth has a frontal area of 8 square feet, is a metal deflector or curtain required at the upper outer edge?
- **Ground truth:** No, the requirement applies only to booths with a frontal area larger than 9 square feet.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "is_metal_deflector_required", "parameters": {"frontal_area": "8"}}
- **✗ VERITROOPER (FAIL):** {"name": "is_metal_deflector_required", "parameters": {"frontal_area": "8"}}

780. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_06`

- **Question:** On what date was § 1902.39(a) suspended indefinitely, effective per the Effective Date Note?
- **Ground truth:** Jan. 20, 1978
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_effective_date_note", "parameters": {"note": "At 43 FR 11196, Mar. 17, 1978, \u00a7 1902.39(a) was suspended indefinitely, effective Jan. 20, 1978."}}
- **✓ VERITROOPER (PASS):** {"name": "get_effective_date", "parameters": {"suspension_date": "January 20, 1978"}}

781. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2155\_07`

- **Question:** What is the 'high level' threshold for cadmium in urine that gives a worker a much greater chance of developing kidney diseases?
- **Ground truth:** greater than 10 micrograms per gram of creatinine (µg/g Cr)
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "high_cadmium_in_urine", "parameters": {"level": "10"}}
- **✓ VERITROOPER (PASS):** {"name": "get_high_level_threshold", "parameters": {"substance": "cadmium in urine", "unit": "micrograms per gram of creatinine", "threshold_value": "10"}}

782. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_01`

HALLUCINATION

- **Question:** What form number must the respondent use to request the award of attorney fees from the ALJ?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Form 20", "parameters": {}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

783. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0478\_01`

HALLUCINATION

- **Question:** What is the maximum allowable air velocity over the open face of a booth during standard (non-electrostatic) spraying operations?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_max_air_velocity", "parameters": {"velocity_type": "standard"}}
- **✗ VERITROOPER (FAIL):** {"name": "max", "parameters": {"a": "100", "b": "60"}}

784. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_06`

HALLUCINATION

- **Question:** What maximum decibel level is required for voice signals to be considered effective under this section?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "effective_decibel_level", "parameters": {"dB": "90"}}
- ✔ **VERITROOPER (PASS):** I don't have enough information to answer the question.

785. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0734\_02`

MODEL ERROR

- **Question:** Based on subparts (i) and (ii), what are the two distinct surrogate-facepiece-based approaches-one for qualitative testing and one for quantitative testing-that can be used in place of modifying the user's actual atmosphere-supplying or powered air-purifying respirator facepiece?
- **Ground truth:** For qualitative fit testing (subpart i), an identical negative pressure air-purifying respirator facepiece with the same sealing surfaces can be used as a surrogate; for quantitative fit testing (subpart ii), a surrogate facepiece with a permanent sampling probe installed can be used to sample air from inside the facepiece.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "fit_testing_of_tight_fitting_atmosphere_supplying_respirators_and_tight_fitting_powered_air_purifying_respirators", "parameters": {"protocol": "OSHA-accepted QNFT protocol", "facepiece": "surrogate facepiece with appropriate filters"}}
- ✔ **VERITROOPER (PASS):** {"name": "surrogate_facepiece_approaches", "parameters": {"qualitative_testing": "using an identical negative pressure air-purifying respirator facepiece with the same sealing surfaces", "quantitative_testing": "installing a permanent sampling probe onto a surrogate facepiece"}}

786. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4247\_00`

- **Question:** Under § 1955.13, if a party believes an administrative law judge is disqualified and wishes to remove him, what two requirements from paragraph (a) must be met regarding timing and supporting documentation, and who rules on the request?
- **Ground truth:** The motion to disqualify and remove the administrative law judge must be filed prior to the time the administrative law judge files his decision, and it must be supported by affidavits setting forth the alleged ground for disqualification. The Chief Administrative Law Judge shall rule upon the motion.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "file a motion to disqualify and remove the administrative law judge", "parameters": {"the_motion_is_filed_prior_to_the_time_the_administrative_law_judge_files_his_decision": "True", "the_motion_must_be_supported_by_affidavits_setting_forth_the_alleged_ground_for_disqualification": "True", "the_Chief_Administrative_Law_Judge": "rules on the motion"}}
- ✖ **VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"question": "Under \u00a7 1955.13, if a party believes an administrative law judge is disqualified and wishes to remove him, what two requirements from paragraph (a) must be met regarding timing and supporting documentation, and who rules on the request?", "requirements": "timing and supporting documentation", "ruling_authority": "Chief Administrative Law Judge"}}

787. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_04`

HALLUCINATION

- **Question:** What form must the lift director complete to record the agreed-upon voice signals?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA 301", "parameters": {"form": "Injury and Illness Incident

Report"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

788. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_01`

HALLUCINATION

- **Question:** At what specific pressure (in psi) must the relief valve on air-pressure containers be set to operate?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "relief valve start-to-discharge pressure setting", "parameters": {"percent of regulator delivery pressure": "200", "Minimum": "300"}}

- ✗ **VERITROOPER (FAIL):** {"name": "get_pressure", "parameters": {"psi": "the requirements of the Code for Unfired Pressure Vessels, Section VIII of the ASME Boiler and Pressure Vessel Code-1968"}}

789. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4065\_02`

- **Question:** To whom do the modification provisions in paragraphs (a) and (b) of § 1926.1434 NOT apply?

- **Ground truth:** They do not apply to modifications made or approved by the U.S. military.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "modification provisions in paragraphs (a) and (b) of § 1926.1434 do not apply to", "parameters": {"x": "the U.S. military"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_exception", "parameters": {"section": "1926.1434", "paragraphs": "['(a)', '(b)']", "exception": "U.S. military"}}

790. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4330\_04`

- **Question:** On whose behalf do field councils act per the introductory text, and according to function (d), what is one way they can assist agencies with inadequate resources?

- **Ground truth:** Per the introductory text, field councils act on behalf of the Secretary or his designees. Per function (d), they promote coordination, cooperation, and sharing of resources and expertise to aid agencies with inadequate or limited resources, for example by conducting inspections and evaluations of the agencies' safety and health programs on request.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "field councils act on behalf of the Secretary", "parameters": {"(a)": "act as a clearinghouse on information and data on occupational accidents, injuries, and illnesses and their prevention", "(b)": "plan, organize and conduct field council meetings or programs which will give technical advice and information on occupational safety and health to representatives of participating agencies and employee organizations", "(c)": "promote improvement of safety and health programs and organizations in each Federal agency represented or participating in council activities", "(d)": "promote coordination, cooperation, and sharing of resources and expertise to aid agencies with inadequate or limited resources"}}

- ✓ **VERITROOPER (PASS):** {"name": "act_on_behalf_and_assist", "parameters": {"behalf_of": "the Secretary or his designees", "assistance_method": "organize and conduct training programs for employee representatives, collateral duty and professional safety and health personnel"}}

791. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_03`

- **Question:** If an employer follows the guidance in appendix E to estimate available heat energy, will OSHA consider the employer in compliance with paragraph (I)(8)(ii)?

- **Ground truth:** Yes; OSHA will deem employers following the guidance in appendix E to be in compliance with paragraph (I)(8)(ii).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "OSHA considers employer in compliance", "parameters": {"compliance": "True"}}

- ✖ **VERITROOPER (FAIL):** {"name": "OSHA_compliance", "parameters": {"guidance_followed": "True"}}

792. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_2159\_05`

- **Question:** Calculate the overall analytical error using $OAE = \pm[|Bias| + (1.96)(CV1(pooled))(100\%)]$ where the absolute bias is 3.5% and the pooled coefficient of variation CV1(pooled) is 0.05.

- **Ground truth:** $\pm 13.3\%$

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "OAE", "parameters": {"Bias": "3", "CV1": "0"}}

- ✓ **VERITROOPER (PASS):** {"name": "OAE", "parameters": {"Bias": "3", "CV1": "0"}}

793. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_05`

HALLUCINATION

- **Question:** What is the minimum compressive load rating that the safety-toe footwear must withstand under standard Z41.1-1967?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_min_compress_load", "parameters": {"standard": "Z41.1-1967"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The evidence only mentions that safety shoes shall conform to the requirements of American National Standard for Men's Safety-Toe Footwear, Z41.1-1967, but it does not provide the specific details of the standard, including the minimum compressive load rating.

794. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_06`

HALLUCINATION

- **Question:** Which specific government office reviews the licensed physician's written medical report for compliance?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA_Review", "parameters": {"report": "Licensed physician"}}

- ✓ **VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question.

795. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4446\_02`

MODEL ERROR

- **Question:** If the Secretary issues a decision upon discretionary review for which judicial review is available, within what time must a petition for review be filed?

- **Ground truth:** Within 60 days after the issuance of the final order.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "time_to_file_petition_for_review", "parameters": {"days": "10"}}

- ✓ **VERITROOPER (PASS):** {"name": "time_to_file_petition", "parameters": {"days": "60"}}

796. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_05`

- **Question:** The cone shall be made of not less than what U.S. gauge sheet steel?

- **Ground truth:** No. 18 U.S. gauge

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_U_S_gauge_sheet_steel", "parameters": {"U_S_gauge": "18"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_cone_gauge", "parameters": {"gauge": "18"}}

797. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4215\_00`

- **Question:** How many total compliance officers (safety and health combined) were approved as revised benchmarks, and on what date did the Assistant Secretary approve them?
- **Ground truth:** A total of 10 compliance officers (7 safety and 3 health) were approved on August 11, 1994.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_compliance_officers", "parameters": {"state": "Maryland", "safety_compliance_officers": "36", "health_compliance_officers": "18", "approval_date": "July 18, 1985"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_compliance_officers", "parameters": {"total_officers": "10", "approval_date": "August 11, 1994"}}

798. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_07`

HALLUCINATION

- **Question:** What is the minimum number of members a safety and health committee must have to be considered 'certified' under this section?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "committee_certification", "parameters": {"members": "2"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The text does not specify the minimum number of members required for a safety and health committee to be considered 'certified'.

799. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0751\_01`

- **Question:** If an employee wears a respirator voluntarily when its use is not required by this section or by the employer, must the employer provide the basic advisory information from appendix D?
- **Ground truth:** Yes, the employer must provide the appendix D advisory information in any written or oral format.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_basic_advisory_information", "parameters": {"required": "true"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_basic_advisory_information", "parameters": {"required_by_section": "false", "required_by_employer": "false"}}

800. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0458\_00`

- **Question:** Under the piping requirements, a line conveying Category 1 flammable liquids leading to a wharf must have a readily accessible block valve. Per the combined requirements of subparts (d) and (e), where must this valve be located, and is this in addition to or instead of the general valve requirement?
- **Ground truth:** It is in addition to the requirement of subpart (d) that piping systems contain a sufficient number of valves to operate the system properly. The block valve must be located on shore near the approach to the wharf and outside of any diked area. Where more than one line is involved, the valves shall be grouped in one location.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "block_valve_location", "parameters": {"location": "on shore near the approach to the wharf and outside of any diked area", "requirement_type": "in addition to"}}
- ✓ **VERITROOPER (PASS):** {"name": "valve_location", "parameters": {"category": "1", "location": "on shore near the approach to the wharf and outside of any diked area", "addition": "in addition to"}}

801. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_02`

HALLUCINATION

- **Question:** How many proof tests must a sling manufacturer perform before a sling is approved for sale?
- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "proof testing", "parameters": {"number of tests": "1"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

802. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_05`

HALLUCINATION

- **Question:** How often must fit testing be repeated under this section's respirator program?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "fit testing", "parameters": {"frequency": "at least annually"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The text does not specify how often fit testing must be repeated under this section's respirator program.

803. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1138\_02`

- **Question:** When a replacement non-critical component is not of the same manufacture and design but is determined by the third-party validation organization to be equivalent by similarity analysis, may the employer recertify without revalidation?
- **Ground truth:** Yes; recertification by the employer without revalidation is permitted when the component is determined equivalent by similarity analysis.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "recertification_by_similarity_analysis", "parameters": {"is_equivalent": "True", "is_same_manufacture_and_design": "False"}}
- ✔ **VERITROOPER (PASS):** {"name": "recertify_without_revalidation", "parameters": {"same_manufacture_and_design": "False", "equivalent_by_similarity_analysis": "True"}}

804. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4450\_00`

- **Question:** On what date was the Patient Protection and Affordable Care Act, Public Law 111-148, signed into law?
- **Ground truth:** March 23, 2010
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "get_date_signed_into_law", "parameters": {"law_name": "Patient Protection and Affordable Care Act, Public Law 111-148", "date_signed": "March 23, 2010"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_date_signed_into_law", "parameters": {"public_law": "111-148"}}

805. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_05`

HALLUCINATION

- **Question:** Within how many hours of a worker being substituted must the voice signal meeting be reconvened?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "reconvene_voice_signal_meeting", "parameters": {"hours": "1"}}
- ✔ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

806. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3809\_01`

- **Question:** Within how many days must the evaluation at the CBD diagnostic center be scheduled?
- **Ground truth:** 30 days
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "CBD diagnostic center", "parameters": {"days": "30"}}
- ✔ **VERITROOPER (PASS):** {"name": "schedule_evaluation", "parameters": {"days": "30"}}

807. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_04`

HALLUCINATION

- **Question:** How long does OSHA have to complete its investigation after a SPA complaint is filed?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "OSHA investigation timeline", "parameters": {"days": "180"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The text only discusses the process of filing a complaint, but it does not mention the timeframe for OSHA to complete its investigation after a complaint is filed.

808. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2632\_02`

MODEL ERROR

- **Question:** If a substance shows a high potency in animals but only a low frequency of occurrence in humans, which sub-category applies?
- **Ground truth:** Sub-category 1A, since high potency in animals presumes potential to produce significant sensitization.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "classify_substance", "parameters": {"substance_name": "substance_name", "human_frequency": "low", "animal_potency": "high"}}
- **✗ VERITROOPER (FAIL):** {"name": "classify_skin_sensitizer", "parameters": {"human_frequency": "low", "animal_potency": "high"}}

809. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3305\_01`

- **Question:** Combining subpart (iii) and subpart (iv), what two requirements must a mobile scaffold stairway-type ladder step satisfy regarding width and tread surface?
- **Ground truth:** It must have a minimum step width of 11 1/2 inches (30 cm) per subpart (iii), and it must have slip-resistant treads on all steps and landings per subpart (iv).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Stairway-type ladders", "parameters": {"(iii)": "Have a minimum step width of 11 1/2 inches (30 cm)", "(iv)": "Have slip-resistant treads on all steps and landings"}}
- **✓ VERITROOPER (PASS):** {"name": "ladder_requirements", "parameters": {"width": "11 1/2 inches", "tread_surface": "slip-resistant"}}

810. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_11`

- **Question:** For a Class 4 glove, what is the AC proof test value (mm)?
- **Ground truth:** 127
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_ac_proof_test_value", "parameters": {"class_of_glove": "4"}}
- **✓ VERITROOPER (PASS):** {"name": "get_AC_proof_test_value", "parameters": {"class_of_glove": "4"}}

811. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_01`

- **Question:** When official notice is taken of a material fact not in the record, what conditions must be satisfied for it to be proper?
- **Ground truth:** The parties must be given adequate notice of the matters noticed and adequate opportunity to show the contrary.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "official_notice", "parameters": {"material_fact": "a material fact not in the record", "conditions": "the parties shall be given adequate notice, at the hearing or by reference in the presiding hearing examiner"}}
- **✓ VERITROOPER (PASS):** {"name": "official_notice", "parameters": {"condition1": "the parties shall be given"}}

adequate notice, at the hearing or by reference in the presiding hearing examiner", "condition2": "the parties shall be given adequate opportunity to show the contrary"}}

812. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3512\_03`

HALLUCINATION

- **Question:** What is the maximum spacing allowed between cleats used to skidproof ramped walkway surfaces?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_max_cleat_spacing", "parameters": {"cm": "14"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

813. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4587\_00`

- **Question:** If a physician detects a medical condition during the examination that is unrelated to occupational exposure, must the employer have it revealed in the written opinion?
- **Ground truth:** No, the employer shall instruct the physician not to reveal specific findings or diagnoses unrelated to occupational exposure in the written opinion.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "physician_written_opinion", "parameters": {"reveal_unrelated_conditions": "False"}}
- **✓ VERITROOPER (PASS):** {"name": "written_opinion", "parameters": {"reveal_unrelated_conditions": "False"}}

814. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4330\_03`

- **Question:** Using function (f) and function (c), how do field councils address both locally peculiar safety problems and agency-level program improvement?
- **Ground truth:** Function (f) directs councils to evaluate safety and health problems peculiar to local conditions and facilitate solutions through council activities, while function (c) directs them to promote improvement of safety and health programs and organizations in each Federal agency represented or participating in council activities.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "§ 1960.87", "parameters": {"f": "To evaluate the safety and health problems peculiar to local conditions and facilitate solutions to these problems through council activities.", "c": "To promote improvement of safety and health programs and organizations in each Federal agency represented or participating in council activities."}}
- **✓ VERITROOPER (PASS):** {"name": "evaluate_and_promote", "parameters": {"safety_problems": "peculiar to local conditions", "program_improvement": "in each Federal agency represented or participating in council activities"}}

815. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_01`

- **Question:** When disconnecting equipment from an energized circuit by means of a conducting wire, which end must be removed first?
- **Ground truth:** The employee shall remove the source end first.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "disconnecting_equipment", "parameters": {"end_to_remove_first": "source end"}}
- **✓ VERITROOPER (PASS):** {"name": "remove_conductor", "parameters": {"end": "source"}}

816. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_02`

HALLUCINATION

- **Question:** What penalty applies if a receptacle is installed that does not conform to the ratings in Table S-5?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "violation_penalty", "parameters": {"violation_type": "receptacle_rating_violation"}}
- **✓ VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question. The evidence only provides the ratings for receptacles and circuits, but it does not mention any penalties for non-conformance. Therefore, I cannot provide a JSON for a function call that answers the question.

817. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4293\_01`

- **Question:** If a representative's participation interferes with a fair and orderly inspection, can the inspector deny that person the right of accompaniment?
- **Ground truth:** Yes; inspectors are authorized to deny the right of accompaniment to any person whose participation interferes with a fair and orderly inspection.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "deny_accompaniment", "parameters": {"representative": "True"}}
- **✓ VERITROOPER (PASS):** {"name": "deny_accompaniment", "parameters": {"participation": "interferes with a fair and orderly inspection"}}

818. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1157\_00`

- **Question:** Under what circumstance does the general gear guarding requirement of paragraph (f)(1) NOT apply?
- **Ground truth:** It does not apply to hand-operated gears used only to adjust machine parts and which do not continue to move after hand power is removed.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "general gear guarding requirement of paragraph (f)(1)", "parameters": {"circumstance": "hand-operated gears used only to adjust machine parts and which do not continue to move after hand power is removed"}}
- **✓ VERITROOPER (PASS):** {"name": "paragraph (f)(1) exception", "parameters": {"type": "hand-operated gears used only to adjust machine parts and which do not continue to move after hand power is removed"}}

819. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0370\_00`

- **Question:** For abrasive cutting-off wheels up to and including 20 inches in diameter, the thickness must not be more than what fraction of their diameter?
- **Ground truth:** one forty-eighth
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "calculate_fraction", "parameters": {"diameter": "20", "thickness": "1/48"}}
- **✓ VERITROOPER (PASS):** {"name": "fraction_of_diameter", "parameters": {"diameter": "20", "fraction": "1/48"}}

820. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_00`

- **Question:** Under § 1902.39, evaluation of the State plan's actual operations must occur for at least how long following publication of the certification in the Federal Register before the Assistant Secretary determines eligibility for an 18(e) determination?
- **Ground truth:** 1 year

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "evaluation_of_state_plan", "parameters": {"duration": "1 year"}}
- ✓ **VERITROOPER (PASS):** {"name": "evaluate_state_plan", "parameters": {"years": "1"}}

821. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2902\_02`

HALLUCINATION

- **Question:** What maximum permissible concentration of dust is allowed in respirator air before it is considered a harmful quantity?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "maximum_use_concentration", "parameters": {"MUC": "1000"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

822. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2784\_00`

MODEL ERROR

- **Question:** For an OSHA regulated substance with exposure monitoring and medical surveillance requirements, if the action level is routinely exceeded in a laboratory, which paragraphs apply?
- **Ground truth:** Paragraphs (d) and (g)(1)(ii) of this section shall apply.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA regulated substance with exposure monitoring and medical surveillance requirements", "parameters": {"action_level_exceeded": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "apply_paragraphs", "parameters": {"action_level_exceeded": "True", "exposure_monitoring": "True", "medical_surveillance": "True"}}

823. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0168\_00`

- **Question:** Under paragraph (a), a party moves for summary decision and serves the motion. Combining this with paragraph (b), where must the opposing party file its affidavits and within what time limit, and to whom must copies be served?
- **Ground truth:** Under paragraph (a), the opposing party may serve opposing affidavits within 10 days after service of the motion. Under paragraph (b), the filing must be with the hearing examiner, and copies must be served in accordance with § 1905.21.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "filing_opposing_affidavits", "parameters": {"location": "with the administrative law judge", "time_limit": "10 days after service of the motion", "service": "on all the parties"}}
- ✓ **VERITROOPER (PASS):** {"name": "serve_documents", "parameters": {"filing_location": "with the hearing examiner", "time_limit": "within 10 days after service of the motion", "service_recipient": "the party who made the motion"}}

824. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4520\_02`

- **Question:** If no timely objection is filed with respect to the findings or preliminary order, what happens, and what limitation applies?
- **Ground truth:** The findings and/or the preliminary order become the final decision of the Secretary, not subject to judicial review.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "outcome_if_no_timely_objection", "parameters": {"findings_or_preliminary_order": "become the final decision of the Secretary", "limitation": "not subject to judicial review"}}
- ✓ **VERITROOPER (PASS):** {"name": "outcome_if_no_timely_objection", "parameters": {"findings_become_final": "True", "preliminary_order_becomes_final": "True", "limitation": "not subject to judicial review"}}

825. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4455\_08`

HALLUCINATION

- **Question:** What standard does the ALJ use to determine whether a complaint was frivolous or brought in bad faith?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "ALJ_determines_frivolous_or_bad_faith", "parameters": {"complaint": "complaint", "standard": "reasonable attorney"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question. The text only mentions that the respondent can request an award of attorney fees not exceeding \$1,000 from the ALJ if the respondent alleges that the complaint was frivolous or brought in bad faith, but it does not specify the standard used by the ALJ to make this determination.

826. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0168\_03`

- **Question:** A hearing examiner denies a summary decision motion under paragraph (c). Combining paragraph (f) with that denial, what conditions must be met before an interlocutory appeal to the Assistant Secretary is allowed, and does such appeal stay the proceeding?
- **Ground truth:** Under paragraph (f), the denial is not subject to interlocutory appeal unless the hearing examiner certifies in writing (1) that the ruling involves an important question of law or policy as to which there is substantial ground for difference of opinion, and (2) that an immediate appeal may materially advance the ultimate termination of the proceeding. Even if allowed, the interlocutory appeal does not stay the proceeding before the hearing examiner unless the Assistant Secretary so orders.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "interlocutory appeal conditions and proceeding stay", "parameters": {"conditions": "The ruling involves an important question of law or policy as to which there is substantial ground for difference of opinion, and an immediate appeal from the ruling may materially advance the ultimate termination of the proceeding.", "stay": "No, unless the Assistant Secretary so orders."}}
- **✓ VERITROOPER (PASS):** {"name": "interlocutory_appeal_conditions", "parameters": {"denial_reason": "paragraph (c)", "certification_requirements": "['important question of law or policy', 'substantial ground for difference of opinion', 'immediate appeal may materially advance ultimate termination']", "stay_proceeding": "false"}}

827. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3386\_03`

- **Question:** When the path to a point of access is not in use, what must be done at the point where the path intersects the warning line erected around the work area?
- **Ground truth:** A rope, wire, chain, or other barricade equivalent in strength and height to the warning line must be placed across the path, or the path must be offset so a person cannot walk directly into the work area.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "intersecting_path_procedure", "parameters": {"procedure": "a rope, wire, chain, or other barricade, equivalent in strength and height to the warning line, shall be placed across the path at the point where the path intersects the warning line erected around the work area, or the path shall be offset such that a person cannot walk directly into the work area"}}
- **✓ VERITROOPER (PASS):** {"name": "place_barricade", "parameters": {"location": "at the point where the path intersects the warning line erected around the work area"}}

828. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3703\_04`

- **Question:** According to Appendix E, what edition of the ILO Guidelines must be used both for classifying chest X-ray film (paragraph (c)) and for classifying digitally-acquired chest X-rays (paragraph (d))?
- **Ground truth:** Both paragraphs (c) and (d) require the revised edition 2011 of the Guidelines for the Use of the ILO International Classification of Radiographs of Pneumoconioses.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_chest_xray", "parameters": {"edition": "2011", "type": "both"}}
- ✓ **VERITROOPER (PASS):** {"name": "ILO Guidelines", "parameters": {"edition": "2011"}}

829. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_2159\_02`

- **Question:** Calculate the total cadmium W using $W = (C)(\text{sample vol, mL})(DF)$ where $C = 0.08 \mu\text{g/mL}$, sample volume = 10 mL, and dilution factor $DF = 5$.
- **Ground truth:** 4 μg
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "W", "parameters": {"C": "0", "sample_vol": "10", "DF": "5"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_W", "parameters": {"C": "0", "sample_vol_mL": "10", "DF": "5"}}

830. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_07`

- **Question:** How often must rubber insulating sleeves be tested after the first issue?
- **Ground truth:** every 12 months thereafter
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "test_frequency", "parameters": {"equipment_type": "Rubber insulating sleeves", "frequency": "every 12 months"}}
- ✓ **VERITROOPER (PASS):** {"name": "test_frequency", "parameters": {"equipment_type": "Rubber insulating sleeves", "frequency": "every 12 months"}}

831. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_06`

- **Question:** The lower edge of the cone shall be rolled to a minimum diameter of what measurement?
- **Ground truth:** one-half inch
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_cone_lower_edge_diameter", "parameters": {"diameter": "one-half inch"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_minimum_diameter", "parameters": {"diameter": "one-half inch"}}

832. PASS · Exception · `OSHA\_29CFR\_full\_2024\_2228\_00`

- **Question:** Although food, beverages, smoking products, and cosmetics are generally prohibited in the regulated area, under what circumstance may these activities be conducted?
- **Ground truth:** These activities may be conducted in the lunchrooms, change rooms and showers required under paragraphs (i)(1)-(i)(3) of this section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "prohibited activities in regulated areas", "parameters": {"exception_areas": "lunchrooms, change rooms and showers"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_circumstance", "parameters": {"activity": "food, beverages, smoking products, and cosmetics", "circumstance": "in the lunchrooms, change rooms and showers required under paragraphs (i)(1)-(i)(3) of this section"}}

833. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2508\_00`

- **Question:** If an employee is using a respirator while working with respirable crystalline silica, does the employee exposure account for the protection provided by the respirator?
- **Ground truth:** No, employee exposure means the exposure that would occur if the employee were not using a respirator.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "employee_exposure", "parameters": {"x": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "employee_exposure", "parameters": {"respirator": "False"}}

834. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0219\_06`

- **Question:** What is the Techstreet fax number?
- **Ground truth:** 734-780-2046
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get fax number", "parameters": {"store": "Techstreet", "phone": "877-699-9277"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_fax_number", "parameters": {"company": "Techstreet"}}

835. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0676\_02`

- **Question:** If there is indication that exposures may have risen over permissible exposure levels since prior monitoring, what monitoring is required?
- **Ground truth:** Periodic monitoring shall be conducted.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "periodic monitoring", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "Periodic monitoring", "parameters": {"possibility": "exposures have risen over permissible exposure levels since prior monitoring"}}

836. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_06`

HALLUCINATION

- **Question:** What penalty applies to an agency that obstructs an OSHA inspector's entry under paragraph (b)?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"number": "8", "letter": "B"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

837. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_3034\_00`

- **Question:** A temporary workplace has 160 employees. Using Table D-65.2, what is the minimum number of toilet seats required?
- **Ground truth:** 4
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Table D-65.2", "parameters": {"Number of employees": "160"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_toilet_seats", "parameters": {"employees": "160"}}

838. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1848\_04`

MODEL ERROR

- **Question:** Combining the friability information in Section I.C with the disposal waste types in Section IV.A, why would vinyl-asbestos floor tile removed during demolition be considered a disposal concern?
- **Ground truth:** Although vinyl-asbestos floor tile is nonfriable and generally does not emit fibers (I.C), building products containing asbestos removed during renovation or demolition are listed as a waste in IV.A.5, and per I.C these materials can emit fibers if cut, sawed, or broken during demolition.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "disposal_concern", "parameters": {"waste_type": "vinyl-asbestos floor tile", "friability": "non-friable", "concern": "emits airborne fibers when subjected to aggressive operations such as sanding or sawing"}}
- ✓ **VERITROOPER (PASS):** {"name": "disposal_concern", "parameters": {"waste_type": "vinyl-asbestos floor tile", "friability": "nonfriable", "demolition": "True"}}

839. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_11`

HALLUCINATION

- **Question:** What is the maximum slot width permitted for a female handle on a metal mesh choker sling?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_max_slot_width", "parameters": {"handle_type": "female", "sling_type": "metal mesh choker sling"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

840. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4330\_01`

- **Question:** Combining function (d) and function (b), what types of activities can field councils use both to aid agencies with limited resources and to give technical advice to participating agencies?
- **Ground truth:** Function (d) allows councils to organize and conduct training programs for employee representatives, collateral duty and professional safety and health personnel, coordinate or promote programs for inspections, or conduct inspections and evaluations. Function (b) covers planning, organizing and conducting field council meetings or programs that give technical advice and information on occupational safety and health to representatives of participating agencies and employee organizations.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "function (d) and function (b)", "parameters": {"d": "To promote coordination, cooperation, and sharing of resources and expertise to aid agencies with inadequate or limited resources.", "b": "To plan, organize and conduct field council meetings or programs which will give technical advice and information on occupational safety and health to representatives of participating agencies and employee organizations."}}
- **✓ VERITROOPER (PASS):** {"name": "organize and conduct training programs", "parameters": {"type": "technical advice and information on occupational safety and health", "target": "employee representatives, collateral duty and professional safety and health personnel"}}

841. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4517\_02`

- **Question:** Based on the 'Who may file' provision in paragraph (a) and the 'Time for filing' provision in paragraph (d), can a person other than the complainant file on his or her behalf, and within what time limit?
- **Ground truth:** Yes. Paragraph (a) allows a complaint to be filed by any person on the complainant's behalf, and paragraph (d) requires this be done within 180 days after the alleged violation of MAP-21 occurs.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Filing of retaliation complaint", "parameters": {"who may file": "A person who believes that he or she has been discharged or otherwise retaliated against by any person in violation of CFPA may file, or have filed by any person on his or her behalf", "time for filing": "Within 180 days after an alleged violation of CFPA occurs"}}
- **✓ VERITROOPER (PASS):** {"name": "file_complaint", "parameters": {"who": "any person on behalf of the complainant", "time_limit": "180 days"}}

842. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_02`

HALLUCINATION

- **Question:** Which specific certified language proficiency test must the operator and signal person pass to demonstrate effective communication?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "language and literacy requirements", "parameters": {"literacy": "relevant to the work", "language": "the language of the certification"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

843. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_04`

- **Question:** Pending a final medical determination, can the employer return an employee to former job status if the initial removal resulted from a final medical determination that differed from the initial physician's findings?
- **Ground truth:** No; in that case the employer shall await a final medical determination.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"option": "return", "exception": "initial removal due to differing final medical determination"}}
- ✓ **VERITROOPER (PASS):** {"name": "employer_options_pending_final_medical_determination", "parameters": {"option": "return", "exception": "initial removal resulted from a final medical determination that differed from the initial physician"}}

844. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1126\_02`

- **Question:** To what presses does the directed component of the inspection program (paragraph (e)(1)(ii)) NOT apply?
- **Ground truth:** Paragraph (e)(1)(ii) does not apply to presses that comply with paragraphs (b)(13) and (14) of this section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (e)(1)(ii)", "parameters": {"iii": "presses that comply with paragraphs (b)(13) and (14) of this section"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_presses_not_applying_to_directed_inspection", "parameters": {"presses": "presses that comply with paragraphs (b)(13) and (14) of this section"}}

845. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1202\_02`

MODEL ERROR

- **Question:** Under paragraphs (d)(1)(iii) and (d)(1)(iv), what two distinct hazards require protective measures, and what is required for each?
- **Ground truth:** Under (d)(1)(iii), arc welding in wet conditions or high humidity requires special protection against electric shock. Under (d)(1)(iv), pressure testing of pipelines requires protecting workers and the public against injury from blowing out of closures or pressure restraining devices and against expulsion of trapped loose dirt.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_hazard_prevention_measures", "parameters": {"hazard_1": "hazard identification and awareness", "hazard_2": "notification of appropriate persons"}}
- ✓ **VERITROOPER (PASS):** {"name": "protective_measures", "parameters": {"hazard1": "electric shock", "requirement1": "special protection", "hazard2": "injury from blowing out of closures or other pressure restraining devices", "requirement2": "protection for workers and the public"}}

846. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_10`

- **Question:** Which ASTM standard is referenced for the Reid Method of vapor pressure measurement?
- **Ground truth:** ASTM D-323-58
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "ASTM D323-68", "parameters": {"Standard Method of Test for Vapor Pressure of Petroleum Products": "Reid Method"}}
- ✓ **VERITROOPER (PASS):** {"name": "ASTM", "parameters": {"D": "323", "year": "58"}}

847. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0362\_00`

- **Question:** Regarding controller contacts carrying the main motor current, under what condition do they NOT need to be copper to carbon or equal?
- **Ground truth:** Where the circuit is broken at two or more points simultaneously (and the requirement also does not apply to the oil-immersed type).

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "controller_contacts_exception", "parameters": {"condition": "the circuit is broken at two or more points simultaneously"}}
- ✓ **VERITROOPER (PASS):** {"name": "controller_contacts_exception", "parameters": {"condition": "the circuit is broken at two or more points simultaneously"}}

848. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4587\_01`

- **Question:** If records of previous medical examinations are otherwise available to the examining physician, must the employer provide the names and addresses of the physicians who provided them?
- **Ground truth:** No, those names and addresses need only be provided if such records are not otherwise available to the examining physician.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_names_and_addresses", "parameters": {"x": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_names_and_addresses", "parameters": {"required": "false"}}

849. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0022\_01`

- **Question:** If no notice of hearing is provided by the Assistant Secretary under paragraph (e), how may an interested person obtain an informal hearing concerning the proposed plan?
- **Ground truth:** By filing particularized written objections within 30 days following publication of the notice in the Federal Register.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "informal hearing", "parameters": {"interested person": "file particularized written objections within 30 days following publication of the notice in the Federal Register"}}
- ✓ **VERITROOPER (PASS):** {"name": "request_informal_hearing", "parameters": {"particularized_written_objections": "filed within 30 days following publication of the notice in the Federal Register"}}

850. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3938\_03`

- **Question:** Per § 1926.1204(e)(2), atmospheric hazards must be continuously monitored. Under what conditions does this requirement NOT apply?
- **Ground truth:** It does not apply if the employer can demonstrate that the equipment for continuously monitoring a hazard is not commercially available, or that periodic monitoring is of sufficient frequency to ensure the atmospheric hazard is being controlled at safe levels.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "continuously monitor atmospheric hazards", "parameters": {"hazards": "not commercially available or periodic monitoring is of sufficient frequency to ensure that the atmospheric hazard is being controlled at safe levels"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_continuous_monitoring", "parameters": {"condition": "equipment for continuously monitoring a hazard is not commercially available or periodic monitoring is of sufficient frequency to ensure that the atmospheric hazard is being controlled at safe levels"}}

851. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1008\_01`

- **Question:** Combining the guidance on rope locations needing particular inspection care (subpart (i)) with the special-care provision in subpart (iii), what specific situations call for particular inspection care?
- **Ground truth:** Particular care must be taken to inspect ropes at locations in contact with equalizer sheaves or other sheaves where rope travel is limited, or with saddles, and particular care must also be taken in the inspection of nonrotating rope.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "rope_inspection", "parameters": {"location": "sections in contact with equalizer sheaves or other sheaves where rope travel is limited, or with saddles", "condition": "idle for a period of a month or more due to shutdown or storage of a crane on which it is installed"}}
- ✓ **VERITROOPER (PASS):** {"name": "inspect_rope", "parameters": {"location": "sections in contact with equalizer sheaves or other sheaves where rope travel is limited, or with saddles", "type": "nonrotating rope"}}

852. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_04`

HALLUCINATION

- **Question:** What form number must an employee use to file a complaint under this part?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA-13", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The evidence provided does not contain enough information to answer the question.

853. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3990\_01`

- **Question:** Per § 1926.1420(c), the operator's reception of signals must be by what type of system?
- **Ground truth:** A hands-free system
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1926.1420(c)", "parameters": {"system type": "hands-free"}}
- ✗ **VERITROOPER (FAIL):** {"name": "answer_question", "parameters": {"section": "1926.1420(c)", "query": "type of system for operator"}}

854. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0734\_01`

- **Question:** Combining the qualitative fit testing method in subpart (i) and the restoration requirement in subpart (iii), if a respirator user's actual facepiece is temporarily converted into a negative pressure respirator with appropriate filters for qualitative fit testing, what condition must be satisfied before workplace use?
- **Ground truth:** The temporary conversion modifications must be completely removed and the facepiece restored to its NIOSH-approved configuration before it can be used in the workplace.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "restoration_requirement", "parameters": {"condition": "any modifications to the respirator facepiece for fit testing must be completely removed, and the facepiece restored to NIOSH-approved configuration"}}
- ✓ **VERITROOPER (PASS):** {"name": "restoration_requirement", "parameters": {"condition": "completely removed and restored to NIOSH-approved configuration"}}

855. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0044\_04`

- **Question:** Under § 1902.39(f), particularized written objections requesting an informal hearing must be filed within how many days following publication of the notice in the Federal Register?
- **Ground truth:** 35 days
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "particularized written objections", "parameters": {"days": "35"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_days_for_filing_written_objections", "parameters": {"days": "35"}}

856. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_02`

- **Question:** For Category 1 flammable liquids, what is the maximum boiling point?
- **Ground truth:** at or below 95 °F (35 °C)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_flammable_liquid_category", "parameters": {"flashpoint": "73", "boiling_point": "95"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_flammable_liquid_category", "parameters": {"category": "1"}}

857. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0884\_01`

MODEL ERROR

- **Question:** Combining the inspection requirement of paragraph (e)(1) and the employee check requirement of paragraph (e)(2)(i), what two conditions concerning the work area must be ensured before energy is

restored?

- **Ground truth:** Per (e)(1), the work area must be inspected to ensure nonessential items have been removed and that machine or equipment components are operationally intact; and per (e)(2)(i), the work area must be checked to ensure all employees have been safely positioned or removed.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"e": "1", "n": "2"}}
- ✔ **VERITROOPER (PASS):** {"name": "ensure_conditions", "parameters": {"condition1": "nonessential items have been removed", "condition2": "all employees have been safely positioned or removed"}}

858. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3280\_01`

- **Question:** Using the definitions of 'Suspension scaffold' and 'Two-point suspension scaffold,' explain what makes a two-point suspension scaffold (swing stage) a specific type of suspension scaffold.
- **Ground truth:** A suspension scaffold is one or more platforms suspended by ropes or other non-rigid means from an overhead structure. A two-point suspension scaffold (swing stage) fits this because it consists of a platform supported by hangers (stirrups) suspended by two ropes from overhead supports, and it is additionally equipped with means to raise and lower the platform to desired work levels.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "Two-point suspension scaffold", "parameters": {"definition": "a suspension scaffold consisting of a platform supported by hangers (stirrups) suspended by two ropes from overhead supports and equipped with means to permit the raising and lowering of the platform to desired work levels"}}
- ✖ **VERITROOPER (FAIL):** {"name": "explain_suspension_scaffold", "parameters": {"scaffold_type": "Two-point suspension scaffold (swing stage)"}}

859. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2250\_00`

MODEL ERROR

- **Question:** A facility uses Universal Precautions for all specimens. A specimen container is being moved from one lab to another lab within the same facility. Is labeling or color-coding required?
- **Ground truth:** No, the labeling/color-coding exemption applies while specimens remain within the facility, provided containers are recognizable as containing specimens.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "labeling_or_color_coding_required", "parameters": {"universal_precautions": "True", "within_facility": "True"}}
- ✖ **VERITROOPER (FAIL):** {"name": "labeling_or_color_coding_required", "parameters": {"universal_precautions": "True", "within_facility": "True"}}

860. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4228\_00`

- **Question:** If a State completes a developmental step by adopting a program component that is identical to the Federal program component, must it submit a full plan supplement?
- **Ground truth:** No, it need only submit documentation (such as a cover page of an implementing directive or notice of promulgation) within 60 days, but must make the underlying documentation available for Federal and public review upon request.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "submitting_documentation_for_identical_plan_changes", "parameters": {"submit_full_supplement": "False"}}
- ✔ **VERITROOPER (PASS):** {"name": "submit_supplement", "parameters": {"submit": "false", "reason": "adopted a program component identical to the Federal program component"}}

861. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4205\_02`

- **Question:** Using the final approval date in paragraph (b) and the date the Assistant Secretary approved the revised staffing requirements in paragraph (c), what do you notice about these two dates?
- **Ground truth:** Both occurred on the same date, July 22, 1985 - the final approval of the plan and the

approval of the revised compliance staffing benchmarks coincided.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_dates", "parameters": {"date1": "July 22, 1985", "date2": "July 22, 1985"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_date_difference", "parameters": {"date1": "July 22, 1985", "date2": "July 22, 1985"}}

862. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0969\_00`

- **Question:** Brakes for stopping the motion of the trolley or bridge shall be sufficient to stop within a distance in feet equal to what percentage of full load speed in feet per minute?
- **Ground truth:** 10 percent
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "brakes_for_stopping_motion", "parameters": {"percentage": "10"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_stopping_distance", "parameters": {"percentage": "10"}}

863. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3285\_02`

- **Question:** If rulemaking on roof bracket scaffold minimum width was not completed by November 25, 1997, was the 12-inch width requirement still stayed?
- **Ground truth:** Yes, the stay lasts until November 25, 1997 or until rulemaking is completed, whichever is later.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_stay_still_in_effect", "parameters": {"rulemaking_completed": "false", "completion_date": "November 25, 1997"}}
- ✗ **VERITROOPER (FAIL):** {"name": "check_rulemaking_status", "parameters": {"completion_date": "November 25, 1997", "rulemaking_status": "not completed"}}

864. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0676\_01`

- **Question:** If a possibility of an IDLH condition or flammable atmosphere has developed since prior monitoring, what type of monitoring must be conducted?
- **Ground truth:** Periodic monitoring shall be conducted.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "periodic_monitoring", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "Periodic monitoring", "parameters": {"type": "monitoring"}}

865. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1186\_00`

- **Question:** When may the upstop or latch allow the center or the handle grips to come closer than 17 inches horizontally behind the closest path of the mower blade(s)?
- **Ground truth:** When manually disengaged.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "upstop_or_latch_allowance", "parameters": {"condition": "manually disengaged"}}
- ✓ **VERITROOPER (PASS):** {"name": "allow_closer_than_17_inches", "parameters": {"condition": "manually disengaged"}}

866. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_04`

MODEL ERROR

- **Question:** At the national level, if employees are not represented by organizations having exclusive recognition on an agency basis or national consultation rights, how are nonmanagement members determined?
- **Ground truth:** They shall be determined through procedures devised by the agency which provide for effective representation of all employees.

