

1. Question type 1: Yes/No questions (Binary Classifications)

Original Question:

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

a) Technique 1: Basic Prompting (Zero-shot)

i. Input: “According to AAOS guidelines, is glucosamine sulfate recommended for knee osteoarthritis?”

ii. Output:

“According to the American Academy of Orthopaedic Surgeons (AAOS) guidelines for the treatment of knee osteoarthritis, glucosamine sulfate is not recommended.

The AAOS specifically recommends against using glucosamine and chondroitin sulfate for patients with symptomatic knee osteoarthritis. This recommendation is based on their review of the scientific evidence, which found insufficient support for the effectiveness of these supplements in treating knee osteoarthritis symptoms.

This position has been consistent across multiple updates of their clinical practice guidelines. While some patients report subjective improvement with these supplements, the AAOS guidelines are based on systematic reviews of controlled clinical trials that did not demonstrate significant benefits compared to placebo.

If you're looking for management options for knee osteoarthritis, the AAOS does recommend several other approaches, including weight loss (if applicable), exercise, physical therapy, NSAIDs, and in appropriate cases, corticosteroid injections or surgical interventions.

I should note that you may want to consult with your healthcare provider for personalized medical advice about managing knee osteoarthritis.”

b) Technique 2: Exemplar-based prompting (Few-shot)

i. Input: “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.\n\nNow, answer the following question:

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

ii. Output:

“Based on the American Academy of Orthopedic Surgeons (AAOS) guidelines, the answer is: No, glucosamine sulfate is not recommended for knee osteoarthritis.

The AAOS guidelines specifically recommend against using glucosamine sulfate for patients with symptomatic knee osteoarthritis. This recommendation stems from multiple clinical trials that failed to demonstrate significant benefits over placebo in terms of pain relief, functional improvement, or disease modification.

While glucosamine supplements are popular and widely available over-the-counter, the scientific evidence does not support their use according to these professional guidelines.”

c) Technique 3: Reasoning-oriented (Chain-of-Thought, CoT)

i. Input: “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?”

ii. Output:

“I'll think step by step to answer whether glucosamine sulfate is recommended for knee osteoarthritis according to the AAOS guidelines.

Step 1: AAOS stands for the American Academy of Orthopaedic Surgeons, which develops evidence-based clinical practice guidelines.

Step 2: AAOS guidelines are based on clinical evidence, with stronger recommendations given to treatments with strong supporting evidence.

Step 3: Glucosamine sulfate has been studied for knee osteoarthritis, but I need to determine the AAOS's specific stance.

Step 4: Based on my knowledge, the AAOS guidelines do not recommend glucosamine sulfate for knee osteoarthritis. In their evidence-based guidelines, they have stated that the evidence does not support the use of glucosamine for knee osteoarthritis.

Step 5: The AAOS specifically issued a "strong" recommendation against glucosamine use, which means they found high-quality evidence that did not support its effectiveness compared

to placebo.

Therefore, glucosamine sulfate is not recommended for knee osteoarthritis according to AAOS guidelines. The AAOS stance is that there is not strong evidence supporting its use for this condition.”

d) Technique 4: Complex Reasoning Variants (Tree-of-Thoughts, ToT)

- i. Input: “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.”

- ii. Output:

“I'll think step by step to answer whether glucosamine sulfate is recommended for knee osteoarthritis according to the AAOS guidelines.

Step 1: AAOS stands for the American Academy of Orthopaedic Surgeons, which develops evidence-based clinical practice guidelines.

Step 2: AAOS guidelines are based on clinical evidence, with stronger recommendations given to treatments with strong supporting evidence.

Step 3: Glucosamine sulfate has been studied for knee osteoarthritis, but I need to determine the AAOS's specific stance.

Step 4: Based on my knowledge, the AAOS guidelines do not recommend glucosamine sulfate for knee osteoarthritis. In their evidence-based guidelines, they have stated that the evidence does not support the use of glucosamine for knee osteoarthritis.

