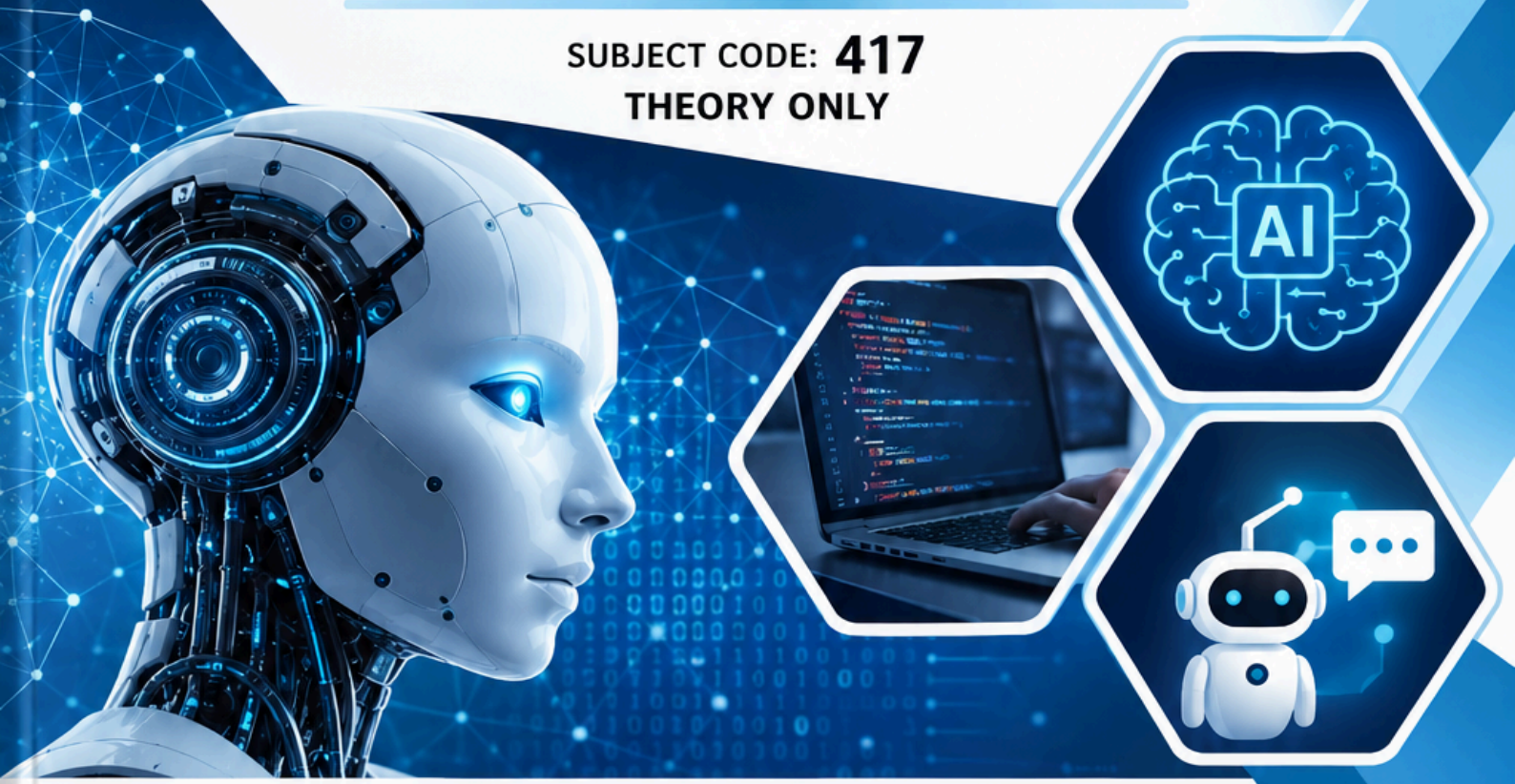
ACADEMIC SESSION 2026-27

CLASS X

ARTIFICIAL INTELLIGENCE

SUBJECT CODE: 417
THEORY ONLY



CONCEPTS

EXAMPLES

EXERCISES

SAMPLE QUESTION PAPERS

By

Mr. Taranjeet Singh
HOD IT-CS

Mr. Ravindra Uniyal
PGT CS

PUBLICATION NOTICE

Class X - Artificial Intelligence (417)
Academic Session 2026-27

By Mr. Taranjeet Singh - HOD IT-CS
Mr. Ravindra Uniyal - PGT CS

Copyright 2026. Rights in original material remain with their lawful owners. Third-party material retains its respective rights. Uses permitted by law are unaffected.

Independent study material. This is not an official CBSE publication. Question-specific practice marking allocations are suggested, unless expressly identified as official.

CONTENTS

Assessment Blueprint and Marking	1
Answer-writing Guide	2
Theory Unit Marks	3
Communication Skills-II	4
Self-Management Skills-II	24
ICT Skills-II	41
Entrepreneurial Skills-II	58
Green Skills-II	72
AI Project Cycle and Ethics	86
Advanced Concepts of AI Modeling	108
Regression and Clustering	112
Evaluating Models	134
Solved Evaluation Numericals	153
Computer Vision	176
Natural Language Processing	192
Solved NLP Numericals	213
Additional Solved Questions: Units 1-3	238
Ten Solved Sample Papers	249

ASSESSMENT BLUEPRINT AND MARKING

Each of the ten new papers is 50 marks, two hours, with 24 objective marks and 26 subjective marks. The full paper offers choices; students answer only the permitted number.

Question group	Answer	Marks
Q1: Employability MCQs	4 of 6	4
Q2-Q5: AI MCQs	5 of 6 in each	20
Q6-Q10: Employability	3 of 5, 2 each	6
Q11-Q16: AI short answers	4 of 6, 2 each	8
Q17-Q21: AI long answers	3 of 5, 4 each	12
Total assessed		50

Subject-specific question offers in each paper

Unit	1-mark items	2-mark items	4-mark items
AI cycle and ethics	5	1	1
Modelling	4	2	2
Evaluation	6	1	1
Computer Vision	4	1	0
NLP	5	1	1

Employability offers: one Communication, two Self-Management, one ICT, one Entrepreneurship and one Green Skills MCQ; one two-mark question from each unit. These are counts offered before choice, not guaranteed marks earned from each unit.

ANSWER-WRITING GUIDE

1-mark questions

Select the correct option or provide the exact fact requested. There is no negative marking.

2-mark questions

Answer every requested part. Use two distinct points when appropriate. A definition and examples may have different mark allocations; follow the question and its answer rubric.

4-mark questions

Address each subpart. For comparisons, use a labelled table. For calculations, show the formula, substitution and final result. Draw a labelled confusion matrix when asked.

Numerical accuracy

Identify the positive class. Check TP, TN, FP and FN before calculating. Keep full precision until the final result. For NLP vectors, fix the vocabulary order and use the stated preprocessing and TF-IDF convention.

Practice marking

Use the suggested allocations printed with each answer to check whether all required elements are present.

Unit allocations from the curriculum

Theory component	Marks
Five Employability Skills units	10
AI cycle and ethical frameworks	7
Advanced modelling	11
Evaluating models	10
Computer Vision	4
Natural Language Processing	8
Total	50

CHAPTER 01 | Communication Skills-II

Concepts and applications

The communication cycle

Communication is the exchange of information and meaning. A sender develops an idea, encodes it into words or signs and sends it through a channel. The receiver decodes the message and returns feedback. Noise is anything that interrupts or distorts this exchange. Feedback shows what was understood; merely sending a message does not prove successful communication.

One process, several channels

In a workshop, a trainer explains a safety instruction aloud and demonstrates it. Speech and demonstration are channels; the learner who repeats the steps provides feedback. A noisy machine can create a physical barrier. An unfamiliar technical word can create a language barrier even in a quiet room.

Verbal, non-verbal and visual forms

Verbal communication uses language: oral communication includes a discussion or telephone call, while written communication includes an email or notice. Non-verbal communication includes posture, facial expression, gesture and appropriate eye contact. Visual communication uses signs, maps, charts or symbols. A presentation can combine all three; a chart is not the same thing as a spoken explanation.

Choosing a suitable method

Choose a channel for the purpose and audience. A written message is useful when a record or detailed instructions are needed. A conversation allows immediate clarification. Tone and body language can reinforce or contradict spoken words. Formal communication follows an authorised structure; informal communication takes place through everyday interaction.

Communication Skills - Infographic Summary

Class X

Artificial Intelligence

Communication Skills

What is Communication?

Communication is the process of sharing information, ideas, thoughts and feelings between people. It helps us express ourselves, understand others and build strong relationships at home, in school and in the workplace.

The Communication Cycle

Communication is a two-way process. It involves a sender, a message, a medium or channel, a receiver and feedback.



Sender

The person who has an idea or information to share.



Message

The information, idea, thought or feeling being shared.



Medium / Channel

The method used to send the message (e.g. face to face, phone, email, chat, etc.).



Receiver

The person who receives and understands the message.



Feedback

The response from the receiver which shows whether the message has been understood.

Feedback

Helps the sender know if the message was understood and allows for further communication.

The 7 Cs of Communication

Good communication is effective communication. The 7 Cs help us communicate in a better way.

1. **Clear** — Be easy to understand. Use simple language.
2. **Concise** — Be brief and to the point.
3. **Concrete** — Be specific and give factual information.
4. **Correct** — Be accurate and truthful.
5. **Coherent** — Be logical and well organised.
6. **Complete** — Provide all necessary information.
7. **Courteous** — Be polite, respectful and considerate.

By Mr. Taranjeet Singh HOD IT-CS, Mr. Ravindra Uniyal PGT CS

Seven checks for an effective message

Clear: use understandable words. Concise: remove unnecessary repetition. Concrete: use specific facts rather than vague claims. Correct: check facts, spelling and grammar. Coherent: arrange ideas logically. Complete: include the information needed for action. Courteous: be respectful. For example, a meeting notice should identify the date, time, venue, purpose and required preparation.

Feedback that helps

Useful feedback is specific, timely, respectful and linked to observable work. "The graph is labelled; add the unit on the vertical axis" tells a learner what to retain and improve. A personal insult or an unexplained "bad work" does not. Active listening means attending, avoiding interruption, asking relevant questions and confirming understanding.

Barriers and remedies

Physical noise can be reduced by changing the location or equipment. Language barriers can be reduced through simpler wording and explanation. Emotional barriers may need calm listening and an appropriate pause. Cultural assumptions require respect and clarification rather than stereotyping. Organisational barriers may require a clear reporting channel. A remedy should address the actual barrier named in the case.

Writing accurately

A complete sentence has a meaningful structure, usually a subject and predicate. Nouns name, pronouns replace nouns, verbs express actions or states, adjectives describe nouns and adverbs modify verbs or other words. Prepositions show relationships; conjunctions join units; interjections express sudden feeling. Use a/an according to sound and the for a specific reference. Check subject-verb agreement, tense and punctuation.

Sentences and paragraphs

A declarative sentence makes a statement; an interrogative asks a question; an imperative gives a request or instruction; an exclamatory expresses strong feeling. A paragraph develops one main idea with supporting detail and a logical sequence. Proofread dates, names and amounts as carefully as grammar.

CHAPTER 01 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A listener restates an instruction in her own words to check its meaning. Which stage does this demonstrate? [1]

(a) Encoding

(b) Noise

(c) Feedback

(d) Channel selection

Answer:

(c) Feedback The response lets the sender check what the receiver understood.

Q2. A notice says that a meeting is postponed but gives no new date. Which quality is chiefly missing? [1]

(a) Grammatical accuracy

(b) Completeness

(c) Courtesy

(d) Brevity

Answer:

(b) Completeness The recipient lacks information needed to act.

Q3. Two students interpret the word "charge" differently in a message. Which barrier is involved? [1]

(a) Physical barrier

(b) Lack of eye contact

(c) Semantic barrier

(d) Network congestion only

Answer:

(c) Semantic barrier Different meanings of a word can distort understanding.

Types of Communication

Choose the communication mode according to purpose, audience and situation.

Mode	Examples	Advantages	Limitations
Oral	Conversation, phone call	Fast feedback and clarification	No automatic permanent record
Written	Email, letter, notice	Permanent record; can be reviewed	Slow response; tone may be misunderstood
Non-verbal	Gestures, facial expression, posture	Expresses feelings without words	Meanings can differ by culture
Visual	Charts, signs, diagrams	Makes patterns and instructions easy to see	Needs clear labels and accessible presentation

Remember: verbal communication uses words (oral or written); visual and non-verbal cues often support the message.

Q4. Which feedback identifies both a strength and an actionable improvement? [1]

(a) Everyone agrees you are wrong.

(b) Do not try presenting again.

(c) Your presentation is useless.

(d) Your chart is clear; label its units to improve interpretation.

Answer:

(d) Your chart is clear; label its units to improve interpretation. Constructive feedback is specific and useful rather than a personal attack.

Q5. A science diagram communicates information through labelled arrows. Its main communication mode is: [1]

(a) Gossip

(b) Tactile

(c) Visual

(d) Oral

Answer:

(c) Visual The diagram conveys meaning through visible symbols and spatial relationships.

Q6. Which sentence uses a courteous request? [1]

(a) Could you please send the revised file by noon?

(b) Only a careless person would delay this.

(c) Send it now, or else.

(d) You never do anything properly.

Answer:

(a) Could you please send the revised file by noon? Courtesy maintains respect while making a clear request.

Q7. A receiver misunderstands a correct spoken message because of loud machinery. The obstacle is mainly: [1]

(a) Non-verbal encouragement

(b) Feedback

(c) Physical noise

(d) A grammatical article

Answer:

(c) Physical noise Environmental sound interferes with the channel.

Q8. Which statement is a complete sentence? [1]

(a) Because of the labels.

(b) After the volunteers.

(c) The volunteers checked every label.

(d) Checking every.

Answer:

(c) The volunteers checked every label. A complete sentence expresses a complete thought with a suitable subject and predicate.

Q9. During a discussion, which action best demonstrates active listening? [1]

(a) Ask a relevant clarifying question after the speaker finishes.

(b) Interrupt at every pause.

(c) Change the topic immediately.

(d) Prepare a reply without attending to the speaker.

Answer:

(a) Ask a relevant clarifying question after the speaker finishes. Clarification after attentive listening checks understanding.

Q10. A sender converts an idea into spoken words. This is: [1]

(a) Decoding

(b) Encoding

(c) Feedback

(d) Archiving

Answer:

(b) Encoding Encoding creates the message; decoding interprets it.

Short-answer questions - 2 marks each

S1. Explain the role of feedback and give one example. [2]

Answer:

(1) Feedback tells a sender how a message was understood. [1]

(2) For example, a learner repeats an instruction to confirm its meaning. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Give two actions that make listening active rather than passive. [2]

Answer:

(1) Attend carefully to the speaker without interrupting or preparing an unrelated reply. [1]

(2) Ask a relevant clarifying question or summarise the message to confirm understanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Distinguish verbal communication from non-verbal communication. [2]

Answer:

(1) Verbal communication uses spoken or written words. [1]

(2) Non-verbal communication conveys meaning through cues such as gestures, facial expressions and posture. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Name one physical and one semantic communication barrier, with an example of each. [2]

Answer:

(1) Loud machinery is a physical barrier because it interferes with hearing. [1]

(2) An unfamiliar technical word is a semantic barrier because its meaning may be misunderstood. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

S5. Explain constructive feedback with one suitable example. [2]

Answer:

(1) Constructive feedback identifies a specific strength or weakness and suggests an improvement. [1]

(2) For example, add a labelled diagram to clarify the process. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. State the purpose of a clear subject line and a polite closing in an email. [2]

Answer:

(1) A clear subject line tells the reader the purpose of an email. [1]

(2) A polite closing states the requested next step without sounding demanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Define a pronoun and illustrate it in a sentence. [2]

Answer:

(1) A pronoun replaces a noun to avoid unnecessary repetition. [1]

(2) In "Meera checked the file; she corrected it", "she" replaces Meera and "it" replaces the file. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. Give one advantage and one limitation of written communication. [2]

Answer:

(1) Written communication provides a record that can be checked later. [1]

(2) However, immediate clarification may be unavailable, so unclear wording can cause misunderstanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Distinguish a complete sentence from a sentence fragment. [2]

Answer:

(1) A complete sentence expresses a complete thought with an appropriate subject and predicate. [1]

(2) A fragment, such as "because the printer stopped", leaves the thought unfinished. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

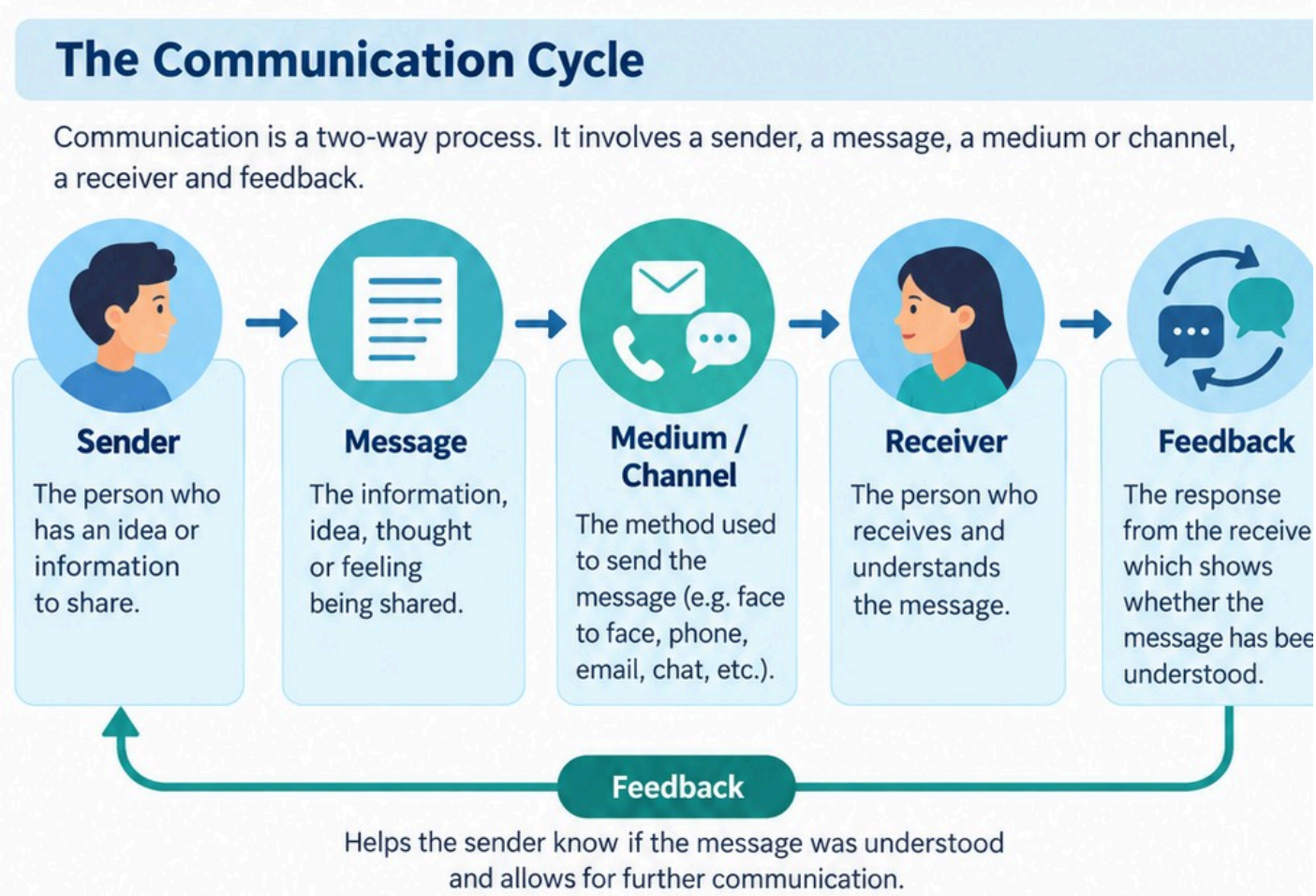
S10. Explain how cultural differences can create a communication barrier. [2]

Answer:

(1) A gesture or expression can have different meanings across cultures. [1]

(2) Using respectful language and checking understanding helps prevent assumptions from turning into misunderstanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.



CHAPTER 01 | Detailed topic lessons

Communication Skills-II: writing, feedback and application

Sentence, phrase and clause

A **sentence** expresses a complete thought and can stand on its own: "The students revised the chapter." A **phrase** is a group of related words that does not by itself express a complete thought: "after the lesson" or "the blue notebook". A **clause** contains a subject and a verb. "Because the bell rang" is a clause but needs a main clause to complete its meaning.

Expression	Classification and reason
under the wooden table	Phrase: no complete subject-verb statement.
The printer is ready.	Sentence: a complete statement.
Although the printer is ready	Dependent clause: the thought remains unfinished.
Although the printer is ready, the file is missing.	Complete sentence: the dependent clause is linked to a main clause.

An imperative can have an **understood subject**: in "Please close the door", the subject is understood to be "you". Do not reject a command as a sentence merely because the word "you" is not printed.

Parts of a sentence

The **subject** tells who or what the sentence is about. The **predicate** tells what is said about the subject and contains the verb. In "Our class completed the assignment", "Our class" is the subject and "completed the assignment" is the predicate.

An **object** receives the action. In "Mira gave Ravi a notebook", Mira is the subject, gave is the verb, a notebook is the **direct object** (what was given), and Ravi is the **indirect object** (to whom it was given). A sentence such as "The baby slept" needs no object.

Kinds of sentences

Kind	Purpose and original example
Declarative	States something: "Our team checked the figures."
Interrogative	Asks a question: "Have you checked the figures?"
Imperative	Gives an instruction or request: "Please check the figures."
Exclamatory	Expresses strong feeling: "What a remarkable result!"

Eight parts of speech

Part	Role and example
Noun	Names a person, place, thing or idea: "The **library** is quiet."
Pronoun	Stands for a noun: "Ritu arrived. **She** sat down."
Verb	Expresses an action or state: "They **work**. He **is** ready."
Adjective	Describes a noun or pronoun: "a **careful** learner".
Adverb	Modifies a verb, adjective or another adverb: "She checked **carefully**."
Preposition	Shows a relationship: "The book is **beside** the bag."
Conjunction	Joins words or clauses: "Read **and** revise."
Interjection	Expresses sudden feeling: "**Oh!** I forgot the file."

Identify a word by its **function in the sentence**, not its spelling alone. In "I **paint** walls", paint is a verb; in "The **paint** is wet", it is a noun.

Articles, agreement and punctuation

Use **a** before a consonant sound, **an** before a vowel sound, and **the** for an identified or particular reference. Write "a university" because the first sound is /y/, and "an hour" because h is silent. "I borrowed a book. The book explains robots" moves from a new reference to a known one.

A singular subject normally takes a singular present-tense verb: "The learner writes"; a plural subject takes "write". Keep tense consistent: "We collected the data and checked it", not "collected and check" when both actions are past. Use capitals at the start of a

sentence, for names, for the pronoun I and appropriate titles. End statements with a full stop, questions with a question mark, and use commas to separate list items or clarify pauses. An apostrophe may show possession or omission; it is not a general plural marker.

Active and passive voice

Active voice places the doer as the subject: "The technician repaired the device." Passive voice places the receiver of the action as the subject: "The device was repaired by the technician." Preserve the **tense and meaning**. Passive voice can be useful when the action or receiver matters more than the doer; active voice often makes instructions more direct.

Constructing a coherent paragraph

Choose one central idea, write a topic sentence, support it with related details, arrange them logically and finish with a suitable closing sentence. Use connecting words only when they show a real relationship. Remove an attractive but unrelated sentence rather than forcing it into the paragraph.

Original model paragraph: A weekly revision plan makes study more manageable. First, I divide the syllabus into small tasks and allot a time to each one. After completing a task, I answer a few questions without looking at the notes. I then record mistakes and revisit them the next day. This routine helps me use feedback rather than repeat the same errors.

Feedback and barrier-remedy matching

"Your explanation was good" is positive but non-specific. "Your explanation defined precision correctly; add the denominator in the formula" is **descriptive and specific**. Effective feedback identifies observable work, explains the needed improvement and avoids judging the learner as a person.

Barrier in a situation	Appropriate response
Traffic noise prevents hearing.	Move to a quieter place or use a suitable communication channel.

Barrier in a situation	Appropriate response
An unfamiliar technical word causes confusion.	Explain it in familiar language and confirm understanding.
An angry listener stops listening.	Pause, use a calm tone and discuss the concern respectfully.
A gesture is interpreted differently across cultures.	Ask what was understood; do not assume a universal meaning.
Unclear reporting channels delay information.	Identify the appropriate contact and a clear communication route.

Additional exam-style practice

C1-Q1. Which expression is a phrase, rather than a complete sentence? [1]

(a) The report is ready.
(b) Under the bright lamp
(c) Please save your work.
(d) Did the team arrive?

C1-Q2. Which pair correctly completes "She joined ___ university and waited for ___ hour"? [1]

(a) an, a
(b) the, a
(c) a, an
(d) an, an

C1-Q3. Define a pronoun and give one example in a sentence. [2]

Model answer: A pronoun is used in place of a noun. In "Neha arrived and she opened the file", **she** refers to Neha.

Suggested practice marking: Definition 1; relevant sentence identifying the pronoun 1.

C1-Q4. Identify the subject and predicate in "The two volunteers organised the books". [2]

Model answer: Subject: **The two volunteers**. Predicate: **organised the books**.

Suggested practice marking: Subject 1; complete predicate 1.

C1-Q5. Rewrite "Bad presentation" as specific feedback. State why your version is more useful. [2]

Model answer: "Your headings were clear; increase the chart labels so the audience can read them." It identifies observable work and a specific improvement.

Suggested practice marking: Specific, respectful feedback 1; explanation of its usefulness 1.

C1-Q6. State two differences between a sentence and a phrase, using an example of each. [2]

Model answer: A sentence expresses a complete thought: "The class is quiet." A phrase does not stand alone as a complete thought: "in the quiet classroom".

Suggested practice marking: Correct sentence with explanation/example 1; correct phrase with explanation/example 1.

MCQ answers and reasons

C1-Q1: (b). The expression has no complete statement.

C1-Q2: (c). University begins with a consonant sound; hour begins with a vowel sound.

IMPORTANT QUESTIONS WITH ANSWERS

Communication Skills-II

Q1. What is communication? Explain the communication cycle. [2]

Answer: Communication is the exchange of information, thoughts or feelings between a sender and a receiver. The cycle is sender → encoding → message → channel → receiver → decoding → feedback; noise may interfere.

Suggested marking: *1 definition + 1 cycle.*

Q2. List the main types of communication and give examples. [4]

Type	Description and example
Oral / verbal	Spoken exchange; classroom discussion
Written	Words in emails, letters and notices
Non-verbal	Gestures, expressions, posture, eye contact
Visual	Graphs, signs, diagrams and charts

Suggested marking: *1 mark per correctly explained type with example.*

Q3. State two advantages and two disadvantages of oral communication. [4]

Advantages of oral communication	Disadvantages
Immediate feedback and clarification	May leave no permanent record
Conveys emotion and tone	Noise, accent or poor listening may lead to misunderstanding

Suggested marking: *1 mark per distinct valid advantage/disadvantage.*

Q4. State two advantages and two disadvantages of written communication. [4]

Advantages of written communication	Disadvantages
Creates a permanent record	Feedback may be delayed
Can be organised and shared precisely	Drafting and reading take time; tone can be misread

Suggested marking: *1 mark per distinct valid advantage/disadvantage.*

Q5. Give two advantages and two limitations of non-verbal communication. [4]

Advantages of non-verbal communication	Limitations
Reinforces spoken meaning	Gestures vary between cultures
Conveys emotion without words	Complex instructions cannot easily be explained alone

Suggested marking: *1 mark per valid advantage/limitation.*

Q6. Explain the 7 Cs of effective communication. [4]

7 C	Meaning
Clear	Easily understood
Concise	Avoid unnecessary words
Concrete	Use facts and examples
Correct	Accurate words and grammar
Coherent	Logical order
Complete	All needed details
Courteous	Respectful and considerate

Suggested marking: *0.5 each for any six accurate Cs plus 1 for the seventh and an appropriate overall explanation; total*

Q7. What is the difference between verbal and non-verbal communication? [2]

Verbal communication	Non-verbal communication
Uses spoken/written words	Uses gestures, facial expressions, eye contact and posture

Suggested marking: *1 +*

Q8. What are communication barriers? Explain four examples. [4]

Answer: Language/semantic barrier: unfamiliar vocabulary or ambiguous words. Physical barrier: noise or poor connectivity. Psychological barrier: anxiety or prejudice. Cultural barrier: different conventions or interpretations.

Suggested marking: *1 mark each for four explained barriers.*

Q9. What is active listening? Give three techniques. [4]

Answer: Active listening means focusing on the speaker and checking the intended meaning. Techniques: maintain appropriate eye contact without interrupting; paraphrase the message;

ask relevant clarification questions.

Suggested marking: *1 definition + 1 each for three techniques.*

Q10. Differentiate phrase, clause and sentence with examples. [3]

Form	Definition and example
Phrase	Group of words without a complete thought; in the classroom
Clause	Group with subject and verb; because students listened
Sentence	Expresses a complete thought; The students listened carefully.

Suggested marking: *1 per correct distinction with example.*

Q11. Explain sentence types with examples. [4]

Answer: Declarative: states a fact ("The class begins at nine."). Interrogative: asks a question ("When does it begin?"). Imperative: gives a direction ("Open the book."). Exclamatory: expresses strong feeling ("What a wonderful idea!").

Suggested marking: *1 each.*

Q12. Why is constructive feedback important? [2]

Answer: It tells the receiver what was effective and what specifically could improve. For example, "Your examples were clear; slow down while explaining the formula" is useful and respectful.

Suggested marking: *1 significance + 1 suitable example.*

CHAPTER 02 | Self-Management Skills-II

Concepts and applications

Managing your own work

Self-management means directing your behaviour, time and effort towards responsible action. It combines self-awareness, self-control, motivation and planning. Working independently does not mean refusing all help: it means taking ownership and seeking appropriate help when it is needed.

Recognising stress

Stress is a response to demands or pressure. A person may notice worry, poor concentration, irritability or difficulty organising work. Identify what is causing the pressure before choosing a response. A heavy workload may call for priorities and a timetable; unclear instructions may call for clarification.

A practical response sequence

First recognise that stress is affecting you. Next identify the likely source. Then choose a constructive method: break work into smaller steps, plan rest, exercise appropriately, use relaxation techniques or speak to a trusted person. Review whether the method helps. Avoid presenting harmful habits or endless avoidance as a solution.

Self-awareness and SWOT

Self-awareness means recognising your abilities, limitations, interests, values and reactions. In a personal SWOT review, strengths and weaknesses describe internal factors; opportunities and threats describe external circumstances. Good memory may be a strength; an available training course may be an opportunity. A weakness is not a permanent identity: it can guide a realistic improvement plan.

Motivation

Intrinsic motivation comes from interest, satisfaction or a personal sense of purpose. Extrinsic motivation comes from an outside reward or recognition. A learner may enjoy understanding a concept and also want a certificate. Build self-motivation by setting manageable goals, recording progress, using constructive self-talk and learning from setbacks.

3.4 Self-Management Skills

Self-management skills help us understand and manage our emotions, thoughts and behaviour. They enable us to handle challenges, stay focused and make healthy choices. Managing stress is an important part of self-management.

Understanding Stress

Stress is the body's natural response to a change or challenge. It can be helpful or harmful depending on its intensity, duration and the way we handle it.

Good Stress (Eustress) A positive and healthy form of stress that motivates us.	Bad Stress (Distress) A negative and harmful form of stress that can affect our well-being.
Motivates and energises Helps us take action and put in effort.	**Causes anxiety and discomfort** Can make us feel worried, tense or overwhelmed.
Improves performance Helps us focus and do our best.	**Reduces performance** Makes it difficult to concentrate and make good decisions.
Builds confidence Gives a sense of achievement and helps us learn.	**Leads to negative health effects** Can cause tiredness, headaches, sleep problems and other health issues.
Short-term and manageable Usually lasts for a short period and goes away after the task is over.	**Long-term and persistent** May continue for a long time if not managed properly.
Feels exciting and meaningful Creates a sense of challenge and keeps us active.	**Feels overwhelming and unpleasant** Creates a sense of loss of control and can affect our mood and behaviour.

Stress-Management Techniques

We can use simple strategies to manage stress and keep our mind and body healthy.

Time Management
Plan your tasks, set priorities and avoid last-minute rush.

Deep Breathing
Take slow, deep breaths to calm your mind and reduce tension.

Regular Exercise
Physical activity helps release stress and improves mood.

Adequate Sleep
Get enough rest to keep your mind fresh and body strong.

Positive Thinking
Focus on the good things, be kind to yourself and look for solutions.

Seek Support
Talk to your family, friends or a trusted teacher when you feel stressed.

