

Type 1 Binary

Our findings indicate that CoT and ToT are the most effective strategies for enhancing reasoning transparency in binary-type medical guideline questions, while Few-shot provides structured consistency, and Interactive prompting offers adaptability for fine-tuning responses. Given that binary medical queries often require not only a definitive answer but also a clear justification, CoT prompting is recommended as the optimal approach, balancing accuracy, interpretability, and efficiency.

Q1Basic Prompting (Zero-shot)

Q: According to AAOS guidelines, is glucosamine sulfate recommended for knee osteoarthritis?

A: According to the **American Academy of Orthopaedic Surgeons (AAOS) Clinical Practice Guideline** for knee osteoarthritis (last updated in 2013), **glucosamine sulfate is not recommended**. The AAOS issued a "**strong" recommendation against its use**", citing a lack of consistent evidence demonstrating significant clinical benefit in reducing pain or improving joint function. This conclusion was based on a comprehensive review of randomized controlled trials (RCTs) that showed no meaningful difference between glucosamine and placebo.

Key Points:

1. **2013 AAOS Guideline:** The recommendation against glucosamine (and chondroitin) is based on high-quality evidence from multiple studies failing to show efficacy.
2. **Strength of Recommendation:** "Strong" (indicating that the benefits of following the recommendation outweigh the risks).
3. **Rationale:** Inconsistent and inconclusive evidence across trials, with many showing no statistically or clinically significant improvement compared to placebo.

Note:

- The AAOS has not updated this specific recommendation in subsequent guidelines (as of the latest available information).
- Some studies, particularly those funded by supplement industries, suggest benefits, but these often lack robust methodology or conflict with independent research.

In summary, **AAOS guidelines do not recommend glucosamine sulfate for knee osteoarthritis** based on current evidence. Patients should consult healthcare providers for personalized advice.

Q2Exemplar-based Prompting (Few-shot)

Q: Below are some examples of Yes/No medical guideline responses:
1. Q: Is acetaminophen recommended for mild headaches? → A: Yes, acetaminophen is commonly recommended for mild headaches.
2. Q: Should antibiotics be used for viral infections? → A: No, antibiotics are ineffective against viruses.
Now, answer the following question:
Q: According to AAOS guidelines, is glucosamine sulfate recommended for knee osteoarthritis?

A: No, the AAOS guidelines do not recommend glucosamine sulfate for knee osteoarthritis due to insufficient evidence of significant clinical benefit.

Explanation:

The American Academy of Orthopaedic Surgeons (AAOS) issued a **strong recommendation** against glucosamine sulfate in their 2013 clinical practice guidelines for

symptomatic knee osteoarthritis. This conclusion was based on meta-analyses of high-quality studies showing no clinically meaningful improvement in pain or function compared to placebo. While some patients report subjective relief, AAOS guidelines emphasize evidence-based efficacy, which remains lacking for glucosamine.

Q3Reasoning-oriented (Chain-of-Thought, CoT)

Q: Think step by step before answering.
1. AAOS guidelines provide recommendations based on clinical evidence.
2. Glucosamine sulfate has been studied for knee osteoarthritis.
3. The AAOS stance is based on whether strong evidence supports its use.
Based on these facts, is glucosamine sulfate recommended for knee osteoarthritis according to AAOS?

A: Based on the provided facts and the AAOS (American Academy of Orthopaedic Surgeons) guidelines:

No, the AAOS does **not recommend glucosamine sulfate** for knee osteoarthritis. This is because the AAOS stance hinges on **strong evidence** supporting a treatment's efficacy, and multiple high-quality studies (including meta-analyses and randomized controlled trials) have shown **insufficient or inconsistent evidence** that glucosamine sulfate provides clinically significant benefits for pain relief or functional improvement in knee osteoarthritis.

The AAOS explicitly states this in its clinical practice guidelines, citing a lack of robust evidence to meet their threshold for recommendation. Thus, their position aligns with the principle of evidence-based decision-making outlined in the facts.