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "determine_nonmanagement_members", "parameters": {"level": "national", "representation": "not represented"}}
- **✗ VERITROOPER (FAIL):** {"name": "determine_nonmanagement_members", "parameters": {"level": "national", "representation": "not represented"}}

867. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1206\_00`

- **Question:** Under paragraph (b)(1)(iv), a cylinder over 30 pounds water weight capacity must have a means to connect a valve protection cap. According to paragraph (b)(2)(iv), how must such a cap be maintained when the cylinder is not in use?
- **Ground truth:** The valve protection cap shall always be in place, hand-tight, except when cylinders are in use or connected for use.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "paragraph (b)(2)(iv)", "parameters": {"cylinder cap maintenance": "hand-tight except when cylinders are in use or connected for use"}}
- **✓ VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "According to paragraph (b)(2)(iv), how must a valve protection cap be maintained when the cylinder is not in use?", "answer": "hand-tight"}}

868. FAIL · Precision · `OSHA\_29CFR\_full\_2024\_2019\_00`

MODEL ERROR

- **Question:** What is the action level concentration of airborne beryllium, calculated as an 8-hour TWA?
- **Ground truth:** 0.1 micrograms per cubic meter of air ($\mu\text{g}/\text{m}^3$)
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "action level", "parameters": {"concentration": "0"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_action_level", "parameters": {"concentration": "0", "time": "8"}}

869. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0003\_07`

- **Question:** What is the OMB control number under which § 1902.3 was approved?
- **Ground truth:** 1218-0004
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_omb_control_number", "parameters": {"section": "1902.3"}}
- **✓ VERITROOPER (PASS):** {"name": "get_omb_control_number", "parameters": {"section_number": "1902.3"}}

870. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_01`

BAD QUESTION

- **Question:** What is the minimum wattage rating required for a heavy-duty lampholder of the medium type?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_wattage_rating", "parameters": {"type": "admedium", "rating": "heavy-duty"}}
- **✗ VERITROOPER (FAIL):** {"name": "get_minimum_wattage_rating", "parameters": {"type": "medium"}}

871. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_07`

HALLUCINATION

- **Question:** Which two forms of contact information must be included for each employee in the COVID-19 log?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_contact_info", "parameters": {"employee_id": "12345",

"contact_info_types": ["phone number", "email address"]}]}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

872. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_01`

- **Question:** Under § 1926.350(a)(6), regulators must be removed and valve protection caps put in place before cylinders are moved - but under what condition is this not required?

- **Ground truth:** Unless the cylinders are firmly secured on a special carrier intended for this purpose.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "regulators removal and valve protection caps placement condition", "parameters": {"condition": "cylinders are secured on a special carrier intended for this purpose"}}

- ✓ **VERITROOPER (PASS):** {"name": "exception_to_removing_regulators", "parameters": {"condition": "cylinders are firmly secured on a special carrier intended for this purpose"}}

873. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4446\_00`

- **Question:** If a person adversely affected by a final order resided in a different circuit than where the violation occurred, in which circuit(s) may they file a petition for review?

- **Ground truth:** Either the circuit in which the violation allegedly occurred or the circuit in which the complainant resided on the date of the violation.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "file_petition_for_review", "parameters": {"circuit": "the circuit in which the violation allegedly occurred or the circuit in which the complainant resided on the date of the violation"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_circuit_for_filing", "parameters": {"circuit_resided": "circuit in which the complainant resided on the date of the violation", "circuit_violation": "circuit in which the violation allegedly occurred"}}

874. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3801\_01`

MODEL ERROR

- **Question:** Combining paragraph (h)(1) and paragraph (h)(2)(ii), what triggers the requirement to provide PPE and what document governs both its use and its removal?

- **Ground truth:** PPE must be provided where airborne exposure exceeds or can reasonably be expected to exceed the TWA PEL or STEL (h)(1), and both its use and its removal are governed by the written exposure control plan required by paragraph (f)(1) (h)(1) and (h)(2)(ii).

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "combining_paragraphs", "parameters": {"h_1": "The employer is not required to pay for non-specialty safety-toe protective footwear (including steel-toe shoes or steel-toe boots) and non-specialty prescription safety eyewear, provided that the employer permits such items to be worn off the job-site.", "h_2": "When the employer provides metatarsal guards and allows the employee, at his or her request, to use shoes or boots with built-in metatarsal protection, the employer is not required to reimburse the employee for the shoes or boots."}}

- ✗ **VERITROOPER (FAIL):** {"name": "answer", "parameters": {"question": "Combining paragraph (h)(1) and paragraph (h)(2)(ii), what triggers the requirement to provide PPE and what document governs both its use and its removal?"}}

875. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_01`

HALLUCINATION

- **Question:** What is the minimum number of training hours committee members must receive under paragraph (b)?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_training_hours", "parameters": {"x": "b"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the

question.

876. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0340\_02`

- **Question:** A vertical lifeline must be provided as part of a personal fall arrest system for each employee on a working platform suspended by two or more wire ropes if failure of one rope would upset the platform - under what condition are vertical lifelines NOT required?
- **Ground truth:** Vertical lifelines are not required if a secondary wire rope suspension is used, provided that each employee is attached to a horizontal lifeline anchored to the platform.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "vertical_lifelines_not_required", "parameters": {"condition": "If a secondary wire rope suspension is used, provided that each employee is attached to a horizontal lifeline anchored to the platform"}}
- ✓ **VERITROOPER (PASS):** {"name": "vertical_lifeline_condition", "parameters": {"condition": "A secondary wire rope suspension is used"}}

877. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_00`

HALLUCINATION

- **Question:** What is the maximum allowable fit-test failure rate before an employer must reassign an employee from asbestos respirator work?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "fit_test_failure_rate", "parameters": {"failure_rate": "10"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the maximum allowable fit-test failure rate before an employer must reassign an employee from asbestos respirator work.

878. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_4366\_03`

MODEL ERROR

- **Question:** If a timely petition for review is filed, when does the ALJ's decision NOT automatically become the final order of the Secretary?
- **Ground truth:** Unless the ARB, within 30 days of the filing of the petition, issues an order notifying the parties that the case has been accepted for review.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "ALJ decision NOT automatically final order", "parameters": {"timely petition for review filed": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "paragraph_b_exception", "parameters": {"timely_petition_for_review": "True"}}

879. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0881\_02`

- **Question:** During a servicing activity such as cleaning a machine, what additional condition makes it qualify as 'servicing and/or maintenance' under this section?
- **Ground truth:** Where the employee may be exposed to the unexpected energization or startup of the equipment or release of hazardous energy.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "servicing_and_or_maintenance", "parameters": {"condition": "the employee may be exposed to the unexpected energization or startup of the equipment or release of hazardous energy"}}
- ✓ **VERITROOPER (PASS):** {"name": "servicing_and/or_maintenance", "parameters": {"activity": "cleaning", "additional_condition": "exposure to unexpected energization or startup of the equipment or release of hazardous energy"}}

880. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_06`

HALLUCINATION

- **Question:** What penalty applies to a worker who performs cutting in a sprinklered building while protection is impaired?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "prohibited_areas", "parameters": {"C": "In sprinklered buildings while such protection is impaired"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

881. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3987\_04`

- **Question:** For brake testing on duty cycle and repetitive lifts where each lift is 90% or more of the maximum line pull, which lifts are excepted from the brake-testing requirement?
- **Ground truth:** The requirement applies to the first lift but not to successive lifts.
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "brake_testing_exception", "parameters": {"lift_type": "successive lifts"}}
- **✓ VERITROOPER (PASS):** {"name": "brake_testing_exception", "parameters": {"lifts_excepted": "successive lifts"}}

882. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2815\_02`

MODEL ERROR

- **Question:** Sections 1926.403, 1926.404, 1926.405, 1926.407, and 1926.408 all fall under which single OMB control number?
- **Ground truth:** All five sections (1926.403, 1926.404, 1926.405, 1926.407, and 1926.408) are assigned OMB control No. 1218-0130.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_omb_control_number", "parameters": {"cfr_citations": "[1926.403, 1926.404, 1926.405, 1926.407, 1926.408]"}}
- **✓ VERITROOPER (PASS):** {"name": "get_omb_control_number", "parameters": {"sections": "[1926.403, 1926.404, 1926.405, 1926.407, 1926.408]"}}

883. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3608\_01`

- **Question:** Combining the requirements in subparts (iii) and (iv), what is the maximum allowed descending velocity and the maximum allowed length of the connection between the carrier/lifeline and the body belt or harness?
- **Ground truth:** The descending velocity must be limited to 7 feet/sec. (2.1 m/sec.) or less, and the connection length shall not exceed 9 inches (23 cm).
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "get_max_velocity_and_connection_length", "parameters": {"velocity": "7", "length": "9"}}
- **✓ VERITROOPER (PASS):** {"name": "get_max_allowed_values", "parameters": {"velocity": "7", "length": "9"}}

884. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_06`

MODEL ERROR

- **Question:** If an employee exposed to electric arc hazards wears nylon clothing, but the employer demonstrates the fabric was treated to withstand the conditions the employee may encounter, is the clothing prohibited?
- **Ground truth:** No; the prohibition does not apply if the employer demonstrates the fabric has been treated

to withstand the conditions or that the employee wears it in a manner that eliminates the hazard.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "prohibits_clothing", "parameters": {"p": "nylon", "x": "True"}}
- ✖ **VERITROOPER (FAIL):** {"name": "check_prohibited_clothing", "parameters": {"clothing_material": "nylon", "treated_to_withstand_conditions": "True"}}

885. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_01`

HALLUCINATION

- **Question:** What is the rated capacity in pounds for a cable laid grommet-hand tucked sling?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_rated_capacity", "parameters": {"sling_type": "cable laid grommet-hand tucked"}}
- ✓ **VERITROOPER (PASS):** This information is not contained in the provided evidence.

886. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3162\_00`

- **Question:** If compressed air is being used to convey materials, does § 1926.306 apply to the special problems this creates?
- **Ground truth:** No, this section does not deal with the special problems created by using compressed air to convey materials.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section_applies", "parameters": {"section": "1926.306", "application": "false"}}
- ✓ **VERITROOPER (PASS):** {"name": "does_not_apply", "parameters": {"section": "1926.306", "situation": "compressed air being used to convey materials"}}

887. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4301\_01`

MODEL ERROR

- **Question:** At an establishment where employees are represented under collective bargaining arrangements, how are nonmanagement committee members appointed?
- **Ground truth:** They shall be appointed from among those recommended by the exclusive bargaining representative.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "appointment_of_nonmanagement_committee_members", "parameters": {"representation_type": "collective bargaining arrangements"}}
- ✖ **VERITROOPER (FAIL):** {"name": "appoint_nonmanagement_members", "parameters": {"method": "collective bargaining arrangements"}}

888. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1984\_00`

- **Question:** Above what exposure level (without regard to respirator use) must the employer provide facilities for employees to vacuum their protective clothing before entering change rooms?
- **Ground truth:** 100 µg/m³
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "vacuuming_clothes", "parameters": {"exposure_level": "100 \u00b5g/m³"}}
- ✓ **VERITROOPER (PASS):** {"name": "vacuum_clothing_facilities", "parameters": {"exposure_level": "100"}}

889. FAIL · Calculation · `OSHA\_29CFR\_full\_2024\_4031\_01`

MODEL ERROR

- **Question:** A boom-attached personnel platform is used on equipment whose rated capacity for the radius and configuration is 24,500 pounds. What is the maximum total weight of the loaded personnel platform

allowed, except during proof testing?

- **Ground truth:** 12,250 pounds

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "calculate_total_weight", "parameters": {"rated_capacity": "24500"}}

- **✗ VERITROOPER (FAIL):** {"name": "calculate_max_weight", "parameters": {"rated_capacity": "24500"}}

890. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_05`

HALLUCINATION

- **Question:** How many days does an administrative law judge have to issue a decision after the post-hearing review?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "issue_decision", "parameters": {"days": "30"}}

- **✓ VERITROOPER (PASS):** I don't have enough information to answer the question.

891. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2913\_03`

HALLUCINATION

- **Question:** What is the maximum allowable negative pressure permitted within the inlet ductwork before additional sealing measures are mandated?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "get_max_negative_pressure", "parameters": {"ductwork": "inlet"}}

- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

892. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0478\_00`

HALLUCINATION

- **Question:** What is the required testing frequency for inspecting filter rolls to ensure proper replacement of filter media?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "filter_roll_inspection_frequency", "parameters": {}}

- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

893. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1748\_02`

- **Question:** Paragraph (g) addresses limiting personnel and disinfecting surfaces for aerosol-generating procedures, but where, according to the note to paragraph (g), are the respirator and other PPE requirements located?

- **Ground truth:** The note to paragraph (g) states that respirator and other PPE requirements during aerosol-generating procedures are contained in paragraph (f)(3) of the section, separate from the personnel-limitation requirement in (g)(1) and the cleaning/disinfecting requirement in (g)(3).

