## Lower Primary Grades 1 to 3

A gentle, playful, mostly-unplugged start: recognise AI, follow rules and patterns by hand, and meet generative AI safely.

### Who it's for

Students, parents and teachers who want a simple picture of what Lower Primary learners should understand, practise, and demonstrate in IJAIO.

### What success looks like

Learners recognise familiar AI examples, follow simple rule-based processes, talk about safety and privacy, and complete visual or unplugged tasks with clear explanations.

### The five AI Pillars

Each Pillar leads with a disposition - the attitude a learner brings, then plain learning outcomes that combine what a learner knows and can do. The same five Pillars return in every band, a little deeper each time.

**Pillar 1 Representation** Disposition. Looks for the rule or pattern behind a behaviour rather than calling it magic.

1.1Explain that AI is a computer program made by people, not alive, with no real feelings.
1.2Describe how computers store words and pictures as numbers, not letters.
1.3Encode and decode a short word using a letter-to-number code sheet.
1.4Sort objects or pictures into groups by a shared feature (early classification).

**Pillar 2 Learning** Disposition. Believes behaviour is learned from examples, not magic; keeps trying and checking.

2.1Explain that AI learns from many examples: many cat pictures help it learn "cat".
2.2Continue a pattern and predict what comes next (the seed of prediction).
2.3Carry out a simple rule by hand, like a robot following instructions.
2.4Find a way through a simple maze by trying and checking (the seed of search).

**Pillar 3 Generation** Disposition. Curious about how new things are made; tells "making" apart from "finding".

3.1Tell apart making something new (generative AI) from finding what already exists (search).
3.2Play "what comes next?" guessing games to experience prediction at work.

**Pillar 4 Interaction & Agency** Disposition. Asks clearly and kindly; designs helpers with care and safety.

4.1Ask an AI helper a clear, kind question (the seed of prompting).
4.2Describe how a helper machine works as a loop: Sense - Think - Act.
4.3Draw and label a simple helper robot: what it senses (its sensors) and the steps it follows.

**Pillar 5 Responsibility** Disposition. Cares for safety and others; checks before trusting and is honest about AI help.

5.1Identify private information - home address, phone number, passwords - and keep it private.
5.2Check important AI answers with a trusted adult, because AI can be wrong.
5.3Choose the safer or kinder action in a simple scenario and explain why.
5.4Say honestly when an AI helper was used, for homework, drawings or answers.

## Upper Primary Grades 4–6

Rules versus learning from data, the first generative-AI mechanics - next-word prediction and prompting - and checking AI for fairness.

### Who it's for

Students, teachers and parents who want to see how Upper Primary learners move from recognising AI ideas to shaping simple solutions and responsible-use habits.

### What success looks like

Learners compare rules with learning from data, improve prompts, spot simple bias or hallucination risks, and present short design ideas with justification.

### The five AI Pillars

Each Pillar leads with a disposition, then plain learning outcomes that combine what a learner knows and can do. The same five Pillars return in every band, a little deeper each time.

**Pillar 1** **Representation** Disposition. Trusts that meaning can be measured; looks for the numbers behind words and pictures.

1.1Turn words into numbers and show that similar things get similar numbers (proto-embeddings).
1.2Group items by their features and explain what makes a group alike.
1.3Collect and label a small set of examples (a mini dataset).
1.4Check whether the sample is fair or leaves some things out.
1.5Hand-tokenise a short sentence and notice that one word is not always one token.
Evidence. Build and label a small picture set, group it by feature, and say whether the set is fair.
Key termsfeaturelabelvector (number-list)

**Pillar 2** **Learning** Disposition. Believes behaviour is learned from data - and that data can be fair or unfair.

2.1Distinguish following rules from learning from data, the idea of machine learning.
2.2Train a simple classifier by giving and refining labelled examples.
2.3Tell fair from unfair training examples.
2.4Group items with no labels, the idea of clustering (unsupervised learning).
2.5Learn from rewards by trial and error, the idea of reinforcement learning.
2.6Describe a single neuron as a tiny decision-maker: inputs ▯ weighted sum ▯ fires (or not).
Evidence. Train a simple classifier from your examples, improve it, and explain one unfair case you fixed.
Key termstraining dataclassifygeneralisation: working on NEW examples