Q4Complex Reasoning Variants (Tree-of-Thoughts, ToT)

Q: Break down the problem into multiple reasoning paths:
- Path 1: Analyze the effectiveness of glucosamine sulfate in clinical studies.
- Path 2: Examine AAOS guidelines and their stance on supplements.
- Path 3: Weigh any conflicting recommendations.
Once each path is explored, determine whether AAOS recommends glucosamine sulfate for knee osteoarthritis.

A: Path 1: Effectiveness of Glucosamine Sulfate in Clinical Studies

Clinical studies on glucosamine sulfate yield mixed results:

- **Positive Findings:** Some European trials (e.g., MOVES trial) and subgroup analyses (e.g., GAIT trial subgroup with moderate-severe pain) suggest modest pain relief and functional improvement, particularly with prescription-grade glucosamine sulfate.
- **Negative Findings:** Larger studies like the GAIT trial in the U.S. found no significant benefit over placebo in the general population. Variability in formulations (e.g., sulfate vs. hydrochloride) and study quality contribute to inconsistent outcomes.

Path 2: AAOS Guidelines on Supplements

The AAOS 2013 clinical practice guidelines for knee osteoarthritis:

- **Explicit Recommendation:** Strongly recommends **against** glucosamine and chondroitin supplementation (grade of recommendation: "Strong" based on high-quality evidence showing no clinically significant benefit).
- **Rationale:** Systematic reviews concluded that evidence for pain reduction or functional improvement was lacking, even in high-quality randomized controlled trials.
- **Current Status:** As of 2023, AAOS has not updated this recommendation,

maintaining its stance.

Path 3: Conflicting Recommendations

Other organizations diverge from AAOS:

- **OARSI** (2023): Conditionally recommends glucosamine sulfate for knee OA, citing potential benefits in specific populations.
- **ESCEO**: Supports its use as part of a first-line pharmacological approach, emphasizing European trials with prescription formulations.
- **Discrepancy Drivers**: Differences in evidence interpretation (e.g., regional study focus, formulation quality, and outcome thresholds).

Conclusion:

The AAOS does not recommend glucosamine sulfate for knee osteoarthritis. Despite conflicting guidelines from other bodies, AAOS maintains its 2013 position due to insufficient evidence of clinical efficacy in rigorously controlled studies. This reflects a conservative, evidence-based approach prioritizing consistent high-quality data over variable regional findings.

Q5 Role-based Prompting (Persona/Role Prompting)

Q: You are a medical doctor specializing in osteoarthritis treatment. Your task is to answer based on AAOS guidelines.
Q: According to AAOS guidelines, is glucosamine sulfate recommended for knee osteoarthritis?

A: According to the American Academy of Orthopaedic Surgeons (AAOS) **2013 Clinical Practice Guideline** on the treatment of knee osteoarthritis, the AAOS issues a **strong recommendation against the use of glucosamine sulfate** (and chondroitin) for patients with symptomatic knee osteoarthritis. This recommendation is based on high-quality meta-analyses of randomized controlled trials that found no significant clinical benefit compared to placebo in reducing pain or improving function.

Key Points:

1. **Evidence Basis:** The guideline reflects a systematic review of evidence showing that glucosamine sulfate does not demonstrate meaningful efficacy in managing knee osteoarthritis symptoms.
2. **Strength of Recommendation:** The "strong recommendation" indicates that the harms/ineffectiveness likely outweigh any perceived benefits.
3. **Formulation Specificity:** The recommendation applies specifically to glucosamine sulfate, as studied in the evaluated trials.

Clinical Implication:

The AAOS advises clinicians to avoid prescribing glucosamine sulfate for knee osteoarthritis. Patients seeking alternative therapies should be guided toward evidence-based interventions (e.g., exercise, weight management, NSAIDs, or intra-articular corticosteroids as appropriate). For the most current guidance, clinicians should verify if updated AAOS guidelines have been released, but as of the latest available data, this recommendation remains in effect.