SMART goals

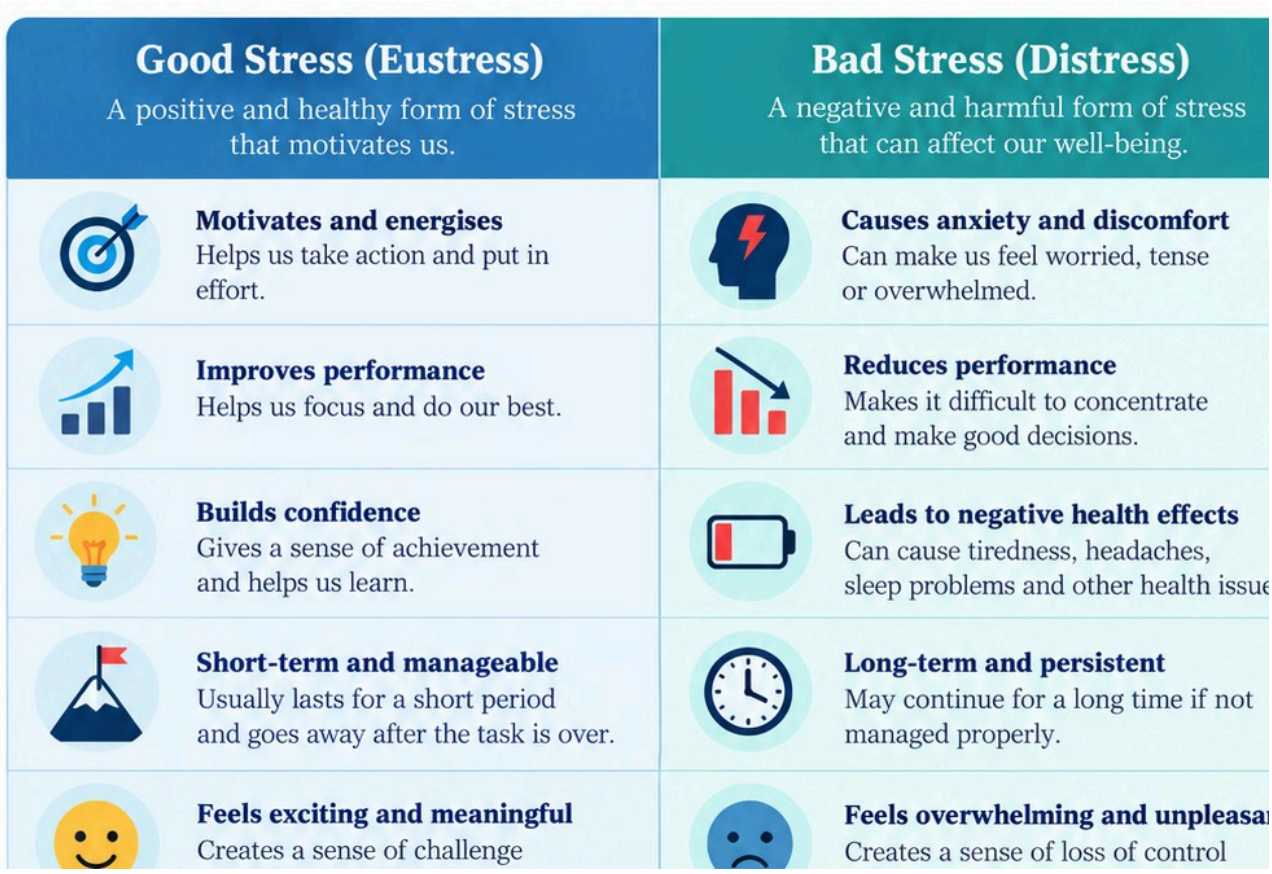
A goal should be Specific, Measurable, Achievable, Relevant and Time-bound. "Practise six confusion-matrix problems by Friday and check every error" states an action, quantity and deadline. "Become perfect soon" does not. Some teaching materials use Realistic for R; follow the wording prescribed by the question or the school text while keeping the idea of a suitable, feasible goal.

Managing time and responsibility

List tasks, identify deadlines, prioritise important work, estimate time and track completion. Reserve time to check the result rather than only to finish it. Self-discipline means following a plan despite distractions. Accountability means accepting responsibility for your decisions, reporting progress honestly and correcting mistakes.

Confidence without overconfidence

Confidence is a reasoned belief that effort and preparation can help you act effectively. It does not mean pretending to know everything. Review evidence of improvement, practise difficult skills and accept useful feedback. Independent workers still communicate risks and ask for clarification before making an unsafe assumption.



CHAPTER 02 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. Riya studies robotics because solving problems interests her, even without prizes. This is: [1]

(a) Financial reward

(b) External punishment

(c) Peer pressure

(d) Intrinsic motivation

Answer:

(d) Intrinsic motivation Her motivation comes from interest in the activity itself.

Q2. A learner completes work only to receive a certificate. This is mainly: [1]

(a) Extrinsic motivation

(b) Self-awareness

(c) Time estimation

(d) Intrinsic curiosity

Answer:

(a) Extrinsic motivation The certificate is an external reward.

Q3. Which goal is most specific and measurable? [1]

(a) Become perfect soon.

(b) Study whenever possible.

(c) Be better than everybody.

(d) Complete ten revision questions each evening for seven days.

Answer:

(d) Complete ten revision questions each evening for seven days. The task, quantity and time period can be checked.

Q4. Before planning revision, a student lists topics she understands and topics she finds difficult. This shows: [1]

(a) Data encryption

(b) Delegation

(c) Avoidance

(d) Self-awareness

Answer:

(d) Self-awareness Recognising strengths and weaknesses supports realistic planning.

Q5. Which action best shows responsible independent work? [1]

(a) Refuse all feedback.

(b) Finish the assigned task and seek clarification when needed.

(c) Hide errors until submission.

(d) Wait for repeated reminders.

Answer:

(b) Finish the assigned task and seek clarification when needed. Independence includes responsibility, not refusing necessary help.

Q6. A student notices tension, identifies unfinished work as its cause, and creates a timetable. This is: [1]

(a) Stress management

(b) Uncontrolled risk taking

(c) File management

(d) Escapism

Answer:

(a) Stress management The student identifies a stressor and takes a constructive step.

Q7. What does the M in a SMART goal stand for? [1]

(a) Measurable

(b) Maximum

(c) Memorable

(d) Mechanical

Answer:

(a) Measurable Progress needs a measure that can be checked.

Q8. A revision plan includes no rest and several impossible daily targets. Which correction is best? [1]

(a) Avoid reviewing progress.

(b) Remove all deadlines.

(c) Add more tasks each day.

(d) Set achievable targets and allow breaks.

Answer:

(d) Set achievable targets and allow breaks. A workable plan should be realistic and sustainable.

Q9. A learner misses a target and reviews the cause before changing the plan. This mainly demonstrates: [1]

(a) Denial

(b) Dependence

(c) Self-regulation

(d) Carelessness

Answer:

(c) Self-regulation Monitoring and adjusting behaviour supports progress.

Q10. Which practice directly supports effective time management? [1]

(a) Prioritise important tasks and allocate time to them.

(b) Switch constantly between unrelated tasks.

(c) Leave difficult work until the last minute.

(d) Start every activity without a deadline.

Answer:

(a) Prioritise important tasks and allocate time to them. Prioritising and scheduling make use of limited time.

Short-answer questions - 2 marks each

S1. Distinguish intrinsic motivation from extrinsic motivation. [2]

Answer:

(1) Intrinsic motivation comes from interest or satisfaction in the activity. [1]

(2) Extrinsic motivation comes from external outcomes such as rewards, marks or recognition. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Give two practical steps for managing revision-related stress. [2]

Answer:

(1) Break revision into realistic tasks and schedule them. [1]

(2) Use a healthy routine, brief relaxation, exercise or appropriate support to manage tension. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

SMART Goals and Stress

Self-management tools for realistic goals and a healthy response to challenges.

SMART	Question to ask	Example
S - Specific	What exactly will I achieve?	Revise the NLP chapter
M - Measurable	How will progress be measured?	Finish 20 practice questions
A - Achievable	Can I reasonably do this?	One-hour sessions
R - Relevant	Does it support my learning goal?	Prepare for the AI exam
T - Time-bound	By when will it be done?	By Friday evening

Good Stress vs Bad Stress

Good stress (eustress)	Bad stress (distress)
May increase focus for a short challenge	May overwhelm attention and concentration
Example: excitement before a presentation	Example: ongoing worry that affects sleep
Manage with preparation and realistic goals	Seek support, rest, exercise and make a manageable plan

S3. Explain self-awareness and state how it helps planning. [2]

Answer:

(1) Self-awareness means recognising personal strengths, weaknesses and emotions. [1]

(2) It helps a learner set realistic goals and choose areas needing improvement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. What makes a goal measurable and time-bound? [2]

Answer:

(1) Measurable means progress can be checked using a clear quantity or criterion. [1]

(2) Time-bound means a deadline or time period is specified. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. Describe two behaviours that demonstrate responsible independent work. [2]

Answer:

(1) Take initiative and complete agreed tasks without repeated reminders. [1]

(2) Monitor progress, accept responsibility and seek clarification when information is genuinely needed. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. Suggest two improvements to an unrealistic revision timetable. [2]

Answer:

(1) Set achievable priorities rather than assigning excessive work each day. [1]

(2) Allocate time for rest and review progress so the timetable can be adjusted. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. List four steps for building self-motivation. [2]

Answer:

(1) Identify personal strengths and areas for improvement. [0.5]

(2) Set clear and achievable goals. [0.5]

(3) Prepare a practical action plan with manageable steps. [0.5]

(4) Continue working and reviewing progress despite difficulties. [0.5]

Suggested marking: 0.5 mark for each of four distinct relevant steps; total 2 marks.

S8. What does SMART stand for? State why time-bound goals help. [2]

Answer:

(1) SMART means Specific, Measurable, Achievable, Realistic and Time-bound. [1]

(2) A deadline helps a learner plan tasks and check whether progress is timely. [1]

Suggested marking: Complete SMART expansion: 1 mark; a relevant explanation of the deadline: 1 mark.

S9. Explain one difference between self-awareness and self-confidence. [2]

Answer:

(1) Self-awareness is understanding personal strengths, limitations and emotions. [1]

(2) Self-confidence is trust in the ability to act or learn, while still recognising areas for improvement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. State two responsible ways of responding to a difficult task. [2]

Answer:

(1) Break the task into manageable steps and begin with a realistic plan. [1]

(2) Seek specific guidance when necessary, then take responsibility for completing and checking the work. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

CHAPTER 02 | Detailed topic lessons

Self-Management Skills-II: self-awareness and self-regulation

Self-awareness: what you need to recognise

Self-awareness is more than listing strengths. Recognise your interests, abilities, values, habits, emotions and areas needing improvement. **Interest** is what attracts you; **ability** is what you can presently do. A student may enjoy public speaking but still need practice to speak clearly. Values influence what the student regards as important, such as honesty or punctuality.

Perspective for study	Evidence and improvement
Internal self-awareness	Notice your motives, feelings and habits. Example: "I rush when I fear missing a deadline."
External perspective / feedback	Consider how your actions affect other people. Example: a partner says rushed messages omit instructions.
Balanced self-review	Compare self-reflection with evidence and trusted feedback; choose an achievable change.

These perspectives help interpret the syllabus wording; the short official curriculum does not prescribe these exact category names. Do not confuse them with **intrinsic and extrinsic motivation**, which describe sources of motivation.

Strength, weakness, interest and ability

Use evidence rather than labels. "I checked all calculations correctly in three tasks" is stronger evidence than "I am perfect". A weakness should lead to a plan: "I lose track of deadlines, so I will maintain a dated checklist." In SWOT, internal strengths/weaknesses differ from external opportunities/threats. An available course is an opportunity; poor time planning is a weakness.

Self-regulation: the missing link between goal and action

Self-regulation means monitoring and directing your thoughts, emotions and behaviour to work towards a goal. It is not the same as suppressing every emotion. Notice a distraction, choose a suitable response, continue the task and review the outcome.

Stage	Example of regulating study
Plan	Choose four short questions and a 25-minute study period.
Act and monitor	Put distracting notifications aside and tick completed questions.
Adjust	When one answer is unclear, mark it for review rather than abandon the whole task.
Reflect	Check the mistakes and change the next session accordingly.

Self-motivation helps you begin and persist; **self-regulation** helps you manage how you act. Motivation says "I want to improve"; regulation turns that wish into monitored action. Working independently includes taking responsibility and knowing when to request necessary clarification, not refusing help in every situation.

Stress management: choose a method for the cause

Stress management helps maintain concentration, balanced decisions and a workable routine. The curriculum names physical exercise, yoga, meditation, enjoyable time with family or friends, holidays and nature walks. These are study topics, not a promise that one activity solves every difficulty. Combine suitable recreation with a realistic workload and enough time to complete work.

For a heavy workload, prioritise tasks and divide large tasks into smaller steps. For unclear instructions, seek clarification. For conflict, communicate calmly. For continuing distress, speak with a trusted adult or appropriate support person. Avoid treating avoidance of all responsibilities as stress management.

SMART goals and independent work

The NCERT sequence uses **Specific, Measurable, Achievable, Realistic and Time-bound**. "I will revise two NLP topics and answer six questions by 6 p.m." can be checked; "I will do better someday" cannot. Some sources use Relevant for R. Use the formulation specified by the question or prescribed source, rather than losing the underlying requirement of a suitable, feasible goal.

To work independently: understand the expected result; arrange resources; schedule the steps; complete them without repeated reminders; verify quality; and accept responsibility for corrections. A realistic plan includes review time, not just writing time.

Additional exam-style practice

C2-Q1. A learner observes that checking messages interrupts revision and switches notifications off during a planned session. This best shows: [1]

(a) Self-regulation
(b) Unsupervised learning
(c) Financial risk
(d) File compression

C2-Q2. Which pair describes sources of motivation, rather than types of data or operating systems? [1]

(a) Internal memory and external disk
(b) Intrinsic and extrinsic
(c) Labelled and unlabelled
(d) Single-user and multi-user

C2-Q3. State the meaning of self-regulation and give one study-related example. [2]

Model answer: Self-regulation means monitoring and adjusting behaviour towards a goal. A learner follows a timetable, checks progress and reduces distractions when work is delayed.

Suggested practice marking: Meaning 1; relevant example 1.

C2-Q4. State four ways to build self-motivation. [2]

Model answer: Set manageable goals; recognise progress; use constructive self-talk; and learn from setbacks instead of giving up.

Suggested practice marking: Any four relevant ways, 0.5 each; do not award separately for synonymous repeats.

C2-Q5. Distinguish an interest from an ability with one connected example. [2]

Model answer: Interest is what attracts a person; ability is a skill the person can demonstrate. Enjoying debates does not automatically mean speaking fluently in public.

Suggested practice marking: Interest correctly explained 0.5; ability correctly explained 0.5; connected example 1.

MCQ answers and reasons

C2-Q1: (a). The action controls a distraction and monitors work.

C2-Q2: (b). These refer to motivation, not self-awareness categories.

IMPORTANT QUESTIONS WITH ANSWERS

Self-Management Skills-II

Q1. What is stress? Differentiate good stress (eustress) and bad stress (distress). [4]

Good stress (eustress)	Bad stress (distress)
Manageable challenge that can motivate and sharpen focus	Excessive or prolonged pressure that harms wellbeing and performance
Example: preparing for a presentation	Example: overwhelming exam worry

Suggested marking: *1 definition + 1 contrast + 1 example each.*

Q2. Explain four common signs of harmful stress. [4]

Answer: Physical: headache or muscle tension. Emotional: irritability or persistent worry. Cognitive: difficulty concentrating. Behavioural: disturbed sleep, withdrawal or procrastination.

Suggested marking: *1 each for four correctly explained signs.*

Q3. Describe the ABC of stress management. [3]

Answer: A - Adversity/activating event: what happened. B - Beliefs: interpretation of the event. C - Consequences: feelings and actions resulting from those beliefs. Reframing an unhelpful belief can change the response.

Suggested marking: *1 each; accept textbook-aligned variations when clearly explained.*

Q4. Suggest four practical ways to manage exam stress. [4]

Answer: Make a realistic revision timetable; take short breaks and practise slow breathing; get adequate sleep and regular physical activity; seek support from a teacher, family member or counsellor if stress becomes overwhelming.

Suggested marking: *1 each for four appropriate practices.*

Q5. What is self-awareness? Why is it important? [2]

Answer: Self-awareness is recognising one's strengths, limitations, emotions and values. It helps learners choose realistic goals and regulate responses in difficult situations.

Suggested marking: *1 definition + 1 importance.*

Q6. Define self-motivation and give two strategies. [3]

Answer: Self-motivation is the internal drive to work towards a goal without repeated external reminders. Strategies include breaking a goal into achievable steps and tracking progress with

small rewards.

Suggested marking: *1 definition + 1 each strategy.*

Q7. What are SMART goals? Explain with an example. [4]

Letter	SMART goal meaning
S	Specific
M	Measurable
A	Achievable
R	Relevant
T	Time-bound

Suggested marking: *2.5 marks for five meanings at 0.5 each + 1.5 for suitable example.*

Q8. What is self-regulation? Give three practical behaviours. [4]

Answer: Self-regulation means controlling attention, emotions and actions to meet goals. Examples: pause before reacting angrily; follow a study schedule despite distractions; review mistakes and adjust study methods.

Suggested marking: *1 definition + 1 each behaviour.*

Q9. Distinguish time management and procrastination. [2]

Answer: Time management means planning and prioritising activities to use available time effectively. Procrastination means unnecessarily delaying an important task despite likely negative consequences.

Suggested marking: *1 each.*

Q10. List four steps for effective time management. [4]

Answer: Prioritise tasks by importance and deadline; split large tasks into smaller steps; allocate realistic time slots; review progress and adjust the plan.

Suggested marking: *1 each.*

CHAPTER 03 | ICT Skills-II

Concepts and applications

Hardware, software and the operating system

Hardware is the physical equipment: processor, memory, storage, keyboard, monitor and other devices. Software is a set of instructions. System software supports the computer; application software serves a user task. An operating system manages resources, files, running programs and interaction with the user. A word processor does not replace these operating-system responsibilities.

Using a desktop environment

A graphical interface provides windows, icons, menus and a pointer. A click usually selects an item; a double-click may open it; dragging can move or select; hovering positions the pointer without a click. Exact actions depend on the application. Use a normal shutdown command so open work can be saved safely.

Files, folders and extensions

A file stores content; a folder organises files and subfolders. A path describes a location. Extensions such as .txt, .pdf, .jpg and .xlsx commonly indicate formats. Renaming an extension does not convert the underlying content. Use meaningful filenames and avoid several unrelated files all called "final".

Copy, move, rename and delete

Copy creates another instance; move changes location; rename changes the name. Delete removes an item from its location, sometimes first placing it in a recycle area. Check the destination and permissions before changing a file. A backup is a separately recoverable copy; a shortcut is only a reference and is not a backup.

Computer care

Keep equipment clean and ventilated, avoid food and liquids nearby, handle cables safely and use suitable power protection. Shut down before appropriate physical cleaning and follow the equipment instructions. Keep software updated from trusted sources. Organise storage and retain backups instead of deleting unfamiliar system files at random.

Unit 1

ICT Skills

ICT skills help us to use computers and digital devices effectively. These skills include understanding operating systems, managing files and folders, and using common tools for study, work and communication.

Types of Operating Systems

An operating system (OS) is system software that manages computer hardware, runs applications and provides a user interface. Major types of operating systems are:

1 Single-user Operating System

Allows one user to use the computer at a time.



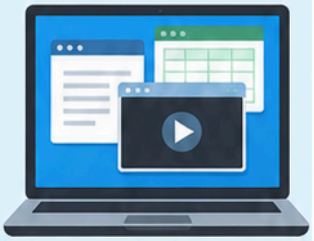
2 Multi-user Operating System

Allows multiple users to use the computer at the same time.



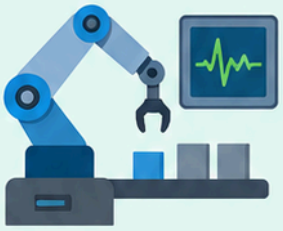
3 Multitasking Operating System

Allows multiple programs to run at the same time.



4 Real-time Operating System

Responds to events in real time, usually used in devices that require quick and accurate control.



File Management

File management helps us to organise, store, find and protect digital data on a computer or other devices.



File

A file stores data such as text, images, videos or programs.



Folder

A folder is used to organise and keep related files together.



Backup

A backup is a copy of important files to keep them safe from loss or damage.



Cloud Storage

Cloud storage saves files on the internet so they can be accessed from anywhere.

Malware types

A virus attaches to a host file or program and spreads when that host runs. A worm can spread across systems without relying on attachment to the same kind of host file. A Trojan disguises harmful behaviour as something apparently useful. Spyware secretly collects information; ransomware attempts to deny access to data and demand payment. Not every unwanted message is itself a virus.

Phishing and account protection

Phishing tries to persuade a person to disclose information or take a harmful action through a deceptive message or site. Verify unexpected requests independently. Use strong, unique passwords, appropriate multifactor authentication and device locking. Do not share OTPs or passwords. Security software is useful but does not make unsafe downloads harmless.

Backup and recovery

Save work regularly and keep a separate, protected backup that can actually be restored. Restrict access according to need. Test a recovery process before an emergency. If a device behaves suspiciously, avoid entering further sensitive information and seek authorised help rather than forwarding suspicious files to others.

Document, spreadsheet and presentation awareness

A word processor creates formatted documents; a spreadsheet arranges data into cells and performs calculations using formulas; presentation software supports a structured visual explanation. Choose the tool for the task. These are revision concepts only: no programming or practical file is included in this book.

CHAPTER 03 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. An unexpected email attachment asks a user to enable unknown macros. What is the safest first response? [1]

(a) Verify the sender and avoid enabling the macros.

(b) Enable them to view the file faster.

(c) Disable antivirus protection.

(d) Forward the attachment to everyone.

Answer:

(a) Verify the sender and avoid enabling the macros. Unexpected active content should not be trusted automatically.

Q2. Which protection helps when a password is stolen? [1]

(a) Renaming the browser

(b) A public password hint

(c) Multifactor authentication

(d) A shorter username

Answer:

(c) Multifactor authentication A second independent factor adds an additional authentication requirement.

Q3. Why should an important file be backed up to a separate reliable location? [1]

(a) To prevent all phishing attacks.

(b) To remove the need for access control.

(c) To guarantee the file is correct.

(d) To recover it after accidental loss or device failure.

Answer:

(d) To recover it after accidental loss or device failure. A backup provides a recoverable copy; it is not a universal security solution.

Q4. Which action changes a file's name without intentionally moving its contents? [1]

(a) Format the drive

(b) Copy

(c) Rename

(d) Compress the folder

Answer:

(c) Rename Renaming changes the file identifier shown to users.

Q5. A program looks useful but secretly performs harmful actions when run. It is a: [1]

(a) Backup utility by definition

(b) Trojan

(c) Spreadsheet formula

(d) Text extension

Answer:

(b) Trojan A Trojan disguises malicious behaviour as legitimate software.

Q6. What does an operating system primarily manage? [1]

(a) Only the appearance of a website

(b) Hardware resources and program execution

(c) Only a user's marks

(d) Only printed documents

Answer:

(b) Hardware resources and program execution The operating system coordinates resources and application processes.

Q7. Which practice is safest on a shared computer? [1]

(a) Share passwords with the next user.

(b) Disable the screen lock.

(c) Sign out after work and do not save personal passwords.

(d) Leave the session unlocked.

Answer:

(c) Sign out after work and do not save personal passwords. Ending the session reduces unauthorised access.

Q8. A folder hierarchy is mainly useful for: [1]

(a) Eliminating malware automatically

(b) Organising related files for retrieval

(c) Making every file executable

(d) Increasing the physical size of a disk

Answer:

(b) Organising related files for retrieval Folders group files in a structured way.

Q9. Which extension conventionally identifies a plain text file? [1]

(a) .txt

(b) .exe

(c) .jpg

(d) .mp3

Answer:

(a) .txt A plain text file commonly uses the .txt extension.

Q10. Which maintenance action is appropriate before physically cleaning a desktop computer? [1]

(a) Pour water onto the keyboard.

(b) Keep it running while removing components.

(c) Shut it down and disconnect power safely.

(d) Open all unknown software first.

Answer:

(c) Shut it down and disconnect power safely. Power should be disconnected before physical cleaning.

Short-answer questions - 2 marks each

S1. State two ways to reduce the risk of unauthorised access to an account. [2]

Answer:

(1) Use a strong unique password and keep it private from other people. [1]

(2) Enable multifactor authentication and sign out properly when using shared computers. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Explain the purpose of a backup and one condition for a useful backup. [2]

Answer:

(1) A backup supports recovery after accidental loss or device failure. [1]

(2) Keep an up-to-date recoverable copy separate from the original and check it periodically. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Distinguish copying a file from moving a file. [2]

Answer:

(1) Copying creates another instance while leaving the original in place. [1]

(2) Moving transfers the file to a new location and removes it from the old location. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. What is a Trojan and why can it mislead a user? [2]

Answer:

(1) A Trojan is malicious software disguised as a legitimate program. [1]

(2) Its useful appearance can persuade a user to run it and permit harmful actions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. Give two safe practices for using a shared computer. [2]

Answer:

(1) Sign out of personal accounts and lock or end the session. [1]

(2) Do not save personal passwords or leave confidential files accessible to others. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. State two functions of an operating system. [2]

Answer:

(1) It manages hardware resources such as memory and storage. [1]

(2) It provides services and an interface for running applications and handling files. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain how a Trojan can threaten a computer. [2]

Answer:

(1) A Trojan disguises itself as useful or legitimate software. [1]

(2) When executed, it can perform harmful actions, such as stealing information or permitting unauthorised access. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. Distinguish copying a file from moving a file. [2]

Answer:

(1) Copying creates another instance while retaining the file at its original location. [1]

(2) Moving transfers the file to another location and removes the original location entry. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. State two precautions when installing software. [2]

Answer:

(1) Obtain software from a trusted source and verify that it is appropriate. [1]

(2) Review requested permissions and keep security protection active during installation and later use. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain the use of a file extension with an example. [2]

Answer:

(1) A file extension helps identify the file type and suitable applications. [1]

(2) For example, .txt commonly identifies a plain-text file that a text editor can open. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Classes of operating systems

Operating systems can be classified by different criteria: number of users, number of tasks or timing requirements. The categories therefore **overlap**. A system may support multiple users and multitasking at the same time; these are not mutually exclusive properties.

Class	Meaning and distinguishing feature
Single-user	Designed to serve one user at a time. It may still run more than one task.
Multi-user	Allows several users to access and use computing resources with separate accounts or sessions.
Single-tasking	Permits one main task at a time in the simplified classification.
Multitasking	Manages several tasks so they can make progress during the same working period.
Real-time	Must respond within specified timing constraints; predictable response time is central.

A person listening to audio while editing a document illustrates **multitasking**, not proof of multiple simultaneous users. A shared system with separate active user sessions illustrates **multi-user** operation. A time-critical controller is judged by whether required deadlines are met, not simply whether its processor is fast.

Desktop components and interfaces

Component	Purpose
Desktop	The main graphical working area after sign-in.
Icon	A small visual symbol for an application, file, folder or command.
Menu	A displayed list of available commands or choices.
Taskbar / dock	Provides access to applications and often indicates running tasks; exact features differ.
Window	An application or document area that can often be moved, resized or switched.
Pointer	An on-screen indicator controlled by a pointing device.

A graphical user interface uses visual elements. A command-line interface uses typed commands. Both can provide access to operating-system functions; an icon is not the same thing as the data file it represents.

Directory and file-system organisation

A file system organises how named files and directories are stored and located. A directory tree has a starting/root directory, folders and subfolders. A **path** identifies a location through that hierarchy. The folder organisation is different from a file format: PDF is a document format, not a folder tree.

Example tree:

```
Study
  AI
    Notes
      Evaluation.pdf
```

```
  Tests
    Paper01.pdf
```

"Study/AI/Notes/Evaluation.pdf" identifies the file within this example hierarchy.

An **absolute path** starts from the system or volume root; a **relative path** is interpreted from a current folder. These are useful elaborations of directory organisation. File extensions help identify formats, but renaming .jpg to .pdf does not convert an image into a PDF document.

Operations and recovery

Operation	Expected effect
Create and save	Make a new file/folder and choose a meaningful name and location.
Copy	Place an additional copy at a destination; the source remains.
Move	Change the location; check that the intended file reached the destination.
Rename	Change the name, not automatically the content or format.
Delete	Remove the item from its current location; some systems temporarily retain it in a recycle area.
Restore	Return a retained deleted item from the recycle area, normally to its previous location.

A recovery outline is: open the recycle area, identify the correct item, use Restore and verify its location. A permanently deleted or non-retained item may not be recoverable there. A recycle area is

not a substitute for backup. A shortcut points to another item; it is not a safe second copy of the original.

A reasoned maintenance schedule

Maintenance protects availability, data and equipment life. Separate **physical care** from **software/data care**. For cleaning, shut down and disconnect power where appropriate, follow the manufacturer instructions, avoid liquid inside devices and keep vents unobstructed.

Illustrative interval	Tasks and purpose
During daily use	Save work, keep food/liquids away and shut down correctly.
Regularly	Back up important files, install trusted updates and run appropriate security scans.
At planned review	Check that backups restore, review storage and access, and arrange safe cleaning.
When a warning appears	Stop risky activity, preserve important evidence and obtain authorised help.

This is an example schedule, not a universal rule about exact days. **Spam** is unwanted messaging; it is not automatically a virus, though links and attachments may be risky. **Temporary files** support tasks and may later become unnecessary. Use an appropriate cleanup method; do not randomly delete system files. Security software reduces risk but does not make unknown downloads safe.

Additional exam-style practice

Digital Safety and Cybersecurity

Everyday precautions for secure use of computers, accounts and online services.

Risk	Warning sign	Safer action
Phishing	Unexpected message asks for passwords or OTP	Verify the sender using an independent trusted channel
Malware	Unknown download or attachment	Use trusted software and keep security updates on
Weak passwords	Same password reused across accounts	Use strong unique passwords and MFA
Data loss	Only one copy of important files	Keep secure backups and test restoration
Privacy exposure	Overshared personal or school data	Limit sharing and check access permissions

Safe workflow

Stop → Check source → Verify independently → Act safely → Report suspicious activity

C3-Q1. A computer runs a word processor and an audio player during one user session. The demonstrated capability is: [1]

(a) Multi-user access only
(b) Multitasking
(c) Permanent file deletion
(d) Image segmentation

C3-Q2. Which action best verifies the usefulness of a backup? [1]

(a) Rename its extension
(b) Keep only its shortcut
(c) Test restoring an appropriate file
(d) Delete the original before checking

C3-Q3. Distinguish multi-user and multitasking operating systems. [2]

Model answer: Multi-user systems support more than one user session. Multitasking systems manage several tasks. One operating system can support both capabilities.

Suggested practice marking: Users distinction 1; tasks distinction 1; do not penalise a correct overlap note.

C3-Q4. State four steps for restoring a file that is still available in the recycle area. [2]

Model answer: Open the recycle area; identify the required file; select Restore; verify that the file is available at the restored location.

Suggested practice marking: Four suitable steps, 0.5 each.

C3-Q5. Why is changing a filename extension not a reliable method of converting a file? [2]

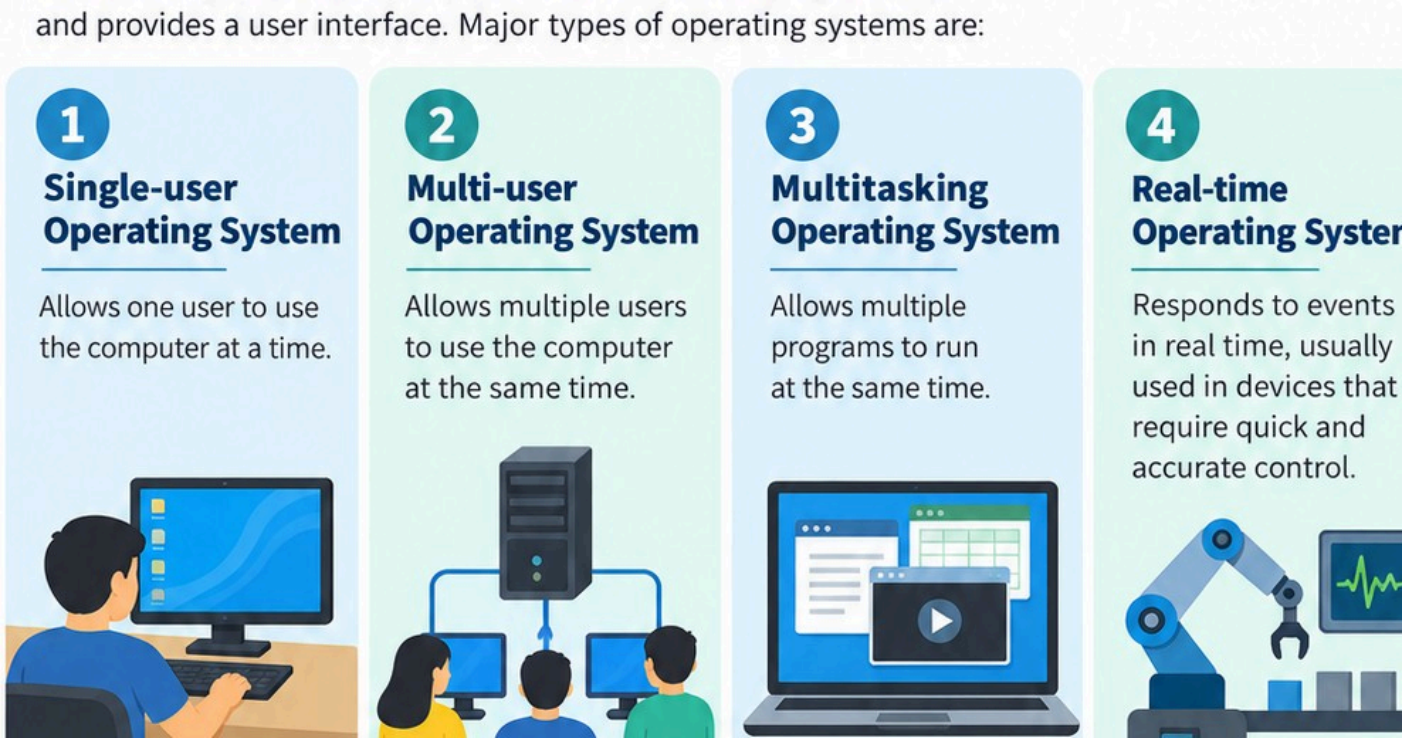
Model answer: The extension is part of the filename. Changing it does not necessarily change the underlying encoding or structure required by the destination format.

Suggested practice marking: Name versus underlying content 1; requirement for actual format conversion 1.

MCQ answers and reasons

C3-Q1: (b). Running several tasks demonstrates multitasking.

C3-Q2: (c). A backup must be recoverable, not merely present.



IMPORTANT QUESTIONS WITH ANSWERS

ICT Skills-II

Q1. What is an operating system? Give two functions. [3]

Answer: An operating system is system software that manages computer resources and provides services to applications. Functions include managing files and coordinating memory, devices and processes.

Suggested marking: *1 definition + 1 per function.*

Q2. Explain four types of operating systems. [4]

Answer: Single-user OS supports one user at a time in its intended mode. Multi-user OS allows multiple users to share resources. Real-time OS responds within specified time limits. Mobile OS is designed for smartphones/tablets and touch interfaces.