**Pillar 3** **Generation** Disposition. Curious about how machines make new things; experiments and improves.

3.1Explain how a chatbot predicts a likely next word.
3.2Explain how an image generator builds a picture from a description.
3.3Write a clear prompt (role + instruction + example) and improve it after seeing the result.
3.4Create something new with a generative tool and refine it through better prompts.
3.5Tell apart generating something new from retrieving (searching for) something that already exists.
Evidence. Improve a prompt across two tries and explain why the second output is better.
Key termsnext-word predictionprompttemperature (the randomness dial)diffusion (painting from noise)

**Pillar 4** **Interaction & Agency** Disposition. Directs the tool with care; designs solutions with people in mind.

4.1Trace how a helper senses the world (its sensors), decides, then acts: perceive ▯ decide ▯ act.
4.2Sketch a solution's architecture (inputs ▯ model ▯ outputs ▯ user).
4.3Name one likely failure case and suggest a fix.
4.4Design a small AI helper for a real classroom or home problem.
Evidence. Design an AI helper, label what it senses and does, and name a failure case and its fix.
Key termsAI agentsense ▯ think ▯ acttool

**Pillar 5** **Responsibility** Disposition. Honest about AI help; fair-minded and caring about effects on people and the planet.

5.1Recognise that generative AI can "hallucinate": sound confident yet be wrong.
5.2Cross-check an AI claim against a reliable source before sharing it.
5.3Acknowledge AI help honestly rather than passing it off as one's own.
5.4Explain that AI uses energy and water to run, so using it has a real-world cost.
Evidence. Spot a likely hallucination, verify it against a source, and name one real-world cost of AI.

## Lower Secondary Grades 7–8

Hands-on mechanics by hand and in simple Python: n-gram language models, tokenisation, a single neuron, the ML pipeline, and AI agents.

### who it's for

Students, teachers and coaches preparing for the first band with explicit model mechanics, coding-related reasoning, and practical AI problem framing.

### what success looks like

Learners explain how simple language models and ML pipelines work, analyse bias and failure points, and reason with short Python-based tasks.

### The five AI Pillars

Each Pillar leads with a disposition, then plain learning outcomes that combine what a learner knows and can do. The same five Pillars return in every band, a little deeper each time.

**Pillar1** **Representation** Disposition. Reasons from how a representation is built rather than treating it as given.

1.1Describe tokenisation: how text is split into tokens, then turned into numbers.
1.2Hand-tokenise a sentence into tokens and IDs.
1.3Explain why a model can miscount the letters in a word.
1.4Inspect an embedding space: similar words sit close together, and analogies are directions.
1.5Represent an image as a grid of number-pixels, the start of how a computer "sees".

**Pillar2** **Learning** Disposition. Treats learning as a workflow to reason about, not a black box.

2.1Explain how a neuron computes a weighted sum and "fires" via an activation, and how a network stacks many.
2.2Tell apart the three ways machines learn: supervised, unsupervised (clustering, e.g. k-means) and reinforcement.
2.3Follow the machine-learning workflow (data → train → test → evaluate) and read a confusion matrix.
2.4Code a single neuron and a simple classifier (k-nearest neighbours or a rule) in Python, with a train/test split.
2.5Explain why dataset quality and balance change what a model learns.

**Pillar3** **Generation** Disposition. Wants to know why generative output looks the way it does.

3.1Explain how a language model (LLM) predicts the next token from probabilities learned over data (n-gram → neural).
3.2Build a bigram/trigram next-token predictor by hand and use it to generate text (be-the-LLM).
3.3Trace how attention re-weights earlier words and changes which token comes next.
3.4Explain how diffusion makes images by reversing noise, steered by a prompt.

**Pillar4** **Interaction & Agency** Disposition. Designs and traces systems; reasons about where they fail.

4.1Use a prompt framework (e.g. CLEAR) with few-shot examples.
4.2Compare the outputs of two prompts and say which works better and why.
4.3Trace how an AI agent runs a Sense → Think → Act loop and calls tools to reach a goal.
4.4Identify where a tool call is needed in an agent's loop, noting that sensors feed the perceive step.