Q6 Interactive/Conversational (Interactive Prompting)

Q: Let's work together to get the best answer.
1. What do AAOS guidelines say about glucosamine sulfate?
2. How does the clinical evidence support or contradict this

recommendation?\\n3. Based on this, do you think AAOS recommends it?\\n\\nProvide an initial answer, and we will refine it together.

A: **Final Answer:**

1. **AAOS Guidelines on Glucosamine Sulfate:**

The American Academy of Orthopaedic Surgeons (AAOS) 2013 clinical practice guideline for knee osteoarthritis **recommends against the use of glucosamine sulfate** (Strength of Recommendation: *Strong*), citing a lack of robust clinical evidence demonstrating meaningful benefit.

2. **Clinical Evidence Overview:**

- **Contradictory Evidence:** Large randomized controlled trials (e.g., the NIH-funded GAIT trial) and meta-analyses often found no significant difference between glucosamine sulfate and placebo in pain relief or functional improvement.
- **Supportive Evidence:** Some European studies (e.g., using patented Rotta-formulated glucosamine sulfate) reported modest benefits, but these results were inconsistent and not replicated in broader populations.
- **Key Limitations:** Heterogeneity in study design, formulation differences (sulfate vs. hydrochloride), and potential industry funding bias contribute to conflicting conclusions.

3. **AAOS Recommendation Status:**

Based on the 2013 guidelines (still current as of the latest update), AAOS **does not recommend glucosamine sulfate** for knee osteoarthritis. While other organizations (e.g., OARSI) offer conditional recommendations, AAOS maintains its stance due to insufficient high-quality evidence meeting their stringent criteria.

Refinements:

- Clarified the distinction between glucosamine formulations (sulfate vs. hydrochloride) and their varying clinical outcomes.
- Highlighted the 2013 guideline's current validity and the rationale for AAOS's strong recommendation against use.
- Contrasted AAOS's position with other guidelines to contextualize the recommendation.

Type_2_Multiple_choice_questions

Based on a comparative analysis of different prompting techniques, the Chain-of-Thought (CoT) approach emerges as the optimal choice for this task. Unlike basic and exemplar-based prompting, which provide concise but shallow explanations, CoT systematically breaks down optimizer performance using a structured evaluation of convergence speed, stability, and trade-offs. While Tree-of-Thoughts (ToT) offers an advanced perspective by considering learning rate decay and divergence minimization, its complexity is more suited for hyperparameter tuning rather than a general optimizer selection. Role-based prompting, though practical for real-world deployment, lacks the theoretical depth required for academic analysis. Therefore, CoT strikes the best balance between logical reasoning, empirical justification, and clarity, making it the most suitable approach for this assignment.

Q1. Basic Prompting (Zero-shot)

Q: Based on training a 3-layer MLP (ReLU activation, batch normalization) on the UCI

Banknote dataset, does Adam optimizer converge faster than SGD?

- A) Yes
- B) No

A:

- A) Yes

Explanation:

Adam optimizer typically converges faster than SGD in scenarios where hyperparameters (like learning rate) are not extensively tuned. Key reasons include:

Adaptive Learning Rates: Adam adjusts learning rates per-parameter, which is advantageous in navigating the loss landscape efficiently, especially with ReLU activations and batch normalization smoothing the optimization terrain.

Robustness to Hyperparameters: Adam's default settings often work well without meticulous tuning, whereas SGD's performance heavily relies on an optimal learning rate. This makes Adam more likely to converge faster out-of-the-box.

Empirical Evidence: Studies (e.g., the Adam paper) and practical experience show Adam's faster initial convergence compared to SGD, particularly in moderate-sized networks like a 3-layer MLP. While SGD with momentum and careful scheduling might match or exceed final performance, Adam generally achieves quicker convergence, aligning with the UCI Banknote dataset's scale and architecture.

Thus, A) Yes is the correct answer.

Q2. Exemplar-Based Prompting (Few-shot)

Q: Example 1:

Task: Image classification (ResNet-18)

Observation: Adam converges faster than SGD due to adaptive learning rates.

