

2026-27 CAIO National Qualifier Practical Competition

Part 1 — AI Fundamentals Challenge (Paper C)

Total: 100 marks

Instructions

- Answer all 35 questions.
- Questions 1-15 are single-choice questions (2 marks each). Choose the one best answer.
- Questions 16-25 are multiple-select questions (2 marks each). Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected; otherwise you receive zero for that question.
- Questions 26-35 are fill-in-the-blank questions (5 marks each). Each has exactly five blanks; write your answer for each blank clearly.
- A non-programmable calculator may be used. Total: 100 marks.

Do not open or begin until instructed by the proctor.

Question 1

Data Preparation

2 Points

Consider the following Python code:

```
import numpy as np
A = np.array([[1, 2, 3], [4, 5, 6]])
B = A[0, :]
C = A[0, :].copy()
B += 10
C += 100
print(A[0, 0], A[0, 1], A[1, 0], A.sum())
```

What is the printed output?

- A. 11 12 4 51
- B. 11 12 104 351
- C. 1 2 4 21
- D. 11 112 4 151
- E. 101 102 4 321

Question 2

Data Preparation

2 Points

Consider the following Python code:

```
import pandas as pd
left = pd.DataFrame({'id': [1, 2, 3], 'v': [10, 20, 30]})
right = pd.DataFrame({'id': [2, 3, 3], 'w': [5, 6, 7]})
m = left.merge(right, on='id', how='left')
print(len(m), m['w'].sum())
```

What is the printed output?

- A. 4 18.0
- B. 3 18.0
- C. 4 13.0
- D. 3 12.0
- E. 4 18

Question 3

Supervised Learning

2 Points

For binary logistic regression with $\sigma(z) = 1/(1+e^{(-z)})$, $z = w \cdot x + b$, the per-example loss is

$$L = -[y \cdot \log \sigma(z) + (1-y) \cdot \log(1-\sigma(z))]$$

The gradient of L with respect to w is:

- A. $(\sigma(z)-y) \cdot x$
- B. $(y-\sigma(z)) \cdot x$
- C. $\sigma(z)(1-\sigma(z)) \cdot x$
- D. $(\sigma(z)-y)$
- E. $-y(1-\sigma(z)) \cdot x$

Question 4

Supervised Learning

2 Points

Training points (features $\rightarrow$ class): (0,0) $\rightarrow$ 0, (4,0) $\rightarrow$ 0, (0,3) $\rightarrow$ 1, (2,2) $\rightarrow$ 1. Euclidean distance.

A 3-NN classifier predicts which class for the test point (1,1)?

- A. Class 1
- B. Class 0
- C. Tie — cannot decide
- D. Class 1, but 1-NN would predict class 0
- E. It depends on the distance metric used

Question 5

Kernel Methods

2 Points

A hard-margin SVM learns the weight vector $w = [3, 4]$ (bias b chosen to maximize the margin).

$$\text{margin width} = 2/\|w\|$$

What is the width of the margin (distance between the two supporting hyperplanes)?

- A. 0.4
- B. 0.2
- C. 2.5
- D. 5.0
- E. 0.8

Question 6

Unsupervised Learning

2 Points

DBSCAN with $\epsilon = 1.5$ and $\text{minPts} = 3$ is run on the points (0,0), (0,1), (1,0), (10,10). Euclidean distance.

How many core points and how many noise points are there?

- A. 3 core, 1 noise
- B. 4 core, 0 noise
- C. 2 core, 2 noise
- D. 3 core, 0 noise (the far point is a border point)
- E. 0 core, 4 noise

Question 7

Evaluation of ML Models

2 Points

A test set has 1000 samples: 950 negative, 50 positive. A classifier predicts "negative" for every sample.

What are its accuracy and its recall on the positive class?