**Pillar5** **Responsibility** Disposition. Verifies before trusting; designs safeguards with people in mind.

5.1Explain why a model can sound confident yet be wrong: its objective is plausibility, not truth.
5.2Explain how deepfakes and synthetic media enable misinformation.
5.3Apply a verification routine to a generative output and flag its risks.
5.4Propose safeguards such as human oversight, explainability and inclusive design.
5.5Trace how bias in a dataset becomes bias in a model's behaviour.

## Upper Secondary Grades 9–12

Implementation and evaluation: the transformer pipeline, sampling and temperature, embeddings and bias, diffusion, and grounded agentic / RAG design - bridging to NOAI and IOAI / IAIO.

### who it's for

Students, teachers and coaches preparing for the most advanced IJAIO band - including learners in Upper Secondary and equivalent applied pathways.

### what success looks like

Learners explain end-to-end generative-AI systems, evaluate models with explicit metrics and trade-offs, and present defensible design choices for agentic or retrieval-grounded systems.

### The five AI Pillars

Each Pillar leads with a disposition, then plain learning outcomes that combine what a learner knows and can do.

**Pillar 1** **Representation** Disposition. Reasons rigorously about representation; surfaces what is hidden inside it.

1.1Compute embedding similarity with cosine similarity and demonstrate vector arithmetic.
1.2Demonstrate the bias baked into embeddings.
1.3Judge the impact of embedded bias on a downstream use.
1.4Explain how computer vision and convolutional neural networks (CNNs) read images as pixel grids.
1.5Explain sub-word tokenisation and how token IDs index the embedding layer.

**Pillar 2** **Learning** Disposition. Practises rigorous, reproducible work; is honest about a model's limits.

2.1Explain neural-network mechanics: forward pass, activation, loss, and gradient-descent / backpropagation intuition.
2.2Distinguish deep learning's layers, pretraining vs fine-tuning vs RLHF, and training vs inference.
2.3Implement a forward pass and a small classifier in Python.
2.4Evaluate the model on accuracy, precision/recall and overfitting, then iterate.
2.5Curate a small dataset for a task, justifying what is included, and document it.

**Pillar 3** **Generation** Disposition. Reasons from the pipeline; weighs usefulness against reliability.

3.1Trace the transformer pipeline end to end: tokens - embeddings - attention - next-token probabilities ⊠ sampling / temperature.
3.2Explain how diffusion generates images by reversing noise, conditioned on a prompt.
3.3Compare outputs at different temperatures and interpret the usefulness-versus-reliability trade-off.
3.4Compare diffusion with GANs as two ways to generate images: iterative denoising versus generator-vs-discriminator.

**Pillar 4** **Interaction & Agency** Disposition. Innovates responsibly; keeps a human in the loop by design.

4.1Construct chain-of-thought reasoning, system prompts and few-shot prompts.
4.2Explain how an agentic system plans, calls tools, observes results and re-plans.
4.3Explain how retrieval-augmented generation (RAG) grounds answers in sources.
4.4Design a multi-step agent or RAG system against a defined metric.
4.5Evaluate the system and surface its limitations.

**Pillar 5** **Responsibility** Disposition. Exercises civic responsibility; defaults to transparency and accountability.

5.1Interpret fairness metrics and error types (false positives / false negatives).
5.2Explain AI governance, accountability and explainable AI.
5.3Explain why hallucination follows from the next-token objective, and the limits of RLHF and retrieval (RAG).
5.4Audit a model or dataset for bias and document it in a model card.
5.5Weigh the societal and environmental costs of a deployment.

### Learning Resources

Given the dynamic changing landscape of AI technologies advancements, a compiled list of online and offline resources will be regularly updated at https://computing.sg/ijaio.

# Format 

No marks are deducted for incorrect answers.