Example 2:

Task: Text generation (Transformer)

Observation: RMSProp reduces oscillations compared to Momentum.

Question: For the banknote task (3-layer MLP, 4 input features), which optimizer is optimal?

- A) SGD (slow but stable)
- B) Adam (fast and stable)
- C) RMSProp (fastest but unstable)
- D) Momentum (moderate speed)

A:For the banknote classification task using a 3-layer MLP with 4 input features, Adam (B) is the optimal optimizer. Here's the reasoning:

Task Similarity: The banknote task is a classification problem, aligning more closely with Example 1 (image classification) than Example 2 (text generation). In Example 1, Adam's adaptive learning rates enabled faster and stable convergence compared to SGD. A 3-layer MLP, while simpler than ResNet-18, still benefits from adaptive optimization.

Adaptive Learning Rates: Adam dynamically adjusts learning rates per-parameter, which is advantageous even in smaller networks. It balances speed and stability by combining momentum-like behavior (for acceleration) and adaptive gradient scaling (to reduce oscillations). This avoids the sluggishness of SGD (A) and the instability risk of RMSProp (C).

Stability vs. Speed: While RMSProp (C) reduced oscillations in the Transformer example, the question labels it as "fastest but unstable" here. For a simpler MLP, Adam's stability (shown in Example 1) outweighs RMSProp's potential instability. Momentum (D) offers moderate speed but lacks Adam's adaptive tuning, making it less optimal.

Answer: B) Adam (fast and stable)

Q3. Reasoning-Oriented (Chain-of-Thought, CoT)

Q:Analyze step-by-step:

1. **Convergence Speed**: Epochs to reach loss=0.05 (Adam:50, SGD:200, RMSProp:25).
2. **Stability**: Gradient oscillation range (Adam: $\pm 5\%$, RMSProp: $\pm 20\%$, SGD: $\pm 15\%$).
3. **Trade-off**: RMSProp converges fastest but risks early divergence.

Question: Which optimizer is optimal overall?

- A) SGD
- B) Adam
- C) RMSProp
- D) None

A:

Step-by-Step Analysis:

Convergence Speed:

RMSProp (25 epochs) is the fastest, followed by Adam (50 epochs), and then SGD (200

epochs). Speed prioritization: C > B > A.

Stability:

Adam ($\pm 5\%$ oscillation) is the most stable, followed by SGD ($\pm 15\%$), and RMSProp ($\pm 20\%$). Stability prioritization: B > A > C.

Trade-off:

While RMSProp (C) converges fastest, its high oscillation range ($\pm 20\%$) and risk of early divergence make it unreliable. Adam (B) balances speed and stability, whereas SGD (A) is too slow and moderately unstable.

Conclusion:

Despite RMSProp's speed, its instability and divergence risk outweigh its benefits. Adam offers the best compromise between convergence speed and stability, making it the optimal overall choice.

Answer: B) Adam

Q4. Complex Reasoning Variants (Tree-of-Thoughts, ToT)

Evaluate strategies for the banknote task (3-layer MLP, batch size=32):

1. **Adam** (learning rate=0.001)
2. **Adam + Learning Rate Decay** ($lr=0.001 \rightarrow 0.0001$)
3. **RMSProp + Momentum** ($lr=0.001, \beta=0.9$)

Question: Which strategy minimizes training divergence?

- A) Strategy 1
- B) Strategy 2
- C) Strategy 3
- D) All

A:

To determine which strategy minimizes training divergence for a 3-layer MLP on the banknote task, we analyze the three strategies:

Adam ($lr=0.001$): Adam adapts learning rates per parameter, combining momentum and squared gradient scaling. While generally robust, a fixed learning rate might risk oscillations or divergence if the initial rate is too high for the problem.

Adam + Learning Rate Decay ($lr=0.001 \rightarrow 0.0001$): Learning rate decay gradually reduces the

step size, mitigating divergence risks by preventing overshooting minima in later training stages. This adds stability on top of Adam's adaptability, addressing potential instability from a fixed high learning rate.