- A. Accuracy 0.95, recall 0.0
- B. Accuracy 0.95, recall 0.95
- C. Accuracy 0.50, recall 0.0
- D. Accuracy 0.95, recall undefined
- E. Accuracy 0.05, recall 0.0

Question 8

AI Search

2 Points

Edges: $S \rightarrow A$ cost 2, $S \rightarrow B$ cost 1, $A \rightarrow G$ cost 4, $B \rightarrow G$ cost 6. True cheapest costs to the goal: $h^*(S)=6$, $h^*(A)=4$, $h^*(B)=6$.

Which heuristic is admissible but NOT consistent?

- A. $h(S)=6$, $h(A)=4$, $h(B)=5$
- B. $h(S)=6$, $h(A)=3$, $h(B)=6$
- C. $h(S)=7$, $h(A)=4$, $h(B)=6$
- D. $h(S)=0$, $h(A)=0$, $h(B)=0$
- E. $h(S)=6$, $h(A)=4$, $h(B)=6$

Question 9

AI Search

2 Points

All step costs are positive but NOT all equal. A student claims breadth-first search will return the cheapest path because "it explores level by level".

Which statement is TRUE?

- A. The student is right: BFS expands nodes in order of increasing path cost.
- B. BFS is optimal here, but only if the goal sits at the shallowest depth.
- C. The student is wrong: BFS levels count steps, not cost — uniform-cost search should be used for cheapest-path optimality.
- D. The student is wrong: BFS is not even complete when step costs differ.
- E. Iterative deepening would fix the problem because it re-expands nodes.

Question 10

AI Search

2 Points

Tree-search A* (no repeated-state checking). Edges: $S \rightarrow A$ cost 1, $S \rightarrow B$ cost 2, $A \rightarrow G$ cost 1, $B \rightarrow G$ cost 1. Heuristic: $h(S)=0$, $h(A)=10$, $h(B)=0$, $h(G)=0$.

Which path does A* return, and is it optimal?

- A. $S \rightarrow B \rightarrow G$, cost 3 — not optimal (the optimum is $S \rightarrow A \rightarrow G$, cost 2)
- B. $S \rightarrow A \rightarrow G$, cost 2 — optimal
- C. $S \rightarrow B \rightarrow G$, cost 3 — optimal
- D. $S \rightarrow A \rightarrow G$, cost 11 — not optimal
- E. $S \rightarrow B \rightarrow G$, cost 4 — not optimal

Question 11

AI Search

2 Points

State space: branching factor b , solution depth d , maximum depth m ($m \geq d$).

Which algorithm has the LOWEST worst-case memory complexity?

- A. Breadth-first search: $O(b^d)$
- B. Depth-first search: $O(b \cdot m)$
- C. Iterative deepening DFS: $O(b \cdot d)$
- D. A*: $O(b^d)$
- E. Uniform-cost search: $O(b^d)$

Question 12

Logical Reasoning

2 Points

Knowledge base $KB = \{ P \rightarrow Q, Q \rightarrow R, \neg R \}$.

Which formula is logically entailed by KB?

- A. P
- B. Q
- C. $\neg Q$
- D. $P \wedge \neg R$
- E. $R \vee P$

Question 13

Logical Reasoning

2 Points $KB = \{ A \vee B, \neg A \vee C, \neg B \vee C, \neg C \}.$ **How many models (satisfying truth assignments) does KB have?**

- A. 0
- B. 1
- C. 2
- D. 3
- E. 4

Question 14

Constraint Satisfaction Problems

2 Points

CSP: variables X (domain {1,2}), Y (domain {1,2,3}), Z (domain {1}). Constraints: $X \neq Z$, $Y > X$.

Using the minimum-remaining-values (MRV) heuristic, which variable is selected first?

- A. Z
- B. X
- C. Y
- D. X or Y (tie)
- E. Any — MRV does not apply before any assignment

Question 15

Reinforcement Learning

2 Points

Q-learning update: $Q(s,a) \leftarrow Q(s,a) + \alpha \cdot [r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)]$. Current $Q(s,a) = 4$, $\alpha = 0.5$, reward $r = 10$, $\gamma = 0.9$, $\max_{a'} Q(s',a') = 8$.