Suggested marking: *1 each for correctly explained types.*

Q3. Differentiate system software and application software. [2]

Answer: System software manages hardware and provides a platform, such as an operating system. Application software performs user tasks, such as a word processor.

Suggested marking: *1 per correct description/example.*

Q4. Explain four safe cyber-security practices. [4]

Answer: Use strong unique passwords and multifactor authentication; install trusted updates; avoid suspicious links and attachments; regularly back up important data.

Suggested marking: *1 each.*

Q5. What is the difference between a file and a folder? [2]

Answer: A file stores data such as a document or image. A folder is a container used to organise files and subfolders.

Suggested marking: *1 each.*

Q6. What are malware and phishing? [2]

Answer: Malware is software designed to harm, disrupt or gain unauthorised access. Phishing is deceptive communication intended to steal credentials or information.

Suggested marking: *1 each.*

CHAPTER 04 | Entrepreneurial Skills-II

Concepts and applications

Entrepreneurship and value creation

Entrepreneurship involves recognising an opportunity, organising resources and taking responsibility for a venture that offers value. Value may come from a useful product, a better service or a more effective way to meet a need. An entrepreneur takes calculated risk, not risk without investigation.

Entrepreneur and employee

An entrepreneur usually makes ownership-level decisions and bears the venture's business risk. An employee performs agreed responsibilities for compensation. Both may be creative, responsible and hardworking. The distinction is about role and risk, not a claim that one group always works harder or earns more.

Functions of an entrepreneur

Important functions include identifying opportunities, planning, arranging resources, making decisions, organising work, managing risk and reviewing performance. Innovation can be an improvement in a process, not only a completely new invention. Delegation shares tasks but does not remove the owner's accountability.

Qualities developed through practice

Initiative helps a person act without waiting for repeated instructions. Perseverance supports sustained effort after setbacks. Creativity suggests alternatives. Communication and leadership help organise people. Responsibility, adaptability and evidence-based decision-making make a venture more reliable. Match the quality to the behaviour described in a case.

From a need to a business idea

Identify a real customer problem, describe the intended customer, compare existing alternatives and test whether people find the solution useful. Estimate costs, likely revenue, resources and practical constraints. A small trial can expose a mistaken assumption before a larger investment. Profit is revenue minus relevant costs, not the total money received.

Risk and learning from failure

A calculated risk has been examined: likely benefits, possible losses and safeguards are considered. Planning does not eliminate uncertainty. If a trial fails, investigate evidence, improve the product or process and avoid repeating an unsupported assumption. Blaming customers or staff without investigation is not responsible risk management.

Common myths

Entrepreneurial ability is not limited to a fixed set of people born with it. Not every venture requires large starting capital or a unique invention. Immediate high profit is not guaranteed. Skills can develop through experience, training and feedback, and a business idea still needs testing.

Role in society and career choices

Enterprises can provide goods and services, generate employment, use local resources and introduce improvements. Social and environmental effects also matter. Entrepreneurship is one career option, not the only respectable option. A responsible venture balances financial survival with fair treatment, legal duties and the needs of its community.

CHAPTER 04 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A repair service starts after its owner notices that customers cannot find local repairs. This shows: [1]

(a) Elimination of business risk

(b) Guaranteed profit

(c) Opportunity identification

(d) Avoidance of customers

Answer:

(c) Opportunity identification An unmet need can become a business opportunity.

Q2. An entrepreneur tests a small batch before investing in large production. This is an example of: [1]

(a) Calculated risk

(b) Price fixing

(c) Ignoring evidence

(d) No risk whatsoever

Answer:

(a) Calculated risk A small trial provides evidence before a larger commitment.

Q3. Which statement about entrepreneurship is correct? [1]

(a) Profit is guaranteed after registration.

(b) An entrepreneur never needs a team.

(c) Skills can be developed through learning and experience.

(d) Only inherited wealth can create an entrepreneur.

Answer:

(c) Skills can be developed through learning and experience. Entrepreneurial abilities are not limited to people born with wealth.

Q4. Revenue from a batch is Rs 2,400 and total cost is Rs 1,800. Profit is: [1]

(a) Rs 1,800

(b) Rs 2,400

(c) Rs 600

(d) Rs 4,200

Answer:

(c) Rs 600 Profit is revenue minus total cost.

Q5. A supplier delivers late. Which response best reflects entrepreneurial responsibility? [1]

(a) Blame customers.

(b) Stop tracking orders.

(c) Communicate honestly and arrange a practical alternative.

(d) Hide the delay indefinitely.

Answer:

(c) Communicate honestly and arrange a practical alternative. Accountability requires a response to the business problem.

Q6. Which is primarily an entrepreneurial function? [1]

(a) Avoiding every decision

(b) Waiting for all needs to disappear

(c) Guaranteeing that no competitor exists

(d) Organising resources to provide a useful product or service

Answer:

(d) Organising resources to provide a useful product or service Entrepreneurs combine resources and take decisions to create value.

Q7. Customer feedback shows demand for a smaller package. Revising the package shows: [1]

(a) Adaptability

(b) Unemployment

(c) Denial

(d) Monopoly by definition

Answer:

(a) Adaptability The business adapts its offering to evidence about customer needs.

Q8. Which distinction is generally valid? [1]

(a) Employees never exercise initiative.

(b) An entrepreneur bears business uncertainty; an employee works under agreed employment terms.

(c) Entrepreneurs never receive income.

(d) Every employee owns the employer's business.

Answer:

(b) An entrepreneur bears business uncertainty; an employee works under agreed employment terms. Ownership and business uncertainty differ from an employment role.

Q9. Why conduct market research before launching a service? [1]

(a) To remove the need for planning

(b) To copy a competitor's protected branding

(c) To ensure every person will buy it

(d) To understand needs, demand and competition

Answer:

(d) To understand needs, demand and competition Research informs decisions; it cannot guarantee sales.

Q10. A failed pilot leads an entrepreneur to improve the design and test again. This shows: [1]

(a) Complete risk avoidance

(b) Persistence with learning

(c) Refusal to use evidence

(d) A guarantee of future profit

Answer:

(b) Persistence with learning Persistence is most useful when accompanied by learning.

Short-answer questions - 2 marks each

S1. Explain calculated risk with a suitable business example. [2]

Answer:

(1) Calculated risk means considering likely costs and benefits before acting under uncertainty. [1]

(2) For example, test a small batch before financing large-scale production. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. State two ways entrepreneurship can benefit a community. [2]

Answer:

(1) Entrepreneurship supplies useful products or services that respond to unmet community needs. [1]

(2) It can create paid work and support local suppliers through business activity. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Distinguish an entrepreneur from an employee. [2]

Answer:

(1) An entrepreneur organises a venture and bears its business uncertainty. [1]

(2) An employee works under agreed terms and receives compensation for assigned responsibilities. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Explain opportunity identification with an example. [2]

Answer:

(1) Opportunity identification recognises an unmet need that a venture could address. [1]

(2) For example, start a repair service where customers lack nearby repairs. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. State two reasons to collect customer feedback. [2]

Answer:

(1) Feedback identifies unmet needs and problems in the current product. [1]

(2) It helps test improvements rather than relying only on the owner's assumptions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. A batch earns Rs 3,600 revenue and costs Rs 2,850 in total. Find profit and distinguish it from revenue. [2]

Answer:

(1) Profit = revenue - total cost = 3,600 - 2,850 = Rs 750. [1]

(2) Revenue is total sales income before deducting costs. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain innovation with a simple business example. [2]

Answer:

(1) Innovation applies a useful new or improved idea to meet a need. [1]

(2) For example, a repair business could introduce convenient scheduled collection for customers. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. State two management functions of an entrepreneur. [2]

Answer:

(1) An entrepreneur coordinates resources and people to deliver the product or service. [1]

(2) The entrepreneur also makes decisions and monitors costs, quality and customer requirements. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Explain why entrepreneurship does not guarantee a profit. [2]

Answer:

(1) Sales, costs and customer preferences are uncertain and may change. [1]

(2) An entrepreneur can reduce risk through planning and testing, but cannot ensure every venture earns profit. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain how market research supports a business decision. [2]

Answer:

(1) Market research gathers evidence about customers, alternatives and unmet needs. [1]

(2) It helps a business assess demand before committing resources to a proposed product or service. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

CHAPTER 04 | Detailed topic lessons

Entrepreneurial Skills-II: functions, myths and career decisions

One venture, several entrepreneurial functions

Consider a fictional bicycle-repair service in an area where commuters wait a long time for repairs. The opportunity is not simply "owning a shop"; it is meeting an unmet service need. Before starting, the entrepreneur checks the type of demand, available skills and resources, costs and possible risks.

Function	Application to the repair service
Identify an opportunity	Ask customers which repairs and opening times they need.
Plan and organise	Arrange tools, workspace, supplies, work sequence and records.
Make decisions	Choose the service range and an affordable operating plan.
Manage calculated risk	Start at a manageable scale and prepare for uncertain demand.
Innovate and improve	Offer a clear booking process or reliable repair-status information.
Coordinate and review	Assign work, check quality, record feedback and correct problems.

Qualities are not the same as functions

A **quality** is a characteristic such as perseverance, initiative, self-confidence, creativity, responsibility or willingness to learn. A **function** is work performed, such as planning or organising. Creativity can help the entrepreneur design a better service; planning is the action that arranges how to deliver it.

Risk-taking should be **calculated**: investigate likely outcomes and limit avoidable loss. Confidence does not mean ignoring evidence; perseverance does not mean repeating a failing method without review. Honesty in prices and promised work supports trust.

Entrepreneurship and society

A useful venture can provide needed goods or services, create employment, use local resources responsibly and introduce improvements. For example, a repair service extends product life and may train assistants. These benefits depend on what the business actually does; merely opening a business does not guarantee social value.

Common myths and their correction

Myth	Reasoned correction
Entrepreneurs are only born.	Skills such as planning, communication and record keeping can be developed.
Every venture needs a huge investment.	Resource needs differ. Some services can begin on a modest scale.
Entrepreneurs never face failure.	Uncertainty exists; learning and adapting are part of responsible enterprise.
The owner has no responsibilities to anyone.	Customers, workers, suppliers and the community are affected by the venture.
Only a completely new invention counts.	Improving how an existing need is met can also create value.

Entrepreneurship as a career option

Potential advantages include initiative, independence in decisions, learning and an opportunity to create value. Challenges include uncertain income, responsibility for losses, varied work and the need to arrange resources. **Wage employment** normally involves agreed work for compensation; **self-employment** involves organising one's own work and bearing its business risk. Neither option guarantees success or failure.

A reasoned career answer considers interests, skills, resources, responsibilities and market need. Do not write that an entrepreneur always earns more than an employee or never works under pressure.

Additional exam-style practice

C4-Q1. Which item is an entrepreneurial function rather than a personal quality? [1]

(a) Perseverance
(b) Self-confidence
(c) Organising resources
(d) Curiosity

C4-Q2. Which action best represents calculated risk? [1]

(a) Borrow without checking demand
(b) Study demand and begin at a manageable scale
(c) Ignore all customer feedback
(d) Assume there can be no loss

C4-Q3. Explain two ways a useful local venture can benefit society. [2]

Model answer: It can meet an unmet need through a useful product or service. It can also create work or training opportunities for other people.

Suggested practice marking: Two distinct explained benefits, 1 each.

C4-Q4. Correct the myth that entrepreneurs are born and cannot be developed. [2]

Model answer: Personal traits differ, but entrepreneurial abilities can be developed through learning and experience. Planning, communication and responsible decision-making can improve with practice.

Suggested practice marking: Reject fixed inborn-only claim with explanation 1; suitable learnable skills/examples 1.

C4-Q5. Give one advantage and one challenge of entrepreneurship as a career. [2]

Model answer: An advantage is greater initiative in business decisions. A challenge is bearing uncertainty and possible losses when income or demand changes.

Suggested practice marking: Relevant advantage 1; relevant challenge 1.

MCQ answers and reasons

C4-Q1: (c). Organising resources is an activity, not a trait.

C4-Q2: (b). It uses evidence and manages uncertainty.

Entrepreneurial Skills

Turn a useful idea into a small, responsible enterprise.

Step	Key question	School-friendly example
Identify need	What problem needs solving?	Students need affordable reusable stationery
Know customers	Who will benefit or buy?	Students, teachers and parents
Plan resources	What materials, time and money are needed?	Suppliers, cost, delivery schedule
Budget	Can selling price cover all costs?	Price based on costs and sustainable margin
Market ethically	How will people learn about it?	Accurate posters and clear product claims
Review feedback	What should improve?	Fix quality and delivery concerns

A business plan typically includes the idea, customer needs, resources, expected costs, price, promotion and risks.

IMPORTANT QUESTIONS WITH ANSWERS

Entrepreneurial Skills-II

Q1. Who is an entrepreneur? Mention three characteristics. [4]

Answer: An entrepreneur identifies an opportunity and organises resources to create a product or service while taking calculated risks. Key traits include creativity, perseverance and responsible decision-making.

Suggested marking: *1 definition + 1 each of three traits.*

Q2. Distinguish entrepreneur and wage employee. [2]

Answer: An entrepreneur starts or operates a venture and bears business risks. A wage employee works for an employer in exchange for agreed remuneration.

Suggested marking: *1 each.*

Q3. Describe four functions of an entrepreneur. [4]

Answer: Identifies customer needs; plans products and operations; organises people, money and materials; manages risks and improves offerings through innovation.

Suggested marking: *1 each.*

Q4. What is market research and why is it useful? [2]

Answer: Market research collects and analyses information about customers, competitors and demand. It helps test whether a business idea solves a real problem.

Suggested marking: *1 definition + 1 use.*

Q5. Explain four myths about entrepreneurship. [4]

Answer: Myth: entrepreneurs are born, not made-skills can be learned. Myth: huge capital is always needed-some ventures start small. Myth: success requires reckless risk-good entrepreneurs manage risk. Myth: entrepreneurs work alone-teams and networks matter.

Suggested marking: *1 each correctly explained myth.*

CHAPTER 05 | Green Skills-II

Concepts and applications

Sustainability

Sustainable development considers both present needs and the ability of people in the future to meet theirs. It requires attention to environmental quality, social well-being and economic viability. An activity is not sustainable merely because it produces income today if it causes serious long-term damage.

Why development choices matter

Water, energy, land and materials are limited or take time to renew. Pollution, waste and unequal access can reduce quality of life. Decisions should consider the full effect of production and consumption, including resources used, waste produced and people affected. A one-time clean-up is useful but does not replace reducing the source of pollution.

Reduce, reuse and recycle

Reduce avoids unnecessary consumption and waste. Reuse keeps a product or component useful for longer. Recycle processes suitable discarded materials into usable material. Repair and careful purchasing can support these actions. Mixing hazardous waste with ordinary recyclable material can make safe recovery difficult.

Actions at home and in learning spaces

Repair leaks, switch off unused equipment, avoid unnecessary single-use items and separate waste correctly. Prefer efficient transport where feasible. Measure use before and after an intervention so that savings are supported by evidence. A claim such as "green" needs a concrete explanation rather than only a label.

Green economy and green work

A green economy aims to improve well-being while reducing environmental harm and resource pressure. Relevant work includes renewable-energy installation, repair, resource-efficient production, ecosystem conservation and safe waste recovery. A job title alone is not enough: the actual methods and outcomes must reduce harm.

Sustainable Development Goals

The UN framework contains 17 interconnected goals. For common classroom cases, relate safe drinking water to Goal 6, clean energy to Goal 7, responsible consumption and production to Goal 12, and climate action to Goal 13. Choose the goal most directly supported by the facts instead of treating every environmental action as the same goal.

Education and shared responsibility

Education builds knowledge, skills and habits for judging resource choices. Government, industry, communities and individuals have different but connected responsibilities. Awareness should lead to a practical, measurable action, such as reducing food waste or tracking a water leak, rather than only a slogan.

A balanced case answer

State the problem, identify the resource or people affected and suggest an action linked to it. Explain how the action can reduce harm while remaining workable. Avoid claiming that recycling has no cost or that every digital activity has no environmental impact.

CHAPTER 05 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A reusable bottle is used many times instead of buying a disposable one each day. This mainly illustrates: [1]

(a) Open burning

(b) Reuse

(c) Uncontrolled extraction

(d) Single-use consumption

Answer:

(b) Reuse Repeated use can reduce the demand for disposable products.

Q2. Which action best supports responsible electronic-waste handling? [1]

(a) Break screens in an open field.

(b) Dump batteries near a water source.

(c) Burn devices with ordinary waste.

(d) Send obsolete devices through an authorised collection route.

Answer:

(d) Send obsolete devices through an authorised collection route. Specialised collection enables safer handling and material recovery.

Q3. Sustainable development aims to meet present needs while: [1]

(a) Stopping all economic activity

(b) Ignoring environmental costs

(c) Using every available resource immediately

(d) Protecting future generations' ability to meet theirs

Answer:

(d) Protecting future generations' ability to meet theirs Present development should not destroy future options.

Q4. Which measure directly reduces avoidable water loss? [1]

(a) Replace all plants with water-intensive species.

(b) Increase tap flow without need.

(c) Repair leaking pipes.

(d) Ignore meter readings.

Answer:

(c) Repair leaking pipes. Leaks waste water without providing useful service.

Q5. Separating dry recyclable paper from wet food waste mainly helps: [1]

(a) Material recovery

(b) Removing the need to collect waste

(c) Increasing litter

(d) Contaminating recyclable waste

Answer:

(a) Material recovery Separation can keep recoverable materials usable.

Q6. Which action reduces energy consumption without changing the intended task? [1]

(a) Switch off unused lights and devices.

(b) Use more devices for the same task.

(c) Leave all equipment on overnight.

(d) Ignore idle equipment.

Answer:

(a) Switch off unused lights and devices. Avoiding unnecessary operation saves energy.

Q7. A product designed to be repaired and used longer supports: [1]

(a) Resource efficiency

(b) Unlimited extraction

(c) Planned waste generation

(d) Uncontrolled disposal

Answer:

(a) Resource efficiency Extending useful life can reduce replacement demand.

Q8. Which statement about renewable resources is most accurate? [1]

(a) They can never be depleted locally.

(b) They can replenish naturally, but still require responsible use.

(c) They always have zero environmental impact.

(d) They are identical to all fossil fuels.

Answer:

(b) They can replenish naturally, but still require responsible use. Renewability does not remove the need for careful use.

Q9. A community reduces waste, sorts recyclables and composts suitable organic waste. These actions support: [1]

(a) A ban on every form of production

(b) Maximum landfill expansion

(c) Responsible consumption and production

(d) Disposal without segregation

Answer:

(c) Responsible consumption and production The actions reduce waste and recover useful material.

Q10. Which decision considers both environmental and economic effects? [1]

(a) Choose only by colour.

(b) Treat waste disposal as irrelevant.

(c) Ignore repairability.

(d) Compare operating cost, durability and resource use before purchase.

Answer:

(d) Compare operating cost, durability and resource use before purchase. Sustainable choices consider several consequences.

Short-answer questions - 2 marks each

S1. Explain sustainable development and give one relevant action. [2]

Answer:

(1) Sustainable development meets present needs without undermining future generations' ability to meet theirs. [1]

(2) Repairing reusable products helps conserve materials and reduce unnecessary replacement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Distinguish reuse from recycling with an example of each. [2]

Answer:

(1) Reuse means using an item again, such as refilling a bottle. [1]

(2) Recycling processes waste into usable material, such as recovering paper fibre. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

S3. Give two suitable measures for reducing electronic waste. [2]

Answer:

(1) Repair or reuse suitable working devices to extend their useful life. [1]

(2) Send unusable electronics through authorised collection or recycling channels instead of mixing them with ordinary waste. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Explain how education can support sustainable behaviour. [2]

Answer:

(1) Education helps people understand resource use and environmental consequences. [1]

(2) It builds the knowledge and skills needed to adopt and evaluate more sustainable practices. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. State two ways a community can reduce unnecessary water use. [2]

Answer:

(1) Detect and repair leaking pipes and taps before water is wasted. [1]

(2) Monitor consumption and adopt water-efficient routines to identify and reduce other avoidable losses. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. State two factors, apart from purchase price, for judging the sustainability of a device. [2]

Answer:

(1) Consider the energy and resources a device consumes throughout everyday use. [1]

(2) Also assess its durability, repairability and options for responsible collection or recycling at end of life. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain responsible consumption with one example. [2]

Answer:

(1) Responsible consumption uses resources carefully and avoids unnecessary waste. [1]

(2) For example, buying a durable, repairable product can reduce repeated replacement and disposal. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. State two benefits of repairing a usable product. [2]

Answer:

(1) Repair can extend the product's useful life and delay disposal. [1]

(2) It can also reduce demand for replacement materials and the resources needed to manufacture a new product. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Define renewable energy and give one example. [2]

Answer:

(1) Renewable energy comes from sources replenished naturally on useful human timescales. [1]

(2) For example, solar panels convert incoming sunlight into electricity for homes or other buildings. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain why sustainability involves more than environmental protection. [2]

Answer:

(1) Sustainability also concerns social wellbeing and economically workable choices. [1]

(2) For example, an environmental measure should consider fair access and the livelihoods of affected people. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

CHAPTER 05 | Detailed topic lessons

Green Skills-II: sustainable choices and reasoned solutions

Three connected dimensions

Sustainable development considers present needs without undermining the ability of later generations to meet theirs. Its **environmental** dimension concerns ecosystems and resources; the **social** dimension concerns people, fairness and basic needs; the **economic** dimension concerns viable livelihoods and responsible production. A solution should examine these together rather than protect one by ignoring the others.

From a local problem to an effective response

Problem and consequence	Practical response and why it helps
Water leaks waste treated water.	Repair leaks and monitor use; stored water cannot compensate for continuing avoidable loss.
Mixed waste makes recovery difficult.	Separate waste at source and use appropriate authorised collection streams.
Unnecessary energy use increases demand.	Switch off unneeded equipment and choose efficient use before assuming more generation is the only answer.
Loss of trees harms habitats and soil.	Protect suitable existing vegetation and restore with locally appropriate species.
Unequal access to basic resources harms people.	Include affected groups when deciding how services and benefits are distributed.

A good answer connects **problem, action and reason**. "Save nature" is too vague. "Repair leaking taps and record weekly water use to check whether wastage decreases" is a testable action. Recycling helps but cannot justify unlimited consumption; avoid, reduce, reuse and repair where suitable.

Environmental citizenship and indigenous knowledge

Environmental citizenship means recognising that personal and collective choices affect the environment and other people. Follow responsible disposal practices, conserve resources, participate in local decisions and report avoidable damage through appropriate channels. Responsibility is not limited to a government or one profession.

Indigenous and local knowledge may offer experience of water storage, crops, local materials or biodiversity. Respect the people and cultural context that maintain it. Assess suitability for the present problem rather than assuming every traditional practice is automatically safe or every new technology automatically better.

Organic farming and conserving resources

Organic farming emphasises ecological processes, soil fertility and appropriate biological methods under the applicable organic system. Compost, crop rotation and suitable biological pest management are illustrative practices. Do not define it as farming without science or claim that all natural substances are harmless. A short exam answer can state the approach and explain benefits such as improved soil management and reduced reliance on specified synthetic inputs.

Sustainable-development goals: a revision connection

The existing chapter introduces sustainable-development goals and examples. For exam application, connect the problem to the goal rather than merely memorising a number: clean-water actions concern water and sanitation; efficient energy use concerns energy; repair and reuse concern responsible consumption; and fair learning opportunities concern education. The current theory syllabus does not require a separate practical submission for this book.

Additional exam-style practice

C5-Q1. Which answer considers environmental, social and economic needs together? [1]

(a) Close every service regardless of people's needs
(b) Provide a viable service while limiting waste and ensuring fair access
(c) Increase output even if water is exhausted
(d) Ignore future resource needs

C5-Q2. Which is the strongest measurable response to excessive water use? [1]

(a) Write "save water" and take no further action
(b) Ignore small leaks
(c) Repair leaks and compare recorded use
(d) Use more disposable bottles

C5-Q3. Define sustainable development and state why future needs matter. [2]

Model answer: Sustainable development meets present needs without undermining future generations' ability to meet theirs. Depleting resources now can remove essential options for people later.

Suggested practice marking: Meaning 1; explanation of future needs 1.

C5-Q4. Explain one environmental problem and one related solution in a community. [2]

Model answer: Mixed household waste makes recovery difficult. Separating recyclable, biodegradable and other waste into appropriate streams supports safer collection and treatment.

Suggested practice marking: Problem 1; linked solution with relevant reason 1.

C5-Q5. State what organic farming emphasises and give two illustrative practices. [2]

Model answer: It emphasises ecological processes and soil fertility. Examples include composting and crop rotation.

Suggested practice marking: Meaning 1; two appropriate practices 0.5 each.

MCQ answers and reasons

C5-Q1: (b). It balances resource use, people and viability.

C5-Q2: (c). The response links an action to evidence.

IMPORTANT QUESTIONS WITH ANSWERS

Green Skills-II

Q1. What is sustainable development? [2]

Answer: Sustainable development meets present needs without undermining future generations' ability to meet their own needs. It balances environmental, social and economic goals.

Suggested marking: *1 definition + 1 balance principle.*

Q2. Differentiate renewable and non-renewable resources. [2]

Answer: Renewable resources replenish naturally over relevant periods, such as sunlight and wind. Non-renewable resources are finite or replenish extremely slowly, such as coal and petroleum.

Suggested marking: *1 each with example.*

Q3. Explain the 3 Rs with examples. [3]

Answer: Reduce: use fewer disposable items. Reuse: refill a water bottle. Recycle: process separated paper into new paper products.

Suggested marking: *1 each.*

Q4. List four steps a school can take to reduce its ecological footprint. [4]

Answer: Conduct an energy audit and switch off unused equipment; fix water leaks; segregate waste and compost organics; promote reusable supplies and responsible transport.

Suggested marking: *1 each.*

Q5. What are green jobs? Give two examples. [3]

Answer: Green jobs contribute to preserving or restoring environmental quality. Examples include solar-panel technician and waste-management specialist.

Suggested marking: *1 definition + 1 each example.*

CHAPTER 06 | Revisiting AI Project Cycle and Ethical Frameworks

Concepts and applications

What makes an AI project purposeful?

Artificial Intelligence concerns systems designed to perform tasks such as recognising patterns, making predictions, using language or selecting actions. An AI project begins with a useful problem, not with the assumption that every problem requires AI. Define what success would look like and whose needs must be considered.

Problem scoping and the 4Ws

Use Who, What, Where and Why to frame the problem: who experiences it, what is happening, where it occurs and why a solution matters. Then write a focused problem statement and a measurable objective. For example, reducing avoidable waiting time is clearer than promising to "make everything smart".

The AI project cycle

Problem scoping defines the task. Data acquisition collects examples. Data exploration checks their quality and patterns. Modelling builds a suitable approach. Evaluation measures results. Deployment puts the evaluated system into use with monitoring. The cycle is iterative: evaluation can lead back to the problem, data or model.

Data acquisition

Acquire data that is relevant, reliable and sufficiently representative of the intended users and conditions. Consider permission, privacy, labelling quality, missing values and collection limits. A very large dataset can still be biased or irrelevant. Document where the data came from and which uses are allowed.

Green Skills and Sustainability

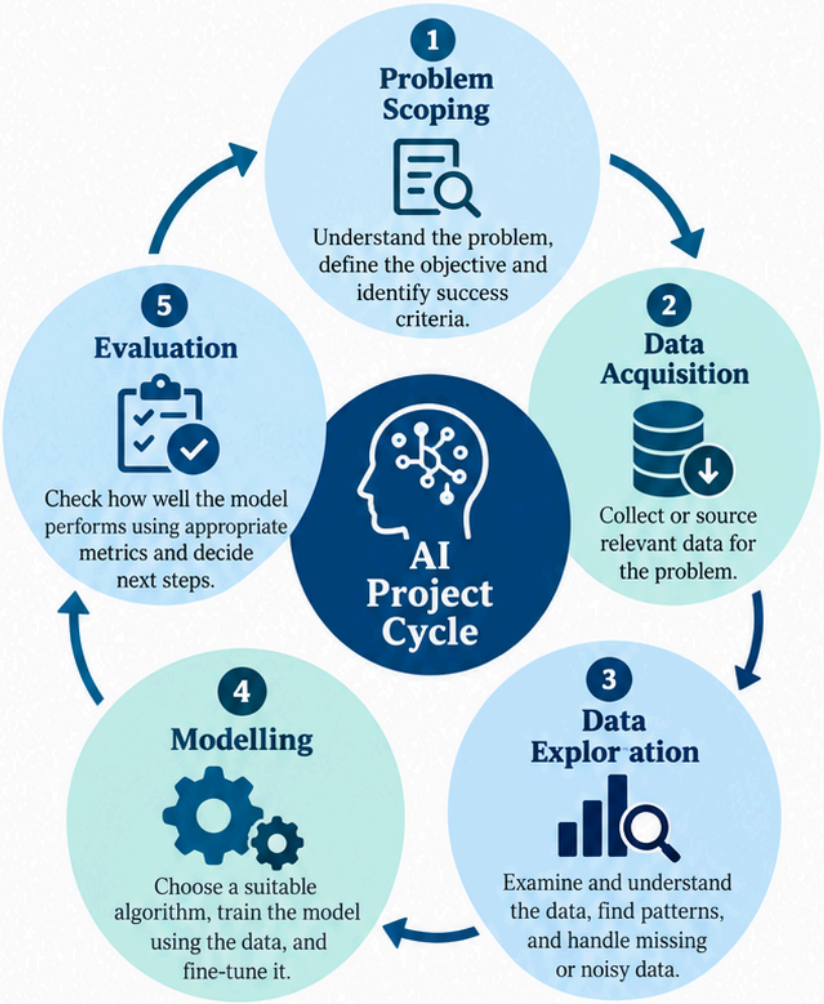
Learn the three Rs, resource conservation and responsible consumption.

Action	Meaning	Practical example
Reduce	Use fewer materials and less energy	Turn off unused lights; avoid unnecessary printing
Reuse	Use a product again when safe	Reuse folders and refill bottles
Recycle	Recover useful material from waste	Segregate paper and recyclable packaging
Conserve	Protect water, energy and other resources	Repair taps and maintain equipment
Responsible disposal	Handle hazardous e-waste safely	Use authorised e-waste collection

Sustainable development meets present needs while considering resources and opportunities for future generations.

AI Project Cycle

The AI Project Cycle is a step-by-step process used to develop an AI solution for a real-world problem. It helps us move from understanding a problem to building and evaluating a model in a systematic way.



Key Takeaways

- The AI Project Cycle provides a structured approach to develop an AI solution.
- Each step is important and may be repeated to improve the results.
- Ethical principles guide us to create AI systems that are fair, safe and beneficial for everyone.

Ethical AI Principles

AI systems can have a significant impact on individuals and society. The following principles help ensure that AI is developed and used responsibly.

Fairness

AI systems should treat all individuals and groups equitably and should not show bias or discrimination.

Privacy

AI systems should protect personal data and respect individual privacy. Data should be collected and used responsibly.

Accountability

People and organisations should be responsible for the development, use and impact of AI systems.

Transparency

AI systems should be understandable. The data, methods and decisions made by AI should be clear and explainable.

Safety

AI systems should be reliable, robust and secure, and should not cause harm to individuals, society or the environment.

Data exploration

Inspect distributions, missing entries, duplicates, unusual values and possible relationships. Tables and visualisations can reveal imbalance or recording errors. Exploration helps choose sensible features and questions; it does not prove that an observed association is a cause. Avoid accidentally using information that would be unavailable when a real prediction is made.

Three domains of application

The Statistical Data domain (data-oriented AI) analyses patterns in datasets, for example estimating demand. Computer Vision interprets visual inputs, for example identifying a damaged package from an image. NLP works with human language, for example categorising a customer message. A single product can combine domains; a spoken visual assistant may use both language and vision.

Data science and big data

Data science uses methods for collecting, organising, analysing and interpreting data. AI or machine learning can support prediction and pattern finding within that work. Big data refers to data whose scale, speed or variety creates processing challenges. The word "big" does not mean every value is correct or that consent is unnecessary.

Why ethical frameworks matter

An ethical framework provides a structured way to examine choices, values and responsibilities. Ask who benefits, who may be harmed, whether data use is justified, whether groups are treated fairly and who can challenge a decision. Ethical review is needed throughout the project, not only after a complaint.

Sector-based and value-based frameworks

Sector-based frameworks address the requirements of a particular area, such as healthcare. Value-based frameworks organise reasoning around values, duties or consequences. These are organising categories, not a licence to ignore other responsibilities. Keep the categories distinct from individual principles such as privacy or fairness.

Three value-based perspectives

A consequence-based or consequentialist view considers expected benefits and harms. A duty-based or deontological view emphasises obligations and rules that should be respected. A virtue-based view considers qualities such as honesty, fairness, responsibility and care. A balanced answer applies a perspective to the facts instead of merely listing its name.

Bioethics

Autonomy respects a person's informed choices. Beneficence seeks benefit. Non-maleficence avoids or minimises harm. Justice concerns fair treatment and allocation. In a hypothetical healthcare AI decision, these principles support explanation, appropriate consent, reliable testing and responsible human review. This is ethical study, not medical advice.

Human-centred and accountable AI

A human-centred system serves people and considers accessibility, dignity and meaningful control. Privacy needs limited and justified collection, secure handling and appropriate retention. Transparency helps people understand data use and relevant decision reasons. Accountability identifies who is responsible and provides a way to review, correct or appeal harmful outcomes.

Fairness and bias

Bias can arise from unrepresentative data, inaccurate labels, unsuitable features or the way predictions are used. Check performance across relevant groups and settings; good overall accuracy can hide poor treatment of a smaller group. Safeguards include better data, appropriate testing, transparent reporting and human review. Removing a group label alone does not guarantee fairness.

CHAPTER 06 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A team identifies affected users and the location of a recurring problem before collecting data. The stage is: [1]

(a) Problem scoping

(b) Classification

(c) Tokenisation

(d) Model deployment

Answer:

(a) Problem scoping Problem scoping defines the problem and its context.

Q2. After defining its objective, a team obtains relevant observations from reliable sources. This is: [1]

(a) Image segmentation

(b) Reward calculation

(c) Data acquisition

(d) Model evaluation

Answer:

(c) Data acquisition Acquisition obtains the data needed for the task.