RMSProp + Momentum ($\beta=0.9$): RMSProp adapts learning rates via gradient magnitude averages, while momentum accelerates convergence. However, Adam (used in Strategies 1 and 2) incorporates similar mechanisms with additional bias correction, often making it more reliable. A fixed learning rate here still risks divergence if poorly tuned.

Conclusion: Strategy 2 (Adam + Learning Rate Decay) is most effective at minimizing divergence. The decaying learning rate provides an extra stabilization layer, reducing the likelihood of unstable updates compared to fixed-rate strategies (1 and 3). While RMSProp + Momentum (Strategy 3) is adaptive, Adam's bias correction and combined momentum make Strategies 1 and 2 inherently more robust, with Strategy 2's decay further enhancing stability.

Answer: B) Strategy 2

Q5. Role-Based Prompting (Persona/Role Prompting)

Act as an AI consultant for a fintech company. Recommend an optimizer for real-time banknote fraud detection (3-layer MLP, UCI dataset):

- A) SGD (slow but low compute cost)
- B) Adam (fast, 99.5% accuracy)
- C) RMSProp (fast but unstable)
- D) Momentum (moderate speed)

A:

Recommendation: B) Adam

Rationale:

For real-time banknote fraud detection using a 3-layer MLP on the UCI dataset, Adam is the optimal choice due to its fast convergence and high accuracy (99.5%), as stated in the problem. Here's why:

Speed vs. Stability:

Adam combines adaptive learning rates (like RMSProp) and momentum, enabling faster training than SGD or standard Momentum.

Unlike RMSProp, Adam is less prone to instability during training, which is critical for reliability in fraud detection systems.

Accuracy:

The 99.5% accuracy benchmark directly aligns with fraud detection's need for precision, where false negatives (missed fraud) are costly.

Real-Time Requirements:

While SGD has low compute cost per iteration, its slow convergence is unsuitable for real-time systems needing rapid model updates.

Adam's efficiency in reaching high accuracy faster offsets its marginally higher compute cost per epoch, especially for a shallow network like a 3-layer MLP.

Practical Tradeoffs:

Momentum (D) offers moderate speed but lacks Adam's adaptive learning rate tuning, risking suboptimal convergence.

RMSProp (C) is fast but unstable, which could compromise reliability in a production environment.

Conclusion: Adam balances speed, stability, and accuracy, making it the best fit for real-time fraud detection where performance and reliability are paramount.

Q6. Interactive/Conversational (Interactive Prompting)

User: "Should I use SGD for stability in banknote classification?"

AI: "Adam matches SGD's stability ($\pm 5\%$ vs. $\pm 15\%$ gradient noise). Proceed with Adam?"

Options:

A) Yes

B) No

A:

Answer:

A) Yes

Explanation:

The AI indicates that Adam offers comparable stability to SGD ($\pm 5\%$ gradient noise for Adam vs. $\pm 15\%$ for SGD), implying Adam is more stable with lower noise. Since stability is the user's priority and Adam achieves this while potentially converging faster (due to adaptive learning rates), proceeding with Adam is optimal for banknote classification.

Type_3_Computational_problems

All methods achieved 100% accuracy, confirming that the model had sufficient mathematical knowledge. ToT was the most structured approach, explicitly validating solutions through

multiple reasoning paths. CoT effectively reduced ambiguity, reinforcing each computational step through logical sequencing. Few-shot ensured format consistency but did not provide additional verification advantages. Interactive prompting was useful for guided problem-solving, especially for step-by-step learning scenarios.