**Lower Primary / Grades 1 to 3 (45 minutes)**
20 MCQ questions [60]
2 structured questions [40]
**Total 100 marks**

**Upper Primary / Grades 4 to 6 (60 minutes)**
30 MCQ questions [60]
2 structured questions [40]
**Total 100 marks**

**Lower Secondary / Grades 7 to 8 (90 minutes)**
20 MCQ questions [40]
3* structured theory problems / practical programming tasks [60]
**Total 100 marks**

**Upper Secondary / ITE / Junior College / Polytechnic / Grades 9 to 12 (90 minutes)**
20 MCQ questions [40]
3* structured theory problems / design challenge / practical programming tasks [60]
**Total 100 marks**

*1 problem / task will be based on an emerging AI technology theme which will be announced two months before IJAIO. This is to keep pace with AI technologies' rapid developments and more significantly, for students and teachers to inculcate the growth mindset of lifelong learning and the critical life skill to effectively and efficiently learn, unlearn and relearn. For example, IJAIO 2026 will be conducted from mid to end January 2026. In mid November 2025, the emerging AI technology theme (eg AI agents) with associated high level overview and scope will be made available at https://computing.sg/ijaio.

**Software (Secondary school division / Secondary 1 to Junior College 2 / Grades 7 to 12 only)**
For practical programming tasks, the language and platform used will be the current stable version of Python 3 and Jupyter Notebook / Google Colab

**Hardware (Secondary school division / Secondary 1 to Junior College 2 / Grades 7 to 12 only)**
A Windows or MacOS laptop is recommended. Candidates using Chromebooks, Android and iOS devices may encounter compatibility issues with editing, testing and executing local Python programs/scripts for practical programming tasks.

Lower Primary / Grades 1 to 3

## Section A · Multiple Choice

Choose the one best answer. There is no penalty for a wrong try.

**Q1** First Principles — 3 marks

Using A=1, B=2, C=3, D=4 ..., write the word "BAD" as numbers.

BAD → ? ? ?

A 214

B 123

C 412

D 241

**Q2** First Principles — 3 marks

A robot reaches a dead-end while trying to get through a maze. The smart thing to do is...

A try the exact same path again

B go back and try a different path

C smash through the wall

D switch itself off

# Section B · Structured Tasks

Write, draw or describe your answer, then reveal the suggested marking points.

### Task 1 · Be the Sorting Robot

[illegible] marks

First Principles

A robot sorts fruit using **only two rules**:
· **Rule 1:** IF the fruit is **yellow**, THEN put it in **Basket A**.
· **Rule 2:** IF the fruit is **red**, THEN put it in **Basket B**.

The fruits are: 🍌 banana (yellow), 🍎 apple (red), 🍋 lemon (yellow), 🍓 strawberry (red), 🍏 green apple (green).

a) a) For each fruit, write which basket it goes in. (Be the robot - follow the rules exactly.)

b) b) The green apple does not match any rule. What should the robot do? Explain your idea.

c) c) Why does a robot need clear rules to do its job well?

Write, draw or describe your answer here...

Show model answer & marks

### Task 2 · Design a Helper Robot (AI Agent)

[illegible] marks

AI Agent AI Ethics

An **AI agent** is a helper that can do three things by itself to reach a goal:
**SENSE** (look or listen) → **THINK** (decide what to do) → **ACT** (do something).

Imagine a friendly helper robot for your classroom whose goal is to **keep the classroom tidy**.

a) a) Draw or describe your helper robot. What can it sense (see/hear)?

b) b) Write its three steps for tidying up: SENSE → THINK → ACT.

c) c) **Think carefully:** Your robot sees a drawing lying on the floor. It is told to throw away rubbish, but this might be someone's special artwork. What should the robot do, and why?

Write, draw or describe your answer here...

Show model answer & marks

## Upper Primary / Grades 4 to 6

# Section A · Multiple Choice

Choose the one best answer. There is no penalty for a wrong try.

**Q1** First Principles — 2 marks

A child sorts [shark, goldfish, eagle, sparrow] into two groups. The most useful FEATURE to group them by is...

- A whether its name is long or short
- B whether it mainly swims or flies
- C which animal you like best
- D the order they were listed in

**Q2** Generative AI — 2 marks

Your first AI image of "a cat" looks too plain. The BEST next step to improve it is to...

- A press generate again with the very same words
- B add detail to your prompt, e.g. "a fluffy ginger cat in a hat, cartoon style"
- C type the word CAT in capitals
- D give it no words at all

Q3 First Principles 2 marks

A computer turns words into number-lists so similar meanings get similar numbers. Which PAIR should be CLOSEST?