Q3. A team plots observations to detect outliers and patterns. This belongs mainly to: [1]

(a) Final publication only

(b) Profit calculation

(c) Password management

(d) Data exploration

Answer:

(d) Data exploration Exploration helps understand the data before modeling.

Q4. A model works badly for a community absent from the training examples. Which concern is central? [1]

(a) A file extension error by definition

(b) Word order only

(c) Representational bias

(d) A larger font size

Answer:

(c) Representational bias Unrepresentative data can create unequal performance.

Q5. Which question belongs to the WHY part of a 4Ws canvas? [1]

(a) Who experiences the problem?

(b) Why is solving this problem valuable?

(c) Where is the problem observed?

(d) What exactly is the problem?

Answer:

(b) Why is solving this problem valuable? WHY explains the value or need for solving the problem.

Q6. Which ethical principle concerns a person's informed control over a decision affecting them? [1]

(a) Clustering

(b) Compression

(c) Regression

(d) Autonomy

Answer:

(d) Autonomy Autonomy concerns informed choice and agency.

Q7. A high-stakes AI recommendation is reviewed by a responsible person before action. This provides: [1]

(a) Anonymity by definition

(b) Automatic infallibility

(c) Human oversight

(d) Data duplication

Answer:

(c) Human oversight Human review supports responsibility and challenge of outputs.

Q8. An AI service states what data it uses and explains important limitations. This supports: [1]

(a) Transparency

(b) Uncontrolled access

(c) Secrecy

(d) Unlimited data retention

Answer:

(a) Transparency Transparency makes relevant processes and limits understandable.

Q9. A clinic checks that an AI allocation system does not disadvantage a particular community. Which principle is most directly involved? [1]

(a) Resolution

(b) Tokenisation

(c) Justice

(d) Encoding

Answer:

(c) Justice Justice concerns fair treatment and distribution.

Q10. Avoiding preventable harm from an AI recommendation is the principle of: [1]

(a) Profit maximisation alone

(b) Non-maleficence

(c) Dimensionality reduction

(d) External motivation

Answer:

(b) Non-maleficence Non-maleficence means avoiding harm.

Q11. A framework compares the likely benefits and harms of an AI action. Its emphasis is: [1]

(a) Alphabetical order

(b) Consequences or utility

(c) File type

(d) Pixel depth

Answer:

(b) Consequences or utility Consequence-based reasoning considers the effects of decisions.

Q12. Which domain is most directly used when a system interprets spoken or written human language? [1]

(a) Hardware cleaning

(b) Natural Language Processing

(c) Disk formatting

(d) Computer Vision only

Answer:

(b) Natural Language Processing NLP concerns computational processing of human language.

Q13. An ethical approach asks whether a decision respects fundamental rights. It is mainly: [1]

(a) Pixel-based

(b) Duty-based (deontological)

(c) Frequency-based

(d) Compression-based

Answer:

(b) Duty-based (deontological) Rights-based reasoning prioritises rights and dignity.

Q14. A team returns to data collection after evaluation reveals missing examples. This shows that the AI project cycle is: [1]

(a) Always one-way

(b) Completed before evaluation

(c) Iterative

(d) Independent of data

Answer:

(c) Iterative Evaluation can reveal a need to revisit earlier stages.

Q15. A decision-maker values honesty, care and integrity when developing AI. This reflects: [1]

(a) Random selection

(b) Term frequency

(c) Virtue-based reasoning

(d) Object detection

Answer:

(c) Virtue-based reasoning Virtue-based reasoning focuses on responsible character and conduct.

Short-answer questions - 2 marks each

S1. Distinguish data acquisition from data exploration. [2]

Answer:

(1) Data acquisition collects relevant observations from appropriate sources. [1]

(2) Data exploration examines and visualises the observations to reveal quality issues and patterns. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Explain why problem scoping should precede model development. [2]

Answer:

(1) Scoping identifies the users, problem and intended outcome. [1]

(2) It guides relevant data collection and prevents building a model for an unclear objective. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. State two safeguards against unfair AI decisions. [2]

Answer:

(1) Use representative data and examine errors across relevant groups. [1]

(2) Provide meaningful human review or an appeal process for important decisions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Define bioethics and name two of its principles. [2]

Answer:

(1) Bioethics applies ethical reasoning to decisions in healthcare and the life sciences. [1]

(2) Autonomy respects informed choices. [0.5]

(3) Justice requires fair treatment and access. [0.5]

Suggested marking: Definition: 1 mark; any two correctly named bioethical principles: 0.5 each. Beneficence and non-maleficence are also valid.

S5. Explain transparency and accountability in an AI service. [2]

Answer:

(1) Transparency makes relevant data use, limitations and decisions understandable. [1]

(2) Accountability means identifiable people or organisations remain responsible for outcomes and corrective action. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. Why is the AI project cycle described as iterative? [2]

Answer:

(1) Evaluation can reveal weaknesses in the data, model or original problem definition. [1]

(2) The team may revisit earlier stages and improve the solution using that feedback. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain the Who and Why questions in problem scoping. [2]

Answer:

(1) Who identifies the people affected by the problem and intended solution. [1]

(2) Why explains the reason for solving it and the benefit the solution should provide. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. Distinguish a sector-based ethical framework from a value-based framework. [2]

Answer:

(1) A sector-based framework addresses needs within a field, such as healthcare. [1]

(2) A value-based framework examines principles such as rights, beneficial outcomes or virtuous conduct. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. What is deployment? Why does monitoring continue afterwards? [2]

Answer:

(1) Deployment makes a tested AI solution available in its intended setting. [1]

(2) Monitoring checks changing conditions and unexpected errors so that the solution can be improved. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain autonomy and non-maleficence in healthcare AI. [2]

Answer:

(1) Autonomy respects a patient's informed choices and control over relevant decisions. [1]

(2) Non-maleficence requires avoiding preventable harm when developing or using a healthcare AI system. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

A complete worked 4Ws canvas

Use one original scenario throughout: a community library wants to reduce queues at its issue desk. An AI project is not justified merely because a queue exists; first identify the problem and whether a simpler organisational change would solve it.

Question	A focused answer
Who?	Readers waiting to borrow books and staff managing the issue desk.
What?	Long, uneven waiting times at particular hours.
Where?	The issue desk during the busiest afternoon period.
Why?	Reducing unnecessary waiting can improve access and staff planning.

Problem statement: Readers at the library issue desk experience long waits during afternoon peaks. The library needs a reliable estimate of demand so it can plan staffing and reduce average waiting time. **Success criterion:** compare waiting time and service quality before and after a carefully reviewed trial. An attractive prediction is not enough if it does not help the real problem.

Stages, outputs and iteration

Stage	Work and useful output for the library
Problem scoping	State users, context, need, constraints and success criteria.
Data acquisition	Collect relevant, permitted records of arrivals, timing and staff availability.
Data exploration	Check missing/duplicate records and visualise busy periods.
Modelling	Choose an appropriate approach for predicting demand; compare against a simple baseline.
Evaluation	Test unseen examples and whether the results serve the intended need.
Deployment	Use the evaluated system responsibly, monitor outcomes and review changes.

The cycle is iterative. A poor result may require better data, changed features or a redefined objective. Return to the relevant earlier stage and evaluate the improved solution.

Data acquisition versus exploration

Possible sources include surveys, observations, sensors and existing records, provided their use is justified and authorised. Choose data relevant to the problem: public library weather records alone do not identify how many desks were available. Representative sampling should include relevant quiet and busy periods rather than one convenient afternoon.

Issue found	Why it matters / suitable response
Missing timing records	May bias the estimate; investigate why values are missing before choosing a treatment.
Duplicate entries	May count the same event more than once; identify and handle true duplicates.
An extreme waiting time	Could be an error or a genuine exceptional event; investigate instead of deleting automatically.
Unbalanced examples	May make the model unreliable in an underrepresented period or group.

A bar chart can compare categories, a line chart can show change over time, a histogram can show a numerical distribution, and a scatter plot can examine two measured quantities. Choose the graph for the question. A visual relationship does not by itself prove causation.

Distinguishing the three application domains

Domain	Input, task and output in an original example
Statistical Data / Data Science	Past daily sales and conditions -> analyse patterns -> estimate tomorrow's demand.
Computer Vision	Parcel photographs -> inspect visual features -> flag visible damage.
Natural Language Processing	Written enquiries -> process language -> assign enquiry categories.

A delivery assistant could combine numerical demand estimates, image inspection and message understanding. Domains describe

the type of information and task; they do not mean that a product must belong to only one domain. Big data highlights scale, speed or variety challenges, not guaranteed quality.

What can influence an ethical decision?

Ethical decisions can be influenced by culture, personal values, intuition, concern for people and other life, and the facts of a situation. Examine these influences fairly. They do not justify favouring relatives or excluding people from another community.

For a scarce-resource decision, ask whether an apparently relevant detail is a fair criterion. A person's immediate need may be relevant; personal friendship with the decision-maker may introduce unfair preference. An ethical framework makes the reasoning explicit so assumptions and possible harms can be questioned.

Framework hierarchy: do not mix categories with principles

Level / approach	Meaning and original application
Sector-based	Tailored to a domain, such as healthcare bioethics.
Value-based	Uses values, duties or consequences to reason about choices.
Consequence-based perspective	Compares likely benefits and harms of an allocation.
Duty-based perspective	Checks rules, rights and obligations that should be respected.
Virtue-based perspective	Examines qualities such as honesty, fairness and care shown by decision-makers.

Bioethics: apply four principles to one case

Fictional case: a clinic uses an AI recommendation to help prioritise follow-up appointments. The output may affect access to care, so technical performance and ethical safeguards both matter. This case is for ethical reasoning, not clinical advice.

Principle	Application to the case
Autonomy	Explain relevant information and respect informed patient choices.
Beneficence	Use the system to promote patient welfare, not merely to demonstrate a new technology.
Non-maleficence	Investigate foreseeable harm, unreliable predictions and unsafe use before relying on the recommendation.
Justice	Assess fair access and group outcomes; do not disadvantage patients on irrelevant grounds.

Human-centred AI adds meaningful human control, accessibility and attention to people's needs. Transparency explains relevant use and limitations; accountability identifies who reviews and corrects errors. A system can be accurate overall and still require action because it performs poorly for one group.

Additional exam-style practice

C6-Q1. Which output most directly belongs to problem scoping? [1]

(a) An unexplained model score
(b) A clear problem statement and success criterion
(c) A list of final model weights
(d) A published prediction without review

C6-Q2. An allocation follows justified eligibility rules even when breaking them would be convenient. The reasoning is primarily: [1]

(a) Duty-based
(b) Pixel-based
(c) Unsupervised
(d) Frequency-based

C6-Q3. Give two reasons to examine data before training a model. [2]

Model answer: Exploration can reveal recording errors or duplicates. It can also reveal imbalance or missing groups that may make predictions unreliable.

Suggested practice marking: Two distinct relevant reasons, 1 each.

C6-Q4. Apply the four bioethical principles to an AI-supported appointment-prioritisation system. [4]

Model answer: Autonomy: give understandable information and respect informed choices. **Beneficence:** use the system to improve patient welfare. **Non-maleficence:** test for foreseeable harm and unsafe errors. **Justice:** check that access and errors do not unfairly disadvantage groups.

Suggested practice marking: Four correctly applied principles, 1 each. Names alone without the requested application are not full answers.

C6-Q5. For the library queue problem, explain four stages from problem scoping through modelling. [4]

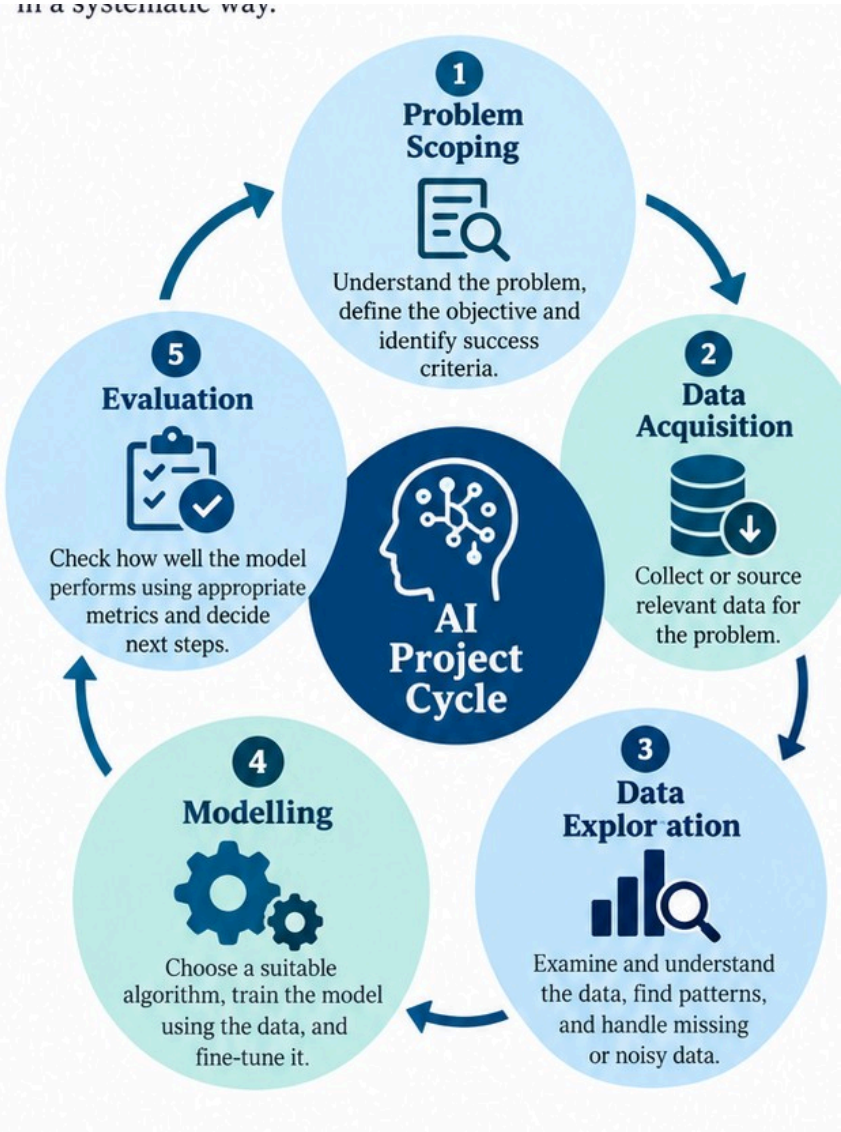
Model answer: Scope the affected readers, busy hours and desired improvement. Acquire relevant arrival and staffing records. Explore them for errors, distributions and busy periods. Build a suitable demand model and compare it with a simple baseline before evaluation.

Suggested practice marking: One correctly applied stage each: scoping 1; acquisition 1; exploration 1; modelling 1.

MCQ answers and reasons

C6-Q1: (b). The problem and intended success must be specified.

C6-Q2: (a). Duty-based reasoning emphasises justified obligations.



IMPORTANT QUESTIONS WITH ANSWERS

AI Project Cycle and Ethics

Q1. Explain the five stages of the AI project cycle. [4]

Answer: Problem scoping: define the challenge and success criterion. Data acquisition: gather suitable data. Data exploration: clean, visualise and understand it. Modelling: train or build an AI solution. Evaluation: compare results with objectives and identify weaknesses.

Suggested marking: *0.5 each stage name + 1.5 for meaningful descriptions of steps.*

Q2. What is the 4Ws problem canvas? [4]

Answer: Who: identify stakeholders. What: specify the problem. Where: identify the context or location. Why: explain the need and expected benefit. Together these frame the project before data collection.

Suggested marking: *1 per accurate W and explanation.*

Q3. Name and explain the three common AI domains. [3]

Answer: Data Science: extracts insights from structured or unstructured data. Computer Vision: analyses visual information such as images. NLP: processes human language such as messages or speech.

Suggested marking: *1 each.*

Q4. Why are fairness, transparency and privacy important in AI? [3]

Answer: Fairness helps avoid unjustly different outcomes. Transparency helps users understand how a system is used and its limitations. Privacy protects personal information through justified collection and safeguards.

Suggested marking: *1 each.*

Q5. Explain four principles in healthcare AI bioethics. [4]

Answer: Autonomy: respect informed patient choice. Beneficence: aim to benefit patients. Non-maleficence: minimise avoidable harm. Justice: ensure equitable access and treatment.

Suggested marking: *1 each.*

Q6. Differentiate rule-based and consequence-based ethical reasoning. [2]

Rule-based / deontological ethics	Consequence-based ethics
Judges actions by duties, rights and rules	Judges actions by outcomes and harms/benefits

Rule-based / deontological ethics	Consequence-based ethics
Example: respect consent regardless of convenience	Example: compare the expected harms of options

Suggested marking: *1 each.*

CHAPTER 07 | Advanced Concepts of Modeling in AI

Concepts and applications

AI, machine learning and deep learning

AI is the broader field. Machine learning is an approach within AI that learns patterns from examples. Deep learning is a subset of machine learning using neural networks with multiple learned layers. An explicit rule system can be AI without being a trained machine-learning model.

Inputs, features and targets

A feature is a measured or encoded input, such as floor area or a word count. A target is the output to be learned, such as a known price or class. Labelled data contains supplied targets; unlabelled data does not. Training fits model parameters; inference uses the trained model to produce an output for an input.

Structured, unstructured and semi-structured data

Structured data follows an organised schema, such as a table with named columns and defined types. Unstructured data, such as free text, audio or photographs, lacks that same fixed table form. Semi-structured data has organising tags or keys without a single rigid table layout, such as JSON records. Unstructured does not mean meaningless.

Traditional programming and learning

In a traditional rule-based approach, a person specifies logic that maps inputs to outputs. In supervised learning, a training process uses input-target examples to learn a mapping. Both need human decisions about purpose and design. A learned model can generalise to examples, but may also learn misleading patterns.

Rule-based versus learning-based modelling

Rule-based systems can suit narrow, stable situations where decisions can be written clearly. Learning-based systems can suit varied patterns where many examples are available. Rules need maintenance when conditions change; learned systems need suitable data, evaluation and monitoring. A hybrid can combine constrained rules with learned components.

Advanced Concepts of Modeling in AI

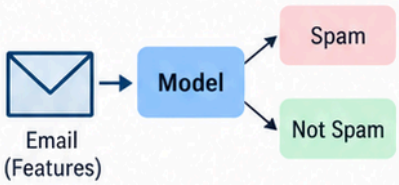
AI models learn patterns from data to make predictions, find hidden structures or take actions. The main types of learning used in AI are Supervised Learning, Unsupervised Learning and Reinforcement Learning. Each type is suitable for different kinds of problems.

Supervised Learning

Learns from **labeled** data to make predictions or classify new data.

- Uses input data with correct output labels.
- Learns the relationship between inputs and outputs.
- Used for classification and regression tasks.

Example:
Predict whether an email is Spam or Not Spam.

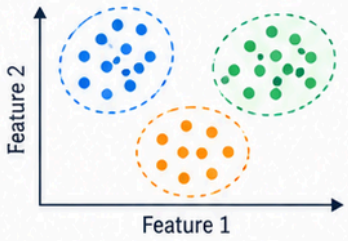


Unsupervised Learning

Finds hidden patterns or structure in unlabeled data.

- Uses input data without labels.
- Discovers natural groupings or patterns.
- Commonly used for clustering, association and dimensionality reduction.

Example:
Group similar customers based on their shopping behaviour.

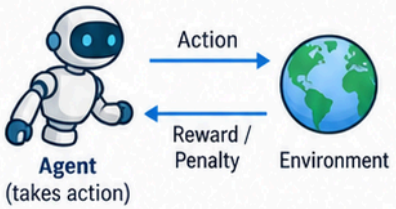


Reinforcement Learning

Learns by interacting with an environment and receiving rewards or penalties.

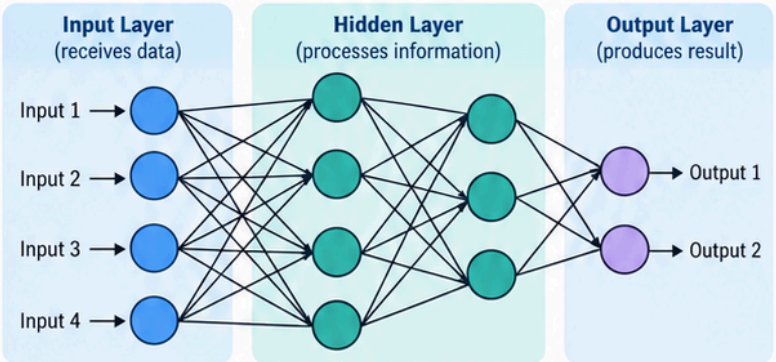
- Learns through trial and error.
- Takes actions, observes results (rewards or penalties) and improves over time.
- Used in decision-making and sequential tasks.

Example:
A robot learns to reach a target by getting a reward for the correct action.



Artificial Neural Network (ANN)

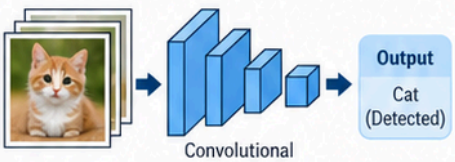
An Artificial Neural Network is inspired by the human brain. It consists of layers of interconnected neurons that process information and learn patterns from data.



Each connection has a weight, and the neurons apply an activation function to pass information to the next layer. The network learns by adjusting the weights during training.

Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a type of neural network specially designed for image data. It uses convolutional layers to automatically learn important features such as edges, shapes and textures.



Key Takeaways

- Supervised learning uses labeled data to predict or classify.
- Unsupervised learning finds hidden patterns in unlabeled data.
- Reinforcement learning learns through interaction and rewards.
- Artificial Neural Networks consist of input, hidden and output layers.
- CNNs are specialized neural networks used for image analysis.

Supervised learning

Supervised learning uses labelled examples. Classification predicts a category: damaged/undamaged, for example. Regression predicts a numerical quantity such as delivery time. Numeric class codes do not turn classification into regression; the nature of the output matters, not merely the way it is written.

Unsupervised learning

Unsupervised learning investigates patterns without supplied target labels. Clustering groups similar examples. Association analysis finds items or events that occur together. Anomaly detection seeks unusual cases relative to a pattern of normality. A discovered group is not automatically meaningful, and a rare example is not automatically harmful.

Reinforcement learning

An agent chooses actions in an environment and receives reward information about outcomes. It learns a strategy, or policy, for selecting actions. Exploration tries alternatives; exploitation uses what has already been learned. This is different from being supplied a correct class label for every example.

Artificial neural networks

An ANN contains connected processing units. Inputs are multiplied by weights, combined with a bias and passed through an activation function. The input layer receives encoded features, hidden layers transform them and the output layer produces task-related scores or predictions. Training changes parameters to improve performance under a chosen objective.

Weighted sums and activation

For two inputs, a simple pre-activation sum is z = w1*x1 + w2*x2 + b. Bias shifts the sum; weights control the contributions of inputs. An activation function transforms the sum, often adding nonlinearity. Do not confuse the pre-activation value with the final output unless the activation rule is specified.

Propagation and training

Forward propagation sends values through the network to produce an output. During training, the output is compared with the target or objective. Backpropagation calculates how parameters influence the loss so an update method can adjust them. It is not a second independent prediction. A biological neuron analogy is useful but is not an exact description of an artificial unit.

CNNs and visual patterns

A Convolutional Neural Network uses learned local filters that are applied across an image. Early features may respond to simple edges; later combinations can represent more complex patterns. Pooling can reduce spatial size. A final prediction stage uses the learned representation. More layers do not automatically guarantee better results.

Model suitability

Choose a model for the task, available data, interpretability needs, computing resources and possible harm. Overly simple models may underfit; overly complex or poorly controlled models may overfit. Evaluation on appropriate unseen data is necessary before trusting apparent training success.

CHAPTER 07

Regression and Clustering

Regression

Regression is a supervised learning technique used when the output to be predicted is a numerical value. It learns a relationship between input features and a continuous target such as price, marks, temperature or delivery time.

Examples: predicting house rent from floor area and location, estimating fuel consumption, forecasting daily sales.

Clustering

Clustering is an unsupervised learning technique that groups similar data points together when no target labels are provided. The aim is to discover natural groups or patterns hidden in the data.

Examples: grouping customers by buying behaviour, arranging news articles by topic, segmenting students by learning patterns.

Regression vs Clustering

Basis	Regression	Clustering
Learning type	Supervised learning	Unsupervised learning
Target/Output	Continuous numerical value	Group/cluster membership
Uses labelled data?	Yes	No
Main purpose	Predict a quantity	Find similar groups
Example	Predict house price	Group customers by shopping habits

Exam Tip

Remember: Regression predicts a number, while clustering forms groups. If the answer mentions labelled data, it points to regression/supervised learning; if it mentions grouping without labels, it points to clustering/unsupervised learning.

CHAPTER 07

Solved Questions: Regression & Clustering

Q1. Differentiate regression and clustering. [2]

Answer: Regression predicts a continuous numerical value using labelled data, such as house price or delivery time. Clustering groups similar data points without labels, such as grouping customers according to buying behaviour.

Suggested marking: 1 mark for regression, 1 mark for clustering.

Q2. A school wants to predict the marks a student may score in the final exam using attendance and periodic-test marks. Which technique is more suitable—regression or clustering? Give a reason. [2]

Answer: Regression is more suitable because the output to be predicted is a numerical value, that is, the final marks. Clustering would only group students into similar clusters and would not directly predict a score.

Suggested marking: correct technique 1, valid reason 1.

Q3. An online store wants to divide its customers into groups such as budget buyers, premium buyers and occasional buyers on the basis of shopping patterns. Which technique should be used? Explain. [2]

Answer: Clustering should be used because the aim is to group similar customers based on their behaviour. There are no predefined labels given in advance, so it is an unsupervised problem.

Suggested marking: correct technique 1, explanation 1.

Q4. Explain regression and clustering with one example each. [4]

Answer: Regression is a supervised learning method used to predict a continuous numerical value. For example, a model can predict the price of a house from area, location and number of rooms. Clustering is an unsupervised learning method used to group similar data points without target labels. For example, customers of a shopping website can be grouped according to purchase history so that similar customers fall in the same cluster.

Suggested marking: regression meaning 1, regression example 1, clustering meaning 1, clustering example 1.

Regression Flow

Features → Model → Numerical Prediction

Example: Area → Model → House Price

Clustering Flow

Data Points → Similarity → Groups

Example: Customers → Clusters

CHAPTER 07 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. Examples contain measurements together with the correct category for each measurement. The learning setting is: [1]

(a) Unsupervised learning

(b) Random sorting

(c) Supervised learning

(d) Data encryption

Answer:

(c) Supervised learning Known target labels guide supervised learning.

Q2. A model estimates the number of minutes needed for a journey. This is usually: [1]

(a) Tokenisation

(b) Clustering

(c) Classification into fixed names

(d) Regression

Answer:

(d) Regression The target is a numerical quantity rather than a category label.

Q3. Unlabelled records are grouped by similarity. Which task is being performed? [1]

(a) Decryption

(b) Label transcription

(c) Supervised regression

(d) Clustering

Answer:

(d) Clustering Clustering seeks groups without supplied group labels.

Q4. An agent changes its actions after receiving rewards from an environment. This is: [1]

(a) File sorting

(b) Rule copying

(c) Reinforcement learning

(d) Image compression

Answer:

(c) Reinforcement learning The agent learns an action strategy from feedback on its actions.

Q5. Which statement correctly relates AI, ML and DL? [1]

(a) Deep learning and AI have no connection.

(b) Machine learning contains every form of intelligence.

(c) Deep learning is a part of machine learning, which is a part of AI.

(d) All AI must use deep learning.

Answer:

(c) Deep learning is a part of machine learning, which is a part of AI. The fields form a nested relationship, but not all AI is learning-based.

Q6. A rule says, "If stock is below ten units, send an alert." No examples are used to learn it. The approach is: [1]

(a) Learned regression by definition

(b) Reinforcement learning

(c) Unsupervised clustering

(d) Rule-based

Answer:

(d) Rule-based The decision rule was explicitly specified.

Q7. Which neural-network layer receives the feature values first? [1]

(a) Footer layer

(b) Input layer

(c) Archive layer

(d) Output layer

Answer:

(b) Input layer The input layer receives the representation supplied to the network.

Q8. What do weights in an artificial neuron mainly represent? [1]

(a) The influence of input signals on a weighted sum

(b) The size of a password

(c) Only the output label's spelling

(d) The number of paper pages

Answer:

(a) The influence of input signals on a weighted sum Weights control the contribution of inputs.

Q9. Why is an activation function used after a neuron's weighted sum? [1]

(a) To transform the signal, often introducing nonlinearity

(b) To rename every training record

(c) To guarantee perfect accuracy

(d) To replace all training examples

Answer:

(a) To transform the signal, often introducing nonlinearity Activation enables useful transformations; nonlinear activations support complex patterns.

Q10. A retailer discovers that two products often occur in the same purchase. This is an example of: [1]

(a) Image segmentation

(b) Manual file naming

(c) Association

(d) Linear text order

Answer:

(c) Association Association discovers co-occurrence relationships.

Q11. Which data is semi-structured? [1]

(a) Records in JSON with keys and nested fields

(b) An unannotated photograph alone

(c) An unrecorded conversation

(d) Only a fixed numerical matrix by definition

Answer:

(a) Records in JSON with keys and nested fields JSON has an organising structure without a single rigid relational table layout.

Q12. A neural network architecture particularly suited to extracting local visual patterns is a: [1]

(a) Convolutional Neural Network

(b) Spell-check dictionary

(c) Folder tree

(d) Password manager

Answer:

(a) Convolutional Neural Network CNNs learn patterns through local filters in image-like inputs.

Q13. In a dataset of houses, floor area is used to predict rent. Floor area is a: [1]

(a) Test score

(b) Final decision rule by itself

(c) Predicted target value

(d) Feature

Answer:

(d) Feature A feature is an input attribute used by the model.

Q14. Which task predicts one of a set of named categories? [1]

(a) Regression of a continuous value

(b) Classification

(c) Frequency normalisation only

(d) File compression

Answer:

(b) Classification Classification selects a category.

Q15. A model flags a sensor reading far outside typical operating behaviour. The task is: [1]

(a) Anomaly detection

(b) Speech synthesis

(c) Alphabetical sorting

(d) Document translation

Answer:

(a) Anomaly detection Anomaly detection looks for unusual observations.

Short-answer questions - 2 marks each

S1. Differentiate supervised learning and unsupervised learning. [2]

Answer:

(1) Supervised learning uses training examples paired with known target labels or values. [1]

(2) Unsupervised learning discovers patterns or groups without supplied target labels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Distinguish classification from regression with an example of each. [2]

Answer:

(1) Classification predicts a category, such as parcel damaged or intact. [1]

(2) Regression predicts a numerical quantity, such as delivery time in minutes. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

S3. Explain association learning with a suitable example. [2]

Answer:

(1) Association learning identifies items or events that frequently occur together. [1]

(2) For example, a shop may find that brushes and paint often appear in one purchase. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. What is anomaly detection? Give an example. [2]

Answer:

(1) Anomaly detection identifies observations that differ markedly from typical patterns in data. [1]

(2) For example, flag an unusually high machine-vibration reading for inspection. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. State the roles of a weight and an activation function in a neuron. [2]

Answer:

(1) A weight controls an input's contribution to the weighted sum. [1]

(2) An activation function transforms that sum, often allowing the network to represent nonlinear relationships. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. Distinguish a rule-based system from a learning-based system. [2]

Answer:

(1) A rule-based system follows explicitly written decision rules. [1]

(2) A learning-based system fits patterns from data, so its behaviour depends on training and evaluation. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Distinguish labelled and unlabelled examples. [2]

Answer:

(1) A labelled example includes an input and the expected target. [1]

(2) An unlabelled example provides the input without a supplied target category or value. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. State the roles of the input and output layers of an ANN. [2]

Answer:

(1) The input layer receives the encoded feature values. [1]

(2) The output layer produces the network's task-specific result, such as class scores. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Distinguish artificial intelligence from machine learning. [2]

Answer:

(1) AI is the broader field of systems performing tasks associated with intelligence. [1]

(2) Machine learning is an AI approach that learns patterns from examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain the relationship between deep learning and machine learning. [2]

Answer:

(1) Deep learning is a subset of machine learning using multilayer neural networks. [1]

(2) It can learn complex representations rather than relying only on manually selected features. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S11. Distinguish structured and unstructured data with examples. [2]

Answer:

(1) Structured data follows a defined arrangement, such as a table of sales records. [1]

(2) Unstructured data lacks that table-like schema, such as photographs or free text. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S12. Explain semi-structured data with one example. [2]

Answer:

(1) Semi-structured data contains organising tags or keys without requiring one fixed table schema. [1]

(2) A JSON event log with named fields is one example. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S13. Define clustering and state why labels are not required. [2]

Answer:

(1) Clustering groups observations that are similar under a chosen representation. [1]

(2) It discovers groups from the data rather than learning supplied correct group labels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S14. Identify the agent and the environment in a learning robot game. [2]

Answer:

(1) The agent is the robot or program selecting actions. [1]

(2) The environment is the game world that responds to those actions with observations and feedback. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S15. State one difference between a feature and a target. [2]

Answer:

(1) A feature is an input attribute, such as parcel weight. [1]

(2) The target is the output to predict, such as the parcel's delivery time. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S16. Explain why an ANN may need an activation function. [2]

Answer:

(1) An activation function transforms a neuron's combined input before passing the signal onward. [1]

(2) Nonlinear activations allow networks to represent relationships that are not merely linear. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S17. State two reasons why a rule-based system may be suitable for a small task. [2]

Answer:

(1) Explicit rules can implement a small set of clear, stable requirements. [1]

(2) Their decisions can also be inspected directly without collecting a large labelled training dataset. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S18. Explain why incorrect training labels can reduce model quality. [2]

Answer:

(1) Labels tell a supervised model which outputs its examples should produce. [1]

(2) Incorrect labels supply misleading learning signals and can produce unreliable decisions on new examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S19. Define a CNN and identify a suitable kind of input. [2]

Answer:

(1) A Convolutional Neural Network uses learned filters to extract local patterns. [1]

(2) Images are suitable inputs because neighbouring pixels often contain related visual information. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S20. Explain why a learned association does not prove causation. [2]

Answer:

(1) An association reports that items or events occur together in the observed data. [1]

(2) Other factors may explain their co-occurrence, so causation requires additional evidence. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Read an actual small dataset

Trip	Distance (km)	Stops	Time (min)
A	3	2	14
B	5	4	24
C	4	1	16

For predicting travel time, each row is an **observation** or example. Distance and stops are **features**; known time is the **target**. Trip is an identifier, not automatically a useful feature. The table is labelled because the target is supplied. If the time column is removed, these records no longer contain that target.