What is the updated $Q(s,a)$?

- A. 10.6
- B. 17.2
- C. 11.2
- D. 13.6
- E. 8.6

Question 16

AI Search

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about search algorithms are TRUE?

- A. A* using graph search with a consistent heuristic never needs to re-expand a node.
- B. Greedy best-first search guarantees an optimal solution.
- C. Uniform-cost search is equivalent to A* with $h(n) = 0$ for all n .
- D. Iterative deepening A* (IDA*) uses asymptotically less memory than A*.
- E. Depth-first search is optimal when all step costs are equal.

Question 17

AI Search

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about uninformed search are TRUE?

- A. Iterative deepening search is complete and optimal when all step costs are equal.
- B. Breadth-first search is optimal when all step costs are equal.
- C. Depth-first search is complete in infinite state spaces.
- D. Uniform-cost search expands nodes in nondecreasing order of path cost $g(n)$.
- E. Breadth-first search uses less memory than depth-first search.

Question 18

Logical Reasoning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about logical entailment are TRUE?

- A. If $KB \models \phi$ then $KB \cup \{\psi\} \models \phi$ for any sentence ψ .
- B. If ϕ is consistent with KB then $KB \models \phi$.
- C. KB is unsatisfiable if and only if $KB \models \perp$ (false).
- D. If $KB \models \phi \vee \psi$ then $KB \models \phi$ or $KB \models \psi$.
- E. Propositional resolution is refutation-complete: $KB \models \phi$ iff $KB \cup \{\neg\phi\}$ derives a contradiction by resolution.

Question 19

Constraint Satisfaction Problems

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about constraint propagation are TRUE?

- A. Forward checking prunes values from the domains of ALL unassigned variables after each assignment.
- B. Maintaining arc consistency (e.g. AC-3) can detect failures that forward checking misses.
- C. The MRV heuristic selects the variable involved in the greatest number of

constraints.

D. Chronological backtracking is complete: it finds a solution if one exists.

E. With forward checking, an unassigned variable's domain can become empty before that variable itself is assigned.

Question 20

Constraint Satisfaction Problems

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about CSPs are TRUE?

- A. A binary CSP with n variables of domain size d has at most d^n complete assignments.
- B. Enforcing arc consistency at every node of the search tree is always faster than plain backtracking.
- C. If every arc in the constraint graph is consistent, the CSP is guaranteed to have a solution.
- D. The least-constraining-value heuristic orders values by how few options they remove for neighboring variables.
- E. Forward checking enforces constraints only between the most recently assigned variable and the unassigned variables.

Question 21

Supervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about supervised learning are TRUE?

- A. 1-NN has zero training error when no two training points share identical features but different labels.
- B. Lasso (L1 regularization) can drive some coefficients to exactly zero, performing feature selection.
- C. In bagging, each base learner is trained on a bootstrap sample (drawn with replacement) the same size as the training set.
- D. Adding more trees to a random forest always reduces its bias.
- E. One-hot encoding is required before a decision tree can use categorical features.

Question 22

Unsupervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about clustering are TRUE?

- A. In DBSCAN, a border point may be assigned to more than one cluster.
- B. Single-linkage agglomerative clustering can discover elongated, chain-like clusters.
- C. The elbow method always pinpoints a single unambiguous optimal k for k -means.
- D. A Gaussian mixture model with diagonal covariances assumes features are uncorrelated within each component.
- E. In agglomerative clustering, the height at which two clusters merge in the dendrogram equals the distance between them.

Question 23

Evaluation of ML Models

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about evaluation are TRUE?

- A. A classifier with 99% accuracy on a dataset that is 99% negative can still have zero recall on the positive class.
- B. Raising the decision threshold always increases the F1 score.
- C. An AUC-ROC of 0.5 indicates performance no better than random guessing.
- D. A classifier that never predicts the positive class has precision 1.0 on that class.
- E. Macro-averaged recall gives equal weight to each class, regardless of class size.

Question 24

Kernel Methods

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about kernels and SVMs are TRUE?