- A cat and car
- B cat and Tuesday
- C cat and kitten
- D cat and bicycle

Q4 AI Agent First Principles 2 marks

You design an AI "lost-property helper" that photographs an item and names its owner. A likely FAILURE to plan for - and a fix - is...

- A it works slowly at busy times; fix: skip the owner check
- B it shows the photo too small; fix: ignore any mistakes
- C it needs many photos to learn; fix: train it on none
- D it confuses two look-alike bags; fix: have a person confirm first

# Section B · Structured Tasks

Write, draw or describe your answer, then reveal the marking guide.

## Task 1 · Be the Next-Word Predictor

20 marks

First Principles Generative AI

A tiny language model has read many sentences. It counted which words come right after the word "play":

**after "play": football → 5, piano → 3, games → 2, outside → 2**

a) The model has just written "I want to play …". Which single word is it MOST likely to pick next? Why?

b) There were 12 counts in total. Write the chance (as "… out of 12") that it picks "piano".

c) Two children give the exact same prompt and get *different* next words. Explain how, using the idea that the model sometimes picks from its likely words instead of always the top one.

Write, draw or describe your answer here...

Show model answer & marks

## Task 2 · Prompt Detective & Fact-Checker

20 marks

Generative AI Ethics & Safety System Design

Maya types this weak prompt into a chatbot: *"tell me about volcanoes"*. The chatbot replies with a long answer that includes the "fact" that the world's tallest volcano is in Singapore.

a) Rewrite Maya's prompt so it is clear and useful. Include a role, a clear instruction, and one constraint (length or audience).

b) The "fact" about Singapore is wrong. What is this kind of confident-but-false AI answer called, and what should Maya do before using it?

c) Maya uses the answer in a project. Write one sentence she should add to be honest about how she used AI.

Write, draw or describe your answer here...

Show model answer & marks

## Lower Secondary (Grades 7 to 8)

# Section A · Multiple Choice

Choose the one best answer. There is no penalty for a wrong try.

**Q1** First Principles · Generative AI — 2 marks

**A tokeniser splits "unhappiness" into ["un","happi","ness"]. This best explains why a language model...**

- A can miscount the letters inside a word
- B always spells every word perfectly
- C treats grammar as more important than meaning
- D reads text one letter at a time

**Q2** Ethics & Safety · Generative AI — 2 marks

**A realistic AI-made video shows a mayor "announcing" a policy they never proposed. Why are such deepfakes so dangerous to public trust?**

- A they look real, so false claims spread before anyone checks
- B they are always obviously low quality
- C they can only copy cartoon characters
- D they are easy for anyone to spot

## Section B · Structured Tasks

Write, draw or describe your answer, then reveal the marking guide.

### Task 1 · Be the Language Model (n-gram from first principles)

20 marks

First Principles Generative AI

A mini language model is trained on just three sentences:

```
the cat sat on the mat
the cat ran to the dog
the dog sat on the log
```

a) List every word that appears immediately after "the", with its count, then write each probability.

b) Starting from "the", the model must pick the most likely next word. Is there a single winner? What does a model do on a tie?

c) Generate a short sentence starting at "cat", always going to a most-frequent next word. On a tie, pick either; stop after 4–5 words. (after cat → sat 1, ran 1; after sat → on; after on → the.)

d) Why does this n-gram method become impossible to store for real language, and what do real LLMs use instead?

Write, draw or describe your answer here...

Show model answer & marks

### Task 2 · One Neuron, from First Principles and in Python

20 marks

First Principles Python

A single neuron takes two 0/1 inputs and outputs 1 (fires) when **w1·x1 + w2·x2 + b ≥ 0**, else 0. Here **w1 = 1, w2 = 1, b = -1.5.**

```
def step(z):
    return 1 if z >= 0 else 0

def predict(x1, x2, w1=1, w2=1, b=-1.5):
    z = w1*x1 + w2*x2 + b
    return step(z)

for x1, x2 in [(0,0),(0,1),(1,0),(1,1)]:
    print(x1, x2, '->', predict(x1, x2))
```

a) Compute the output for (0,0), (0,1), (1,0), (1,1). Which everyday logic gate is this?

b) Change **only the bias b** so the neuron fires when *at least one* input is 1 (an OR gate). Give the new value.

c) Run the code to confirm your answers, then describe what "learning" would change.