A collection of examples is a **dataset**. Training examples are used to fit parameters; a held-out test set checks performance on examples not used for fitting. Inference applies a trained model to a new input. A label may be a category or, in a broader supervised-learning use of the term, a target value; the output type determines classification versus regression.

A decision tree as an interpretable model

A decision tree follows tests from a root through branches to a leaf outcome. An internal node asks a question; a branch represents an answer; a leaf gives a decision or prediction. Example: "Is the parcel damaged?" If yes, send for inspection; if no, ask whether the address is readable, then route accordingly. This is a hand-written rule tree. A learned decision tree chooses splits from training data. A diagram with branches is not automatically a trained model.

Classification, regression, clustering and association

Task	Example and defining output
Classification	Predict "late" or "on time": output is a category.
Regression	Predict journey duration: output is a numerical quantity.
Clustering	Group similar journeys without supplying group labels.
Association	Find items or events that frequently occur together.

Numbered category codes do not make a task regression: predicting class 0 or class 1 remains classification. Clustering creates groups from similarity; association identifies co-occurrence relationships. The observation that notebooks and pens are often bought together does not prove that buying one causes the other.

Reinforcement learning: interaction and rewards

An **agent** selects an **action** in an **environment**, observes the outcome and receives a **reward**. Its **policy** is a strategy for choosing actions. A route-finding agent may gain a reward for reaching a goal and incur a negative reward for wasting time. It learns through outcomes rather than receiving a correct target label for every possible action.

.

Biological analogy and artificial computation

Term	What it means in this lesson
Biological neuron	A nerve cell that receives and transmits signals.
Dendrites and synapses	Dendrites receive signals; synapses are junctions through which cells communicate.
Axon	Carries the outgoing signal from a biological neuron.
Artificial neuron	A computational unit that combines weighted inputs and bias and applies an activation rule.
Weight	Scales an input's contribution; a larger magnitude generally means a stronger contribution at that unit.
Bias parameter	Shifts the combined input before activation. It is not the same as unfair social bias.

Worked weighted sum and activation

Let x1=3, x2=2, w1=0.4, w2=0.7 and b=-0.3. Then **z=w1x1+w2x2+b=0.4(3)+0.7(2)-0.3=2.3**. If the stated rule outputs 1 when z is at least 2 and 0 otherwise, the output is 1. A different activation rule could give a different output from the same z. Always separate the weighted sum from the activation result.

An activation function transforms the combined input. Nonlinear activation functions enable a network to model relationships that stacked linear operations alone cannot represent. At this level, use the specified function or threshold; do not invent one from the input values.

Layers, propagation and learning

Component	Role
Input layer	Receives features that represent the example.
Hidden layer(s)	Transform combinations of features into intermediate representations.
Output layer	Produces task-related scores, values or decisions.
Forward propagation	Passes inputs through the network to compute an output.
Backpropagation	During training, calculates how parameters affect loss so an update method can improve them.

A neural model does not become accurate merely by adding layers. It needs useful data, an appropriate objective, training and independent evaluation. Learning adjusts parameters; inference uses the resulting model. A numerical bias parameter is part of the calculation; societal bias concerns unfair patterns or effects.

ANN, CNN and recognition applications

An ANN is a broad neural-network model. A CNN uses shared local filters, commonly to identify patterns in images. Early filters may respond to edges; later representations can combine them into more complex features. Pooling may reduce spatial size before a prediction stage. Not every ANN is a deep network, and not every deep-learning task is an image task.

Application	What the system determines
Anomaly detection	Whether an observation is unusual relative to the learned normal pattern; an unusual case still needs interpretation.
Object classification	Which category an observed object or image belongs to.
Object identification	Which object is present; the intended level of identification should be stated in the question.
Digit recognition	Which handwritten or printed digit an image represents.

Additional exam-style practice

C7-Q1. For x1=3, x2=2, w1=0.4, w2=0.7 and b=-0.3, the weighted sum is: [1]

(a) 1.1
(b) 2.6
(c) 2.3
(d) 3.0

C7-Q3. Distinguish a bias parameter from unfair bias in AI decisions. [2]

Model answer: A bias parameter shifts an artificial neuron's weighted sum. Unfair bias is a systematic pattern that can disadvantage people or groups in decisions.

Suggested practice marking: Computational meaning 1; ethical meaning 1.

C7-Q5. Explain the input, hidden and output layers of an ANN and distinguish training from inference. [4]

Model answer: The input layer receives encoded features. Hidden layers transform weighted combinations of those features. The output layer produces task-related predictions or scores. Training adjusts parameters using examples and an objective; inference applies the learned model to an input.

Suggested practice marking: Input 1; hidden 1; output 1; training/inference distinction 1.

MCQ answers and reasons

C7-Q1: (c). The bias is added after the two weighted contributions.

IMPORTANT QUESTIONS WITH ANSWERS

Advanced AI Modelling

Q1. Differentiate artificial intelligence, machine learning and deep learning. [3]

AI	Machine Learning	Deep Learning
Broad field of intelligent systems	Learns patterns from data	ML using multilayer neural networks

Suggested marking: *1 each.*

Q2. Compare supervised, unsupervised and reinforcement learning. [3]

Answer: Supervised learning learns input-to-target examples. Unsupervised learning finds patterns in unlabelled data. Reinforcement learning learns actions through rewards or penalties from an environment.

Suggested marking: *1 each.*

Q3. Differentiate classification and regression. [2]

Answer: Classification predicts a discrete category (spam/not spam). Regression predicts a continuous numerical value (house price).

Suggested marking: *1 each.*

Q4. What is clustering? Give an example. [2]

Answer: Clustering groups similar data points without predefined class labels, such as grouping customers by purchasing patterns.

Suggested marking: *1 definition + 1 example.*

Q5. Explain input, hidden and output layers of an ANN. [3]

Answer: Input layer receives features; hidden layers transform signals through weighted connections and activation functions; output layer produces predictions or scores.

Suggested marking: *1 each.*

Q6. Define neuron, weight, bias and activation function. [4]

Answer: Neuron: basic computational unit receiving inputs. Weight: parameter indicating input influence. Bias: adjustable offset added before activation. Activation function: mapping that introduces or controls output response, often nonlinearly.

Suggested marking: *1 each.*

Q7. What is an overfitted model? [2]

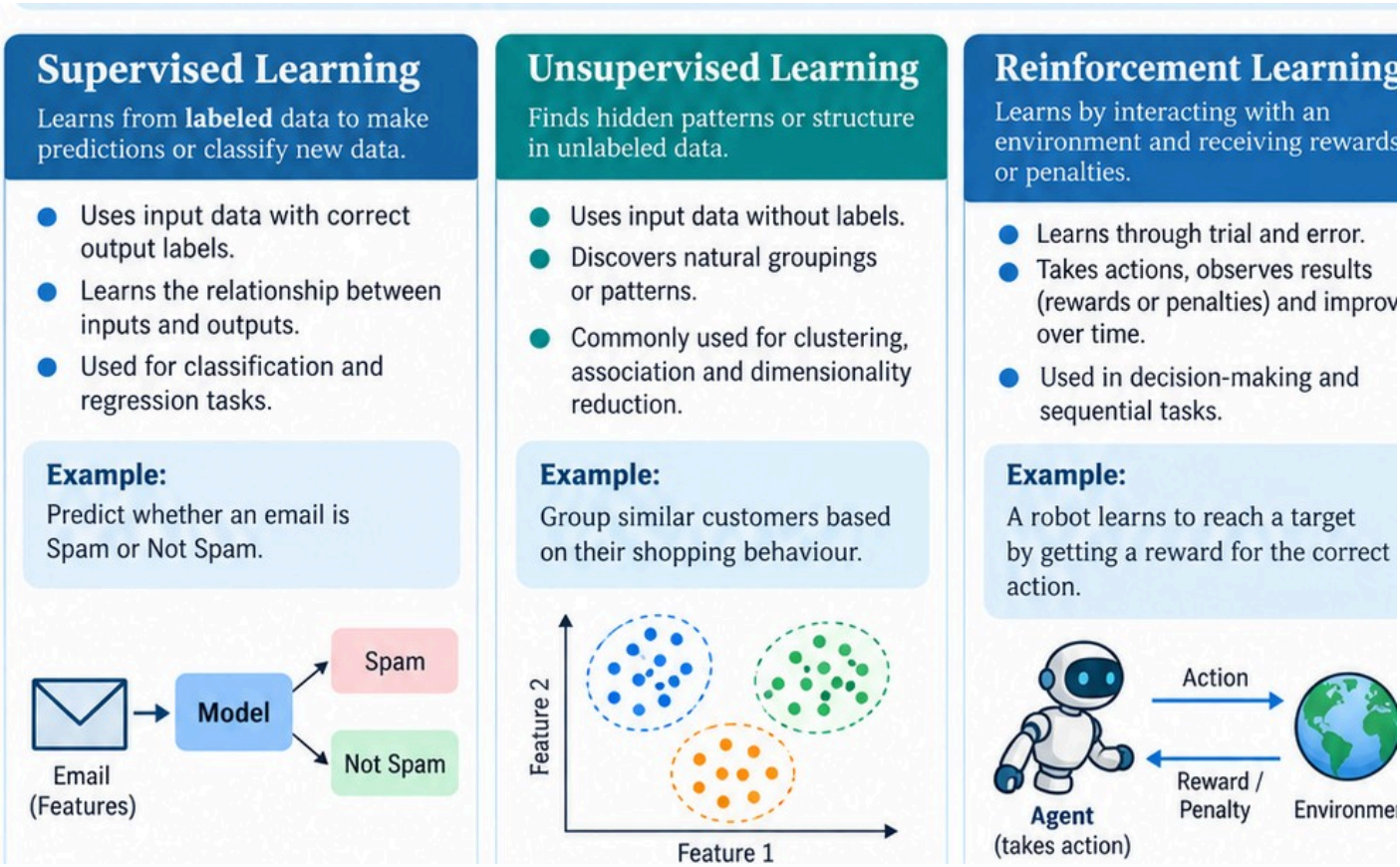
Answer: It fits training examples-including noise-too closely and performs worse on unseen cases. It can be detected by a large train-test performance gap.

Suggested marking: *1 explanation + 1 symptom.*

Q8. What is underfitting? [2]

Answer: Underfitting occurs when a model is too simple or poorly trained to capture important patterns, so performance is weak on both training and test sets.

Suggested marking: *1 definition + 1 symptom.*



CHAPTER 08 | Evaluating Models

Concepts and applications

Four-mark calculations and tables are fully worked in the Numerical Studio; use the contents page or PDF bookmarks to reach that section.

Why evaluation is necessary

A model must be judged against the purpose it is meant to serve. Training success is not enough: the same examples helped determine the model. Evaluation on suitable unseen data estimates how well it generalises and reveals different types of error.

Training and testing data

Training data is used to learn parameters. Test data is held apart to estimate performance. Reusing test examples during model fitting can inflate results through leakage. An 80:20 split is a common classroom illustration, not a compulsory universal ratio. For 750 examples, it gives 600 for training and 150 for testing.

Underfitting and overfitting

An underfitted model misses important relationships and often performs poorly on both training and test data. An overfitted model captures details or noise that do not generalise, often showing much better training than test performance. Suitable complexity, representative data and appropriate evaluation help; changing a split without investigation is not a guaranteed cure.

Name the positive class first

Positive means the condition chosen for evaluation: a faulty product, spam or a specified category. It does not necessarily mean desirable. In this book matrices use actual classes as rows and predicted classes as columns. Read the labels because other presentations may transpose this arrangement.

The four outcomes

TP: actually positive and predicted positive. FN: actually positive but predicted negative. FP: actually negative but predicted positive. TN: actually negative and predicted negative. "False" means the prediction was wrong; the second word names the predicted class. A false alarm is FP; a missed positive is FN.

Evaluating Models

After training a model, we need to check how well it performs. Model evaluation helps us understand how many predictions are correct and what types of errors it makes.

1 Confusion Matrix

A confusion matrix shows the number of correct and incorrect predictions made by a classification model.

		Predicted Class: Predicted Positive (Positive)	Predicted Class: Predicted Negative (Negative)
Actual Class	Actual Positive (Positive)	**TP** (True Positive) Correctly predicted positive	**FN** (False Negative) Actual positive predicted negative
	Actual Negative (Negative)	**FP** (False Positive) Actual negative predicted positive	**TN** (True Negative) Correctly predicted negative

2 Evaluation Metrics

Common metrics used to evaluate a classification model are:

Accuracy

Proportion of all correct predictions (both positive and negative).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Proportion of predicted positive cases that are actually positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

Proportion of actual positive cases that are correctly predicted.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 Score

Harmonic mean of Precision and Recall. Useful when we want a balance between both.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3 Worked Example

A model is tested on 100 images to detect whether an email is spam (positive) or not spam (negative). The confusion matrix is given below.

		Predicted Class: Predicted Positive (Spam)	Predicted Class: Predicted Negative (Not Spam)
Actual Class	Actual Positive (Spam)	40	10
	Actual Negative (Not Spam)	5	45

Metric Calculations

$$\text{Accuracy} = \frac{40 + 45}{40 + 45 + 5 + 10} = \frac{85}{100} = 0.85\ (85\%)$$

$$\text{Precision} = \frac{40}{40 + 5} = \frac{40}{45} = 0.89\ (88.9\%)$$

$$\text{Recall} = \frac{40}{40 + 10} = \frac{40}{50} = 0.80\ (80\%)$$

$$\text{F1 Score} = 2 \times \frac{0.89 \times 0.80}{0.89 + 0.80} = 0.84\ (84\%)$$

By Mr. Taranjeet Singh HOD IT-CS, Mr. Ravindra Uniyal PGT CS

Accuracy and error

Accuracy = (TP + TN)/N, where N = TP + TN + FP + FN. Classification error rate = (FP + FN)/N = 1 - accuracy. Multiply by 100 for a percentage. The number of wrong predictions is FP + FN, not the error percentage. A separate relative numerical-error formula is explained only when a question defines it; it is not a universal regression accuracy score.

Precision and recall

Precision = TP/(TP + FP): among predicted positives, what share are correct? Recall = TP/(TP + FN): among actual positives, what share were detected? Precision is sensitive to false alarms; recall is sensitive to missed positives. Use the case's costs of errors to justify which should be prioritised.

F1 score

F1 = 2PR/(P + R), using precision P and recall R on the same scale. Equivalently, F1 = 2TP/(2TP + FP + FN). Use the direct count formula to avoid rounding early. F1 balances precision and recall through a harmonic mean, not a simple arithmetic mean.

Interpreting results

In a dataset with 990 ordinary items and 10 faulty ones, a model predicting every item as ordinary has 99% accuracy but detects none of the faulty items. Recall for faults is 0. High accuracy alone therefore does not prove usefulness, fairness or safety. Check the confusion matrix and the importance of each error.

Zero denominators and rounding

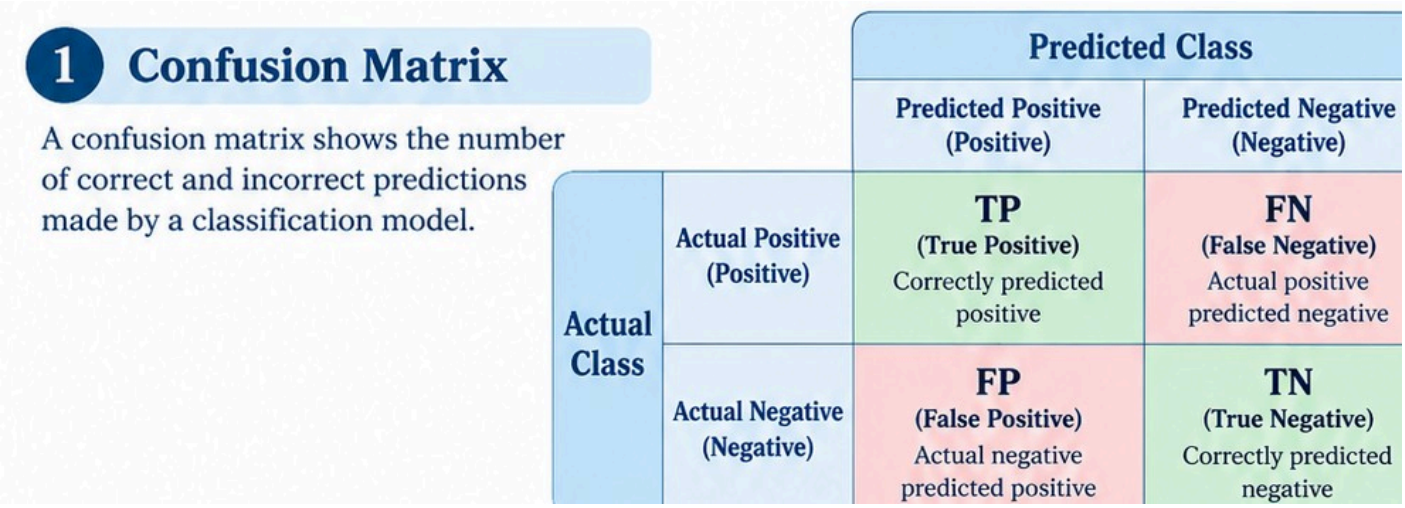
If no positives are predicted, precision has denominator zero and is undefined under the direct formula. Report the issue instead of dividing by zero; software may apply a stated convention. The standard practice problems here avoid that ambiguity. Preserve fractions or full decimals in intermediate steps and round only the final percentage.

Fair evaluation

Use relevant, representative examples and examine performance across important groups or conditions. Keep results honest and identify limitations. Bias may be reduced through improved data, label review, appropriate metrics and review of how predictions affect people. No single metric proves a system is ethical.

Four-mark numerical methods

A four-metric problem can assign one mark each for accuracy, precision, recall and F1. A matrix-based problem can instead allocate two marks for the labelled cells, one for accuracy and one for the number of wrong predictions. Follow the actual subparts; do not assume every four-mark question has the same split. Detailed worked cases follow this chapter.



CHAPTER 08 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. With "defective" defined as positive, a faulty item is predicted to be good. The outcome is: [1]

(a) False negative

(b) False positive

(c) True positive

(d) True negative

Answer:

(a) False negative The actual positive was missed.

Q2. With "spam" defined as positive, a legitimate message is blocked as spam. The outcome is: [1]

(a) True positive

(b) False negative

(c) False positive

(d) True negative

Answer:

(c) False positive A negative item was incorrectly predicted positive.

Q3. Which denominator belongs to precision? [1]

(a) TN + FN

(b) All cases in every situation

(c) TP + FN

(d) TP + FP

Answer:

(d) TP + FP Precision conditions on predicted positives.

Q4. Which denominator belongs to recall? [1]

(a) TN only

(b) TP + FP

(c) TN + FP

(d) TP + FN

Answer:

(d) TP + FN Recall conditions on actual positives.

Q5. A model memorises training details but performs poorly on new examples. This suggests: [1]

(a) Underfitting only

(b) A guaranteed fair model

(c) Overfitting

(d) Ideal generalisation

Answer:

(c) Overfitting The learned behaviour does not generalise well.

Q6. A model performs poorly on both training and test data because it is too simple. This suggests: [1]

(a) No model error

(b) Ideal complexity

(c) Perfect calibration

(d) Underfitting

Answer:

(d) Underfitting The model fails to capture important relationships.

Q7. Why keep a test set separate from training? [1]

(a) To avoid ever collecting data

(b) To ensure all predictions become correct

(c) To make labels unnecessary

(d) To estimate performance on unseen examples

Answer:

(d) To estimate performance on unseen examples Held-out examples give a more realistic evaluation of generalisation.

Q8. A model predicts the majority class for every record and obtains a high overall score. What should be examined next? [1]

(a) Only the file name

(b) Class-specific errors and recall

(c) Only the font used in the report

(d) Only the number of columns

Answer:

(b) Class-specific errors and recall Class imbalance can conceal failure on a smaller class.

Q9. F1 is the harmonic mean of: [1]

(a) TP and TN

(b) Precision and recall

(c) Accuracy and file size

(d) Training and test counts

Answer:

(b) Precision and recall F1 combines precision and recall in one score.

Q10. Which operation risks data leakage? [1]

(a) Reserving unseen examples

(b) Reporting all confusion-matrix counts

(c) Documenting the positive class

(d) Using the final test results repeatedly to choose the model

Answer:

(d) Using the final test results repeatedly to choose the model Repeated selection using the final test information contaminates evaluation.

Q11. If actual positives are rare and missing them is costly, which metric is especially important? [1]

(a) The vocabulary length

(b) Recall for the positive class

(c) Only overall accuracy

(d) The model's file extension

Answer:

(b) Recall for the positive class Recall shows how many actual positives are detected.

Q12. Which pair counts all incorrect binary classifications? [1]

(a) FP + FN

(b) TP + TN

(c) TP + FP

(d) TN + FN

Answer:

(a) FP + FN False positives and false negatives are the errors.

Q13. A model has no predicted positives. Its precision formula has: [1]

(a) No dependence on predictions

(b) The same denominator as accuracy

(c) A value of one automatically

(d) A zero denominator

Answer:

(d) A zero denominator TP + FP is zero; a reporting convention is needed for undefined precision.

Q14. Which report is most transparent? [1]

(a) State the dataset, metrics, limitations and group-wise results.

(b) Hide errors in small groups.

(c) Show only the most favourable number.

(d) Present training accuracy as a guarantee.

Answer:

(a) State the dataset, metrics, limitations and group-wise results. Meaningful evaluation includes context and limitations.

Q15. An 80:20 split of 250 examples gives how many test examples? [1]

(a) 50

(b) 200

(c) 20

(d) 80

Answer:

(a) 50 The test share is 20 percent of 250.

Short-answer questions - 2 marks each

S1. Explain why high training accuracy is insufficient evidence of a useful model. [2]

Answer:

(1) A model may memorise training details rather than learn general patterns. [1]

(2) Performance must also be checked on suitable unseen examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Distinguish a false positive from a false negative. [2]

Answer:

(1) A false positive predicts the positive class for an actually negative example. [1]

(2) A false negative predicts the negative class for an actually positive example. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Distinguish precision from recall using their denominators. [2]

Answer:

(1) Precision divides TP by TP + FP, the number of predicted positives. [1]

(2) Recall divides TP by TP + FN, the number of actual positives. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Give two reasons to report more than overall accuracy. [2]

Answer:

(1) Class imbalance can hide poor detection of a smaller but important class. [1]

(2) False positives and false negatives may also have different consequences for users. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. Explain train-test split and its purpose. [2]

Answer:

(1) Divide data into examples used for training and a separate set used for evaluation. [1]

(2) The held-out set estimates performance on unseen data. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. Explain underfitting and contrast it with overfitting. [2]

Answer:

(1) Underfitting occurs when a model misses important patterns and performs poorly even in training. [1]

(2) Overfitting fits training details but generalises poorly. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain why false negatives matter in an early-warning system. [2]

Answer:

(1) A false negative misses an actual positive event that needed attention. [1]

(2) Missing a real warning condition may delay a response, so recall deserves particular attention. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. State two practices that make an evaluation report transparent. [2]

Answer:

(1) Identify the test dataset, metrics and conditions under which results were obtained. [1]

(2) Report relevant limitations and error patterns rather than showing only the most favourable score. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. What does F1 measure, and when can it be useful? [2]

Answer:

(1) F1 is the harmonic mean of precision and recall. [1]

(2) It is useful when both false alarms and missed positives matter and neither precision nor recall should be ignored. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain one reason for checking performance across user groups. [2]

Answer:

(1) A high overall score can hide poor performance for a smaller group. [1]

(2) Group-wise evaluation helps reveal unequal errors that may require better data or design. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Evaluating Models: error conventions and interpretation

Classification accuracy versus prediction error

For classification, accuracy is the fraction of examples whose predicted class is correct. A wrong class is an error regardless of how its class code is written. For a numerical prediction, an exercise may instead compare how far the prediction is from the actual value. These are **different evaluation conventions**.

Let actual value be A and predicted value be P. Absolute error = |P-A|. For A > 0, relative error = |P-A|/A. If a question defines a score as 1 - relative error, use that stated convention. This score differs from confusion-matrix accuracy and is not a universal regression metric.

Quantity	Original worked example
Actual / predicted	A=25,000; P=24,250.
Absolute error	|24,250-25,000|=750.
Relative error	750/25,000=0.03=3%.
Stated 1-error score	1-0.03=0.97=97%.

If A=0, division by A is undefined. If error exceeds A, 1-error can be negative. These limitations show why a formula must be identified before applying it; do not force every prediction score into a confusion matrix.

Mean of per-example scores: do not change the question

Actual	Predicted	Relative error	1-error score
100	90	0.10	0.90
200	220	0.10	0.90
400	380	0.05	0.95

If asked to average the three displayed scores: (0.90+0.90+0.95)/3=0.9167, or approximately **91.67%**. Computing 1-(10+20+20)/(100+200+400)=92.86% gives a differently weighted calculation. It is not the arithmetic mean requested. Follow the stated method.

Build a confusion matrix from observations

Take **defective** as the positive class. For eight items, actual labels are [1,0,1,0,1,0,0,1] and predicted labels are [1,1,0,0,1,0,1,1]. Here 1 means defective and 0 means not defective. Count outcomes one item at a time rather than infer them from the number of predicted positives.

Actual / Predicted	Positive	Negative
Positive	TP=3	FN=1
Negative	FP=2	TN=2

Total=8; correct predictions=3+2=5; wrong predictions=2+1=3. Accuracy=5/8=62.50%. Precision=3/(3+2)=60%. Recall=3/(3+1)=75%. F1=2TP/(2TP+FP+FN)=6/9=66.67%. A transposed matrix is equivalent only when row and column labels are changed consistently.

Choose the metric for the consequences

Concern	Reasoned emphasis
Missing actual positives is particularly costly.	Inspect recall and false negatives, while also reporting the other errors.
False positive alerts cause substantial disruption.	Inspect precision and false positives.
Both precision and recall matter.	F1 summarises their harmonic balance, but still report the underlying counts.
One class is very rare.	Do not rely on overall accuracy alone; inspect per-class outcomes.

No metric replaces reasoning about use. A high recall system can still generate many false alarms. An overall accuracy of 99% can hide failure on a rare class. For example, if 10 of 1,000 cases are positive and a model predicts all cases negative, TN=990 and FN=10: accuracy is 99%, but recall is 0%. Precision has a zero denominator because there are no predicted positives; it must not be invented as 100%.

Training, testing and honest evaluation

For 750 examples, an illustrative 80:20 split gives **600 training and 150 test examples**. It is a choice, not a compulsory ratio for every project. Test data must not be used to fit the model or repeatedly choose a model while still being advertised as an untouched final test. Representative splitting and relevant grouping/time constraints matter.

Underfitting means the model fails to capture useful relationships and often performs poorly on both sets. **Overfitting** means it learns training-specific patterns that generalise poorly. A large training-test performance gap is evidence to investigate, not a full diagnosis without context.

Ethical evaluation: bias, transparency and accuracy

Bias may arise from who is represented, how targets are labelled and how results are reported. Check errors by relevant groups, document the evaluation method and disclose limitations. Transparency makes assumptions and results understandable; accuracy measures a particular aspect of correctness. Neither automatically guarantees the other. Review impacts in hiring, healthcare or finance rather than equating a prediction with a final decision.

Additional exam-style practice

C8-Q1. Actual price is 25,000 and predicted price is 24,250. Using |P-A|/A, the error rate is: [1]

(a) 30%
(b) 3%
(c) 97%
(d) 750%

C8-Q2. For 750 records, an 80:20 split gives training and test counts of: [1]

(a) 600 and 150
(b) 150 and 600
(c) 650 and 100
(d) 500 and 250

C8-Q3. Why can 99% accuracy be misleading when only 1% of examples are positive? [2]

Model answer: Predicting every example negative gives 99% accuracy but misses every positive. Recall is zero, so the system may fail its important detection task.

Suggested practice marking: Explain all-negative result 1; explain missed positives/recall consequence 1.

C8-Q4. A matrix has TP=3, FN=1, FP=2 and TN=2. Draw it with actual classes on rows. Find accuracy and the number of errors. [4]

Model answer: The labelled matrix has row Positive: TP=3, FN=1; row Negative: FP=2, TN=2. Total=8. Accuracy=(TP+TN)/8=5/8=62.50%. Errors=FP+FN=2+1=3.

Suggested practice marking: Correctly positioned four labelled cells 0.5 each (=2); accuracy formula/substitution 0.5 and result 0.5; error formula/substitution 0.5 and result 0.5.

C8-Q5. For TP=28, FP=7, FN=12 and TN=53, find accuracy, precision, recall and F1. [4]

Model answer: N=100. Accuracy=(28+53)/100=81%. Precision=28/(28+7)=80%. Recall=28/(28+12)=70%. F1=2(28)/(56+7+12)=56/75=74.67%.

Suggested practice marking: Each metric: correct formula with substitution 0.5 and result 0.5. Total 4; equivalent correctly labelled fractions are acceptable.

MCQ answers and reasons

C8-Q1: (b). The stated exercise uses relative numerical error.

C8-Q2: (a). The remaining 20% is 150.

IMPORTANT QUESTIONS WITH ANSWERS

Evaluating Models

Q1. Define TP, TN, FP and FN for a positive class. [4]

Result	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Suggested marking: *1 each.*

Q2. Given TP=30, TN=50, FP=10, FN=10, calculate accuracy, precision, recall and F1. [4]

Answer: N=100. Accuracy=(TP+TN)/N=80%. Precision=TP/(TP+FP)=30/40=75%. Recall=TP/(TP+FN)=30/40=75%. F1=2PR/(P+R)=75%.

Suggested marking: *1 per correct formula/substitution and value for four metrics.*

Q3. Explain why accuracy can be misleading on imbalanced data. [2]

Answer: If 99% of examples belong to one class, always predicting that majority class gets 99% accuracy but detects none of the minority examples. Examine recall, precision and the confusion matrix.

Suggested marking: *1 explanation + 1 illustration or corrective metric.*

Q4. Define precision and recall, and identify when each matters. [4]

Answer: Precision = TP/(TP+FP), measuring reliability of positive predictions; important when false alarms are costly. Recall = TP/(TP+FN), measuring detection of actual positives; important when missed cases are costly.

Suggested marking: *1 per formula + 1 per context.*

Q5. What is a train-test split, and why is it necessary? [2]

Answer: A training set teaches the model patterns; a separate test set estimates performance on unseen data. Separation helps detect overfitting and reduces misleading evaluation.

Suggested marking: *1 method + 1 purpose.*

Q6. Distinguish false positive and false negative in a spam filter. [2]

Answer: FP: a legitimate email is incorrectly marked as spam. FN: a spam email is incorrectly allowed into the inbox.

Suggested marking: *1 each.*

FOUR-MARK NUMERICAL STUDIO

Solved training examples | Evaluation and NLP

Read the task before calculating

Evaluation: name the positive class, identify all four outcomes, calculate N, use the correct denominator and round at the end. NLP: fix preprocessing and vocabulary order, count repeated terms, distinguish TF from DF, and identify whether IDF means a ratio or its logarithm.

Reference matrix arrangement

Actual / Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

This layout uses actual class on rows. A correctly labelled transposed layout contains the same information. Confusion over the axes is a common reason for swapping FP and FN.

What a four-mark solution needs

For a four-metric question, show the formula, substitution and result for accuracy, precision, recall and F1. For a matrix question, draw and label the matrix before calculating the requested values. In NLP answers, include the requested vocabulary, vectors or frequency tables and formulas.

CM-01 | Confusion matrix and metrics

CM-01. The chosen positive class has TP=40, FP=5, FN=10 and TN=45. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 40	FN = 10
Negative	FP = 5	TN = 45

Total N = 40 + 5 + 10 + 45 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (40 + 45)/100 x 100 = 85.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 40/(40 + 5) x 100 = 88.89%. [1]

(c) Recall = TP/(TP + FN) x 100 = 40/(40 + 10) x 100 = 80.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 80/(80 + 5 + 10) x 100 = 84.21%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-02 | Confusion matrix and metrics

CM-02. For a binary classifier, TP=50, FP=10, FN=10 and TN=45. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 50	FN = 10
Negative	FP = 10	TN = 45

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (50 + 45)/115 x 100 = 82.61%. [1]

Wrong predictions = FP + FN = 10 + 10 = 20. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-03 | Confusion matrix and metrics

CM-03. The chosen positive class has TP=54, FP=6, FN=9 and TN=51. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 54	FN = 9
Negative	FP = 6	TN = 51

Total N = 54 + 6 + 9 + 51 = 120.

(a) Accuracy = (TP + TN)/N x 100 = (54 + 51)/120 x 100 = 87.50%. [1]

(b) Precision = TP/(TP + FP) x 100 = 54/(54 + 6) x 100 = 90.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 54/(54 + 9) x 100 = 85.71%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 108/(108 + 6 + 9) x 100 = 87.80%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-04 | Confusion matrix and metrics

CM-04. For a binary classifier, TP=63, FP=7, FN=21 and TN=109. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 63	FN = 21
Negative	FP = 7	TN = 109

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (63 + 109)/200 x 100 = 86.00%. [1]

Wrong predictions = FP + FN = 7 + 21 = 28. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-05 | Confusion matrix and metrics

CM-05. The chosen positive class has TP=28, FP=12, FN=7 and TN=53. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 28	FN = 7
Negative	FP = 12	TN = 53

Total N = 28 + 12 + 7 + 53 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (28 + 53)/100 x 100 = 81.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 28/(28 + 12) x 100 = 70.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 28/(28 + 7) x 100 = 80.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 56/(56 + 12 + 7) x 100 = 74.67%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-06 | Confusion matrix and metrics

CM-06. For a binary classifier, TP=81, FP=9, FN=18 and TN=92. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 81	FN = 18
Negative	FP = 9	TN = 92

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (81 + 92)/200 x 100 = 86.50%. [1]

Wrong predictions = FP + FN = 9 + 18 = 27. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-07 | Confusion matrix and metrics

CM-07. The chosen positive class has TP=24, FP=16, FN=6 and TN=74. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 24	FN = 6
Negative	FP = 16	TN = 74

Total N = 24 + 16 + 6 + 74 = 120.