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "paragraph_location", "parameters": {"note_to_paragraph_g": "paragraph (f)(3)"}}

- **✓ VERITROOPER (PASS):** {"name": "get_location_of_respirator_requirements", "parameters": {"paragraph": "g", "location": "paragraph (f)(3)"}}

894. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_04`

- **Question:** What is the upper flashpoint limit for Category 3 flammable liquids?
- **Ground truth:** 140 °F (60 °C)
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_flashpoint_limit", "parameters": {"category": "Category 3", "flashpoint_limit": "100"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_flashpoint_limit", "parameters": {"category": "3"}}

895. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0370\_02`

- **Question:** What diameter threshold distinguishes the two thickness-to-diameter ratios for abrasive cutting-off wheels?
- **Ground truth:** 20 inches
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_diameter_threshold", "parameters": {"diameter": "20"}}
- ✓ **VERITROOPER (PASS):** {"name": "diameter_threshold", "parameters": {"diameter": "20"}}

896. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1599\_05`

HALLUCINATION

- **Question:** What is the minimum branch-circuit conductor ampacity required for a cord- and plug-connected arc welder receptacle?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_ampacity", "parameters": {"circuit_rating": "15 or 20", "receptacle_rating": "15"}}
- ✗ **VERITROOPER (FAIL):** {"name": "Table S-4", "parameters": {}}

897. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2437\_00`

- **Question:** If the employer and employee reach an agreement that is consistent with the recommendations of only one of the three physicians (not the third physician), must the employer still act on the third physician's findings?
- **Ground truth:** No; the agreement consistent with at least one of the three physicians' recommendations overrides the requirement to act on the third physician's findings.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "act_consistent_with_third_physician", "parameters": {"consistent": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer", "parameters": {"question": "If the employer and employee reach an agreement that is consistent with the recommendations of only one of the three physicians (not the third physician), must the employer still act on the third physician's findings?"}}

898. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1126\_00`

- **Question:** Under the directed component of the inspection program, which inspections are exempt from the requirement to maintain certification records specified by paragraph (e)(1)(i)(C)?
- **Ground truth:** Inspections of the clutch/brake mechanism, antirepeat feature, and single-stroke mechanism conducted under the directed component of the inspection program are exempt from the requirement to maintain certification records specified by paragraph (e)(1)(i)(C).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph", "parameters": {"e": "1", "i": "ii", "C": "Inspections of the clutch/brake mechanism, antirepeat feature, and single-stroke mechanism conducted under the directed component of the inspection program"}}
- ✓ **VERITROOPER (PASS):** {"name": "inspect_component", "parameters": {"component": "clutch/brake mechanism, antirepeat feature, and single-stroke mechanism", "program": "directed component of the"}}

899. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1138\_01`

- **Question:** If a non-critical subsystem is replaced by one of the same manufacture and design as the original, is revalidation required?
- **Ground truth:** No; recertification by the employer without revalidation is permitted in that case.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "revalidation", "parameters": {"required": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "recertification_byemployer_without_revalidation", "parameters": {"component": "non-critical", "replacement": "same manufacture and design as the original"}}

900. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1695\_01`

MODEL ERROR

- **Question:** Both paragraph (a)(2)(ii) (portable equipment) and paragraph (c)(2) (test instruments) describe what to do upon finding a defect that might expose an employee to injury. What common requirement do both impose?
- **Ground truth:** In both cases the defective or damaged item shall be removed from service, and no employee may use it until repairs and tests necessary to render the equipment safe have been made.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "requirement_imposed_by_OSHA", "parameters": {"requirement": "removal from service and repairs"}}
- ✗ **VERITROOPER (FAIL):** {"name": "common_requirement", "parameters": {"paragraphs": "['(a)(2)(ii)', '(c)(2)']"}}

901. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1695\_02`

- **Question:** Grounding-type equipment is used in a highly conductive work location inundated with water. Combining paragraphs (a)(3) and (a)(4), what two requirements apply to the flexible cord and equipment?
- **Ground truth:** The flexible cord used with grounding-type equipment must contain an equipment grounding conductor (per (a)(3)(i)), and the portable electric equipment and flexible cords used in highly conductive work locations must be approved for those locations (per (a)(4)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Grounding-type equipment requirements", "parameters": {"requirement1": "The flexible cord used with grounding-type equipment shall contain an equipment grounding conductor.", "requirement2": "Portable electric equipment and flexible cords used in highly conductive work locations shall be approved for those locations."}}
- ✓ **VERITROOPER (PASS):** {"name": "apply_requirements", "parameters": {"cord_type": "flexible", "equipment_type": "grounding-type", "work_location": "highly conductive", "location_condition": "inundated with water"}}

902. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1622\_00`

- **Question:** Under the open wiring on insulators rules, what is the maximum distance for supporting a conductor smaller than No. 8 from a tap, and what is the maximum length for the flexible nonmetallic tubing permitted to enclose conductors in dry locations?
- **Ground truth:** Supports must be within 152 mm (6 in.) from a tap or splice (paragraph (ii)(A)), and the flexible nonmetallic tubing must be in continuous lengths not exceeding 4.57 m (15.0 ft) (paragraph (iii)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "open_wiring_on_insulators", "parameters": {"distance_from_tap": "152", "tubing_length": "4.57"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_open_wiring_rules", "parameters": {"tap_support_distance": "152", "tubing_length": "4570"}}

903. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_02`

HALLUCINATION

- **Question:** Within how many days after filing must OSHA notify the named respondent of the complaint?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "notify_respondent", "parameters": {"days": "20"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

904. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_1434\_05`

- **Question:** Under what conditions may an employer make broad incident heat energy estimates covering multiple system areas?
- **Ground truth:** Provided the employer uses reasonable assumptions about the energy-exposure distribution throughout the system and the estimates represent the maximum employee exposure for those areas.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "make_broad_estimates", "parameters": {"condition_1": "The employer uses reasonable assumptions about the energy-exposure distribution throughout the system", "condition_2": "The estimates represent the maximum employee exposure for those areas"}}
- ✓ **VERITROOPER (PASS):** {"name": "make_incident_heat_energy_estimates", "parameters": {"assumptions": "reasonable", "distribution": "reasonable", "exposure": "maximum"}}

905. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1769\_02`

BAD QUESTION

- **Question:** How many years must the COVID-19 log be preserved after this section is no longer in effect?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_log_preservation_years", "parameters": {"x": "3"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

906. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0012\_01`

- **Question:** When an alleged violation of standards has occurred, what means may be used to provide prompt notice to the employer including abatement requirements?
- **Ground truth:** By such means as the issuance of a written citation to the employer and posting of the citation at or near the site of the violation.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Provides for prompt notice to employers and employees when an alleged violation of standards has occurred", "parameters": {"means": "such as the issuance of a written citation to the employer and posting of the citation at or near the site of the violation"}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_prompt_notice", "parameters": {"means": "issuance of a written citation to the employer and posting of the citation at or near the site of the violation"}}

907. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2542\_01`

- **Question:** Combining the secondary power supply requirement in (ii) with the component-design principle in (iv), what dual constraints apply to a signal-generating system that must operate during power failure?
- **Ground truth:** It must be provided with an automatically activated secondary power supply adequate to simultaneously power all connected emergency equipment (unless batteries are the primary power source), while also being designed with the minimum number of components necessary to function, using components that do not require frequent servicing such as lubrication or cleaning.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "signal_generating_system", "parameters": {"requirement_1":

"automatically activated secondary power supply", "requirement_2": "minimum number of components necessary to make it function as intended"}}

- ✓ **VERITROOPER (PASS):** {"name": "constraint_application", "parameters": {"constraint1": "automatically activated secondary power supply", "constraint2": "minimum number of components necessary"}}

908. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_00`

HALLUCINATION

- **Question:** Within how many days after the alleged violation must an employee file a discrimination complaint under the Pipeline Safety Act?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "file_discrimination_complaint", "parameters": {"days": "180"}}

- ✓ **VERITROOPER (PASS):** I don't have enough information to answer the question.

909. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_02`

HALLUCINATION

- **Question:** How often, in days or months, must pressure hoses and couplings be inspected?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "inspect_pressure_hoses_and_couplings", "parameters": {"frequency": "regular intervals"}}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

910. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1126\_01`

- **Question:** Are inspections of the clutch/brake mechanism, antirepeat feature, and single-stroke mechanism conducted under the general component exempt from the certification record requirement?

- **Ground truth:** No. Inspections of the clutch/brake mechanism, antirepeat feature, and single-stroke mechanism conducted under the general component of the inspection program are not exempt from the requirement to maintain certification records.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "check_exemption", "parameters": {"component": "general", "mechanism": "clutch/brake mechanism, antirepeat feature, and single-stroke mechanism"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "Are inspections of the clutch/brake mechanism, antirepeat feature, and single-stroke mechanism conducted under the general component exempt from the certification record requirement?", "answer": "No"}}

911. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2129\_01`

HALLUCINATION

- **Question:** What is the PEL (permissible exposure limit) for cadmium mentioned in subsection (vi)?

- **Ground truth:** UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_PEL", "parameters": {"vi": "5"}}

- ✓ **VERITROOPER (PASS):** This information is not provided in the given evidence.

912. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_03`

- **Question:** What edition of the American Chemical Society's Safety in Academic Chemistry Laboratories is referenced?

- **Ground truth:** 4th edition, 1985

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "American Chemical Society", "parameters": {"title": "Safety in Academic Chemistry Laboratories", "edition": "4th"}}

- ✓ **VERITROOPER (PASS):** {"name": "get_edition", "parameters": {"book_title": "Safety in Academic Chemistry Laboratories", "publisher": "American Chemical Society"}}

913. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1738\_00`

HALLUCINATION

- **Question:** What is the maximum allowable distance, in feet, that defines an AIIR's negative pressure boundary?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_distance", "parameters": {"x": "6"}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

914. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_4251\_00`

MODEL ERROR

- **Question:** If an interested person submits a petition to participate as a party 35 days after the notice was published in the Federal Register, does it meet the timing requirement of paragraph (b)(1)?
- **Ground truth:** No, because the petition must be submitted within 30 days after the notice has been published in the Federal Register.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "check_timing_requirement", "parameters": {"days": "35"}}
- ✗ **VERITROOPER (FAIL):** {"name": "check_timing_requirement", "parameters": {"days_after_notice": "35", "required_days": "30"}}

915. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_04`

HALLUCINATION

- **Question:** How often must agencies update the Material Safety Data Sheets made available to committees?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "update_MSDS", "parameters": {"frequency": "annually"}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

916. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4247\_01`

- **Question:** Comparing paragraph (a) and paragraph (b) of § 1955.13, when an administrative law judge deems himself disqualified versus when he addresses a witness's refusal to answer, what action does each provision authorize him to take?
- **Ground truth:** Under paragraph (a), if an administrative law judge deems himself disqualified, he shall withdraw by notice on the record directed to the Chief Administrative Law Judge. Under paragraph (b), if a witness refuses to answer a question after being directed, the administrative law judge may make such orders with regard to the refusal as are just and proper, including the striking of all testimony previously given by such witness on related matters.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compare_provisions", "parameters": {"paragraph_a": "An administrative law judge who deems himself disqualified shall withdraw by notice on the record directed to the Chief Administrative Law Judge.", "paragraph_b": "If a witness or party refuses to answer a question after being so directed, or refuses to obey an order to provide or permit discovery, the administrative law judge may make such orders with regard to the refusal as are just and proper, including the striking of all testimony previously given by such witness on related matters."}}
- ✓ **VERITROOPER (PASS):** {"name": "compare_provisions", "parameters": {"paragraph_a": "withdraw by notice on the record directed to the Chief Administrative Law Judge", "paragraph_b": "make such orders with

regard to the refusal as are just and proper, including the striking of all testimony previously given by such witness on related matters"}}}

917. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0227\_02`

- **Question:** Which document, and from which seller, is IBR approved for §§ 1910.34, 1910.35, 1910.36, and 1910.37?
- **Ground truth:** IFC-2009, International Fire Code (copyright 2009), available for purchase from the International Code Council, Chicago District Office.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "IFC-2009", "parameters": {"seller": "International Code Council"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_document", "parameters": {"document_name": "IFC-2009", "seller": "International Code Council"}}

918. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0751\_00`

- **Question:** If an employee's use of a respirator reveals they have not retained the requisite skill, but the workplace and respirator type are unchanged, is retraining triggered?
- **Ground truth:** Yes, because inadequacies in the employee's knowledge or use of the respirator indicating they have not retained the requisite understanding or skill is itself a triggering condition.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "retraining", "parameters": {"employee": "employee", "workplace": "unchanged", "respirator_type": "unchanged", "trigger": "inadequacies in an affected employee"}}
- ✓ **VERITROOPER (PASS):** {"name": "retraining_triggered", "parameters": {"reason": "inadequacies in the employee"}}

919. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1037\_00`

HALLUCINATION

- **Question:** What is the maximum angle of loading permitted for a basket hitch configuration before it must be re-rated?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "get_angle_of_loading", "parameters": {"angle": "90"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_angle_of_loading", "parameters": {"configuration": "basket hitch"}}

920. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1748\_00`

- **Question:** When an aerosol-generating procedure is performed on a person with suspected or confirmed COVID-19, under paragraph (g)(1) the employer limits attendance to essential personnel, and under paragraph (g)(2) where must the procedure be performed if such a space is available?
- **Ground truth:** Under paragraph (g)(1) only employees essential for patient care and procedure support may be present, and under paragraph (g)(2) the procedure must be performed in an existing AIIR (airborne infection isolation room), if available.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (g)(2)", "parameters": {"location": "an existing AIIR"}}
- ✓ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "When an aerosol-generating procedure is performed on a person with suspected or confirmed COVID-19, under paragraph (g)(1) the employer limits attendance to essential personnel, and under paragraph (g)(2) where must the procedure be performed if such a space is available?", "answer": "in an existing AIIR"}}

921. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4294\_04`

- **Question:** Combining paragraph (d)(1) and paragraph (d)(4), who receives copies of reports and of no-inspection notifications at the establishment level?
- **Ground truth:** Under (d)(1), a copy of each report received is sent to the appropriate establishment safety and health committee; under (d)(4), a copy of each no-inspection notification is provided to the appropriate certified safety and health committee where established under Executive Order 12196.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (d)(1) and paragraph (d)(4)", "parameters": {"recipient": "the appropriate establishment safety and health committee"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_recipients_of_copies", "parameters": {"paragraph": "d", "items": ["reports", "no-inspection notifications"]}}

922. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0040\_02`

- **Question:** Per § 1902.35, evaluation conducted for 18(e) determination purposes will be based on the criteria set forth in which sections?
- **Ground truth:** §§ 1902.37 and 1902.38
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "evaluation_for_18e_determination", "parameters": {"criteria_sections": "[\u0027\u00a71902.37\u0027, \u0027\u00a71902.38\u0027]"}}
- ✓ **VERITROOPER (PASS):** {"name": "", "parameters": {"sections": "[1902.37, 1902.38]"}}

923. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_02`

HALLUCINATION

- **Question:** What is the minimum efficiency rating, expressed as a percentage, required for the HEPA filters provided for air-purifying respirators?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "HEPA_filter_efficiency", "parameters": {"percentage": "99"}}
- ✗ **VERITROOPER (FAIL):** HEPA filters are required to be 99.97% efficient.

924. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1441\_00`

- **Question:** Under paragraph (n)(5), when must employees test lines and equipment to verify the absence of nominal voltage before installing a ground?
- **Ground truth:** In all cases unless a previously installed ground is present.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "test_lines_and_equipment", "parameters": {"condition": "unless a previously installed ground is present"}}
- ✓ **VERITROOPER (PASS):** {"name": "test_lines_and_equipment", "parameters": {"condition": "unless a previously installed ground is present"}}

925. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0358\_08`

- **Question:** A floating cone must actuate a limit switch if a force of how many pounds is applied on the edge closest to the hinge?
- **Ground truth:** 2 pounds
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "floating type safety cone", "parameters": {"force": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "actuate_limit_switch", "parameters": {"force": "2"}}

926. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0772\_02`

MODEL ERROR

- **Question:** Based on paragraphs (a)(7), (a)(8), and (a)(9), what is the total maximum number of squeezes performed before a subject is declared unable to taste saccharin per paragraph (a)(11)?
- **Ground truth:** Thirty squeezes - ten in the first set, ten in the second set, and ten in the third set. If saccharin is not tasted after these 30 squeezes, the subject is unable to taste saccharin and may not perform the saccharin fit test.
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "calculate_total_squeezes", "parameters": {"x": "10", "y": "10", "z": "10"}}
- **✓ VERITROOPER (PASS):** {"name": "add", "parameters": {"a": "10", "b": "10", "c": "10"}}

927. PASS · Precision · `OSHA\_29CFR\_full\_2024\_3130\_06`

- **Question:** Below what viscosity (in Saybolt Universal Seconds at 100 °F) and below what flashpoint must the Tag Closed Tester (ASTM D-56-69) be used?
- **Ground truth:** less than 45 Saybolt Universal Seconds at 100 °F and a flashpoint below 175 °F (79.4 °C)
- **✓ Baseline (vanilla RAG) (PASS):** {"name": "Tag Closed Tester", "parameters": {"viscosity": "45", "flashpoint": "175"}}
- **✓ VERITROOPER (PASS):** {"name": "Tag Closed Tester", "parameters": {"viscosity": "45", "flashpoint": "175"}}

928. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4297\_04`

HALLUCINATION

- **Question:** How frequently are the scheduled inspections under paragraph (c) conducted during an agency's program evaluation?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "paragraph", "parameters": {"c": "as deemed necessary by the Secretary"}}
- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

929. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0484\_07`

HALLUCINATION

- **Question:** What gauge of wire cable is required when containers are suspended rather than placed on metal supports?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "wire_cable_gauge", "parameters": {"suspension": "True"}}
- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question about the gauge of wire cable required when containers are suspended rather than placed on metal supports.

930. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_2625\_01`

HALLUCINATION

- **Question:** Which standardized animal test method must be used to obtain the 'positive results from an appropriate animal test' for respiratory sensitizer classification?
- **Ground truth:** UNANSWERABLE
- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "A.4.2.1.3.1", "parameters": {"2": "recognized and validated"}}

animal models for the testing of respiratory hypersensitivity are not available", "3": "the potential of a substance to cause sensitization by inhalation in humans"}}

- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

931. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0643\_02`

MODEL ERROR

- **Question:** Based on the provisions for documenting responses and retaining reports, what must an employer document regarding each audit finding, and which reports must be kept available?

- **Ground truth:** The employer must promptly determine and document an appropriate response to each finding and document that deficiencies have been corrected, while retaining the two most recent compliance audit reports.

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "documenting_responses", "parameters": {"action_taken": "what was done and why", "corrective_action": "the corrective action to be taken noted", "deficiency_identified": "each deficiency identified is addressed"}}

- ✓ **VERITROOPER (PASS):** {"name": "document_response_and_retain_reports", "parameters": {"response": "that deficiencies have been corrected", "reports": "the two (2) most recent compliance audit reports"}}

932. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0455\_02`

- **Question:** If a pit has any part within a Division 1 or 2 classified area and is NOT provided with mechanical ventilation, how must all areas within the pit be classified?

- **Ground truth:** All areas within the pit shall be classified Division 1.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_pit_area", "parameters": {"division": "1"}}

- ✓ **VERITROOPER (PASS):** {"name": "classify_pit_areas", "parameters": {"has_mechanical_ventilation": "False", "division": "1"}}

933. PASS · Cause & Effect · `OSHA\_29CFR\_full\_2024\_0150\_00`

- **Question:** Under § 1905.10(b)(7)(i), what circumstances must an applicant establish facts about that caused their inability to comply with a standard by its effective date?

- **Ground truth:** The applicant must establish facts showing that they are unable to comply because of unavailability of professional or technical personnel or of materials and equipment needed to come into compliance with the standard, or because necessary construction or alteration of facilities cannot be completed by the effective date.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "establish_facts", "parameters": {"applicant": "applicant", "facts": "unavailability of professional or technical personnel or of materials and equipment needed to come into compliance with the standard or because necessary construction or alteration of facilities cannot be completed by the effective date"}}

- ✓ **VERITROOPER (PASS):** {"name": "establish_circumstances", "parameters": {"circumstances": "unavailability of professional or technical personnel or of materials and equipment needed to come into compliance with the standard or because necessary construction or alteration of facilities cannot be completed by the effective date"}}

934. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3172\_00`

MODEL ERROR

- **Question:** For a horizontal overhead belt whose lower part is 8 feet above the floor, located over a passageway, with a center-to-center pulley distance of 12 feet, traveling 1,000 feet per minute, and 10 inches in width, must it be guarded for its entire length under paragraph (e)(2)(ii)?

- **Ground truth:** Yes. Although it travels less than 1,800 feet per minute (failing condition (a)), it must be guarded for its entire length because the center-to-center distance is 10 feet or more (condition (b), here 12

feet) and the belt is 8 inches or more in width (condition (c), here 10 inches). Any one of the conditions being met requires full-length guarding.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "guard_horizontal_overhead_belt", "parameters": {"distance_from_floor": "8", "location": "passageway", "pulley_distance": "12", "traveling_speed": "1000", "belt_width": "10"}}
- ✔ **VERITROOPER (PASS):** {"name": "guard_required", "parameters": {"feet_above_floor": "8", "over_passageway": "True", "center_to_center_pulley_distance": "12", "feet_per_minute": "1000", "width_in_inches": "10"}}

935. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0035\_02`

- **Question:** Under which section of the Occupational Safety and Health Act of 1970 does the Assistant Secretary determine whether to grant final approval to State plans?
- **Ground truth:** section 18(e)
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "section 18", "parameters": {"x": "e"}}
- ✔ **VERITROOPER (PASS):** {"name": "get_section", "parameters": {"section_number": "18", "subsection": "e"}}

936. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3162\_02`

- **Question:** If a contractor wants to place a shutoff valve between the air receiver and its safety valve for maintenance, is this permitted?
- **Ground truth:** No, no valve of any type shall be placed between the air receiver and its safety valve or valves.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "shutoff valve placement", "parameters": {"location": "between the air receiver and its safety valve", "context": "for maintenance"}}
- ✔ **VERITROOPER (PASS):** {"name": "is_permitted", "parameters": {"action": "place a shutoff valve between the air receiver and its safety valve"}}

937. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0227\_03`

- **Question:** Comparing the materials in paragraph (z) and paragraph (aa), which source provides material approved for OSHA tank regulations versus fire code sections?
- **Ground truth:** Paragraph (z) - Underwriters Laboratories - provides standards approved for § 1910.106(b)(1)(iii)(a)(1) (tanks), while paragraph (aa) - International Code Council - provides the IFC-2009 approved for §§ 1910.34, 1910.35, 1910.36, and 1910.37.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "compare_sources", "parameters": {"x": "Underwriters Laboratories (UL)", "y": "International Code Council"}}
- ✔ **VERITROOPER (PASS):** {"name": "compare_sources", "parameters": {"source1": "Underwriters Laboratories (UL)", "source2": "International Code Council"}}

938. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1332\_00`

MODEL ERROR

- **Question:** Considering the dry deck storage requirement to minimize accidental rolling and the log decks 'Stops' provision, what two distinct kinds of devices/arrangements does the evidence require to prevent logs from rolling in these respective areas?
- **Ground truth:** For dry deck storage (paragraph (v)(c)), logs shall be arranged to minimize the chance of accidentally rolling from the deck; for log decks (paragraph (3)(ii) 'Stops'), log decks shall be provided with adequate stops, chains, or other safeguards to prevent logs from rolling down the deck onto the carriage or its runway.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "Log decks - Stops", "parameters": {"arrangements": "adequate"}}

stops, chains, or other safeguards"}}}

- ✖ **VERITROOPER (FAIL):** {"name": "log_decks_Stops", "parameters": {"stops": "adequate stops, chains, or other safeguards", "location": "log decks"}}

939. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3641\_02`

MODEL ERROR

- **Question:** Both the glove bag work practices and the negative pressure glove box work practices specify a minimum staffing level. How many persons are required to perform each operation?

- **Ground truth:** At least two persons are required for Class I glovebag removal operations, and at least two persons are required to perform the negative pressure glove box removal.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "glove_bag_work_practices", "parameters": {"persons_required": "2"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_minimum_staffing_level", "parameters": {"operation": "glove bag work practices"}}

940. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_0478\_02`

HALLUCINATION

- **Question:** What is the minimum size or capacity required for the water-filled metal container used to store discarded filter pads?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "get_minimum_size", "parameters": {"unit": "gallons"}}

- ✓ **VERITROOPER (PASS):** I don't have enough information to answer the question.

941. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_00`

- **Question:** Compressed gas cylinders must be secured in an upright position at all times. Under what circumstance is an exception to this allowed?

- **Ground truth:** If necessary, for short periods of time while cylinders are actually being hoisted or carried.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "compressed_gas_cylinders_exception", "parameters": {"circumstance": "for short periods of time while cylinders are actually being hoisted or carried"}}

- ✓ **VERITROOPER (PASS):** {"name": "exception_to_securing_cylinders", "parameters": {"circumstance": "for short periods of time while cylinders are actually being hoisted or carried"}}

942. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4450\_01`

- **Question:** On what date was the Health Care and Education Reconciliation Act of 2010, Public Law 111-152, signed into law?

- **Ground truth:** March 30, 2010

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "date_signed_into_law", "parameters": {"date": "March 30, 2010"}}

- ✖ **VERITROOPER (FAIL):** {"name": "get_date_signed_into_law", "parameters": {"public_law": "111-152"}}

943. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_00`

HALLUCINATION

- **Question:** How many days does an agency have to provide requested information to a committee under paragraph (a)?

- **Ground truth:** UNANSWERABLE

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "agency_responsibilities", "parameters": {"days": "15"}}

- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

944. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_2160\_00`

MODEL ERROR

- **Question:** Why was the standard deviation of the 0.2 ng/mL cadmium standard (0.75 abs-sec), rather than the reagent blank, used to calculate the detection limits by the IUPAC method?
- **Ground truth:** Because the near lack of a blank signal (five of the six reagent blank peak area readings were zero) does not satisfy a strict interpretation of the IUPAC method for determining the detection limits.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_detection_limits", "parameters": {"sd": "0", "m": "51"}}
- ✗ **VERITROOPER (FAIL):** {"name": "calculate_detection_limits", "parameters": {"std_dev": "0", "slope": "51"}}

945. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3653\_06`

HALLUCINATION

- **Question:** Which OSHA form must employers use to document the negative exposure assessment referenced for Class I asbestos work?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "OSHA Form", "parameters": {"Form ID": "None", "Form Name": "Employers must use a written opinion from a qualified person, not an OSHA form, to document the negative exposure assessment for Class I asbestos work."}}
- ✓ **VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

946. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2019\_02`

- **Question:** Within how many hours must a CBD diagnostic center have the capacity to transfer BAL samples to a laboratory for diagnostic testing?
- **Ground truth:** 24 hours
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "CBD diagnostic center transfer BAL samples", "parameters": {"hours": "24"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_time_for_transferring_BAL_samples", "parameters": {"hours": "24"}}

947. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0163\_02`

- **Question:** If the presiding hearing examiner has not ordered otherwise, which party proceeds first at a hearing?
- **Ground truth:** The party applicant for relief proceeds first.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Order of proceeding", "parameters": {"applicant": "party applicant for relief"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_party_to_proceed_first", "parameters": {"party": "applicant"}}

948. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2444\_00`

- **Question:** What is the column zone temperature in the GC analysis conditions for MDA?
- **Ground truth:** 220 degrees C
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "GC conditions", "parameters": {"Zone temperatures": "Column\u2014220 degrees C"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_gc_condition", "parameters": {"condition": "column zone"}}

temperature"}}}

949. PASS · Exception · `OSHA\_29CFR\_full\_2024\_0340\_01`

- **Question:** Access to and egress from a working platform must be provided by stairs, ladders, platforms and runways - under what circumstance does this requirement NOT apply?
- **Ground truth:** It does not apply to working platforms that land directly on a safe surface.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "access_to_working_platform", "parameters": {"requirement_applies": "false", "circumstance": "the platform lands directly on a safe surface"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_access_requirement", "parameters": {"circumstance": "when the platform lands directly on a safe surface"}}

950. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2632\_01`

- **Question:** If there is evidence in humans that a substance can lead to sensitization by skin contact in a substantial number of persons, is it classified as a skin sensitizer?
- **Ground truth:** Yes, the substance is classified as a skin sensitizer.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "classify_skin_sensitizer", "parameters": {"evidence_in_humans": "True"}}
- ✓ **VERITROOPER (PASS):** {"name": "classify_skin_sensitizer", "parameters": {"evidence_in_humans": "True"}}

951. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_03`

HALLUCINATION

- **Question:** What penalty applies to an agency that fails to provide relevant information to a committee?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "agency_responsibilities", "parameters": {"a": "penalty for failing to provide relevant information"}}
- ✓ **VERITROOPER (PASS):** I don't have enough information to answer the question.

952. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1194\_03`

HALLUCINATION

- **Question:** Who in management must sign off to authorize cutting or welding in a restricted area?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "designated management representative", "parameters": {}}
- ✓ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

953. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4065\_01`

- **Question:** When the manufacturer rejects a proposed modification after reviewing its technical safety merits, when may the employer still treat the situation as a manufacturer refusal to review under paragraph (a)(2)?
- **Ground truth:** When the manufacturer rejects the proposal but does not explain the reasons for the rejection in writing.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "paragraph (a)(2)", "parameters": {"manufacturer_rejects_proposal": "True", "manufacturer_explains_rejection": "False"}}
- ✓ **VERITROOPER (PASS):** {"name": "manufacturer_refusal_to_review", "parameters": {"reason": "manufacturer rejects the proposal but does not explain the reasons for the rejection in writing"}}

954. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2016\_03`

- **Question:** Using the 1980 RTECS listing counts described in Appendix B, how many of the 168,096 total listings are NOT names of different chemicals with associated toxicity data?
- **Ground truth:** 122,940. Of the 168,096 total listings, 45,156 are names of different chemicals with their associated toxicity data; the remaining 122,940 are synonyms ($168,096 - 45,156 = 122,940$).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_synonyms", "parameters": {"total_listings": "168096", "different_chemicals": "45156"}}
- ✗ **VERITROOPER (FAIL):** {"name": "subtract", "parameters": {"a": "168096", "b": "45156"}}

955. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0882\_02`

- **Question:** Per the periodic inspection requirement in paragraph (c)(6)(i)(D) when tagout is used, the review must include the elements set forth in another paragraph. According to that referenced paragraph (c)(7)(ii), how many tag limitations must be reviewed/trained?
- **Ground truth:** Six limitations of tags must be reviewed (paragraphs (A) through (F) of (c)(7)(ii)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "training", "parameters": {"D": "6"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_tag_limitations", "parameters": {"paragraph": "(c)(7)(ii)", "tag_limitations": "6"}}

956. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0411\_09`

- **Question:** In Table F-1 (males), what is the age correction value at 2000 Hz for age 32?
- **Ground truth:** 7
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1 (males)", "frequency": "2000", "age": "32"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_age_correction_value", "parameters": {"table": "Table F-1", "gender": "males", "frequency": "2000 Hz", "age": "32"}}

957. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1186\_01`

- **Question:** Which wheel drive disengaging controls are exempt from the requirement to move opposite to the direction of vehicle motion in order to disengage the drive?
- **Ground truth:** Deadman controls.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "wheel drive disengaging controls", "parameters": {"type": "deadman controls"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_exempt_controls", "parameters": {"control_type": "deadman controls"}}

958. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4517\_00`

- **Question:** Under paragraph (b) regarding the nature of filing and paragraph (c) regarding place of filing, if a complainant who cannot speak English wishes to file orally, what two accommodations does the rule provide and where should the complaint be directed?
- **Ground truth:** Per paragraph (b), the complaint may be filed orally (OSHA will reduce it to writing) and OSHA will accept the complaint in any language if the complainant cannot file in English. Per paragraph (c), it should be filed with the OSHA office responsible for enforcement activities in the geographical area where the complainant resides or was employed, but may be filed with any OSHA officer or employee.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "Filing of retaliation complaint", "parameters": {"b": "No particular form of complaint is required. A complaint may be filed orally or in writing. Oral complaints will be reduced to writing by OSHA. If the complainant is unable to file the complaint in English, OSHA will accept the complaint in any language.", "c": "The complaint should be filed with the OSHA office responsible for enforcement activities in the geographical area where the complainant resides or was employed, but may be

filed with any OSHA officer or employee. Addresses and telephone numbers for these officials are set forth in local directories and at the following Internet address: <http://www.osha.gov>."}}

- ✓ **VERITROOPER (PASS)**: {"name": "filing_accommodations", "parameters": {"language_accommodation": "OSHA will accept the complaint in any language", "location_accommodation": "The complaint should be filed with the OSHA office responsible for enforcement activities in the geographical area where the complainant resides or was employed"}}

959. PASS · Precision · `OSHA\_29CFR\_full\_2024\_2809\_08`

- **Question**: What is the NFPA number for the Safety Standard for Laboratories in Health Related Institutions?

- **Ground truth**: NFPA, 56c, 1980

- ✓ **Baseline (vanilla RAG) (PASS)**: {"name": "get_nfpa_number", "parameters": {"institution": "Laboratories in Health Related Institutions"}}

- ✓ **VERITROOPER (PASS)**: {"name": "NFPA", "parameters": {"number": "56c", "year": "1980"}}

960. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_1259\_03`

HALLUCINATION

- **Question**: What is the penalty for permitting a person to ride on a powered hand truck not designed for riding?

- **Ground truth**: UNANSWERABLE

- ✗ **Baseline (vanilla RAG) (FAIL)**: {"name": "penalty_for_riding_on_powered_hand_truck", "parameters": {"penalty": "shall not be permitted"}}

- ✓ **VERITROOPER (PASS)**: The evidence does not contain enough information to answer the question.

961. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4060\_02`

- **Question**: When hoisting an employee in a chimney operation, may a boatswain's chair be used if the employer has not demonstrated that a personnel platform is infeasible?

- **Ground truth**: No; the employee must be in a personnel platform unless the employer can demonstrate that use of a personnel platform is infeasible.

- ✓ **Baseline (vanilla RAG) (PASS)**: {"name": "s", "parameters": {"1": "False"}}

- ✗ **VERITROOPER (FAIL)**: {"name": "is_boatswains_chair_allowed", "parameters": {"feasible": "False"}}

962. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1962\_01`

- **Question**: Combining paragraph (k) and subpart (k)(1), at what point must an exposed employee first be provided the opportunity for examinations and tests, and who bears the cost?

- **Ground truth**: At the time of initial assignment, or upon institution of medical surveillance, and the examinations must be provided without cost to the employee.

- ✓ **Baseline (vanilla RAG) (PASS)**: {"name": "medical_surveillance_program", "parameters": {"provides": "examinations and tests", "at": "the time of initial assignment, or upon institution of medical surveillance", "cost": "without cost to the employee"}}

- ✓ **VERITROOPER (PASS)**: {"name": "medical_surveillance", "parameters": {"employee": "exposed employee", "examinations_and_tests": "in accordance with paragraph (k)", "cost": "without cost to the employee"}}

963. FAIL · Cause & Effect · `OSHA\_29CFR\_full\_2024\_2160\_01`

WRONG CALC/QUOTE

- **Question:** What were the consequences for analytical bias and OAE resulting from the average recovery of 118.2% at 0.1 times and 104.0% in the range of 0.5 to 2.0 times the TWA PEL target concentration in the flame AAS analysis?
- **Ground truth:** The analytical bias found over the tested concentration range was +4.0% and the OAE was ±6.0%.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "calculate_consequences", "parameters": {"recovery_1": "118", "recovery_2": "104"}}
- ✖ **VERITROOPER (FAIL):** {"name": "calculate_consequences", "parameters": {"analytical_bias_1": "118.2", "analytical_bias_2": "104.0", "oae_1": "6.0", "oae_2": "6.0"}}

964. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_3312\_06`

- **Question:** If wind screens are to be used during high winds, what condition must first be met?
- **Ground truth:** The scaffold must be secured against the anticipated wind forces imposed; otherwise wind screens shall not be used.
- ✔ **Baseline (vanilla RAG) (PASS):** {"name": "condition_for_wind_screens", "parameters": {"condition": "The scaffold is secured against the anticipated wind forces imposed."}}
- ✔ **VERITROOPER (PASS):** {"name": "condition_for_wind_screens", "parameters": {"condition": "the scaffold is secured against the anticipated wind forces imposed"}}

965. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3806\_00`

HALLUCINATION

- **Question:** What specific airborne exposure level (in µg/m³) triggers the requirement to provide this information to the physician?
- **Ground truth:** UNANSWERABLE
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "information_to_physician", "parameters": {"actual_or_representative_exposure_level": "5"}}
- ✔ **VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

966. FAIL · Conditional · `OSHA\_29CFR\_full\_2024\_2784\_03`

MODEL ERROR

- **Question:** If an OSHA health standard specifies prohibition of eye and skin contact, must a laboratory observe it even though only the permissible exposure limit requirement generally applies?
- **Ground truth:** Yes, prohibition of eye and skin contact where specified by any OSHA health standard shall be observed.
- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "§ 1910.1450", "parameters": {"a": "2", "i": "True"}}
- ✔ **VERITROOPER (PASS):** {"name": "answer_question", "parameters": {"question": "If an OSHA health standard specifies prohibition of eye and skin contact, must a laboratory observe it even though only the permissible exposure limit requirement generally applies?", "answer": "Yes"}}

967. PASS · Exception · `OSHA\_29CFR\_full\_2024\_4366\_01`

- **Question:** Although an order of reinstatement is effective during ARB review, under what circumstance can that effectiveness be stayed?
- **Ground truth:** Unless the ARB grants a motion by the respondent to stay that order based on exceptional circumstances.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "stay_effectiveness_of_reinstatement_order", "parameters": {"circumstance": "exceptional circumstances"}}
- ✗ **VERITROOPER (FAIL):** {"name": "circumstance_for_stay", "parameters": {"circumstance": "exceptional circumstances"}}

968. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3187\_05`

- **Question:** Cylinders must be kept far enough from welding or cutting operations so sparks, hot slag, or flame will not reach them. What must be provided when this is impractical?
- **Ground truth:** When this is impractical, fire resistant shields shall be provided.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "provide_fire_resistant_shields", "parameters": {}}
- ✓ **VERITROOPER (PASS):** {"name": "provide_fire_resistant_shields", "parameters": {"impractical": "True"}}

969. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1597\_02`

- **Question:** Receptacles and cord connectors having grounding contacts must have those contacts effectively grounded, except in which two situations?
- **Ground truth:** Except for receptacles mounted on portable and vehicle-mounted generators in accordance with paragraph (g)(3), and replacement receptacles installed in accordance with paragraph (b)(2)(iv).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "receptacles_and_cord_connectors_exception", "parameters": {"situation1": "receptacles mounted on portable and vehicle-mounted generators", "situation2": "replacement receptacles installed in accordance with paragraph (b)(2)(iv) of this section"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_situations", "parameters": {"situation1": "receptacles mounted on portable and vehicle-mounted generators", "situation2": "replacement receptacles installed in accordance with paragraph (b)(2)(iv) of this section"}}

970. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1597\_00`

- **Question:** Receptacles installed on 15- and 20-ampere branch circuits must be of the grounding type, but under what exception is this not required?
- **Ground truth:** Except as permitted for replacement receptacles in paragraph (b)(2)(iv) of this section.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "receptacles_installed_on_15_and_20_ampere_branch_circuits_must_be_of_the_grounding_type_but_under_what_exception_is_this_not_required", "parameters": {"exception": "replacement receptacles in paragraph (b)(2)(iv) of this section"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_grounding_type_receptacles", "parameters": {"receptacle_type": "replacement", "condition": "as permitted for replacement receptacles in paragraph (b)(2)(iv) of this section"}}

971. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1798\_00`

- **Question:** When implementing a respiratory protection program under 29 CFR 134, which subparagraph is excluded from the requirement to comply with paragraphs (b) through (d)?
- **Ground truth:** Paragraph (d)(1)(iii) is excluded.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "respiratory protection program", "parameters": {"paragraph": "(d)(1)(iii)"}}
- ✓ **VERITROOPER (PASS):** {"name": "respiratory protection program", "parameters": {"requirement": "(d)(1)(iii)"}}

972. PASS · Precision · `OSHA\_29CFR\_full\_2024\_1431\_02`

- **Question:** During the in-service test under paragraph (j)(2)(iii)(E), what voltage per meter applies for a fiberglass tool and for how long?
- **Ground truth:** 246,100 volts per meter (75,000 volts per foot) for 1 minute
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "in-service test under paragraph (j)(2)(iii)(E)", "parameters": {"voltage_per_meter": "246100", "duration": "1 minute", "material": "fiberglass"}}
- ✓ **VERITROOPER (PASS):** {"name": "in_service_test", "parameters": {"voltage_per_meter": "246100", "duration": "1"}}

973. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1964\_02`

WRONG CALC/QUOTE

- **Question:** According to paragraphs (l)(2)(ii) and (l)(6), what must the sign at hazardous operations areas state, and what restriction applies to any additional statements placed near that sign?
- **Ground truth:** Under (l)(2)(ii) the sign must read: 'DANGER VINYL CHLORIDE MAY CAUSE CANCER WEAR RESPIRATORY PROTECTION AND PROTECTIVE CLOTHING IN THIS AREA AUTHORIZED PERSONNEL ONLY'. Under (l)(6), no statement may appear on or near it that contradicts or detracts from the effect of any required warning, information, or instruction.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "sign_at_hazardous_operations_areas", "parameters": {"text": "No Smoking or Open Flames", "restriction": "No other warning or caution signs are permitted in the area"}}
- ✓ **VERITROOPER (PASS):** {"name": "sign_at_hazardous_operations_areas", "parameters": {"required_statement": "DANGER VINYL CHLORIDE MAY CAUSE CANCER WEAR RESPIRATORY PROTECTION AND PROTECTIVE CLOTHING IN THIS AREA AUTHORIZED PERSONNEL ONLY", "restriction": "No statement shall appear on or near any required sign, label, or instruction which contradicts or detracts from the effect of any required warning, information, or instruction."}}

974. PASS · Precision · `OSHA\_29CFR\_full\_2024\_4630\_00`

- **Question:** Under § 1991.109, what is the maximum reasonable attorney fee the ALJ may award to a respondent if a complaint was frivolous or brought in bad faith?
- **Ground truth:** \$1,000
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_attorney_fee", "parameters": {"award_amount": "1000"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_attorney_fee", "parameters": {"fee_type": "maximum reasonable attorney fee", "complaint_type": "frivolous or brought in bad faith"}}

975. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4485\_06`

HALLUCINATION

- **Question:** What penalty does an employer face if found to have retaliated against a seaman under SPA?
- **Ground truth:** UNANSWERABLE
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "penalty_for_retaliation", "parameters": {"amount": "litigation costs, expert witness fees, and reasonable attorney fees"}}
- ✓ **VERITROOPER (PASS):** I don't have enough information to answer the question.

976. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4303\_02`

HALLUCINATION

- **Question:** Which agency office is responsible for compiling the accident, injury, and illness data referenced in paragraph (a)?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "Office of Federal Agency Programs", "parameters": {}}

- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer this question.

977. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1264\_02`

- **Question:** Using paragraphs (d)(2) and (d)(4), what two distinct hazards related to picker/opener machinery must be addressed - one concerning cleanout holes and one concerning material in the opener line?

- **Ground truth:** Per (d)(2), cleanout holes within reaching distance of the fan or picker beater must have securely fastened covers that are not opened while the machine is in motion. Per (d)(4), all textile opener lines must be equipped with magnetic separators, tramp iron separators, or other means for the removal of foreign ferrous material.

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "picker opener machinery hazard identification", "parameters": {"cleanout_holes": "covered with securely fastened covers", "opener_line": "equipped with magnetic separators, tramp iron separators, or other means for the removal of foreign ferrous material"}}

- **✓ VERITROOPER (PASS):** {"name": "address_hazards", "parameters": {"cleanout_holes": "unsecured covers", "material_in_opener_line": "foreign ferrous material"}}

978. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_4403\_06`

HALLUCINATION

- **Question:** What is the deadline for filing objections to the findings and preliminary orders?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "filing objections", "parameters": {"deadline": "30 days"}}

- **✓ VERITROOPER (PASS):** The evidence does not contain enough information to answer the question.

979. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2508\_04`

- **Question:** If an individual's legally permitted scope of practice allows them to independently provide health care services required by paragraph (i), do they qualify as a PLHCP?

- **Ground truth:** Yes, that individual qualifies as a physician or other licensed health care professional (PLHCP).