Q1: Basic Prompting (Zero-shot)

Q: Solve for x in the equation:
$$3x + 5 = 14$$

A: To solve the equation $3x+5=14$:

1. **Subtract 5 from both sides** to isolate the term with x :

$$3x+5-5=14-5 \Rightarrow 3x=9$$

2. **Divide both sides by 3** to solve for x :

$$3x/3=9/3 \Rightarrow x=3$$

Verification: Substitute $x=3$ back into the original equation:

$$3(3)+5=9+5=14 \checkmark$$

Answer: 3

Q2: Exemplar-based Prompting (Few-shot)

Q: Example responses:
1. Q: Solve for x : $2x + 4 = 10$.
A: $2x = 6 \rightarrow x = 3 \checkmark$

2. Q: Solve for x : $5x - 3 = 12$.
A: $5x = 15 \rightarrow x = 3 \checkmark$

Now solve: $3x + 5 = 14$.

A: A: $3x = 9 \rightarrow x = 3 \checkmark$

Check: $3(3) + 5 = 9 + 5 = 14$. Correct!

Answer:

$$3x + 5 = 14$$

$$3x = 9$$

$$x = 3 \checkmark$$

Q3: Reasoning-oriented (Chain-of-Thought, CoT)

Q: Think step by step before solving:
1. Start with the equation: $3x + 5 = 14$.
2. Subtract 5 from both sides: $3x = 9$.
3. Divide both sides by 3: $x = 3$.
What is the value of x ?

A: To solve the equation $3x+5=14$:

1. **Subtract 5 from both sides** to isolate the term with x :

$$3x+5-5=14-5 \Rightarrow 3x=9$$

2. **Divide both sides by 3** to solve for x :

$$3x/3=9/3 \Rightarrow x=3$$

Verification: Substitute $x=3$ back into the original equation:

$$3(3)+5=9+5=14$$

The left side equals the right side, confirming the solution is correct.

Answer: 3

Q4: Complex Reasoning Variants (Tree-of-Thoughts, ToT)

Q: Break the solution into multiple paths:
- Path 1: Isolate x by first subtracting 5 from both sides.
- Path 2: Solve for x by dividing both sides by 3.
- Path 3: Verify the solution by plugging x back into the original equation.
After analyzing these paths, determine the correct value of x .

A: **Solution:**

- **Path 1: Isolate xx by subtracting 5 from both sides**

Starting with the equation:

$$3x+5=14$$

Subtract 5 from both sides:

$$3x+5-5=14-5 \Rightarrow 3x=9$$

- **Path 2: Solve for xx by dividing both sides by 3**

Divide both sides of $3x=9$ by 3:

$$3x/3=9/3 \Rightarrow x=3$$

- **Path 3: Verify the solution by substitution**

Substitute $x=3$ back into the original equation:

$$3(3)+5=9+5=14$$

This matches the right-hand side of the equation, confirming the solution is correct.

Final Answer:

$$3$$

Q5: Role-based Prompting (Persona/Role Prompting)

Q: You are a mathematics tutor helping students solve linear equations. Provide a step-by-step explanation to solve for x in the equation:
 $3x + 5 = 14$.

A: To solve the equation $3x+5=14$ for xx , follow these steps:

1. **Subtract 5 from both sides** to isolate the term with xx :

$$3x+5-5=14-5$$

Simplifying both sides:

$$3x=9$$

2. **Divide both sides by 3** to solve for xx :

$$3x/3=9/3$$

This simplifies to:

$$x=3$$

Verification: Substitute $x=3$ back into the original equation:

$$3(3)+5=9+5=14$$

The left side equals the right side, confirming the solution is correct.

Answer: $x=3$

Q6: Interactive/Conversational (Interactive Prompting)

Q: Let's solve this together step by step.
 1. What happens if we subtract 5 from both sides?
 2. What is $14 - 5$?
 3. Now, what do we get when we divide by 3?
 Provide an initial response, and we will refine it together.

A: Let's work through this step by step:

1. **Subtract 5 from both sides:**

$$3x+5-5=14-5$$

Simplifies to:

$$3x=9$$

2. **Calculate $14-5$:**

$$14-5=9$$

3. **Divide by 3:**

$$3 \times 3 = 9$$

Result:

$$x = 3$$

Final Answer: $x = 3$

Let me know if you'd like further clarification or adjustments!