(a) Accuracy = (TP + TN)/N x 100 = (24 + 74)/120 x 100 = 81.67%. [1]

(b) Precision = TP/(TP + FP) x 100 = 24/(24 + 16) x 100 = 60.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 24/(24 + 6) x 100 = 80.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 48/(48 + 16 + 6) x 100 = 68.57%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-08 | Confusion matrix and metrics

CM-08. For a binary classifier, TP=48, FP=8, FN=12 and TN=132. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 48	FN = 12
Negative	FP = 8	TN = 132

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (48 + 132)/200 x 100 = 90.00%. [1]

Wrong predictions = FP + FN = 8 + 12 = 20. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-09 | Confusion matrix and metrics

CM-09. The chosen positive class has TP=35, FP=5, FN=15 and TN=45. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 35	FN = 15
Negative	FP = 5	TN = 45

Total N = 35 + 5 + 15 + 45 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (35 + 45)/100 x 100 = 80.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 35/(35 + 5) x 100 = 87.50%. [1]

(c) Recall = TP/(TP + FN) x 100 = 35/(35 + 15) x 100 = 70.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 70/(70 + 5 + 15) x 100 = 77.78%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-10 | Confusion matrix and metrics

CM-10. For a binary classifier, TP=90, FP=15, FN=10 and TN=85. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 90	FN = 10
Negative	FP = 15	TN = 85

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (90 + 85)/200 x 100 = 87.50%. [1]

Wrong predictions = FP + FN = 15 + 10 = 25. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-11 | Confusion matrix and metrics

CM-11. The chosen positive class has TP=56, FP=14, FN=24 and TN=106. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 56	FN = 24
Negative	FP = 14	TN = 106

Total N = 56 + 14 + 24 + 106 = 200.

(a) Accuracy = (TP + TN)/N x 100 = (56 + 106)/200 x 100 = 81.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 56/(56 + 14) x 100 = 80.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 56/(56 + 24) x 100 = 70.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 112/(112 + 14 + 24) x 100 = 74.67%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-12 | Confusion matrix and metrics

CM-12. For a binary classifier, TP=42, FP=18, FN=8 and TN=132. [4]

(a) Construct a labelled confusion matrix. [2]

(b) Calculate accuracy. [1]

(c) Calculate the number of wrong predictions. [1]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 42	FN = 8
Negative	FP = 18	TN = 132

Matrix: 2 marks = 0.5 for each correctly identified and placed outcome. A transposed matrix is acceptable when actual/predicted headings and all cells are consistent.

Accuracy = (TP + TN)/N x 100 = (42 + 132)/200 x 100 = 87.00%. [1]

Wrong predictions = FP + FN = 18 + 8 = 26. [1]

For each of the last two parts: 0.5 for the correct formula/method and 0.5 for the correct result. Suggested practice allocation.

CM-13 | Confusion matrix and metrics

CM-13. The chosen positive class has TP=44, FP=6, FN=11 and TN=39. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 44	FN = 11
Negative	FP = 6	TN = 39

Total N = 44 + 6 + 11 + 39 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (44 + 39)/100 x 100 = 83.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 44/(44 + 6) x 100 = 88.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 44/(44 + 11) x 100 = 80.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 88/(88 + 6 + 11) x 100 = 83.81%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-14 | Confusion matrix and metrics

CM-14. The chosen positive class has TP=18, FP=12, FN=2 and TN=68. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 18	FN = 2
Negative	FP = 12	TN = 68

Total N = 18 + 12 + 2 + 68 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (18 + 68)/100 x 100 = 86.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 18/(18 + 12) x 100 = 60.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 18/(18 + 2) x 100 = 90.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 36/(36 + 12 + 2) x 100 = 72.00%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-15 | Confusion matrix and metrics

CM-15. The chosen positive class has TP=62, FP=8, FN=18 and TN=12. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 62	FN = 18
Negative	FP = 8	TN = 12

Total N = 62 + 8 + 18 + 12 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (62 + 12)/100 x 100 = 74.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 62/(62 + 8) x 100 = 88.57%. [1]

(c) Recall = TP/(TP + FN) x 100 = 62/(62 + 18) x 100 = 77.50%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 124/(124 + 8 + 18) x 100 = 82.67%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-16 | Confusion matrix and metrics

CM-16. The chosen positive class has TP=9, FP=1, FN=6 and TN=84. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 9	FN = 6
Negative	FP = 1	TN = 84

Total N = 9 + 1 + 6 + 84 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (9 + 84)/100 x 100 = 93.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 9/(9 + 1) x 100 = 90.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 9/(9 + 6) x 100 = 60.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 18/(18 + 1 + 6) x 100 = 72.00%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-17 | Confusion matrix and metrics

CM-17. The chosen positive class has TP=25, FP=25, FN=5 and TN=45. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 25	FN = 5
Negative	FP = 25	TN = 45

Total N = 25 + 25 + 5 + 45 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (25 + 45)/100 x 100 = 70.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 25/(25 + 25) x 100 = 50.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 25/(25 + 5) x 100 = 83.33%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 50/(50 + 25 + 5) x 100 = 62.50%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-18 | Confusion matrix and metrics

CM-18. The chosen positive class has TP=50, FP=10, FN=20 and TN=20. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 50	FN = 20
Negative	FP = 10	TN = 20

Total N = 50 + 10 + 20 + 20 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (50 + 20)/100 x 100 = 70.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 50/(50 + 10) x 100 = 83.33%. [1]

(c) Recall = TP/(TP + FN) x 100 = 50/(50 + 20) x 100 = 71.43%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 100/(100 + 10 + 20) x 100 = 76.92%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-19 | Confusion matrix and metrics

CM-19. The chosen positive class has TP=72, FP=18, FN=3 and TN=7. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 72	FN = 3
Negative	FP = 18	TN = 7

Total N = 72 + 18 + 3 + 7 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (72 + 7)/100 x 100 = 79.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 72/(72 + 18) x 100 = 80.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 72/(72 + 3) x 100 = 96.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 144/(144 + 18 + 3) x 100 = 87.27%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-20 | Confusion matrix and metrics

CM-20. The chosen positive class has TP=7, FP=13, FN=3 and TN=77. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 7	FN = 3
Negative	FP = 13	TN = 77

Total N = 7 + 13 + 3 + 77 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (7 + 77)/100 x 100 = 84.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 7/(7 + 13) x 100 = 35.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 7/(7 + 3) x 100 = 70.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 14/(14 + 13 + 3) x 100 = 46.67%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-21 | Confusion matrix and metrics

CM-21. The chosen positive class has TP=36, FP=4, FN=24 and TN=36. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 36	FN = 24
Negative	FP = 4	TN = 36

Total N = 36 + 4 + 24 + 36 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (36 + 36)/100 x 100 = 72.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 36/(36 + 4) x 100 = 90.00%. [1]

(c) Recall = TP/(TP + FN) x 100 = 36/(36 + 24) x 100 = 60.00%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 72/(72 + 4 + 24) x 100 = 72.00%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CM-22 | Confusion matrix and metrics

CM-22. The chosen positive class has TP=80, FP=5, FN=10 and TN=5. Calculate accuracy, precision, recall and F1 score. Show working and round percentages to two decimal places. [4]

Worked solution

Labelled confusion matrix

Actual / Predicted	Positive	Negative
Positive	TP = 80	FN = 10
Negative	FP = 5	TN = 5

Total N = 80 + 5 + 10 + 5 = 100.

(a) Accuracy = (TP + TN)/N x 100 = (80 + 5)/100 x 100 = 85.00%. [1]

(b) Precision = TP/(TP + FP) x 100 = 80/(80 + 5) x 100 = 94.12%. [1]

(c) Recall = TP/(TP + FN) x 100 = 80/(80 + 10) x 100 = 88.89%. [1]

(d) F1 = 2TP/(2TP + FP + FN) x 100 = 160/(160 + 5 + 10) x 100 = 91.43%. [1]

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

CHAPTER 09 | Computer Vision

Concepts and applications

Seeing, processing and interpreting

Computer Vision extracts useful information from images or videos. A camera senses light; a computer represents the captured data and processes it to produce task-related information. Image processing changes or improves an image, for example adjusting contrast. Computer Vision aims to interpret its contents. A system can use both.

Applications

Examples include visual product search, document reading, vehicle counting, image-based quality inspection and assistance with medical-image review. Identify the visual input and the required output in a case. A system that only analyses typed feedback is not primarily a vision task.

Pixels and resolution

A digital image is represented on a grid of pixels. A width of 320 and height of 240 gives 76,800 spatial pixels. Higher resolution provides more samples but does not automatically solve poor lighting, blur or model bias. Specify the units: pixels are not centimetres or file bytes.

Grayscale and RGB

A common 8-bit grayscale representation stores one intensity from 0 to 255 at each pixel. RGB uses separate red, green and blue channel values, often 8 bits each. An RGB image has three channel values per pixel, but the same spatial pixel count as a grayscale image of equal width and height.

Classification and detection

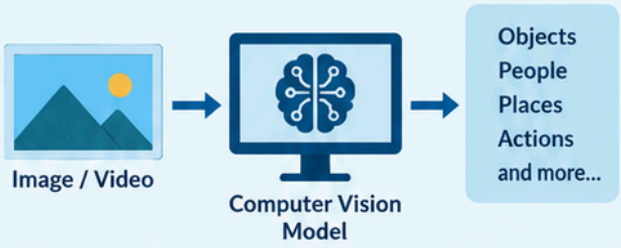
Image classification assigns a label to an image, such as "bicycle". Object detection identifies and localises objects, often with labelled bounding boxes. An image can contain several detections. A box gives an approximate region, not a precise boundary for every object pixel.

1 Computer Vision

Teaching computers to understand and make sense of visual information from the world, such as images and videos.

What is Computer Vision?

Computer Vision is a branch of Artificial Intelligence that enables computers to see, understand and interpret images or videos, similar to the way humans do. It uses algorithms and machine learning models to extract useful information from visual data.

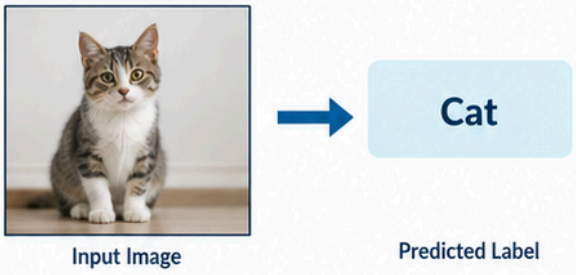


Major Tasks of Computer Vision

Computer vision can perform different tasks depending on what information we want to obtain from an image. The four major tasks are:

1 Image Classification

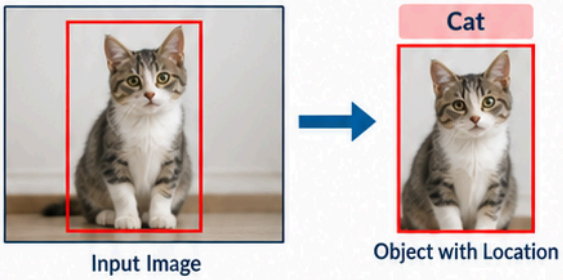
Identifies what object or scene is present in an image and assigns a single label to the whole image.



The model looks at the entire image and predicts the most likely class, such as cat, dog, car, etc.

2 Image Localization

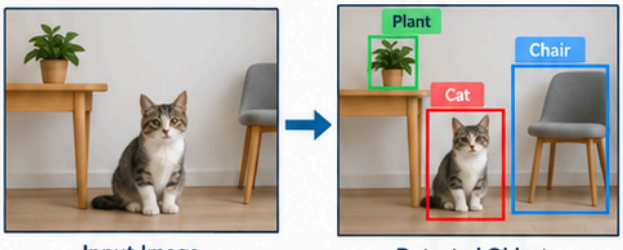
Identifies what object is present in the image and finds its location using a bounding box.



The model predicts the class (e.g., cat) and the location of the object using a bounding box.

3 Object Detection

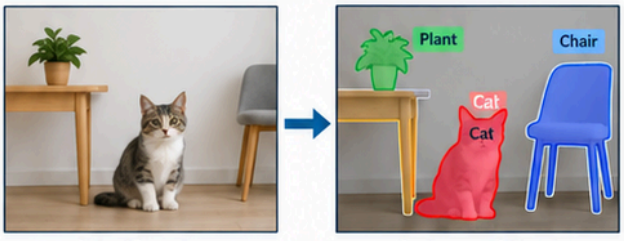
Identifies and locates multiple objects of different classes in an image using bounding boxes.



The model finds all the objects in the image, their classes and their locations using bounding boxes.

4 Image Segmentation

Identifies each object at the pixel level and separates different objects by coloring their regions.



The model assigns a class label to each pixel, so that every object is separated with a different color, giving precise boundaries.

Segmentation

Semantic segmentation assigns class labels at pixel level. Instance segmentation also distinguishes separate objects of the same class. Two nearby cars can share a semantic class but have different instance identities. State the output difference when asked to compare detection and segmentation.

Image features

Edges, shapes, textures and colour patterns help recognise objects. A feature representation describes the relevant visual properties used for the recognition task.

Limitations and responsible use

Illumination, viewpoint, occlusion, blur and unrepresentative data can affect performance. Face images raise privacy concerns. Test the system in realistic conditions and do not treat an image prediction as infallible.

CHAPTER 09 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A system assigns the label "bicycle" to an entire photograph. The task is: [1]

(a) Image classification

(b) Document translation

(c) Speech recognition

(d) Instance segmentation

Answer:

(a) Image classification A category is assigned to the whole image.

Q2. A system draws a separate bounding box around each visible bus. The task is: [1]

(a) Only image storage

(b) Object detection

(c) Sentiment analysis

(d) Text generation

Answer:

(b) Object detection Detection identifies and locates objects.

Q3. A system assigns a class label to each pixel but does not distinguish two cars of the same class. This is: [1]

(a) Speech synthesis

(b) Instance segmentation

(c) Semantic segmentation

(d) File compression

Answer:

(c) Semantic segmentation Semantic segmentation labels pixel classes without separate instances.

Q4. Which task gives different pixel masks to two separate people in one photograph? [1]

(a) Instance segmentation

(b) Colour inversion only

(c) Tokenisation

(d) Whole-image classification

Answer:

(a) Instance segmentation Instance segmentation separates individual objects at pixel level.

Q5. A digital image's smallest addressable picture element is a: [1]

(a) Pixel

(b) Token

(c) Record label

(d) Paragraph

Answer:

(a) Pixel A pixel stores image information at a spatial position.

Q6. An 8-bit grayscale channel has how many possible values? [1]

(a) 256

(b) 8

(c) 255

(d) 16

Answer:

(a) 256 Eight bits represent 2 to the power 8 values, from 0 to 255.

Q7. Which sequence best describes a vision pipeline? [1]

(a) Sensor, interpretation, creation of reality

(b) Interpretation, real scene, sensing device

(c) Stored answer, scene, sensor

(d) Real scene, sensing device, interpretation

Answer:

(d) Real scene, sensing device, interpretation The sensor captures visual data that the system processes.

Q8. Which task is optical character recognition designed to perform? [1]

(a) Predict future weather from time alone

(b) Extract text characters from an image

(c) Cluster music without labels

(d) Recognise every sound in a recording

Answer:

(b) Extract text characters from an image OCR converts visible characters into machine-readable text.

Q9. Why may a vision model fail on very dark images? [1]

(a) Useful visual detail may be lost or differ from training examples.

(b) Resolution must always become zero.

(c) Dark images cannot contain pixels.

(d) All labels are automatically deleted.

Answer:

(a) Useful visual detail may be lost or differ from training examples. Visibility and data distribution can affect recognition.

Q10. A grayscale pixel value of 0 usually represents: [1]

(a) A missing file

(b) White

(c) Pure red

(d) Black

Answer:

(d) Black Under the usual 8-bit convention, zero is black and 255 is white.

Q11. Which representation typically has three colour channels? [1]

(a) A word-count vector

(b) A single grayscale channel

(c) A file path

(d) RGB image

Answer:

(d) RGB image RGB stores red, green and blue components.

Q12. Adjusting brightness without deciding what objects are present is mainly: [1]

(a) Object understanding by definition

(b) Sentiment analysis

(c) Image processing

(d) Language translation

Answer:

(c) Image processing Pixel transformation need not include semantic interpretation.

Short-answer questions - 2 marks each

S1. Distinguish image classification from object detection. [2]

Answer:

(1) Image classification assigns a category to an entire image. [1]

(2) Object detection identifies objects and locates them, commonly using bounding boxes. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Distinguish semantic segmentation from instance segmentation. [2]

Answer:

(1) Semantic segmentation assigns a class label to each pixel. [1]

(2) Instance segmentation also separates individual objects belonging to the same class. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Define a pixel and image resolution. [2]

Answer:

(1) A pixel is an individual picture element containing image information. [1]

(2) Resolution describes image dimensions, commonly as width and height in pixels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Explain the usual 0 to 255 range of an 8-bit grayscale image. [2]

Answer:

(1) Eight bits allow 2 to the power 8, or 256, values. [1]

(2) Under the usual convention, 0 is black and 255 is white. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. Distinguish an RGB image from a grayscale image. [2]

Answer:

(1) An RGB image usually stores separate red, green and blue channel values. [1]

(2) A grayscale image stores one intensity value per pixel. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. State two reasons a visual model may fail in a new environment. [2]

Answer:

(1) Lighting, viewing angle or obstruction may hide important features. [1]

(2) The new images may differ from those represented in its training data. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Distinguish image processing from computer vision. [2]

Answer:

(1) Image processing changes or improves pixel data, for example adjusting brightness. [1]

(2) Computer vision interprets visual content, for example identifying and locating a vehicle. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. Define object localisation and explain how it differs from classification. [2]

Answer:

(1) Localisation identifies where an object is, often through coordinates or a bounding box. [1]

(2) Classification identifies its category without necessarily supplying its position. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Explain optical character recognition with a use case. [2]

Answer:

(1) Optical character recognition identifies text characters in an image and converts them into machine-readable text. [1]

(2) It can extract printed invoice details for later processing. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

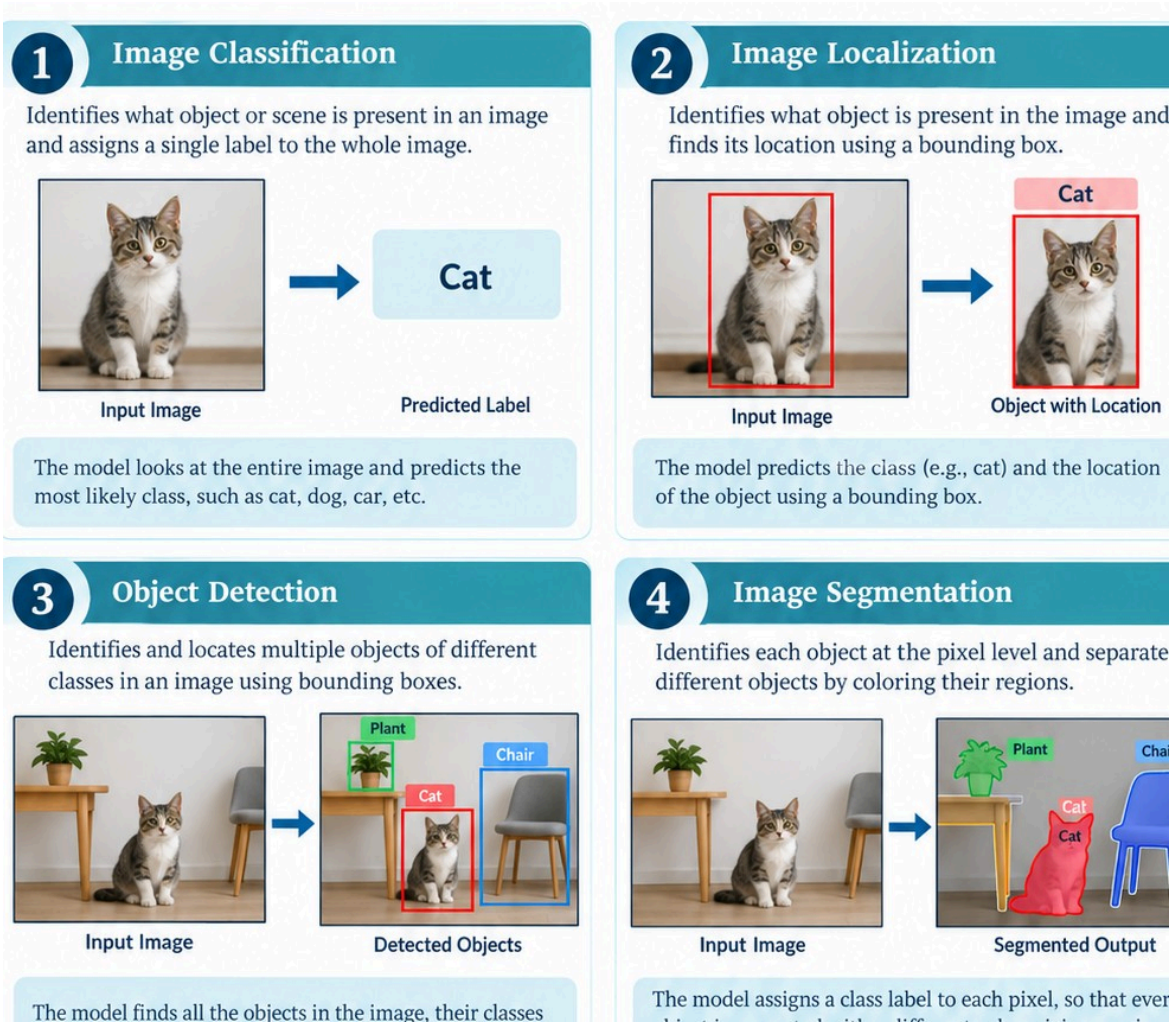
S10. State two data precautions for a face-recognition service. [2]

Answer:

(1) Collect only necessary face data for a clear, justified purpose. [1]

(2) Protect access and retention, and evaluate errors across representative conditions and intended user groups. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.



Computer Vision: representations and task distinctions

From an image to a task output

A camera supplies digital image data. A vision system prepares a suitable representation, extracts or learns useful visual features and produces a task-specific result. Features may concern edges, colour, texture or shapes. In a learned system, some useful features can be learned from examples rather than selected completely by hand.

Changing contrast or resizing an image is **image processing**; deciding whether it contains a crack is a **vision interpretation task**. A computer operates on represented values; it does not need to experience the image as a person does.

Pixels, resolution and channel values

A pixel is an element of a digital image grid. Spatial resolution is commonly stated as width by height in pixels. An image with width 160 and height 120 has **19,200 spatial pixels**. More pixels can represent finer spatial detail, but resolution alone does not guarantee a sharp or useful image.

Representation	Values at each spatial pixel
8-bit grayscale	One intensity value, normally 0 to 255; 0 is black and 255 is white in the usual display convention.
8-bit-per-channel RGB	Three channel values: red, green and blue, each normally 0 to 255.

Representation	Values at each spatial pixel
Binary image	Two possible values representing two categories, commonly black and white; encoding must be specified.

For the 160 x 120 RGB image, there are still **19,200 spatial pixels**, but 57,600 channel samples. In a simple uncompressed 8-bit-per-channel representation this uses 57,600 bytes for pixel data, excluding headers and metadata. A compressed file need not have that size. This calculation is an explanatory extension of image representation.

RGB example	Usual displayed colour
(255,0,0)	Red
(0,255,0)	Green
(0,0,255)	Blue
(0,0,0)	Black
(255,255,255)	White
(128,128,128)	A middle gray

Classification, localisation, detection and segmentation

Task	What is returned for a photograph
Image classification	One or more class labels for the image; e.g., whether a bicycle is present.
Classification + localisation	A class and a location, commonly a bounding box, for the target object.
Object detection	Locations and labels for one or more objects in the image.
Semantic segmentation	A class label for each pixel; objects of the same class are not necessarily separated into individual instances.
Instance segmentation	A separate pixel mask for each identified object instance.

In a photograph containing three bicycles, classification may say "bicycle"; detection may draw three labelled boxes; semantic segmentation may label bicycle pixels; instance segmentation distinguishes the three individual bicycle masks. A bounding box is rectangular; a segmentation mask can follow an object's outline.

Features, limitations and applications

A feature helps distinguish patterns relevant to the task. Edge arrangements may help identify shape; colour distributions may help separate regions; texture may help find surface irregularities. These clues can be affected by lighting, viewpoint, occlusion and background. A model should be tested under the intended conditions rather than only on neat demonstration images.

Applications include inspecting products for visible defects, reading digits from images, identifying objects in transport scenes and estimating plant condition from photographs. The output needs

suitable checking, especially when it influences safety, access or people's rights.

Additional exam-style practice

C9-Q1. A system returns the category of a single object and a box around it. The task is: [1]

(a) Bag of Words
(b) Classification plus localisation
(c) Text normalisation
(d) Train-test splitting

C9-Q2. How many spatial pixels are present in a 160 x 120 RGB image? [1]

(a) 57,600
(b) 480
(c) 19,200
(d) 280

C9-Q3. Distinguish object detection and instance segmentation. [2]

Model answer: Object detection gives object labels and locations, commonly boxes. Instance segmentation gives separate pixel-level masks for individual object instances.

Suggested practice marking: Detection 1; instance segmentation 1.

C9-Q4. State two differences between common 8-bit grayscale and RGB representations. [2]

Model answer: Grayscale uses one intensity channel per pixel. RGB uses three colour channels per pixel; each channel normally has a value from 0 to 255.

Suggested practice marking: Grayscale representation 1; RGB representation 1.

C9-Q5. Explain why high image resolution alone does not guarantee reliable recognition. [2]

Model answer: Resolution supplies spatial samples, not guaranteed useful evidence. Blur, poor lighting, occlusion or unsuitable training data can still prevent reliable recognition.

Suggested practice marking: Resolution versus useful evidence 1; relevant limitation with reasoning 1.

MCQ answers and reasons

C9-Q1: (b). The output combines classification with location.

C9-Q2: (c). 160 multiplied by 120 is 19,200.

IMPORTANT QUESTIONS WITH ANSWERS

Computer Vision

Q1. Differentiate image classification, object detection, and segmentation. [3]

Task	Output
Image classification	A label for the whole image
Object detection	Labels and bounding boxes
Segmentation	Pixel-level regions or masks

Suggested marking: *1 each.*

Q2. What is a pixel? Explain resolution. [2]

Answer: A pixel is the smallest addressable picture element in a raster image. Resolution describes image dimensions in pixels, such as 1920 × 1080.

Suggested marking: *1 each.*

Q3. What are grayscale and RGB images? [2]

Answer: A grayscale image stores intensity values for shades of gray. An RGB image uses red, green and blue channels to represent colours.

Suggested marking: *1 each.*

Q4. Explain four responsible uses or risks of facial-recognition systems. [4]

Answer: Identify a legitimate purpose; obtain appropriate consent/legal authority; restrict biometric retention and access; test accuracy and demographic bias before deployment.

Suggested marking: *1 each.*

Q5. Distinguish object localization from object detection. [2]

Answer: Localization identifies the location of a primary object, often using one bounding box. Detection identifies and locates multiple object instances and their classes.

Suggested marking: *1 each.*

CHAPTER 10 | Natural Language Processing

Language as an AI input and output

NLP develops methods for computers to analyse, interpret or generate human language. Language contains ambiguity and depends on context. Recognising words alone does not establish the speaker's intention.

Why NLP is useful

It supports communication through text and speech, helps interpret requests and enables useful language-based services.

Applications

Application	What the system does
Automatic captions	Turns spoken content into readable subtitles.
Voice assistants	Interprets a spoken request to perform a task.
Translation	Produces a message in another language.
Sentiment analysis	Estimates whether expressed opinion is positive, negative or neutral.
Keyword extraction	Selects informative terms from text.
Text classification	Assigns a text category, such as topic or spam status.

Class X Artificial Intelligence

Natural Language Processing (NLP)



What is NLP?

Natural Language Processing (NLP) is a branch of Artificial Intelligence that helps computers understand, interpret and generate human language. It enables machines to work with text or speech, such as reading documents, analysing sentiment, answering questions and more.

1 Bag-of-Words Representation

The bag-of-words model represents a document as a collection of word counts, ignoring grammar and the order of words. Each document is converted into a vector of word frequencies from a fixed vocabulary.

2 Example Sentences

Document 1 (D1)	The cat sits on the mat.
Document 2 (D2)	The dog sits on the log.
Document 3 (D3)	The cat and the dog are friends.

3 Tokenization

Each sentence is split into individual words (tokens), usually by removing punctuation and converting to lowercase.

D1 tokens	the, cat, sits, on, the, mat
D2 tokens	the, dog, sits, on, the, log
D3 tokens	the, cat, and, the, dog, are, friends

4 Vocabulary

The vocabulary is the list of all unique words from the documents.

Vocabulary (V)	and, are, cat, dog, friends, log, mat, on, sits, the (10 unique words)

5 Word-Frequency Vectors

Each document is represented by a vector showing the frequency of each vocabulary word in that document.

Document	Vocabulary Words (V)									
	and	are	cat	dog	friends	log	mat	on	sits	the
D1: The cat sits on the mat.	0	0	1	0	0	0	1	1	1	2
D2: The dog sits on the log.	0	0	0	1	0	1	0	1	1	2
D3: The cat and the dog are friends.	1	1	1	1	1	0	0	0	0	2

6 TF-IDF Pipeline (Overview)

TF-IDF (Term Frequency – Inverse Document Frequency) gives higher weight to words that are important in a document but not common in all documents.

1. Preprocessing
- Lowercase text
- Remove punctuation
- Tokenize into words

→

2. Compute Term Frequency (TF)
- Count how many times each word appears in a document

→

3. Compute Inverse Document Frequency (IDF)
- Gives higher weight to rare words across documents

→

4. Compute TF-IDF
- TF-IDF = TF × IDF
- Result: a weighted vector for each document

By Mr. Taranjeet Singh HOD IT-CS, Mr. Ravindra Uniyal PGT CS

Text Processing and Normalisation

Text processing prepares language data for analysis. In this chapter the main operations are normalisation followed by a numerical representation such as Bag of Words.

Corpus: the collection of documents being processed. A document is one text within that collection.

Operation	Purpose and example
Sentence segmentation	Separates a passage into sentences. "Birds fly. Fish swim." gives two sentences.
Tokenisation	Divides a sentence into units such as words, numbers or punctuation, according to the stated rule.
Stop-word and symbol handling	Removes selected common words, symbols or numbers when they are unnecessary for the task.
Common case	Converts case consistently: "School" and "school" can share one vocabulary entry.
Stemming	Trims word forms by rules and may produce a non-dictionary stem: studies -> studi.
Lemmatisation	Finds a suitable dictionary base form: studies -> study.

Apply only the operations requested. Keep repeated tokens when counting frequencies. Stemming and lemmatisation are alternative reduction methods; both need not be applied.

TF, DF, IDF and TF-IDF

Quantity	Meaning / formula
TF(w,d)	Raw count of word w in document d.
DF(w)	Number of documents containing w at least once.
N	Total number of documents in the corpus.
IDF(w)	For this chapter, the ratio N / DF(w).
TF-IDF(w,d)	TF(w,d) x log10(IDF(w)).

Worked calculation

Suppose a corpus contains 3 documents. A word occurs twice in D1 and appears in 2 documents overall. TF = 2, DF = 2, IDF = 3/2 = 1.5. Its weight in D1 is 2 x log10(1.5) = 0.3522 (approximately).

Interpreting weights

A word frequent in one document but less widespread across the corpus can receive a larger weight. If it occurs in all documents, IDF = 1 and its logarithmic weight is zero.

Uses of TF-IDF

Weighted terms support document classification, topic analysis and retrieval of relevant documents. Low weights can help identify common, less discriminative terms for filtering; filtering remains a task-specific choice.

CHAPTER 10 | QUESTIONS WITH IMMEDIATE ANSWERS

Objective questions - 1 mark each

Q1. A system assigns a complaint to "delivery", "payment" or "product". This is: [1]

(a) Regression of pixel values

(b) Text classification

(c) Face detection

(d) Image segmentation

Answer:

(b) Text classification The system assigns a text category.

Q2. What is removed when a process deletes a specified list of common words? [1]

(a) Stop words

(b) All proper nouns automatically

(c) Every meaningful term

(d) All punctuation only

Answer:

(a) Stop words Stop-word removal uses a chosen list of common terms.

Q3. A Bag-of-Words vector normally preserves: [1]

(a) Term counts in a fixed vocabulary order

(b) The complete original sentence order

(c) Every grammatical relationship

(d) Tone of voice automatically

Answer:

(a) Term counts in a fixed vocabulary order The positions refer to vocabulary entries, not original word order.

Q4. Which choice describes lemmatisation? [1]

(a) Deleting random letters

(b) Assigning words random colours

(c) Counting image pixels

(d) Mapping an inflected form to an appropriate dictionary base form

Answer:

(d) Mapping an inflected form to an appropriate dictionary base form Lemmatisation aims at meaningful base forms using linguistic information.

Q5. A menu-driven bot responds through fixed paths. It is mainly a: [1]

(a) Script bot

(b) Learning-based bot by definition

(c) Regression dataset

(d) Vision sensor

Answer:

(a) Script bot The response paths are explicitly designed.

Q6. A sentence "Wonderful, the train is late again" expresses frustration despite a positive word. The difficulty is: [1]

(a) Sarcasm and context

(b) A missing file extension

(c) Image size

(d) Disk capacity

Answer:

(a) Sarcasm and context The intended sentiment differs from a literal word-level reading.

Q7. Which NLP stage concerns the grammatical structure of a sentence? [1]

(a) Colour segmentation

(b) Model compression

(c) Pixel normalisation

(d) Syntactic analysis

Answer:

(d) Syntactic analysis Syntax concerns how words are structurally combined.

Q8. A text system resolves what "she" refers to using the previous sentence. This requires: [1]

(a) Single-pixel classification

(b) Hardware cleaning

(c) Discourse integration

(d) Alphabetical sorting

Answer:

(c) Discourse integration Meaning can depend on relationships across sentences.