Write, draw or describe your answer here...

Show model answer & marks

**Upper Secondary / ITE / JC / Polytechnic (Grades 9 to 12)**

## Section A · Multiple Choice

Choose the one best answer. There is no penalty for a wrong try.

**Q1** First Principles — 2 marks

Word embeddings place king=(4,4), man=(3,1), woman=(1,1). king − man + woman gives…

- A (8, 6)
- B (2, 4)
- C (6, 2)
- D (0, 0)

**Q2** Generative AI — 2 marks

The transformer's attention mechanism mainly lets the model…

- A store images as raw pixels
- B connect itself directly to the live internet
- C weigh which other tokens are most relevant
- D delete its own training data

Q3 First Principles 2 marks

A CNN classifies X-rays. An early convolutional filter slides over the pixel grid and fires wherever a sharp light-to-dark change occurs. What has it most likely learned?

A The overall class label directly

B The stored image file format

C Edges / local intensity changes

D Global colour balance only

Q4 First Principles 2 marks

A tiny network outputs 0.8 for an example whose true label is 1. After one gradient-descent step its output for that example becomes 0.85. What does this step demonstrate?

A Inference adjusting its own weights

B The activation function being replaced

C Overfitting being corrected

D Weights moving to reduce the loss

Q5 First Principles 2 marks

A company takes a general open-weights language model and keeps training it only on its own legal contracts so it writes in-house clauses well. This process is best called:

- A Pretraining a model from scratch
- B Inference-time prompting only
- C RLHF reward modelling here
- D Fine-tuning a pretrained model

Q6 First Principles 2 marks

A fraud classifier scores 99% accuracy, but only 1% of transactions are fraud and it labels everything 'legitimate'. Which metric best exposes the real problem?

- A Accuracy on the full set
- B Number of model layers
- C Recall on the fraud class
- D Training-set size used

Q7 Generative AI 2 marks

At one decoding step, candidate logits are 'help'=2.0, 'hello'=1.0, 'hex'=0.0, and the model picks 'help' every time it reaches this state. Which setting matches this?

A High temperature random sampling

B Uniform random token sampling

C Re-tokenising the prompt differently each time

D Greedy / argmax decoding step

Q8 Generative AI 2 marks

A text-to-image diffusion model is asked for 'a red bicycle in snow'. Starting from pure random noise, what does it do across its denoising steps?

A Removes noise gradually toward the prompt

B Adds more noise until a structure appears

C Looks up the closest stored photo online

D Sharpens a photo you uploaded yourself

Q9 Generative AI 2 marks

A team runs one brainstorming prompt at temperature 0.2 and at 1.1. Versus 0.2, the 1.1 outputs are more varied but include more off-topic, wrong lines. This illustrates:

A More reliability, less creativity

B More creativity, less reliability

C Faster speed, but higher cost

D Shorter context, longer outputs

Q10 Interaction & Agency 2 marks

A maths tutor model gives wrong final answers when asked directly, but improves sharply when the prompt adds 'show each step before the final answer'. Which technique explains the gain?

A Lowering the temperature value

B Expanding the context window size

C Chain-of-thought step prompting

D Switching to greedy decoding

Q11 Interaction & Agency 2 marks

**A travel system gets 'What's the weather where my 3pm meeting is?'. It checks the calendar tool, finds the city, calls a weather API, then writes the reply. Which loop is this?**

- A Generate one token at a time only
- B Retrieve some documents then stop
- C Fine-tune on each new request
- D Plan, call tools, observe, respond

Q12 Interaction & Agency 2 marks

**A RAG study-helper retrieves textbook passages before answering. To measure whether retrieval is actually helping accuracy, which evaluation is most informative?**

- A Count the model's total parameters
- B Compare accuracy with vs without it
- C Measure the average response length
- D Check the model's training cut-off

Q13 Ethics & Society 2 marks

**A loan model's false-positive rate (approving applicants who later default) is twice as high for one neighbourhood as another at equal incomes. Most appropriate governance response?**