- A. The polynomial kernel $(x \cdot z + c)^1$ is equivalent to a linear kernel (up to the constant c).
- B. A smaller γ in the RBF kernel gives each training point a wider sphere of influence.
- C. The Gram (kernel) matrix of a valid kernel is always positive semidefinite.
- D. The RBF kernel can be expanded as a finite-degree polynomial in the input features.
- E. Increasing the SVM regularization parameter C always increases the number of support vectors.

Question 25

Recommender Systems

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about recommender systems are TRUE?

- A. Content-based filtering recommends items similar to ones the user has liked before.
- B. User-based collaborative filtering predicts ratings from users with similar taste profiles.
- C. Matrix factorization learns latent vector representations for both users and items.
- D. Collaborative filtering needs item content features (genre, text) to compute recommendations.
- E. A brand-new user with no ratings can receive personalized collaborative-filtering recommendations.

Question 26

Data Preparation

5 Points

Sensor Log Aggregation (fill in the 5 blanks)

readings.csv columns: ts (string 'YYYY-MM-DD HH:MM'), sensor, value.

Goal: daily mean value per sensor, forward-filling days with no readings.

Complete the script:

```
import pandas as pd
df = pd.read_csv('readings.csv')
df['ts'] = pd._____(df['ts'])
df = df.drop_duplicates(_____)
df = df._____('ts')
daily = df.set_index('ts').groupby('sensor')['value']._____('D').mean()
daily = daily.______() # forward-fill days with no readings
```

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 27

Supervised Learning

5 Points

Gini Split Selection (fill in the 5 blanks; 2 decimals)

8 samples. Feature X and class y:

X: 1 2 3 | 4 5 | 6 7 8

y: 0 0 0 | 1 1 | 1 1 1

Consider the candidate splits at 3.5 and 5.5. Round to 2 decimals.

Gini(parent) = _____
 Split at 3.5 → Gini(left) = _____ (right side is pure)
 Split at 5.5 → weighted Gini = _____
 Gini gain at 5.5 = _____
 Better split threshold = _____

Write your answer for each blank below:

- (a) _____
 (b) _____
 (c) _____
 (d) _____
 (e) _____

Question 28

Supervised Learning

5 Points

Naive Bayes by Hand (fill in the 5 blanks; 2 decimals)

6 training emails, Laplace smoothing $\alpha = 1$.

Spam: "buy cheap", "buy now", "cheap deal"

Ham: "see now", "see deal", "lunch now"

Vocabulary: {buy, cheap, now, deal, see, lunch}. Test email: "buy now".

Round to 2 decimals.

P(buy | spam) = _____
 P(now | spam) ≈ _____
 P(buy | ham) ≈ _____
 P(spam | "buy now") ≈ _____ (normalized posterior)
 Predicted class = _____

Write your answer for each blank below:

- (a) _____
 (b) _____
 (c) _____
 (d) _____
 (e) _____

Question 29

Supervised Learning

5 Points

One Perceptron Update (fill in the 5 blanks)

Perceptron: predict 1 if $w \cdot x + b > 0$, else 0.

Start: $w = [0.5, -1.0]$, $b = 0.2$. Training example: $x = (1, 2)$, $y = 1$.

Update rule on a mistake: $w \leftarrow w + y \cdot x$, $b \leftarrow b + y$.

$z = w \cdot x + b =$ _____
 predicted class = _____
 updated $w =$ _____ (e.g. $[1.5, 1.0]$)
 updated $b =$ _____
 new z on the same example = _____

Write your answer for each blank below:

- (a) _____
 (b) _____
 (c) _____
 (d) _____
 (e) _____

Question 30

AI Search

5 Points

Uniform-Cost Search Trace (fill in the 5 blanks)

Edges: $S \rightarrow A$ cost 2, $S \rightarrow B$ cost 5, $A \rightarrow B$ cost 1, $A \rightarrow G$ cost 6, $B \rightarrow G$ cost 2.