Q9. A smart assistant interprets "Could you open the window?" as a request. This uses: [1]

(a) Image thresholding

(b) Folder creation

(c) Raw character count only

(d) Pragmatic interpretation

Answer:

(d) Pragmatic interpretation Pragmatics considers intended meaning in context.

Q10. Document frequency of a term means: [1]

(a) Number of documents containing it

(b) Number of letters in it

(c) Only its count in the first document

(d) Its total count across all documents

Answer:

(a) Number of documents containing it Each containing document contributes once to DF.

Q11. Which information is needed before comparing two Bag-of-Words vectors? [1]

(a) Only the number of pages

(b) The same ordered vocabulary for both

(c) The colour of the text

(d) A different vocabulary for each vector

Answer:

(b) The same ordered vocabulary for both Vector positions must refer to the same terms.

Q12. A word appears in all N documents. Under the school convention, its IDF ratio N/DF is: [1]

(a) 1

(b) N squared

(c) 0

(d) Undefined

Answer:

(a) 1 Here DF equals N, so the ratio is one.

Q13. Why can a basic Bag-of-Words representation confuse opposite statements? [1]

(a) It stores only images.

(b) It always removes every negative word.

(c) It discards word order and much grammatical structure.

(d) It cannot store zeroes.

Answer:

(c) It discards word order and much grammatical structure. Order-sensitive meaning can be lost even if all words are retained.

Q14. Which task selects the important terms that represent a document's subject? [1]

(a) Screen locking

(b) Instance segmentation

(c) Pixel interpolation

(d) Keyword extraction

Answer:

(d) Keyword extraction Keyword extraction identifies useful content terms.

Q15. A term occurs three times in one document. Using raw-count TF, its term frequency there is: [1]

(a) The number of documents only

(b) 0

(c) 1

(d) 3

Answer:

(d) 3 Raw-count TF records repeated occurrences in that document.

Short-answer questions - 2 marks each

S1. Distinguish tokenisation from stop-word removal. [2]

Answer:

(1) Tokenisation splits text into units such as words. [1]

(2) Stop-word removal deletes terms from a specified common-word list after appropriate processing. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S2. Explain one use and one limitation of Bag of Words. [2]

Answer:

(1) It converts text into fixed-length count vectors that a model can process. [1]

(2) It usually discards word order and grammatical relationships. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S3. Distinguish stemming from lemmatisation. [2]

Answer:

(1) Stemming applies rules that trim words and may create non-dictionary stems. [1]

(2) Lemmatisation aims to return an appropriate dictionary base form. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S4. Distinguish a script bot from a smart bot. [2]

Answer:

(1) A script bot follows predefined response paths. [1]

(2) A smart bot can use NLP and learned patterns to handle varied inputs, but may still make errors. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S5. Distinguish term frequency and document frequency. [2]

Answer:

(1) Term frequency counts occurrences of a word within one document. [1]

(2) Document frequency counts the number of documents containing that word. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S6. Explain why sentiment analysis must consider context. [2]

Answer:

(1) Positive words may occur in sarcastic or negative statements. [1]

(2) Context and negation can change the intended sentiment beyond individual word meanings. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S7. Explain semantic analysis with one example. [2]

Answer:

(1) Semantic analysis examines meaning in words and sentences. [1]

(2) For example, it distinguishes a financial bank from a river bank using relevant information from the text. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S8. Explain discourse integration with one example. [2]

Answer:

(1) Discourse integration connects meaning across sentences rather than treating each independently. [1]

(2) For example, it identifies whom "she" refers to from the preceding sentence. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S9. Explain pragmatic analysis with one example. [2]

Answer:

(1) Pragmatic analysis interprets intended meaning using the situation and purpose. [1]

(2) For example, "Could you close the door?" is understood as a request, not merely an ability question. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

S10. Explain why a common word can receive a low TF-IDF weight. [2]

Answer:

(1) A word appearing in many documents has a high document frequency. [1]

(2) Its inverse-frequency contribution is therefore small, reducing its ability to distinguish one document from others. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

4 Vocabulary

The vocabulary is the list of all unique words from the documents.

Vocabulary (V)	and, are, cat, dog, friends, log, mat, on, sits, the (10 unique words)

5 Word-Frequency Vectors

Each document is represented by a vector showing th frequency of each vocabulary word in that document

Document	Vocabulary Words (V)									
	and	are	cat	dog	friends	log	mat	on	sits	the
D1: The cat sits on the mat.	0	0	1	0	0	0	1	1	1	2
D2: The dog sits on the log.	0	0	0	1	0	1	0	1	1	2
D3: The cat and the dog are friends.	1	1	1	1	1	0	0	0	0	2

Five Stages of Language Analysis

Stage	What it examines
Lexical analysis	Identifies words or tokens. A lexicon is the vocabulary available in a language or language resource.
Syntactic analysis	Examines grammatical structure and relationships, such as subject and verb.
Semantic analysis	Examines the meaning of words and their combinations.
Discourse integration	Connects meaning across sentences; for example, resolves who "she" refers to using the preceding sentence.
Pragmatic analysis	Interprets intended meaning using the situation; "Could you close the door?" can be a request.

Script Bots and Smart Bots

Aspect	Script bot	Smart bot
Response method	Predefined rules or dialogue paths.	NLP and data-based patterns can handle varied wording.
Suitable use	Narrow, predictable enquiries.	More flexible language interaction.
Limitation	Unexpected wording can break the flow.	Needs suitable data and evaluation; can still make errors.

Bag of Words - Worked Vector Example

An NLP numerical: preprocessing, vocabulary, vectors and interpretation.

Documents: D1 = "AI helps students"; D2 = "students learn AI"; D3 = "AI helps AI". Lowercase all words and keep repeats.

Vocabulary in alphabetical order: ai, helps, learn, students

Document	ai	helps	learn	students
D1	1	1	0	1
D2	1	0	1	1
D3	2	1	0	0

How to award four practice marks

Step	Suggested marks
Correct preprocessing and distinct vocabulary	1
Fixed vocabulary order applied to every document	1
Correct D1 and D2 vectors	1
Correct D3 count vector (including repeated "AI")	1

Limitation: Bag of Words usually ignores word order and therefore loses meaning such as who did what to whom.

Questions with Immediate Answers

R1. Define a lexicon and sentence segmentation. [2]

Answer: A lexicon supplies the vocabulary of a language or resource. Sentence segmentation separates a passage into individual sentences.

Suggested marking: 1 mark for each correct definition.

R2. Name the five stages of NLP and explain any three. [4]

Answer: Lexical analysis, syntactic analysis, semantic analysis, discourse integration and pragmatic analysis. Lexical analysis identifies words or tokens; syntactic analysis examines grammar; semantic analysis examines meaning.

Suggested marking: 1 for the complete list; 1 each for three correct explanations.

R3. State two uses of TF-IDF. [2]

Answer: It helps rank relevant documents in information retrieval and supplies weighted word features for document classification.

Suggested marking: 1 mark for each distinct use.

R4. Why should all document vectors use the same vocabulary order? [2]

Answer: Each position must represent the same word in every vector. This allows the counts to be compared consistently across documents.

Suggested marking: 1 for fixed word positions; 1 for consistent comparison.

Bag of Words: Building Vectors

A Bag-of-Words representation uses a shared vocabulary to convert text into numerical features. A count vector records occurrences; a binary vector records only presence or absence. Use the representation specified in the question.

Method

1. Apply the stated preprocessing.
2. Form one vocabulary of distinct words.
3. Fix its order for every document.
4. Count each vocabulary word in each document, including zeros.

Worked corpus

D1: "learn ai learn"
D2: "ai helps people"
D3: "people learn"
All text is already lowercase. Remove no stop words and do not stem. Use vocabulary [learn, ai, helps, people].

Document	learn	ai	helps	people
D1	2	1	0	0
D2	0	1	1	1
D3	1	0	0	1

The vocabulary contains each word once; its document count may exceed 1. Standard BoW does not preserve the original word order. The frequency and weighting calculations continue on the following page.

Document	learn	ai	helps	people
D1	2	1	0	0
D2	0	1	1	1
D3	1	0	0	1

The count vectors are D1=[2,1,0,0], D2=[0,1,1,1] and D3=[1,0,0,1]. Every vector has the same four positions, including zeros. The vocabulary contains each term once; a document vector records how often it occurs in that document. BoW normally loses word order: rearranging the same words without changing counts gives the same count vector.

A **binary** vector would indicate presence or absence, changing D1 to [1,1,0,0]. That is not the count vector requested above. If a question fixes the vocabulary order, retain it. If it allows you to choose an order, state the order and use it consistently for all documents.

TF, DF, IDF and TF-IDF: keep the conventions distinct

Using raw-count TF, learn has TF = 2 in D1. Its DF = 2 because it appears in D1 and D3; repetitions within one document do not increase DF. Here N = 3. Record the IDF ratio N/DF, then calculate TF-IDF = TF x log10(N/DF).

Term	DF	N/DF	log10(N/DF)
learn	2	1.5	0.1761
ai	2	1.5	0.1761
helps	1	3	0.4771
people	2	1.5	0.1761

Worked Weights and Further Practice

For vocabulary [learn, ai, helps, people], the TF-IDF vectors are:
D1: [0.3522, 0.1761, 0, 0]
D2: [0, 0.1761, 0.4771, 0.1761]
D3: [0.1761, 0, 0, 0.1761].

C10-Q1. Resolving a pronoun using the previous sentence involves which stage? [1]

Options: (a) Discourse integration; (b) Pixel normalisation; (c) Object detection; (d) File indexing.

Answer: (a) Discourse integration.

Suggested marking: 1 for option (a).

C10-Q2. A word occurs five times in D1, twice in D2 and not in D3. What is its DF? [1]

Options: (a) 7; (b) 5; (c) 3; (d) 2.

Answer: (d) 2. Only two documents contain it.

Suggested marking: 1 for option (d).

Further Questions with Answers

C10-Q3. Distinguish syntactic and semantic analysis. [2]

Answer: Syntactic analysis examines grammatical structure and word relationships. Semantic analysis examines the meaning conveyed by words and their combinations.

Suggested marking: 1 for each distinction.

C10-Q4. Give the three count vectors for the worked corpus and state one BoW limitation. [4]

Answer: Vocabulary: [learn, ai, helps, people]. D1 = [2,1,0,0]; D2 = [0,1,1,1]; D3 = [1,0,0,1]. BoW does not preserve the original word order.

Suggested marking: 1 for each correct vector; 1 for the limitation.

C10-Q5. Find TF of learn in D1, its DF, IDF and TF-IDF weight. [4]

Answer: TF = 2; DF = 2; IDF = 3/2 = 1.5. TF-IDF = 2 x log10(1.5) = 0.3522 approximately.

Suggested marking: TF 1; DF 1; IDF 1; formula with substitution 0.5 and result 0.5.

IMPORTANT QUESTIONS WITH ANSWERS

Natural Language Processing

Q1. Name five stages of NLP and explain any three. [4]

Answer: Stages: lexical, syntactic, semantic, discourse integration and pragmatic analysis. Lexical analysis identifies tokens; syntax examines grammar; semantics examines meaning.

Suggested marking: List 1; three correct explanations 1 each.

Q2. What is text normalization? Name three operations. [4]

Answer: Text normalization prepares text in a consistent form. Operations may include case folding, punctuation handling, tokenization, stop-word handling, stemming or lemmatization.

Suggested marking: *1 definition + 1 each of three valid operations.*

Q3. Construct Bag-of-Words vectors for "AI helps school" and "school uses AI". [4]

Answer: After lowercasing, fix vocabulary order [ai, helps, school, uses]. D1 counts [1,1,1,0]. D2 counts [1,0,1,1]. A count vector records term frequencies, not word order.

Suggested marking: *1 vocabulary + 1 each document vector + 1 interpretation.*

Q4. Construct vectors for "data data helps" and "data helps AI". [4]

Answer: Vocabulary in order [ai, data, helps]. First document: [0,2,1]. Second document: [1,1,1]. The repeated word "data" contributes count 2 in the first vector.

Suggested marking: *1 vocabulary + 1 each vector + 1 repeated-token explanation.*

Q5. What is document frequency (DF)? How does it differ from term frequency (TF)? [2]

Answer: TF counts how often a term occurs in a particular document. DF counts how many documents in the collection contain that term at least once.

Suggested marking: *1 each.*

Q6. A corpus has four documents; "ai" occurs in two. Find IDF ratio and explain TF-IDF. [4]

Answer: N=4, DF(ai)=2. IDF = N/DF = 2. TF-IDF = TF x log10(IDF). For TF=1, the weight is log10(2), approximately 0.3010.

Suggested marking: N and DF 1; IDF ratio 1; weighting formula 1; explanation or calculation 1.

Q7. Differentiate stemming and lemmatization. [2]

Answer: Stemming heuristically cuts word endings and may produce a non-dictionary root. Lemmatization uses linguistic information to produce a valid base form.

Suggested marking: *1 each.*

Q8. Compare script bots and NLP-based chatbots. [2]

Answer: A script bot follows predefined dialogue rules and options. An NLP-based bot interprets flexible language patterns and can respond to varied wording, but may still make errors.

Suggested marking: *1 each.*

Q9. Explain why Bag of Words cannot distinguish "dog bites man" and "man bites dog". [2]

Answer: Both sentences have the same word-count vector when the same vocabulary is used. Bag of Words ignores token order and therefore loses who acted on whom.

Suggested marking: *1 same vector + 1 consequence.*

NLP-01 | Vocabulary and count vectors

NLP-01. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "green buses save fuel"

Document 2: "clean buses save time"

Document 3: "green travel saves time"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [green, buses, save, fuel, clean, time, travel, saves]. [1]

(b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
green	1	0	1
buses	1	1	0
save	1	1	0
fuel	1	0	0
clean	0	1	0
time	0	1	1
travel	0	0	1
saves	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-02 | Vocabulary and count vectors

NLP-02. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "robots sort small boxes"

Document 2: "small robots label boxes"

Document 3: "robots sort sort labels"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [robots, sort, small, boxes, label, labels]. [1]

(b) Document 3 vector: [1, 2, 0, 0, 0, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
robots	1	1	1
sort	1	0	2
small	1	1	0
boxes	1	1	0
label	0	1	0

Term	Doc 1	Doc 2	Doc 3
labels	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-03 | Vocabulary and count vectors

NLP-03. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "solar lamps light lanes"

Document 2: "solar panels charge lamps"

Document 3: "bright lamps light homes"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [solar, lamps, light, lanes, panels, charge, bright, homes]. [1]

(b) Document 3 vector: [0, 1, 1, 0, 0, 0, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
solar	1	1	0

Term	Doc 1	Doc 2	Doc 3
lamps	1	1	1
light	1	0	1
lanes	1	0	0
panels	0	1	0
charge	0	1	0
bright	0	0	1
homes	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-04 | Vocabulary and count vectors

NLP-04. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "clean water helps plants"

Document 2: "plants need clean air"

Document 3: "clean air helps people"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [clean, water, helps, plants, need, air, people]. [1]

(b) Document 3 vector: [1, 0, 1, 0, 0, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
clean	1	1	1
water	1	0	0
helps	1	0	1
plants	1	1	0
need	0	1	0
air	0	1	1
people	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-05 | Vocabulary and count vectors

NLP-05. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "smart bins collect waste"

Document 2: "teams sort waste"

Document 3: "smart teams clean bins"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [smart, bins, collect, waste, teams, sort, clean]. [1]

(b) Document 3 vector: [1, 1, 0, 0, 1, 0, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
smart	1	0	1
bins	1	0	1
collect	1	0	0
waste	1	1	0
teams	0	1	1
sort	0	1	0

Term	Doc 1	Doc 2	Doc 3
clean	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-06 | Vocabulary and count vectors

NLP-06. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "local trains save time"

Document 2: "local buses save fuel"

Document 3: "trains connect towns"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [local, trains, save, time, buses, fuel, connect, towns]. [1]

(b) Document 3 vector: [0, 1, 0, 0, 0, 0, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
local	1	1	0

Term	Doc 1	Doc 2	Doc 3
trains	1	0	1
save	1	1	0
time	1	0	0
buses	0	1	0
fuel	0	1	0
connect	0	0	1
towns	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-07 | Vocabulary and count vectors

NLP-07. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "readers share useful books"

Document 2: "useful books teach ideas"

Document 3: "readers discuss ideas"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [readers, share, useful, books, teach, ideas, discuss]. [1]

(b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
readers	1	0	1
share	1	0	0
useful	1	1	0
books	1	1	0
teach	0	1	0
ideas	0	1	1
discuss	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-08 | Vocabulary and count vectors

NLP-08. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "sensors track warm rooms"

Document 2: "sensors track cool rooms"

Document 3: "fans cool warm rooms"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [sensors, track, warm, rooms, cool, fans]. [1]

(b) Document 3 vector: [0, 0, 1, 1, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
sensors	1	1	0
track	1	1	0
warm	1	0	1
rooms	1	1	1
cool	0	1	1
fans	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-09 | Vocabulary and count vectors

NLP-09. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "careful drivers check brakes"

Document 2: "drivers check lights"

Document 3: "careful checks prevent faults"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [careful, drivers, check, brakes, lights, checks, prevent, faults]. [1]

(b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
careful	1	0	1
drivers	1	1	0

Term	Doc 1	Doc 2	Doc 3
check	1	1	0
brakes	1	0	0
lights	0	1	0
checks	0	0	1
prevent	0	0	1
faults	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

NLP-10 | Vocabulary and count vectors

NLP-10. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4]

Document 1: "artists mix blue paint"

Document 2: "blue paint covers walls"

Document 3: "artists paint paint signs"

(a) List the vocabulary in first-appearance order. [1]

(b) Construct the count vector for Document 3 in that order. [1]

(c) Explain how the vector is useful to a text classifier. [1]

(d) State one order-related limitation of this representation. [1]

Worked solution

(a) Vocabulary: [artists, mix, blue, paint, covers, walls, signs]. [1]

(b) Document 3 vector: [1, 0, 0, 2, 0, 0, 1]. [1]

Term	Doc 1	Doc 2	Doc 3
artists	1	0	1
mix	1	0	0
blue	1	1	0
paint	1	1	2
covers	0	1	0
walls	0	1	0
signs	0	0	1

The table shows the count at every vocabulary position. The vector is read down the Document 3 column in the listed term order. Other documents are shown as checking support, not additional assessed subparts.

(c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

TF-01 | TF, DF and IDF tables

TF-01. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. **[4]**

Document 1: "green buses save fuel"

Document 2: "clean buses save time"

Document 3: "green travel saves time"

(a) Find the raw term frequency of each word in Document 1. [1]

(b) Construct the document-frequency (DF) table for the corpus. [1]

(c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1]

(d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. [1]

Worked solution

N = 3 documents. TF means raw occurrence count, while DF counts documents containing a term.

Term	TF in D1	DF	IDF ratio N/DF
green	1	2	3/2 = 1.5
buses	1	2	3/2 = 1.5
save	1	2	3/2 = 1.5
fuel	1	1	3/1 = 3
clean	0	1	3/1 = 3

Term	TF in D1	DF	IDF ratio N/DF
time	0	2	3/2 = 1.5
travel	0	1	3/1 = 3
saves	0	1	3/1 = 3

(a) TF values for Document 1 are shown in the second column. Zero is recorded for a vocabulary term absent from that document. [1]

(b) The third column counts the documents in which each term appears at least once. Repeated occurrences within one document add only one to DF. [1]

(c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

TF-02 | TF, DF and IDF tables

TF-02. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. **[4]**

Document 1: "solar lamps light lanes"

Document 2: "solar panels charge lamps"

Document 3: "bright lamps light homes"

(a) Find the raw term frequency of each word in Document 1. [1]

(b) Construct the document-frequency (DF) table for the corpus. [1]

(c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1]

(d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. [1]

Worked solution

N = 3 documents. TF means raw occurrence count, while DF counts documents containing a term.

Term	TF in D1	DF	IDF ratio N/DF
solar	1	2	3/2 = 1.5
lamps	1	3	3/3 = 1
light	1	2	3/2 = 1.5
lanes	1	1	3/1 = 3
panels	0	1	3/1 = 3
charge	0	1	3/1 = 3
bright	0	1	3/1 = 3
homes	0	1	3/1 = 3

(a) TF values for Document 1 are shown in the second column. Zero is recorded for a vocabulary term absent from that document. [1]

(b) The third column counts the documents in which each term appears at least once. Repeated occurrences within one document add only one to DF. [1]

(c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

TF-03 | TF, DF and IDF tables

TF-03. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. **[4]**

Document 1: "smart bins collect waste"

Document 2: "teams sort waste"

Document 3: "smart teams clean bins"

(a) Find the raw term frequency of each word in Document 1. [1]

(b) Construct the document-frequency (DF) table for the corpus. [1]

(c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1]

(d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. [1]

Worked solution

N = 3 documents. TF means raw occurrence count, while DF counts documents containing a term.

Term	TF in D1	DF	IDF ratio N/DF
smart	1	2	3/2 = 1.5
bins	1	2	3/2 = 1.5
collect	1	1	3/1 = 3
waste	1	2	3/2 = 1.5
teams	0	2	3/2 = 1.5
sort	0	1	3/1 = 3
clean	0	1	3/1 = 3

(a) TF values for Document 1 are shown in the second column. Zero is recorded for a vocabulary term absent from that document. [1]

(b) The third column counts the documents in which each term appears at least once. Repeated occurrences within one document add only one to DF. [1]

(c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

TF-04 | TF, DF and IDF tables

TF-04. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. **[4]**

Document 1: "readers share useful books"

Document 2: "useful books teach ideas"

Document 3: "readers discuss ideas"

(a) Find the raw term frequency of each word in Document 1. [1]

(b) Construct the document-frequency (DF) table for the corpus. [1]

(c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1]

(d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. [1]

Worked solution

N = 3 documents. TF means raw occurrence count, while DF counts documents containing a term.

Term	TF in D1	DF	IDF ratio N/DF
readers	1	2	3/2 = 1.5
share	1	1	3/1 = 3
useful	1	2	3/2 = 1.5
books	1	2	3/2 = 1.5

Term	TF in D1	DF	IDF ratio N/DF
teach	0	1	3/1 = 3
ideas	0	2	3/2 = 1.5
discuss	0	1	3/1 = 3

(a) TF values for Document 1 are shown in the second column. Zero is recorded for a vocabulary term absent from that document. [1]

(b) The third column counts the documents in which each term appears at least once. Repeated occurrences within one document add only one to DF. [1]

(c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

TF-05 | TF, DF and IDF tables

TF-05. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. **[4]**

Document 1: "careful drivers check brakes"

Document 2: "drivers check lights"

Document 3: "careful checks prevent faults"

(a) Find the raw term frequency of each word in Document 1. [1]

(b) Construct the document-frequency (DF) table for the corpus. [1]

(c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1]

(d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. [1]

Worked solution

N = 3 documents. TF means raw occurrence count, while DF counts documents containing a term.

Term	TF in D1	DF	IDF ratio N/DF
careful	1	2	3/2 = 1.5
drivers	1	2	3/2 = 1.5
check	1	2	3/2 = 1.5
brakes	1	1	3/1 = 3
lights	0	1	3/1 = 3
checks	0	1	3/1 = 3
prevent	0	1	3/1 = 3
faults	0	1	3/1 = 3

(a) TF values for Document 1 are shown in the second column. Zero is recorded for a vocabulary term absent from that document. [1]

(b) The third column counts the documents in which each term appears at least once. Repeated occurrences within one document add only one to DF. [1]

(c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

WEIGHT-01 | Numerical TF-IDF

WEIGHT-01. In a corpus of 4 documents, "solar" occurs in 2 different documents and appears 3 times in Document A. Use raw TF and IDF ratio = N/DF. Use log10(2) = 0.3010. **[4]**

(a) State TF and DF. [1]

(b) Calculate the IDF ratio. [1]

(c) Write and substitute into the TF-IDF formula. [1]

(d) Calculate the TF-IDF weight. [1]

Worked solution

(a) TF = 3; DF = 2. [1: 0.5 each]

(b) IDF ratio = N/DF = 4/2 = 2. [1]

(c) TF-IDF = TF x log10(N/DF) = 3 x log10(2). [1]

(d) Using the supplied logarithm: 3 x 0.3010 = 0.9030. [1]

The result is a term weight, not a percentage. Using an unrounded logarithm may produce a slightly different last decimal; accept a result consistent with the supplied approximation.

WEIGHT-02 | Numerical TF-IDF

WEIGHT-02. In a corpus of 5 documents, "repair" occurs in 1 different documents and appears 2 times in Document A. Use raw TF and IDF ratio = N/DF. Use log10(5) = 0.6990. **[4]**

(a) State TF and DF. [1]

(b) Calculate the IDF ratio. [1]

(c) Write and substitute into the TF-IDF formula. [1]

(d) Calculate the TF-IDF weight. [1]

Worked solution

(a) TF = 2; DF = 1. [1: 0.5 each]

(b) IDF ratio = N/DF = 5/1 = 5. [1]

(c) TF-IDF = TF x log10(N/DF) = 2 x log10(5). [1]

(d) Using the supplied logarithm: 2 x 0.6990 = 1.3980. [1]

The result is a term weight, not a percentage. Using an unrounded logarithm may produce a slightly different last decimal; accept a result consistent with the supplied approximation.

WEIGHT-03 | Numerical TF-IDF

WEIGHT-03. In a corpus of 3 documents, "clean" occurs in 2 different documents and appears 2 times in Document A. Use raw TF and IDF ratio = N/DF. Use log10(1.5) = 0.1761. **[4]**

(a) State TF and DF. [1]

(b) Calculate the IDF ratio. [1]

(c) Write and substitute into the TF-IDF formula. [1]

(d) Calculate the TF-IDF weight. [1]

Worked solution

(a) TF = 2; DF = 2. [1: 0.5 each]

(b) IDF ratio = N/DF = 3/2 = 1.5. [1]

(c) TF-IDF = TF x log10(N/DF) = 2 x log10(1.5). [1]

(d) Using the supplied logarithm: 2 x 0.1761 = 0.3522. [1]

The result is a term weight, not a percentage. Using an unrounded logarithm may produce a slightly different last decimal; accept a result consistent with the supplied approximation.

WEIGHT-04 | Numerical TF-IDF

WEIGHT-04. In a corpus of 6 documents, "use" occurs in 6 different documents and appears 4 times in Document A. Use raw TF and IDF ratio = N/DF. Use log10(1) = 0.0000. **[4]**

(a) State TF and DF. [1]

(b) Calculate the IDF ratio. [1]

(c) Write and substitute into the TF-IDF formula. [1]

(d) Calculate the TF-IDF weight. [1]

Worked solution

(a) TF = 4; DF = 6. [1: 0.5 each]

(b) IDF ratio = N/DF = 6/6 = 1. [1]

(c) TF-IDF = TF x log10(N/DF) = 4 x log10(1). [1]

(d) Using the supplied logarithm: 4 x 0.0000 = 0.0000. [1]

The result is a term weight, not a percentage. Using an unrounded logarithm may produce a slightly different last decimal; accept a result consistent with the supplied approximation.

A term appearing in every document has IDF ratio 1 and log10(1)=0. Its weight is therefore zero under this unsmoothed school convention, even when its TF is large.

VECTOR CHECK | Count versus presence

V-01. Use vocabulary [lamp, lights, room]. Document A is "lamp lights room" and Document B is "lamp lamp lights room". **[4]**

(a) Write the count vector for A. [1]

(b) Write the count vector for B. [1]

(c) Write the binary presence vector for B. [1]

(d) Explain why the two representations differ. [1]

Worked solution

(a) A count vector = [1, 1, 1]. [1]

(b) B count vector = [2, 1, 1]. [1]

(c) B binary vector = [1, 1, 1]. [1]

(d) Count encoding preserves repeated occurrences; binary encoding records only whether a term appears at least once. [1]

VECTOR CHECK | Fixed preprocessing

V-02. Document A is "The small fan cools the room". Document B is "A fan cools a small room". Lowercase and remove exactly the stop words {the, a}. Use first-appearance vocabulary order. **[4]**

(a) Write the cleaned token lists. [1]

(b) Write the common vocabulary. [1]

(c) Write both count vectors. [1]

(d) State whether BoW preserves the remaining word order. [1]

Worked solution

(a) A: [small, fan, cools, room]; B: [fan, cools, small, room]. [1: 0.5 each]

(b) Vocabulary = [small, fan, cools, room]. [1]

(c) A = [1, 1, 1, 1]; B = [1, 1, 1, 1]. [1: 0.5 each]

(d) No. Standard count-based BoW ignores the order, so these differently ordered token lists produce identical vectors. [1]

W01. Define the AI project cycle and outline its five-stage formulation. [4]

(1) An AI project cycle is an iterative process that organises the development and evaluation of a proposed AI solution. [1]

(2) Problem scoping identifies the need and success criteria; data acquisition gathers relevant examples from suitable sources. [1]

(3) Data exploration examines patterns and quality; modelling selects and trains or builds an appropriate approach to the task. [1]

(4) Evaluation checks performance against the objectives, revealing errors that may require revisiting earlier stages of the cycle. [1]

W02. Give two reasons why the AI project cycle is useful. [2]

(1) It links data and modelling choices to a clearly scoped problem. [1]

(2) It provides checkpoints for detecting mistakes and improving the solution. [1]

W03. Explain big data and one reason its quality still matters. [2]

(1) Big data has scale, speed or variety that creates processing challenges. [1]

(2) Large volume does not remove errors or bias; unsuitable data can still produce misleading results. [1]

W04. Name the three AI application domains and give one application of each. [4]

(1) The three application domains are Statistical Data or Data Science, Computer Vision, and Natural Language Processing. [1]

(2) A data-oriented model can use recorded sales and relevant conditions to estimate the demand for a product. [1]

(3) A Computer Vision system can inspect a photograph of a parcel to identify visible signs of damage. [1]

(4) An NLP system can examine a written enquiry and route it to the department matching its meaning. [1]

W05. Explain data science and how AI can support it. [2]

(1) Data science collects, analyses and interprets data to investigate questions. [1]

(2) Machine learning can support prediction or pattern discovery within that work. [1]

W06. Define Computer Vision and give one application. [2]

(1) Computer Vision develops methods that enable computers to interpret information from digital images or video. [1]

(2) An image-inspection system can identify visibly damaged products. [1]

W07. Define NLP and give one application. [2]

(1) NLP develops methods for computers to process and interpret human language in text or speech. [1]

(2) A message classifier can route enquiries according to their topic. [1]

W08. Distinguish AI from machine learning. [2]

(1) AI is the broader field of systems that perform intelligent tasks. [1]

(2) Machine learning is an approach within AI that learns patterns from examples. [1]

W09. Distinguish machine learning from deep learning. [2]

(1) Machine learning includes methods that learn from data. [1]

(2) Deep learning is a subset using neural networks with multiple learned layers. [1]

W10. What is an ethical framework, and why use one in AI? [2]

(1) It is a structured approach for examining values, duties and effects of a decision. [1]

(2) It helps identify benefits, risks and safeguards before acting on an AI output. [1]

W11. Distinguish sector-based and value-based ethical frameworks. [2]

(1) A sector-based framework addresses the needs of a particular field, such as healthcare. [1]

(2) A value-based framework reasons from values, duties or consequences across decisions. [1]

W12. Explain why fairness is important in AI decisions. [2]

(1) Biased decisions can disadvantage people for reasons unrelated to the legitimate task. [1]

(2) Fair evaluation checks relevant groups and provides a way to review harmful outcomes. [1]

W13. Explain human-centred AI and one safeguard it supports. [2]

(1) Human-centred AI is designed around people's needs, dignity and meaningful control. [1]

(2) A review or appeal process helps people challenge a harmful automated recommendation. [1]

W14. Apply consequence-based, duty-based and virtue-based perspectives to allocating a public benefit, and suggest one safeguard. [4]

(1) A consequence-based perspective compares the likely benefits and harms of the allocation for the affected people. [1]

(2) A duty-based perspective checks whether the process respects eligibility rules, individual rights and relevant obligations. [1]

(3) A virtue-based perspective expects the decision-makers to act honestly, fairly and with care for vulnerable applicants. [1]

(4) Provide clear eligibility criteria and an accessible human review process so that an incorrect or unfair outcome can be challenged. [1]

W15. Name and explain the four bioethical principles. [4]

(1) Autonomy means respecting a person's informed choices and providing understandable information before decisions affecting that person. [1]

(2) Beneficence means using decisions and interventions to promote the welfare and legitimate interests of the people affected. [1]

(3) Non-maleficence means avoiding preventable harm and reducing foreseeable risks when developing or applying a healthcare system. [1]

(4) Justice means treating people fairly and allocating benefits, burdens and limited healthcare resources using defensible criteria. [1]

W16. Why is autonomy relevant to an AI-supported healthcare decision? [2]

(1) People should receive understandable information and retain meaningful choice. [1]

(2) An automated recommendation should not replace appropriate informed consent and human judgement. [1]

W17. Explain beneficence in an AI-supported healthcare context. [2]

(1) Beneficence means seeking a genuine benefit for the person affected. [1]

(2) A useful application should be evaluated for whether it improves the intended outcome, not only whether it is technically impressive. [1]

W18. Explain justice in allocating a healthcare resource. [2]

(1) Justice requires fair criteria and consideration of relevant need. [1]

(2) Check whether the system systematically disadvantages a group and provide appropriate review. [1]

W19. Explain structured, unstructured and semi-structured data with examples. State the basis of this distinction. [4]

(1) Structured data follows a predefined schema, as in a spreadsheet with fixed columns for names and scores. [1]

(2) Unstructured data lacks that fixed tabular organisation; photographs or free-form reviews are common examples of this category. [1]

(3) Semi-structured data includes organising keys or tags without requiring one rigid table, as in JSON event records. [1]

(4) The distinction concerns how information is organised, not whether it has meaning, value or possible uses for AI. [1]

W20. Compare traditional programming with supervised machine learning. [2]

(1) Traditional programming applies rules written by a person to the inputs. [1]

(2) Supervised learning fits a mapping using examples with known targets. [1]

W21. Compare a rule-based model with a learning-based model. [2]

(1) A rule-based model follows explicitly supplied logic. [1]

(2) A learning-based model fits patterns from data and requires suitable training and evaluation. [1]