A Ignore it since overall accuracy stays high

B Raise the decision threshold for everyone

C Audit the disparity and assign accountability

D Hide the metric from the regulators entirely

Q14 Ethics & Society 2 marks

**A support LLM confidently invents a refund policy that does not exist. The team added RLHF and a RAG knowledge base, yet rare fabrications persist. Best explanation?**

A It predicts plausible text, not verified facts

B The model is broken and must be retrained

C RLHF removes all hallucination by design

D RAG guarantees only true outputs always

Q15 Ethics & Society

2 marks

**Before deploying a city-wide face-matching CCTV tool, a board reads its model card. Which combination should most influence the go/no-go decision?**

- A Headline accuracy and inference speed only
- B Public download count and brand reputation
- C Model parameter size and training cost only
- D Per-group errors, intended use, societal cost

Q16 First Principles

2 marks

**Two headline embeddings are a=(3,4) and b=(4,3). A junior says they mean the same thing because both have length 5. Why is that reasoning flawed?**

$$\cos = (12+12)/(5*5) = 24/25 = 0.96$$

- A Equal magnitude need not mean equal direction
- B Length always equals meaning in embeddings
- C Cosine ignores both vectors entirely here
- D Only negative vectors carry any meaning

# Section B · Structured Tasks

Write, draw or describe your answer, then reveal the marking guide.

## Task 1 · Sampling, Greedy Decoding & Temperature

20 marks

First Principles | Generative AI | Python

At one step, a language model assigns these probabilities over a 4-token vocabulary for the next word:

```
{ "sunny": 0.50, "cloudy": 0.30, "rainy": 0.15, "banana": 0.05 }
```

a) Under greedy decoding, which token is emitted? State the rule you used.

b) Explain, in terms of this distribution, why raising the temperature increases the chance of emitting "rainy" or even "banana."

c) Complete the Python to (i) return the greedy token and (ii) sample a token by the probabilities. Describe an experiment to show sampling gives varied outputs.

d) Why might a support bot use low temperature while a brainstorming tool uses higher temperature?

Write, draw or describe your answer here...

Show model answer & marks

## Task 2 · Meaning as Vectors - Embeddings & Bias

20 marks

First Principles | Generative AI | Ethics & Safety

Three words have tiny 2-D embeddings: **cat = (2, 3), dog = (3, 2), car = (5, 0)**. Cosine similarity measures the angle between vectors: cos(u,v) = (u·v) / (|u| |v|); closer to 1 = more similar.

```
import math
vecs = {"cat":(2,3), "dog":(3,2), "car":(5,0)}

def cosine(u, v):
    dot = u[0]*v[0] + u[1]*v[1]
    return dot / (math.hypot(*u) * math.hypot(*v))

print(round(cosine(vecs['cat'], vecs['dog']), 3))   # 0.923
print(round(cosine(vecs['cat'], vecs['car']), 3))   # 0.555
```

a) Compute the cosine similarity of cat–dog and cat–car. Which pair is more similar? Does that match meaning?

b) Use cosine() to find cat's nearest neighbour.

c) A real model places "nurse" closer to "woman" and "engineer" closer to "man." Explain in first-principles terms where this bias comes from, and one consequence if it powers a hiring tool.

Write, draw or describe your answer here...

Show model answer & marks

### Task 3 · Build & Evaluate a Classifier in Python

20 marks

First Principles · Python · Skill

You are given a tiny labelled dataset and starter Python for a one-neuron classifier. Complete the forward pass, split the data, and judge how well it really generalises.

```
w = [1.5, -2.0];  b = 0.5
z = w[0]*x[0] + w[1]*x[1] + b
out = 1 if z >= 0 else 0     # step activation
```

a) Complete the forward pass. With w = [1.5, −2.0], b = 0.5 and input x = [2, 1], compute $z = w \cdot x + b$, then the neuron's output after the step activation (fires if $z \geq 0$). Show z.

b) You split 100 rows 80/20. The model scores 100% on the training rows but 60% on the held-out test rows. Name what is happening and give one concrete fix.

c) The test confusion matrix is TP = 8, FP = 2, FN = 6, TN = 4. Compute precision and recall. For a disease screen, which should you prioritise, and why?

Write, draw or describe your answer here...

Show model answer & marks