Step 5: The AAOS specifically issued a "strong" recommendation against glucosamine use, which means they found high-quality evidence that did not support its effectiveness compared to placebo.

Therefore, glucosamine sulfate is not recommended for knee osteoarthritis according to AAOS guidelines. The AAOS stance is that there is not strong evidence supporting its use for this condition.”

e) Technique 5: Role-based Prompting (Persona/Role Prompting)

- i. Input: “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?"

ii. Output:

"As a medical professional referencing the American Academy of Orthopaedic Surgeons (AAOS) guidelines, I can inform you that glucosamine sulfate is not recommended for the treatment of knee osteoarthritis according to AAOS guidelines.

The AAOS guidelines specifically recommend against using glucosamine and chondroitin sulfate, either alone or in combination, for patients with symptomatic knee osteoarthritis. This recommendation is based on high-quality evidence that these supplements do not provide clinically significant improvement in patient-reported outcomes.

It's important to note that while these supplements are widely available over the counter and some patients report subjective benefits, the current clinical evidence does not support their use as an effective treatment for knee osteoarthritis according to AAOS guidelines.

I should emphasize that anyone dealing with knee osteoarthritis should consult with their own healthcare provider for personalized treatment recommendations. Guidelines are important, but individual cases may have specific considerations that affect treatment decisions."

f) Technique 6: Interactive/Conversational (Interactive Prompting)

- i. Input: "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? 3. Based on this, do you think AAOS recommends it? Provide an initial answer, and we will refine it together"

ii. Output:

"I'll help you analyze the AAOS (American Academy of Orthopaedic Surgeons) guidelines regarding glucosamine sulfate and the related clinical evidence.

1. What do AAOS guidelines say about glucosamine sulfate?

The AAOS guidelines do not recommend glucosamine sulfate for the treatment of knee osteoarthritis. In their most recent clinical practice guidelines for knee osteoarthritis treatment, the AAOS issued a "strong" recommendation against using glucosamine sulfate. This recommendation was based on their systematic review of available clinical evidence, which they determined did not support its efficacy for symptom relief or disease modification in knee

osteoarthritis.

2. How does the clinical evidence support or contradict this recommendation?

The clinical evidence regarding glucosamine sulfate is mixed:

Evidence supporting the AAOS recommendation:

- Multiple large randomized controlled trials (including the GAIT trial) found no significant benefits of glucosamine over placebo for pain relief or functional improvement
- Meta-analyses have shown inconsistent results, with many showing minimal to no clinically significant benefit
- Methodological issues in positive studies, including industry funding and potential publication bias
- Heterogeneity in formulations used across studies (glucosamine sulfate vs. glucosamine hydrochloride)

Evidence potentially contradicting the AAOS recommendation:

- Some European studies have shown modest benefits for pain relief and possibly slowing joint space narrowing
- Some meta-analyses suggest small but statistically significant improvements in pain and function
- Subgroup analyses suggest certain patients might benefit more than others
- Safety profile is generally favorable with minimal adverse effects

3. Based on this, do you think AAOS recommends it?

Based on the information above, the AAOS does not recommend glucosamine sulfate for knee osteoarthritis treatment. Their strong recommendation against its use indicates that they believe the quality of evidence does not support clinical benefit that would outweigh costs and potential risks, even though those risks are minimal.

This position contrasts somewhat with other organizations like EULAR (European League Against Rheumatism), which has taken a more permissive stance on glucosamine based on

some positive European studies.

Would you like me to delve deeper into any particular aspect of the evidence or recommendations?"

2. Question type 2: Multiple-choice questions (Reasoning & knowledge application)

Original Question: In a banknote classification task (UCI Banknote Authentication Dataset, 4 input features), which optimizer (SGD, Adam, RMSProp) balances convergence speed and stability best?

a) Technique 1: Basic Prompting (Zero-shot)

- i. Input: “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

”

- ii. Output:

To answer this question accurately, I’ll need to analyze what’s typically observed when comparing Adam and SGD optimizers on a 3-layer MLP with ReLU activation and batch normalization for the UCI Banknote dataset.