Uniform-cost search (graph search), ties broken deterministically.

After expanding S , the frontier holds $A(2)$ and $B(5)$. Expanded next: _____
 B's path cost is then improved from 5 to _____
 The node expanded after that is _____
 Returned optimal path cost = _____
 Returned optimal path = _____

Write your answer for each blank below:

- (a) _____
 (b) _____
 (c) _____
 (d) _____
 (e) _____

Question 31

AI Search

5 Points

BFS vs DFS Expansion Order (fill in the 5 blanks)

Edges: $S \rightarrow A$, $S \rightarrow B$, $A \rightarrow D$, $B \rightarrow C$, $C \rightarrow G$, $D \rightarrow G$. Goal: G.

BFS: queue; neighbors enqueued alphabetically.

DFS: stack; neighbors pushed in reverse-alphabetical order (so the alphabetically-first neighbor is popped first).

BFS node expansion order (until G is popped): _____

DFS node expansion order (until G is popped): _____

Nodes expanded by BFS before popping G: _____

Nodes expanded by DFS before popping G: _____

Which algorithm popped the goal first: _____

Write your answer for each blank below:

(a) _____

(b) _____

(c) _____

(d) _____

(e) _____

Question 32

Logical Reasoning

5 Points

Model Counting (fill in the 5 blanks)

$KB = \{ A \rightarrow B, B \rightarrow C \}$ over variables $\{A, B, C\}$.

Total truth assignments: _____

Models of KB: _____

Is $(A \rightarrow C)$ entailed by KB? (yes/no): _____

Models in which $\neg A$ holds: _____

Models in which $(B \wedge \neg C)$ holds: _____

Write your answer for each blank below:

(a) _____

(b) _____

(c) _____

(d) _____

(e) _____

Question 33

Constraint Satisfaction Problems

5 Points

Backtracking with Forward Checking (fill in the 5 blanks)

Variables X, Y, Z, each domain {R, G, B}. Constraints: $X \neq Y$, $Y \neq Z$.

Backtracking with MRV (ties $\rightarrow$ alphabetical) and forward checking.

Assign X = R. Forward checking reduces Y's domain to _____ (e.g. G,B)

MRV now selects variable _____

Assign Y = G. Forward checking reduces Z's domain to _____

Complete solutions with X = R: _____

Total complete solutions: _____

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 34

Evaluation of ML Models

5 Points

Cost-Sensitive Comparison (fill in the 5 blanks)

Test set: 1000 samples. Costs: false positive = 1, false negative = 10.

Classifier A: TP=80, FP=20, FN=30, TN=870.

Classifier B: TP=60, FP=5, FN=50, TN=885.

Total cost of A = _____

Total cost of B = _____

Accuracy of A = _____

Accuracy of B = _____

Better classifier under this cost model: _____

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 35

Kernel Methods

5 Points

Polynomial Kernel by Hand (fill in the 5 blanks)

$k(x, z) = (x \cdot z + 1)^2$. Explicit map for 2-D inputs:

$\phi(x) = [1, \sqrt{2} \cdot x_1, \sqrt{2} \cdot x_2, x_1^2, \sqrt{2} \cdot x_1 x_2, x_2^2]$.

Take $x = (1, 2)$, $z = (2, 0)$. Hint: $\sqrt{2} \approx 1.414$.

$x \cdot z = \underline{\hspace{2cm}}$

$k(x, z) = \underline{\hspace{2cm}}$

Dimension of the explicit feature space = $\underline{\hspace{2cm}}$

The $\sqrt{2} \cdot x_1 x_2$ component of $\phi(x) \approx \underline{\hspace{2cm}}$

$\|\phi(x)\|^2 = k(x, x) = \underline{\hspace{2cm}}$

Write your answer for each blank below:

(a) $\underline{\hspace{2cm}}$

(b) $\underline{\hspace{2cm}}$

(c) $\underline{\hspace{2cm}}$

(d) $\underline{\hspace{2cm}}$

(e) $\underline{\hspace{2cm}}$