- **✓ Baseline (vanilla RAG) (PASS):** {"name": "PLHCP", "parameters": {"i": "True"}}

- **✓ VERITROOPER (PASS):** {"name": "is_PLHCP", "parameters": {"individual_scope": "true"}}

980. FAIL · Negative (hallucination traps) · `OSHA\_29CFR\_full\_2024\_neg\_3991\_03`

HALLUCINATION

- **Question:** How long after agreeing on voice signals must the workers document their agreement in writing?

- **Ground truth:** UNANSWERABLE

- **✗ Baseline (vanilla RAG) (FAIL):** {"name": "document_agreement", "parameters": {"duration": "30", "unit": "calendar days"}}

- **✓ VERITROOPER (PASS):** The provided evidence does not contain enough information to answer the question.

981. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_4228\_04`

- **Question:** When a significant Federal program change would have an adverse impact on the 'at least as effective' status of a State program if no parallel State modification were made, is State adoption required?

- **Ground truth:** Yes, State adoption of a change in response to the Federal program change shall be required.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "is_state_adoption_required", "parameters": {"change_type": "significant Federal program change"}}
- ✓ **VERITROOPER (PASS):** {"name": "evaluate_federal_program_change", "parameters": {"impact": "adverse", "status": "at least as effective"}}

982. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2190\_00`

- **Question:** If initial monitoring reveals an employee's exposure is at the action level but does not exceed the TWA, how often must the employer repeat monitoring for that employee?
- **Ground truth:** At least every year.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "periodic monitoring and monitoring frequency", "parameters": {"i": "every year"}}
- ✓ **VERITROOPER (PASS):** {"name": "repeat_monitoring", "parameters": {"monitoring_frequency": "every year"}}

983. FAIL · Exception · `OSHA\_29CFR\_full\_2024\_3987\_03`

WRONG CALC/QUOTE

- **Question:** If a warning (tag-out or do not operate) sign is on the equipment or starting control, when may the operator activate the switch or start the equipment?
- **Ground truth:** Only after the sign has been removed by a person authorized to remove it, or after the operator has verified that (A) no one is servicing, working on, or otherwise in a dangerous position on the machine, and (B) the equipment has been repaired and is working properly.
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "activate_switch_or_start_equipment", "parameters": {"sign_removed": "True", "verified_safe": "True"}}
- ✗ **VERITROOPER (FAIL):** {"name": "activate_switch_or_start_equipment", "parameters": {"warning_sign": "tag-out or do not operate", "authorized_removal": "True", "verification": "True"}}

984. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2239\_00`

- **Question:** According to the record-keeping requirements, both a copy of the physician's written opinion and the cytologic examination slide and written description for the most recent 10 years must be kept. Under which subparts are these two items respectively required?
- **Ground truth:** The copy of the physician's written opinion is required under paragraph (i)(b), and the cytologic examination slide and written description for the most recent 10 years is required under paragraph (ii)(g).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "record_keeping_requirements", "parameters": {"copy_of_physician_written_opinion": "(m)(2)(i)", "cytologic_examination_slide_and_written_description": "(m)(2)(ii)(G)"}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_record_keeping_requirements", "parameters": {"item": "copy of the physician"}}

985. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3110\_00`

MODEL ERROR

- **Question:** Under the marine service station rules, if a tank is installed on a pier (rather than on shore) and is located at an elevation producing gravity head on the dispensing unit, what two requirements must be met regarding maximum stored quantity and the tank outlet?
- **Ground truth:** Per paragraph (k)(2)(ii), the quantity stored on the pier may not exceed 1,100 gallons (4,163.5 L) aggregate capacity, and per paragraph (k)(2)(iv), the tank outlet must be equipped with a pressure control valve positioned adjacent to and outside the tank block valve, adjusted so liquid cannot flow by gravity in case of piping or hose failure.

- ✖ **Baseline (vanilla RAG) (FAIL):** {"name": "marine_service_station_tank_requirements", "parameters": {"elevation_producing_gravity": "True", "stored_quantity": "1100", "tank_outlet_requirement": "pressure control valve"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_marine_service_station_requirements", "parameters": {"tank_location": "pier", "elevation": "producing gravity head"}}

986. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3985\_03`

- **Question:** For which equipment on rails are rail clamps and rail stops NOT required?
- **Ground truth:** Portal cranes.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "rail clamps and rail stops exception", "parameters": {"equipment": "portal cranes"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_rail_clamps_and_stops", "parameters": {"equipment": "portal cranes"}}

987. PASS · Precision · `OSHA\_29CFR\_full\_2024\_0807\_03`

- **Question:** For Class 1 equipment, what is the maximum use voltage AC rms?
- **Ground truth:** 7,500
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "get_max_use_voltage", "parameters": {"class_equipment": "1"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_max_use_voltage", "parameters": {"class_of_equipment": "1"}}

988. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4269\_01`

- **Question:** Combining the policy statement referenced from section 2(b) with the exclusion in section 3(5), why does the Act's general policy not automatically protect State and local government employees?
- **Ground truth:** Although section 2(b) sets forth the policy of assuring every working man and woman safe and healthful working conditions, State and local government agencies are excluded from the definition of 'employer' in section 3(5), so public employees are not automatically covered and can only be protected under an approved State plan via section 18.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "section_3_5_exclusion", "parameters": {"policy_statement": "section 2(b)", "exclusion": "State and local government agencies are excluded from the definition of \\"}}
- ✖ **VERITROOPER (FAIL):** {"name": "get_policy_exclusion_reason", "parameters": {"policy_statement": "section 2(b)", "exclusion_section": "section 3(5)", "employee_type": "State and local government employees"}}

989. PASS · Calculation · `OSHA\_29CFR\_full\_2024\_3034\_03`

- **Question:** A site has 300 employees (more than 200). Combining required toilet seats and urinals per Table D-65.2, what is the total number of these facilities?
- **Ground truth:** 12
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_total_facilities", "parameters": {"number_of_employees": "300"}}
- ✓ **VERITROOPER (PASS):** {"name": "calculate_toilet_facilities", "parameters": {"number_of_employees": "300", "toilet_seats_per_n_employees": "50", "urinals_per_n_employees": "50"}}

990. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_0075\_01`

- **Question:** If a violation is abated in 1 working day, how long must the citation remain posted?
- **Ground truth:** For 3 working days, since the citation must remain posted until abatement or for 3 working days, whichever is later.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_posting_duration", "parameters": {"abatement_duration": "1", "citation_type": "abatement"}}

{"abatement_time": "1", "posting_time": "3"}}

- ✓ **VERITROOPER (PASS):** {"name": "max", "parameters": {"a": "3", "b": "1"}}

991. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4313\_00`

- **Question:** Under § 1960.59, when an agency provides occupational safety and health training, which group's training must include both introductory and specialized courses enabling participation in workplace safety and health inspections, and according to the same provision, what existing arrangements may this training requirement not alter?

- **Ground truth:** Per paragraph (b), employees who are representatives of employee groups, such as recognized labor organizations, must receive both introductory and specialized courses and materials enabling them to assist in conducting workplace safety and health inspections. The same paragraph states that nothing in it shall be construed to alter training provisions provided by law, Executive Order, or collective bargaining arrangements.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "§ 1960.59 Training of employees and employee representatives", "parameters": {"group": "employees of the agency who are representatives of employee groups, such as labor organizations which are recognized by the agency", "arrangements": "training provisions provided by law, Executive Order, or collective bargaining arrangements"}}

- ✓ **VERITROOPER (PASS):** {"name": "answer_cross_reference_question", "parameters": {"section_number": "1960.59", "group": "employees of the agency who are representatives of employee groups", "existing_arrangements": "law, Executive Order, or collective bargaining arrangements"}}

992. PASS · Exception · `OSHA\_29CFR\_full\_2024\_3985\_00`

- **Question:** The crane level indicator requirement does not apply to which types of equipment?

- **Ground truth:** Portal cranes, derricks, floating cranes/derricks and land cranes/derricks on barges, pontoons, vessels or other means of flotation.

- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "crane_level_indicator_requirement", "parameters": {"equipment_types": "[portal cranes, derricks, floating cranes/derricks, land cranes/derricks on barges, pontoons, vessels or other means of flotation]"}}

- ✓ **VERITROOPER (PASS):** {"name": "exception_to_crane_level_indicator_requirement", "parameters": {"equipment_types": "[portal cranes, derricks, floating cranes/derricks, land cranes/derricks on barges, pontoons, vessels or other means of flotation]"}}

993. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_0913\_03`

MODEL ERROR

- **Question:** What is the content focus of the program required in (g)(1) versus the training required in (g)(3)?

- **Ground truth:** The (g)(1) educational program familiarizes employees with the general principles of fire extinguisher use and the hazards involved with incipient stage fire fighting, whereas the (g)(3) training covers the use of the appropriate fire fighting equipment for designated employees.

- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "content_focus", "parameters": {"x": "program required in (g)(1)", "y": "training required in (g)(3)"}}

- ✓ **VERITROOPER (PASS):** {"name": "compare_program_content_focus", "parameters": {"g_1": "general principles of fire extinguisher use and the hazards involved with incipient stage fire fighting", "g_3": "use of the appropriate equipment"}}

994. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2204\_03`

- **Question:** According to the section on monocytosis and the section on a serious peripheral blood finding, what proportion of benzene-affected individuals show persistent monocytosis, and what proportion of those who ultimately develop acute myelogenous leukemia show the acquired Pelger-Huet anomaly?

- **Ground truth:** About 20 percent of individuals reported to have mild but persisting abnormalities from benzene exposure show persistent monocytosis, while only about two percent of patients who ultimately develop acute myelogenous leukemia show the acquired Pelger-Huet anomaly.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "calculate_proportion", "parameters": {"x": "20", "y": "2"}}
- ✓ **VERITROOPER (PASS):** {"name": "get_proportion", "parameters": {"monocytosis": "20", "leukemia": "2"}}

995. PASS · Conditional · `OSHA\_29CFR\_full\_2024\_2190\_04`

- **Question:** If periodic monitoring is to be discontinued because exposures are below the action level, what evidence is required?
- **Ground truth:** At least two consecutive measurements taken at least 7 days apart indicating exposures are below the action level.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "discontinue_periodic_monitoring", "parameters": {"exposures": "below action level", "evidence": "at least two consecutive measurements taken at least 7 days apart"}}
- ✓ **VERITROOPER (PASS):** {"name": "terminate_monitoring", "parameters": {"exposures": "below action level", "consecutive measurements": "at least two", "days apart": "at least 7"}}

996. FAIL · Cross-Reference · `OSHA\_29CFR\_full\_2024\_2239\_02`

MODEL ERROR

- **Question:** Using the transfer of records provisions, what must the successor employer do when the original employer ceases to do business, and what additional regulation governs the transfer?
- **Ground truth:** Whenever the employer ceases to do business, the successor employer shall receive and retain all records required to be maintained by paragraph (m) (paragraph (4)(i)). The employer must also comply with any additional requirements involving transfer of records set forth in 29 CFR 1910.1020(h) (paragraph (4)(ii)).
- ✗ **Baseline (vanilla RAG) (FAIL):** {"name": "transfer of records", "parameters": {"paragraph": "29 CFR 1910.1020(h)"}}
- ✓ **VERITROOPER (PASS):** {"name": "transfer_of_records", "parameters": {"action": "receive and retain all records required to be maintained by paragraph (m) of this section", "additional_regulation": "29 CFR 1910.1020(h)"}}

997. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_1349\_00`

- **Question:** According to the starting requirements for a chain saw, what two conditions must be met regarding both the starting location relative to the fueling area and the chain brake?
- **Ground truth:** The chain saw shall be started at least 10 feet (3 m) from the fueling area (per (v)) and shall be started with the chain brake engaged (per (vii)).
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "chain saw starting requirements", "parameters": {"starting location relative to fueling area": "at least 10 feet (3 m) away", "chain brake": "engaged"}}
- ✓ **VERITROOPER (PASS):** {"name": "start_chain_saw", "parameters": {"distance_from_fueling_area": "10", "chain_brake_engaged": "True"}}

998. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_4138\_02`

- **Question:** Under Option (3) rails and Option (4)(D), what additional safeguard must each system include beyond simply mounting the crane?
- **Ground truth:** Option (3) rails requires that rail clamps and rail stops are used unless the system is designed to prevent movement during operation by other means. Option (4)(D) requires that means are installed to prevent the crane/derrick from passing the forward or aft end of the wire rope attachments. Both require stops/clamps or means to limit travel along the system.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "rails and centerline cable system", "parameters": {"Option (3)":

"rails", "Option (4)(D)": "means are installed to prevent the crane/derrick from passing the forward or aft end of the wire rope attachments"}}
- ✓ **VERITROOPER (PASS):** {"name": "rail clamps and rail stops", "parameters": {"Option (3)": "rails", "Option (4)(D)": "means are installed to prevent the crane/derrick from passing the forward or aft end of the wire rope attachments"}}

999. PASS · Exception · `OSHA\_29CFR\_full\_2024\_1947\_00`

- **Question:** Although this section applies to the handling or use of vinyl chloride or polyvinyl chloride, in what case does it NOT apply?
- **Ground truth:** It does not apply to the handling or use of fabricated products made of polyvinyl chloride.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "exception_to_applicability", "parameters": {"section_applies_to": "handling or use of vinyl chloride or polyvinyl chloride", "exception": "the handling or use of fabricated products made of polyvinyl chloride"}}
- ✓ **VERITROOPER (PASS):** {"name": "exception_to_application", "parameters": {"case": "fabricated products made of polyvinyl chloride"}}

1000. PASS · Cross-Reference · `OSHA\_29CFR\_full\_2024\_3047\_00`

- **Question:** According to paragraph (i), one option for clean-up is to meet all requirements of paragraphs (b) through (o). What alternative does paragraph (ii) provide when the clean-up is done on plant property using plant or workplace employees?
- **Ground truth:** Under paragraph (ii), as an alternative to meeting all requirements of paragraphs (b) through (o), such employees shall have completed the training requirements of 29 CFR 1926.35, 1926.59, and 1926.103, and other appropriate safety and health training made necessary by the tasks they are expected to perform, such as personal protective equipment and decontamination procedures; additionally, all equipment used shall be in serviceable condition and inspected prior to use.
- ✓ **Baseline (vanilla RAG) (PASS):** {"name": "clean-up alternative", "parameters": {"option": "Where the clean-up is done on plant property using plant or workplace employees, such employees shall have completed the training requirements of the following: 29 CFR 1910.38, 1910.134, 1910.1200, and other appropriate safety and health training made necessary by the tasks they are expected to perform such as personal protective equipment and decontamination procedures."}}
- ✗ **VERITROOPER (FAIL):** {"name": "get_alternative_clean_up_option", "parameters": {"option": "clean-up is done on plant property using plant or workplace employees"}}