A) Yes – Adam optimizer typically converges faster than SGD.

This is the correct answer because:

1. The Adam optimizer combines the benefits of AdaGrad and RMSprop by maintaining per-parameter learning rates that are adapted based on both the first and second moments of the gradients.
2. For small datasets like the UCI Banknote dataset (which contains only 1,372 instances with 4 features), Adam’s adaptive learning rate approach provides significant advantages in convergence speed.
3. When using ReLU activation functions and batch normalization (as specified in your question), the training process becomes more stable, but Adam still maintains its speed advantage over SGD.
4. SGD, while effective, requires more careful tuning of learning rates and often needs more epochs to reach the same performance level as Adam for this particular architecture and dataset.
5. The 3-layer architecture is relatively simple, which plays to Adam’s strengths in quickly finding good parameter values without excessive fine-tuning.

The empirical evidence consistently shows that Adam converges in fewer iterations than standard SGD for this particular configuration.

b) Technique 2: Exemplar-based prompting (Few-shot)

Input: “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)

Options:

- A. Surgery
- B. Statin therapy
- C. Antibiotics
- D. Physical therapy”

Output:

1 Based on the examples and the characteristics of the banknote task, I believe option B) Adam (fast and stable) would be the optimal optimizer.

2 The banknote task appears to be a classification problem with a relatively simple architecture (3-layer MLP) and low-dimensional input (4 features). From the examples provided:

1. Example 1 shows that Adam converges faster than SGD in an image classification task due to its adaptive learning rates.
2. Example 2 indicates that RMSProp reduces oscillations compared to Momentum in a text generation task.

3 Adam combines the benefits of RMSProp (handling sparse gradients well) and Momentum (smoothing oscillations). For a small MLP with 4 input features:

- The problem is likely well-conditioned compared to deeper networks or more complex tasks
- Adam’s adaptive learning rates would help reach convergence quickly

- The stability mentioned in option B matches the requirement for a classification task where consistent performance is important
- The low-dimensional input space (only 4 features) would benefit from Adam's efficient parameter updates

4 While SGD might eventually reach good results, it would be slower than necessary. RMSProp might be faster in some cases but is described as unstable, which is undesirable for a classification task. Momentum offers moderate speed but doesn't provide the adaptive benefits of Adam.

c) Technique 3: Reasoning-oriented (Chain-of-Thought, CoT)

Input: "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:±5%, RMSProp:±20%, SGD:±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

Output:

Based on the information provided about the banknote classification task, I'll analyze each optimizer's performance to determine which one offers the best balance between convergence speed and stability.

Let's analyze the data step-by-step:

Convergence Speed:

Adam: 50 epochs to reach loss=0.05

SGD: 200 epochs to reach loss=0.05

RMSProp: 25 epochs to reach loss=0.05

Stability (measured by gradient oscillation range):

Adam: $\pm 5\%$

SGD: $\pm 15\%$

RMSProp: $\pm 20\%$

Trade-off Analysis:

RMSProp converges fastest (25 epochs) but has the highest gradient oscillation ($\pm 20\%$), risking early divergence

SGD is the slowest (200 epochs) with moderate stability ($\pm 15\%$)

Adam takes a moderate time to converge (50 epochs) but offers the best stability ($\pm 5\%$)

When evaluating the optimal optimizer, we need to consider both factors. While RMSProp provides the fastest convergence, its high oscillation range makes it less reliable for this task. SGD is too slow despite moderate stability. Adam provides a good balance – its convergence speed is reasonably fast ($4\times$ faster than SGD), and it offers the best stability with minimal oscillations.

For a banknote classification task where reliability is crucial, the stability factor should be weighted heavily. Adam's excellent stability combined with reasonable convergence speed makes it the optimal choice.

Therefore, the answer is B) Adam.

d) Technique 4: Complex Reasoning Variants (Tree-of-Thoughts, ToT)

Input: "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, $\beta=0.9$)

”

Output:

Based on the characteristics of the UCI Banknote Authentication Dataset (which has 4 features and is a binary classification task) and the optimization strategies you've outlined, I'll analyze which approach would likely minimize training divergence.

For a relatively small dataset like the UCI Banknote Authentication with only 4 input features, the choice of optimizer can significantly impact convergence stability.

Strategy 2 (Adam + Learning Rate Decay) is likely to minimize training divergence because:

Adam provides adaptive learning rates for each parameter, which helps with the potentially different scales of the 4 banknote features.

The learning rate decay (0.001→0.0001) is particularly valuable in this context as it allows for:

- Initial fast learning with the 0.001 rate

- Gradually reducing step sizes as training progresses

- More stable convergence in later epochs

- Reduced oscillations around the minimum

While Strategy 1 (basic Adam) would converge quickly, it might oscillate near the optimum with the fixed learning rate. Strategy 3 (RMSProp + Momentum) would likely perform well but may not match the stability of Adam with learning rate decay for this particular task.

The answer is B) Strategy 2.

e) Technique 5: Role-based Prompting (Persona/Role Prompting)

Input: “In a banknote classification task (UCI Banknote Authentication Dataset, 4 input features), which optimizer (SGD, Adam, RMSProp) balances convergence speed and stability best? 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)

”

Output:

For your banknote classification task using the UCI Banknote Authentication Dataset with a 3-layer MLP, I recommend option B: Adam optimizer.

Adam is the best choice for your real-time banknote fraud detection system for several reasons:

- The 99.5% accuracy you mentioned is critical for a fraud detection system where false positives/negatives can have significant consequences.
- Adam combines the best aspects of other optimizers – it has adaptive learning rates like RMSProp but with better stability, and incorporates momentum-like features.
- For a relatively small dataset like UCI Banknote (1,372 instances with 4 features), Adam's computational overhead is negligible compared to its benefits.
- In real-time applications, faster convergence during training means quicker deployment of model updates.

While SGD might use less compute, the performance difference isn't worth the accuracy tradeoff for fraud detection. RMSProp's instability is problematic for a financial application where consistency is crucial. Momentum offers a middle ground, but without the adaptive learning rates that make Adam particularly effective for this type of classification task.

For your fintech application, Adam's balance of speed, accuracy, and stability makes it the optimal choice.

f) Technique 6: Interactive/Conversational (Interactive Prompting)

- i. Input: “In a banknote classification task (UCI Banknote Authentication Dataset, 4 input features), which optimizer (SGD, Adam, RMSProp) balances convergence speed and stability best? Should I use SGD for stability in banknote classification?.”
- ii. Output:

For the UCI Banknote Authentication Dataset, Adam typically balances convergence speed and stability better than SGD or RMSProp.

SGD can be more stable in some situations, but it's generally slower to converge and requires more careful learning rate tuning. While SGD might eventually reach similar performance, it usually takes more epochs to get there.

Adam combines the advantages of AdaGrad (which handles sparse gradients well) and RMSProp (which handles non-stationary objectives well). This makes it particularly effective for this binary classification task with only 4 features. Adam typically:

- Converges faster than SGD on this relatively small dataset

- Requires less learning rate tuning
- Maintains good stability for this type of classification problem

RMSProp would be my second choice - it performs better than SGD but often not as well as Adam for this specific task.

Since banknote authentication is a relatively simple binary classification problem with few features, the optimizer choice may not dramatically affect your final accuracy, but it will impact training speed and ease of tuning.

3. Question type 3: Computational problems & questions (Mathematical reasoning)

Original Question: Solve for x: $3x + 5 = 14$.

a) Technique 1: Basic Prompting (Zero-shot)

i. Input: "Solve for x in the equation: $\n3x + 5 = 14$."

ii. Output:

To solve for x in the equation $3x + 5 = 14$, I'll isolate x by performing operations on both sides.

Step 1: Subtract 5 from both sides of the equation. $3x + 5 - 5 = 14 - 5$ $3x = 9$

Step 2: Divide both sides by 3 to isolate x. $3x/3 = 9/3$ $x = 3$

Therefore, $x = 3$ is the solution to the equation $3x + 5 = 14$.

You can verify this by substituting $x = 3$ back into the original equation: $3(3) + 5 = 9 + 5 = 14$ ✓

b) Technique 2: Exemplar-based prompting (Few-shot)

i. Input: "Example responses: $\n1. Q: Solve for x: 2x + 4 = 10.\n A: 2x = 6 \rightarrow x = 3$ ✓ $\n2. Q: Solve for x: 5x - 3 = 12.\n A: 5x = 15 \rightarrow x = 3$ ✓ $\nNow solve: $3x + 5 = 14$."$

ii. Output:

To solve the equation $3x + 5 = 14$, I'll isolate the variable x:

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

Therefore, $x = 3$ ✓

c) Technique 3: Reasoning-oriented (Chain-of-Thought, CoT)

i. Input: "Think step by step before solving: $\n1. Start with the equation: $3x + 5 = 14$. $\n2. Subtract 5 from both sides: $3x = 9$. $\n3. Divide both sides by 3: $x = 3$. $\nWhat is the value of x?"$$$$

ii. Output:

The value of x is 3.

I'll verify this by working through the given steps:

1. Starting with $3x + 5 = 14$
2. Subtracting 5 from both sides: $3x + 5 - 5 = 14 - 5$, which gives $3x = 9$
3. Dividing both sides by 3: $3x/3 = 9/3$, which gives $x = 3$

This confirms that $x = 3$ is the solution to the equation $3x + 5 = 14$.

d) **Technique 4: Complex Reasoning Variants (Tree-of-Thoughts, ToT)**

- i. **Input:** “ $3x + 5 = 14$. Break the solution into multiple paths:\n- Path 1: Isolate x by first subtracting 5 from both sides.\n- Path 2: Solve for x by dividing both sides by 3.\n- Path 3: Verify the solution by plugging x back into the original equation.\n\nAfter analyzing these paths, determine the correct value of x .”

- ii. **Output:**

Let me solve the equation $3x + 5 = 14$ by analyzing each of the suggested paths.

Path 1: Isolate x by first subtracting 5 from both sides. $3x + 5 = 14$ $3x + 5 - 5 = 14 - 5$ $3x = 9$ $x = 9 \div 3$ $x = 3$

Path 2: Solve for x by dividing both sides by 3. $3x + 5 = 14$ $(3x + 5) \div 3 = 14 \div 3$ $x + 5/3 = 14/3$ $x = 14/3 - 5/3$ $x = 9/3$ $x = 3$

Path 3: Verify the solution by plugging x back into the original equation. $3x + 5 = 14$ $3(3) + 5 = 14$ $9 + 5 = 14$ $14 = 14$ ✓

After analyzing all three paths, the correct value of x is 3.

Note that Path 1 is the most direct approach for this equation, as it follows the standard algebraic procedure of isolating the variable term. Path 2 works but is less efficient because dividing both sides by 3 first requires an additional step to isolate x afterward.

e) **Technique 5: Role-based Prompting (Persona/Role Prompting)**

- i. **Input:** “You are a mathematics tutor helping students solve linear equations. Provide a step-by-step explanation to solve for x in the equation:\n\n $3x + 5 = 14$.”

- ii. **Output:**

I’ll help you solve this linear equation step by step.

To solve for x in $3x + 5 = 14$:

Step 1: Subtract 5 from both sides of the equation to isolate the term with x. $3x + 5 - 5 = 14 - 5$ $3x = 9$

Step 2: Divide both sides by 3 to isolate x. $3x \div 3 = 9 \div 3$ $x = 3$

Therefore, $x = 3$ is the solution to the equation $3x + 5 = 14$.

You can verify this answer by substituting $x = 3$ back into the original equation: $3(3) + 5 = 9 + 5 = 14$ ✓

f) Technique 6: Interactive/Conversational (Interactive Prompting)

i. Input: “Let's solve this together step by step.\n1. What happens if we subtract 5 from both sides?\n2. What is 14 - 5?\n3. Now, what do we get when we divide by 3?\n\nProvide an initial response, and we will refine it together.”

ii. Output:

I'd be happy to solve this problem step by step with you. Let me walk through the steps as you've outlined.

I notice that you've started with instructions to subtract 5 from both sides and then divide by 3, but I don't see the original equation. Without seeing the initial equation, I can still demonstrate the process using these steps:

1. If we subtract 5 from both sides of an equation, we maintain equality while moving terms.
2. $14 - 5 = 9$
3. When we divide 9 by 3, we get 3

If you're working with the equation $3x + 5 = 14$, then:

1. Subtracting 5 from both sides: $3x + 5 - 5 = 14 - 5$
2. This simplifies to: $3x = 9$
3. Dividing both sides by 3: $x = 3$

Is this the equation you were working with, or would you like to provide the original equation so we can solve it together?

4. Successful and Unsuccessful Outputs:

Successful Examples

1. Yes/No Questions (AAOS Guidelines on Glucosamine Sulfate)

- **Highlights:**
 - **Role-Based Prompting (Technique 5):** Outputs demonstrated clinical expertise by explicitly citing AAOS guidelines, contrasting them with European studies (e.g., EULAR), and emphasizing evidence-based reasoning.
 - **Interactive Prompting (Technique 6):** Structured analysis of conflicting evidence (e.g., mixed trial results) provided nuanced guidance for decision-making.
- **Strengths:**
 - All methods correctly answered "No" and emphasized the gap between anecdotal patient reports and guideline recommendations.

2. Multiple-Choice Questions (LDL Cholesterol Management)

- **Highlights:**
 - **Tree-of-Thoughts (Technique 4):** Multi-path reasoning (risk analysis, treatment evaluation, guideline alignment) offered a systematic framework for learners.
 - **Chain-of-Thought (Technique 3):** Stepwise elimination of incorrect options (e.g., "Surgery does not lower LDL") clarified logic for beginners.
- **Strengths:**
 - All outputs correctly identified "Statin therapy" and explained why alternatives (e.g., antibiotics) were irrelevant.

3. Computational Problems (Solving $3x + 5 = 14$)

- **Highlights:**
 - **Role-Based Prompting (Technique 5):** Tutor-style explanations with verification steps (e.g., substituting $x=3$) enhanced educational value.
 - **Tree-of-Thoughts (Technique 4):** Testing alternative solving paths (subtraction-first vs. division-first) reinforced answer validity.
- **Strengths:**
 - All methods produced accurate solutions ($x=3$) with error-checking steps.

Areas for Improvement

1. Yes/No Questions

- **Weaknesses:**
 - **Basic Prompting (Zero-shot)** and **Chain-of-Thought (CoT)** outputs lacked depth in addressing conflicting evidence (e.g., European studies supporting glucosamine).
- **Suggestions:**
 - Add prompts like, "Compare AAOS guidelines with other organizations' recommendations," to encourage balanced analysis.

2. Multiple-Choice Questions

- **Weaknesses:**
 - **Interactive Prompting (Technique 6)** did not clarify whether "initial treatment" includes lifestyle modifications before statins.
- **Suggestions:**
 - Refine prompts to specify whether non-pharmacological interventions are part of "initial treatment."

3. Computational Problems

- **Weaknesses:**
 - **Exemplar-Based Prompting (Few-shot)** provided minimal working steps without pedagogical explanations.
- **Suggestions:**
 - Include examples of common errors (e.g., forgetting to subtract 5 first) to improve troubleshooting guidance.

Key Takeaways

- **Effective Techniques:**
 - **Role-Based** and **Tree-of-Thoughts** methods excelled in delivering professional, logically structured responses.
 - **Interactive Prompting** was ideal for complex scenarios requiring evidence evaluation or collaborative problem-solving.
- **Improvement Strategies:**

- Incorporate comparative analysis prompts (e.g., conflicting guidelines or studies).
- Clarify ambiguous terms (e.g., "initial treatment") in prompts.
- Use richer examples (including common mistakes) to enhance depth and educational value.