W22. Explain labelled and unlabelled data. [2]

(1) Labelled data contains input examples paired with supplied target outputs or correct category labels. [1]

(2) Unlabelled data contains examples without those supplied targets for learning. [1]

W23. Explain supervised, unsupervised and reinforcement learning with examples, and compare the feedback available. [4]

(1) Supervised learning uses input-target examples, such as parcel images paired with their correct damaged or intact labels. [1]

(2) Unsupervised learning discovers structure without supplied target labels, such as grouping customers with similar purchasing characteristics. [1]

(3) Reinforcement learning improves action choices through interaction and rewards, as when a game agent learns a successful strategy. [1]

(4) The feedback differs: supplied correct targets guide supervision, unsupervised methods lack those targets, and reinforcement methods use outcome-based rewards. [1]

W24. Distinguish classification and regression. [2]

(1) Classification predicts a named category, such as a product being faulty or acceptable. [1]

(2) Regression predicts a numerical quantity, such as the expected waiting time. [1]

W25. Distinguish clustering and association. [2]

(1) Clustering groups examples that are similar according to their represented features, such as customer groups. [1]

(2) Association discovers co-occurrence relationships, such as products frequently purchased together. [1]

W27. Explain agent, action, reward and policy in reinforcement learning. [4]

(1) The agent is the decision-making system that interacts with an environment, for example the controller of a game character. [1]

(2) An action is an available choice made by the agent, such as moving left, waiting or selecting a route. [1]

(3) Reward communicates how desirable an outcome is and helps compare the results of different action sequences. [1]

(4) A policy is the agent's strategy for selecting actions in the situations it encounters while pursuing better outcomes. [1]

W28. Explain the role of weights and bias in an artificial neuron. [2]

(1) Weights scale how input values contribute to the combined signal. [1]

(2) Bias shifts the pre-activation sum before the activation function is applied. [1]

W29. Define an artificial neuron and an activation function. [2]

(1) An artificial neuron combines weighted inputs and a bias. [1]

(2) An activation function transforms that sum to produce its output, often introducing nonlinearity. [1]

W30. Explain synapse and axon in the biological analogy used for ANN. [2]

(1) A biological synapse is a connection through which a neuron can receive a signal from another cell. [1]

(2) An axon carries its outgoing signal; artificial networks use computational connections. [1]

W31. Distinguish forward propagation and backpropagation. [2]

(1) Forward propagation passes input values through layers to obtain an output. [1]

(2) Backpropagation computes how parameters affect loss, supporting parameter updates during training. [1]

W32. Describe an ANN and the roles of its three layer categories. [4]

(1) An artificial neural network is a collection of connected computational units whose adjustable parameters help it learn patterns. [1]

(2) The input layer receives the numerical features representing an example, such as measurements extracted from an image. [1]

(3) Hidden layers combine weighted inputs and activation functions to construct intermediate representations relevant to the task. [1]

(4) The output layer produces the required prediction or scores, such as the probabilities or category scores for possible classes. [1]

W33. Explain overfitting with an example. [2]

(1) An overfitted model learns details that do not generalise well beyond training data. [1]

(2) A digit recogniser may perform well on familiar training images but poorly on independently collected handwriting. [1]

W34. Explain underfitting with an example. [2]

(1) An underfitted model is too limited to represent important data relationships. [1]

(2) A very simple trend model may miss a curved relationship and perform poorly on both training and test examples. [1]

W35. What does a confusion matrix show, and why is it useful? [2]

(1) It counts combinations of actual and predicted classes. [1]

(2) It separates correct predictions from false alarms and missed positives, supporting suitable evaluation metrics. [1]

W36. Taking spam as the positive class, explain TP, TN, FP and FN. [4]

(1) A true positive is an email that is actually spam and is correctly predicted to be spam. [1]

(2) A true negative is a legitimate email correctly predicted not to belong to the spam class. [1]

(3) A false positive is a legitimate email incorrectly predicted to be spam, creating an unnecessary spam alert. [1]

(4) A false negative is an actual spam email incorrectly predicted to be legitimate, so the system misses it. [1]

W37. Write the four classification-metric formulas. [4]

(1) Accuracy = (TP + TN)/(TP + TN + FP + FN). [1]

(2) Precision = TP/(TP + FP). [1]

(3) Recall = TP/(TP + FN). [1]

(4) F1 = 2TP/(2TP + FP + FN), or 2PR/(P + R). [1]

W38. Explain the purpose of a train-test split. [2]

(1) Separate data is used for fitting the model and for evaluating it. [1]

(2) The held-out test data helps estimate performance on examples not used to fit the model. [1]

W39. Why might an 80:20 split be used, and is it compulsory? [2]

(1) It leaves a larger share for learning while reserving some examples for testing. [1]

(2) It is an illustrative choice, not a compulsory universal rule; suitability depends on the data and task. [1]

W40. Explain why both false positives and false negatives matter. [2]

(1) False positives trigger incorrect positive decisions, such as unnecessary alerts. [1]

(2) False negatives miss real positive cases; their relative harm depends on the use case. [1]

W41. Distinguish the classification error rate from accuracy. [2]

(1) Error rate = (FP + FN)/N, the share of wrong predictions. [1]

(2) Accuracy = (TP + TN)/N = 1 - error rate, the share of correct predictions. [1]

W42. Explain bias in AI evaluation and one effect on decisions. [2]

(1) Bias can arise when data or evaluation methods systematically misrepresent relevant people or conditions. [1]

(2) A system can then allocate opportunities or services unfairly despite acceptable overall accuracy. [1]

W43. Suggest four ways to reduce biased AI outcomes. [4]

(1) Collect representative examples and check the labels, particularly for groups that are missing or poorly represented. [1]

(2) Evaluate errors separately across relevant groups and conditions rather than relying only on one overall accuracy score. [1]

(3) Review the features, decision rules and intended use, then test whether proposed changes reduce unjustified differences. [1]

(4) Provide clear reporting, human review and a way to correct harmful outcomes instead of treating every prediction as final. [1]

TEN SOLVED SAMPLE PAPERS

Time: 2 hours | Maximum marks: 50

Section A - Objective: 24 marks
Q1: Attempt any 4 of 6 parts.
Q2-Q5: Attempt any 5 of 6 parts in each group.

Section B - Subjective: 26 marks
Q6-Q10: Attempt any 3 questions; 2 marks each.
Q11-Q16: Attempt any 4 questions; 2 marks each.
Q17-Q21: Attempt any 3 questions; 4 marks each.

Attempt 15 numbered questions/groups: all 5 objective groups and 10 subjective questions. Attempt questions in the correct order within each section. There is no negative marking.

Each question is followed by its answer. For exam practice, cover the answer while attempting the question, then check the working and scoring points.

SAMPLE PAPER 01 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). A listener restates an instruction in her own words to check its meaning. Which stage does this demonstrate? **[1]**

(a) Noise

(b) Encoding

(c) Feedback

(d) Channel selection

ANSWER: (c) Feedback The response lets the sender check what the receiver understood.

Q1 (ii). Riya studies robotics because solving problems interests her, even without prizes. This is: **[1]**

(a) Peer pressure

(b) Financial reward

(c) External punishment

(d) Intrinsic motivation

ANSWER: (d) Intrinsic motivation Her motivation comes from interest in the activity itself.

Q1 (iii). A learner completes work only to receive a certificate. This is mainly: **[1]**

(a) Time estimation

(b) Intrinsic curiosity

(c) Self-awareness

(d) Extrinsic motivation

ANSWER: (d) Extrinsic motivation The certificate is an external reward.

Q1 (iv). An unexpected email attachment asks a user to enable unknown macros. What is the safest first response? **[1]**

(a) Enable them to view the file faster.

(b) Disable antivirus protection.

(c) Forward the attachment to everyone.

(d) Verify the sender and avoid enabling the macros.

ANSWER: (d) Verify the sender and avoid enabling the macros. Unexpected active content should not be trusted automatically.

Q1 (v). A repair service starts after its owner notices that customers cannot find local repairs. This shows: **[1]**

(a) Avoidance of customers

(b) Elimination of business risk

(c) Guaranteed profit

(d) Opportunity identification

ANSWER: (d) Opportunity identification An unmet need can become a business opportunity.

Q1 (vi). A reusable bottle is used many times instead of buying a disposable one each day. This mainly illustrates: **[1]**

(a) Uncontrolled extraction

(b) Single-use consumption

(c) Reuse

(d) Open burning

ANSWER: (c) Reuse Repeated use can reduce the demand for disposable products. Q2 - Answer any FIVE of SIX

Q2 (i). Which denominator belongs to precision? **[1]**

(a) TP + FN

(b) TN + FN

(c) TP + FP

(d) All cases in every situation

ANSWER: (c) TP + FP Precision conditions on predicted positives.

Q2 (ii). A system assigns a complaint to "delivery", "payment" or "product". This is: **[1]**

(a) Face detection

(b) Text classification

(c) Regression of pixel values

(d) Image segmentation

ANSWER: (b) Text classification The system assigns a text category.

Q2 (iii). A system assigns the label "bicycle" to an entire photograph. The task is: **[1]**

(a) Image classification

(b) Speech recognition

(c) Document translation

(d) Instance segmentation

ANSWER: (a) Image classification A category is assigned to the whole image.

Q2 (iv). A system draws a separate bounding box around each visible bus. The task is: **[1]**

(a) Only image storage

(b) Object detection

(c) Sentiment analysis

(d) Text generation

ANSWER: (b) Object detection Detection identifies and locates objects.

Q2 (v). What is removed when a process deletes a specified list of common words? **[1]**

(a) Stop words

(b) Every meaningful term

(c) All punctuation only

(d) All proper nouns automatically

ANSWER: (a) Stop words Stop-word removal uses a chosen list of common terms.

Q2 (vi). A team identifies affected users and the location of a recurring problem before collecting data. The stage is: **[1]**

(a) Tokenisation

(b) Problem scoping

(c) Classification

(d) Model deployment

ANSWER: (b) Problem scoping Problem scoping defines the problem and its context. Q3 - Answer any FIVE of SIX

Q3 (i). After defining its objective, a team obtains relevant observations from reliable sources. This is: **[1]**

(a) Reward calculation

(b) Image segmentation

(c) Model evaluation

(d) Data acquisition

ANSWER: (d) Data acquisition Acquisition obtains the data needed for the task.

Q3 (ii). Unlabelled records are grouped by similarity. Which task is being performed? **[1]**

(a) Label transcription

(b) Supervised regression

(c) Clustering

(d) Decryption

ANSWER: (c) Clustering Clustering seeks groups without supplied group labels.

Q3 (iii). A neuron receives x1=2, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 3

(b) 4

(c) 5

(d) 2

ANSWER: (a) 3 z = 0.5 x 2 + 1.5 x 2 - 1 = 3.

Q3 (iv). A model estimates the number of minutes needed for a journey. This is usually: **[1]**

(a) Clustering

(b) Classification into fixed names

(c) Tokenisation

(d) Regression

ANSWER: (d) Regression The target is a numerical quantity rather than a category label.

Q3 (v). A model works badly for a community absent from the training examples. Which concern is central? [1]

(a) Representational bias

(b) A larger font size

(c) A file extension error by definition

(d) Word order only

ANSWER: (a) Representational bias Unrepresentative data can create unequal performance.

Q3 (vi). Which denominator belongs to recall? [1]

(a) TP + FN

(b) TN + FP

(c) TN only

(d) TP + FP

ANSWER: (a) TP + FN Recall conditions on actual positives. Q4 - Answer any FIVE of SIX

Q4 (i). A Bag-of-Words vector normally preserves: [1]

(a) Tone of voice automatically

(b) Every grammatical relationship

(c) Term counts in a fixed vocabulary order

(d) The complete original sentence order

ANSWER: (c) Term counts in a fixed vocabulary order The positions refer to vocabulary entries, not original word order.

Q4 (ii). A team plots observations to detect outliers and patterns. This belongs mainly to: [1]

(a) Profit calculation

(b) Final publication only

(c) Data exploration

(d) Password management

ANSWER: (c) Data exploration Exploration helps understand the data before modeling.

Q4 (iii). A classifier reports TP=54 and FP=6. What is its precision, rounded to two decimal places? **[1]**

(a) 90.00%

(b) 10.00%

(c) 87.50%

(d) 85.71%

ANSWER: (a) 90.00% Precision = 54/(54+6) x 100 = 90.00%.

Q4 (iv). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "green travel saves time"? **[1]**

(a) 4

(b) 3

(c) 5

(d) 6

ANSWER: (a) 4 Tokens are green, travel, saves, time; repeated occurrences count separately.

Q4 (v). In a corpus of 6 documents, the term "green" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 3

(b) 6

(c) 2

(d) 0.3333

ANSWER: (a) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q4 (vi). An image is 120 pixels wide and 80 pixels high. How many pixels does it contain? **[1]**

(a) 9600

(b) 200

(c) 400

(d) 28800

ANSWER: (a) 9600 Pixel count = width x height = 120 x 80 = 9600; RGB channels do not triple the number of spatial pixels. Q5 - Answer any FIVE of SIX

Q5 (i). With "defective" as positive, a faulty item is predicted to be good. The outcome is: **[1]**

(a) False positive

(b) True negative

(c) False negative

(d) True positive

ANSWER: (c) False negative The actual positive was missed.

Q5 (ii). With "spam" as positive, a legitimate message is blocked as spam. The outcome is: **[1]**

(a) True negative

(b) True positive

(c) False negative

(d) False positive

ANSWER: (d) False positive A negative item was incorrectly predicted positive.

Q5 (iii). Examples contain measurements together with the correct category for each measurement. The learning setting is: **[1]**

(a) Random sorting

(b) Data encryption

(c) Unsupervised learning

(d) Supervised learning

ANSWER: (d) Supervised learning Known target labels guide supervised learning.

Q5 (iv). Use relative error = |actual - predicted|/actual. A numerical prediction is 440 when the actual value is 500. What is the relative error? **[1]**

(a) 0.012

(b) 1.12

(c) 0.88

(d) 0.12

ANSWER: (d) 0.12 |500-440|/500=0.12. This is relative numerical error, not a confusion-matrix count.

Q5 (v). A system assigns a class label to each pixel but does not distinguish two cars of the same class. This is: [1]

(a) Instance segmentation

(b) File compression

(c) Semantic segmentation

(d) Speech synthesis

ANSWER: (c) Semantic segmentation Semantic segmentation labels pixel classes without separate instances.

Q5 (vi). Which question belongs to the WHY part of a 4Ws canvas? [1]

(a) Where is the problem observed?

(b) Why is solving this problem valuable?

(c) Who experiences the problem?

(d) What exactly is the problem?

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (b) Why is solving this problem valuable? WHY explains the value or need for solving the problem. Section B | Two-mark model answers

Q6. Explain the role of feedback and give one example. [2]

ANSWER: (1) Feedback tells a sender how a message was understood. [1] (2) For example, a learner repeats an instruction to confirm its meaning. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Distinguish intrinsic motivation from extrinsic motivation. [2]

ANSWER: (1) Intrinsic motivation comes from interest or satisfaction in the activity. [1] (2) Extrinsic motivation comes from external outcomes such as rewards, marks or recognition. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. State two ways to reduce the risk of unauthorised access to an account. [2]

ANSWER: (1) Use a strong unique password and keep it private from other people. [1] (2) Enable multifactor authentication and sign out properly when using shared computers. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Explain calculated risk with a suitable business example. [2]

ANSWER: (1) Calculated risk means considering likely costs and benefits before acting under uncertainty. [1] (2) For example, test a small batch before financing large-scale production. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Explain sustainable development and give one relevant action. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Sustainable development meets present needs without undermining future generations' ability to meet theirs. [1] (2) Repairing reusable products helps conserve materials and reduce unnecessary replacement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Distinguish data acquisition from data exploration. [2]

ANSWER: (1) Data acquisition collects relevant observations from appropriate sources. [1] (2) Data exploration examines and visualises the observations to reveal quality issues and patterns. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Differentiate supervised learning and unsupervised learning. [2]

ANSWER: (1) Supervised learning uses training examples paired with known target labels or values. [1] (2) Unsupervised learning discovers patterns or groups without supplied target labels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Distinguish classification from regression with an example of each. [2]

ANSWER: (1) Classification predicts a category, such as parcel damaged or intact. [1] (2) Regression predicts a numerical quantity, such as delivery time in minutes. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

Q14. Explain why high training accuracy is insufficient evidence of a useful model. [2]

ANSWER: (1) A model may memorise training details rather than learn general patterns. [1] (2) Performance must also be checked on suitable unseen examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Distinguish image classification from object detection. [2]

ANSWER: (1) Image classification assigns a category to an entire image. [1] (2) Object detection identifies objects and locates them, commonly using bounding boxes. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Distinguish tokenisation from stop-word removal. [2]

Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Tokenisation splits text into units such as words. [1] (2) Stop-word removal deletes terms from a specified common-word list after appropriate processing. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A voice-based public information service works reliably for one accent but often rejects other local accents. [4] (a) Identify the data-related fairness concern. [1] (b) Suggest one data improvement. [1] (c) Suggest one evaluation improvement. [1] (d) Suggest one user safeguard. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Training data dominated by one accent may produce unequal recognition and restrict access for other speakers. [1]	1
(b)	Collect appropriately consented examples covering the accents and speaking conditions of the intended user population. [1]	1
(c)	Measure errors separately for relevant accent groups; a strong overall score can conceal poor subgroup results. [1]	1
(d)	Provide human assistance or another usable access method, and let users report or challenge incorrect rejections. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A sorting service predicts whether a photographed parcel is damaged using labelled examples. [4] (a) Name the learning category. [1] (b) Name the prediction task. [1] (c) Identify suitable input data. [1] (d) Explain why separate test examples are needed. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Supervised learning is suitable because the training examples pair parcel images with known damage labels. [1]	1
(b)	The prediction task is classification: damaged and intact are categories rather than a continuous measurement. [1]	1
(c)	Use clear parcel images or relevant image features, with accurate labels representing different packaging and lighting conditions. [1]	1
(d)	Separate unseen test images estimate generalisation and reveal whether the model merely memorised its training examples. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A club manually writes rules to answer fixed membership queries. Later, users ask many paraphrased questions. [4] (a) Identify the initial approach. [1] (b) Explain how it gives an answer. [1] (c) State one limitation. [1] (d) Suggest a suitable improvement. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	The original assistant uses a rule-based approach because people explicitly write its conditions and responses. [1]	1
(b)	It selects an answer by matching the user's input to predefined patterns, keywords or decision paths. [1]	1
(c)	An unfamiliar paraphrase may match no rule, so a valid membership query could receive an unsuitable response. [1]	1

Subpart	Model answer	Marks
(d)	Add evaluated NLP-based intent recognition and a clear fallback to human help when the request remains uncertain. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. A damaged-parcel detection model has TP=54, FP=6, FN=9 and TN=51. Calculate accuracy, precision, recall and F1 score. Show formula and substitution for each. Round percentages to two decimal places. [4]

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	(TP + TN)/N x 100 = (54 + 51)/120 x 100 = 87.50%. [1]	1
Precision	TP/(TP + FP) x 100 = 54/(54 + 6) x 100 = 90.00%. [1]	1
Recall	TP/(TP + FN) x 100 = 54/(54 + 9) x 100 = 85.71%. [1]	1
F1	2TP/(2TP + FP + FN) x 100 = 108/(108 + 6 + 9) x 100 = 87.80%. [1]	1

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "green buses save fuel" Document 2: "clean buses save time" Document 3: "green travel saves time" (a) List the vocabulary in first-appearance order. [1] (b) Construct the count vector for Document 3 in that order. [1] (c) Explain how the vector is useful to a text classifier. [1] (d) State one order-related limitation of this representation. **[1]**

ANSWER: Fixed-vocabulary Bag-of-Words count table

Vocabulary term	Document 1	Document 2	Document 3
green	1	0	1
buses	1	1	0
save	1	1	0
fuel	1	0	0
clean	0	1	0
time	0	1	1
travel	0	0	1
saves	0	0	1

(a)/(b) (a) Vocabulary: [green, buses, save, fuel, clean, time, travel, saves]. [1]
(b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1, 1]. [1]

(c) (c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) (d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

SAMPLE PAPER 02 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). A notice says that a meeting is postponed but gives no new date. Which quality is chiefly missing? **[1]**

(a) Courtesy

(b) Completeness

(c) Grammatical accuracy

(d) Brevity

ANSWER: (b) Completeness The recipient lacks information needed to act.

Q1 (ii). A learner completes work only to receive a certificate. This is mainly: **[1]**

(a) Self-awareness

(b) Time estimation

(c) Intrinsic curiosity

(d) Extrinsic motivation

ANSWER: (d) Extrinsic motivation The certificate is an external reward.

Q1 (iii). Which goal is most specific and measurable? **[1]**

(a) Become perfect soon.

(b) Complete ten revision questions each evening for seven days.

(c) Study whenever possible.

(d) Be better than everybody.

ANSWER: (b) Complete ten revision questions each evening for seven days. The task, quantity and time period can be checked.

Q1 (iv). Which protection helps when a password is stolen? **[1]**

(a) Multifactor authentication

(b) A shorter username

(c) A public password hint

(d) Renaming the browser

ANSWER: (a) Multifactor authentication A second independent factor adds an additional authentication requirement.

Q1 (v). An entrepreneur tests a small batch before investing in large production. This is an example of: **[1]**

(a) Price fixing

(b) Ignoring evidence

(c) No risk whatsoever

(d) Calculated risk

ANSWER: (d) Calculated risk A small trial provides evidence before a larger commitment.

Q1 (vi). Which action best supports responsible electronic-waste handling? **[1]**

(a) Burn devices with ordinary waste.

(b) Send obsolete devices through an authorised collection route.

(c) Break screens in an open field.

(d) Dump batteries near a water source.

ANSWER: (b) Send obsolete devices through an authorised collection route. Specialised collection enables safer handling and material recovery. Q2 - Answer any FIVE of SIX

Q2 (i). A neuron receives x1=3, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 4.5

(b) 3.5

(c) 2.5

(d) 5.5

ANSWER: (b) 3.5 z = 0.5 x 3 + 1.5 x 2 - 1 = 3.5.

Q2 (ii). Which choice describes lemmatisation? **[1]**

(a) Assigning words random colours

(b) Deleting random letters

(c) Counting image pixels

(d) Mapping an inflected form to an appropriate dictionary base form

ANSWER: (d) Mapping an inflected form to an appropriate dictionary base form Lemmatisation aims at meaningful base forms using linguistic information.

Q2 (iii). Which task gives different pixel masks to two separate people in one photograph? [1]

(a) Tokenisation

(b) Instance segmentation

(c) Colour inversion only

(d) Whole-image classification

ANSWER: (b) Instance segmentation Instance segmentation separates individual objects at pixel level.

Q2 (iv). Which ethical principle concerns a person's informed control over a decision affecting them? [1]

(a) Autonomy

(b) Compression

(c) Clustering

(d) Regression

ANSWER: (a) Autonomy Autonomy concerns informed choice and agency.

Q2 (v). Which denominator belongs to recall? [1]

(a) TN + FP

(b) TP + FN

(c) TN only

(d) TP + FP

ANSWER: (b) TP + FN Recall conditions on actual positives.

Q2 (vi). A rule says, "If stock is below ten units, send an alert." No examples are used to learn it. The approach is: [1]

(a) Rule-based

(b) Reinforcement learning

(c) Unsupervised clustering

(d) Learned regression by definition

ANSWER: (a) Rule-based The decision rule was explicitly specified. Q3 - Answer any FIVE of SIX

Q3 (i). A sentence "Wonderful, the train is late again" expresses frustration despite a positive word. The difficulty is: **[1]**

(a) A missing file extension

(b) Image size

(c) Sarcasm and context

(d) Disk capacity

ANSWER: (c) Sarcasm and context The intended sentiment differs from a literal word-level reading.

Q3 (ii). A classifier reports TP=63 and FP=7. What is its precision, rounded to two decimal places? **[1]**

(a) 10.00%

(b) 75.00%

(c) 90.00%

(d) 86.00%

ANSWER: (c) 90.00% Precision = 63/(63+7) x 100 = 90.00%.

Q3 (iii). A menu-driven bot responds through fixed paths. It is mainly a: **[1]**

(a) Script bot

(b) Vision sensor

(c) Regression dataset

(d) Learning-based bot by definition

ANSWER: (a) Script bot The response paths are explicitly designed.

Q3 (iv). An agent changes its actions after receiving rewards from an environment. This is: **[1]**

(a) Reinforcement learning

(b) Image compression

(c) Rule copying

(d) File sorting

ANSWER: (a) Reinforcement learning The agent learns an action strategy from feedback on its actions.

Q3 (v). An 8-bit grayscale channel has how many possible values? **[1]**

(a) 256

(b) 8

(c) 255

(d) 16

ANSWER: (a) 256 Eight bits represent 2 to the power 8 values, from 0 to 255.

Q3 (vi). An image is 140 pixels wide and 90 pixels high. How many pixels does it contain? **[1]**

(a) 37800

(b) 12600

(c) 460

(d) 230

ANSWER: (b) 12600 Pixel count = width x height = 140 x 90 = 12600; RGB channels do not triple the number of spatial pixels. Q4 - Answer any FIVE of SIX

Q4 (i). An AI service states what data it uses and explains important limitations. This supports: **[1]**

(a) Unlimited data retention

(b) Transparency

(c) Uncontrolled access

(d) Secrecy

ANSWER: (b) Transparency Transparency makes relevant processes and limits understandable.

Q4 (ii). Which question belongs to the WHY part of a 4Ws canvas? **[1]**

(a) Why is solving this problem valuable?

(b) Who experiences the problem?

(c) What exactly is the problem?

(d) Where is the problem observed?

ANSWER: (a) Why is solving this problem valuable? WHY explains the value or need for solving the problem.

Q4 (iii). In a corpus of 6 documents, the term "robots" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 0.3333

(b) 3

(c) 2

(d) 6

ANSWER: (b) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q4 (iv). A digital image's smallest addressable picture element is a: **[1]**

(a) Record label

(b) Token

(c) Pixel

(d) Paragraph

ANSWER: (c) Pixel A pixel stores image information at a spatial position.

Q4 (v). Use relative error = |actual - predicted|/actual. A numerical prediction is 484 when the actual value is 550. What is the relative error? **[1]**

(a) 0.012

(b) 0.88

(c) 0.12

(d) 1.12

ANSWER: (c) 0.12 |550-484|/550=0.12. This is relative numerical error, not a confusion-matrix count.

Q4 (vi). A high-stakes AI recommendation is reviewed by a responsible person before action. This provides: **[1]**

(a) Anonymity by definition

(b) Human oversight

(c) Automatic infallibility

(d) Data duplication

ANSWER: (b) Human oversight Human review supports responsibility and challenge of outputs. Q5 - Answer any FIVE of SIX

Q5 (i). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "robots sort sort labels"? **[1]**

(a) 4

(b) 3

(c) 6

(d) 5

ANSWER: (a) 4 Tokens are robots, sort, sort, labels; repeated occurrences count separately.

Q5 (ii). A model memorises training details but performs poorly on new examples. This suggests: **[1]**

(a) Overfitting

(b) Ideal generalisation

(c) A guaranteed fair model

(d) Underfitting only

ANSWER: (a) Overfitting The learned behaviour does not generalise well.

Q5 (iii). A model performs poorly on both training and test data because it is too simple. This suggests: **[1]**

(a) No model error

(b) Ideal complexity

(c) Underfitting

(d) Perfect calibration

ANSWER: (c) Underfitting The model fails to capture important relationships.

Q5 (iv). A model works badly for a community absent from the training examples. Which concern is central? **[1]**

(a) A larger font size

(b) Representational bias

(c) Word order only

(d) A file extension error by definition

ANSWER: (b) Representational bias Unrepresentative data can create unequal performance.

Q5 (v). Why keep a test set separate from training? [1]

(a) To ensure all predictions become correct

(b) To make labels unnecessary

(c) To estimate performance on unseen examples

(d) To avoid ever collecting data

ANSWER: (c) To estimate performance on unseen examples Held-out examples give a more realistic evaluation of generalisation.

Q5 (vi). Which statement correctly relates AI, ML and DL? [1]

(a) Machine learning contains every form of intelligence.

(b) Deep learning and AI have no connection.

(c) All AI must use deep learning.

(d) Deep learning is a part of machine learning, which is a part of AI.

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (d) Deep learning is a part of machine learning, which is a part of AI. The fields form a nested relationship, but not all AI is learning-based. Section B | Two-mark model answers

Q6. Give two actions that make listening active rather than passive. [2]

ANSWER: (1) Attend carefully to the speaker without interrupting or preparing an unrelated reply. [1] (2) Ask a relevant clarifying question or summarise the message to confirm understanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Give two practical steps for managing revision-related stress. [2]

ANSWER: (1) Break revision into realistic tasks and schedule them. [1] (2) Use a healthy routine, brief relaxation, exercise or appropriate support to manage tension. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. Explain the purpose of a backup and one condition for a useful backup. [2]

ANSWER: (1) A backup supports recovery after accidental loss or device failure. [1] (2) Keep an up-to-date recoverable copy separate from the original and check it periodically. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. State two ways entrepreneurship can benefit a community. [2]

ANSWER: (1) Entrepreneurship supplies useful products or services that respond to unmet community needs. [1] (2) It can create paid work and support local suppliers through business activity. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Distinguish reuse from recycling with an example of each. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Reuse means using an item again, such as refilling a bottle. [1] (2) Recycling processes waste into usable material, such as recovering paper fibre. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

Q11. Explain why problem scoping should precede model development. [2]

ANSWER: (1) Scoping identifies the users, problem and intended outcome. [1] (2) It guides relevant data collection and prevents building a model for an unclear objective. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Explain association learning with a suitable example. [2]

ANSWER: (1) Association learning identifies items or events that frequently occur together. [1] (2) For example, a shop may find that brushes and paint often appear in one purchase. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. What is anomaly detection? Give an example. [2]

ANSWER: (1) Anomaly detection identifies observations that differ markedly from typical patterns in data. [1] (2) For example, flag an unusually high machine-vibration reading for inspection. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Distinguish a false positive from a false negative. [2]

ANSWER: (1) A false positive predicts the positive class for an actually negative example. [1] (2) A false negative predicts the negative class for an actually positive example. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Distinguish semantic segmentation from instance segmentation. [2]

ANSWER: (1) Semantic segmentation assigns a class label to each pixel. [1] (2) Instance segmentation also separates individual objects belonging to the same class. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain one use and one limitation of Bag of Words. [2] Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) It converts text into fixed-length count vectors that a model can process. [1] (2) It usually discards word order and grammatical relationships. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A recycling centre wants to identify visitors through facial images even though anonymous counting would meet its objective. [4] (a) Identify one privacy concern. (b) Suggest a less intrusive design. [1] (c) State one retention safeguard. [1] (d) State who should remain accountable. **[1]** [1]

ANSWER:

Subpart	Model answer	Marks
(a)	Identifiable facial data exceeds what anonymous counting requires, creating unnecessary privacy and possible misuse risks. [1]	1
(b)	Count visits without identifying individuals, using a less intrusive design suited to the stated purpose. [1]	1
(c)	Define a justified retention period, restrict access to necessary records, and delete information when no longer needed. [1]	1
(d)	The deploying organisation and authorised people remain responsible for safeguards, monitoring, responding to complaints and correcting failures. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. An ANN classifies recorded sound features into three environmental categories. [4] (a) Explain an ANN briefly. (b) State the input layer's role. [1] (c) State the hidden layers' role. [1] (d) State the output layer's role. **[1]** [1]

ANSWER:

Subpart	Model answer	Marks
(a)	An ANN is a network of connected computational units that learns patterns by adjusting parameters from examples. [1]	1

Subpart	Model answer	Marks
(b)	Its input layer receives numerical sound features representing the recording, such as useful measurements of the signal. [1]	1
(c)	Hidden layers combine weighted signals and activation functions, developing intermediate representations that help distinguish the categories. [1]	1
(d)	The output layer produces category scores or a final classification, which must be checked against correct labels during evaluation. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A store analyses unlabelled purchase records for groups of similar customers and for products bought together. [4] (a) Name the broad learning category. [1] (b) Identify the customer-grouping task. [1] (c) Identify the co-purchase task. [1] (d) Explain one limitation of the findings. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	This is unsupervised learning because correct customer-group and product-relationship labels are not supplied to the system. [1]	1
(b)	Clustering groups customers with similar purchasing characteristics according to the selected features and method used. [1]	1
(c)	Association learning identifies product combinations that frequently appear together in the available transaction records. [1]	1

Subpart	Model answer	Marks
(d)	Such patterns do not establish causation or automatically justify decisions; business use should consider evidence quality and possible unfairness. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. For a library-return reminders task, define the condition being detected as positive. Out of 200 examples, 63 actual positives were predicted positive; 21 actual positives were missed; 7 actual negatives were predicted positive; and 109 actual negatives were predicted negative. [4] (a) Construct and label the confusion matrix. [2] (b) Calculate accuracy with working. [1] (c) Find the number of wrong predictions with working. **[1]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	See given computation	1
Precision	See given computation	1
Recall	See given computation	1
F1	See given computation	1

Suggested marking: 1 mark per metric.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "robots sort small boxes" Document 2: "small robots label boxes" Document 3: "robots sort sort labels" (a) Find the raw term frequency of each word in Document 1. [1] (b) Construct the document-frequency (DF) table for the corpus. [1] (c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1] (d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. **[1]**

ANSWER: Term frequency, document frequency and IDF ratio

Vocabulary term	TF in D1	DF	IDF ratio (3/DF)
robots	1	3	1
sort	1	2	1.5
small	1	2	1.5
boxes	1	2	1.5
label	0	1	3
labels	0	1	3

(a) Raw TF counts are in column 2. (b) DF counts the documents containing each word. (c) The IDF ratio is N/DF with N=3. (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)).

(c) (c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

SAMPLE PAPER 03 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). Two students interpret the word "charge" differently in a message. Which barrier is involved? **[1]**

(a) Physical barrier

(b) Lack of eye contact

(c) Semantic barrier

(d) Network congestion only

ANSWER: (c) Semantic barrier Different meanings of a word can distort understanding.

Q1 (ii). Which goal is most specific and measurable? **[1]**

(a) Complete ten revision questions each evening for seven days.

(b) Be better than everybody.

(c) Become perfect soon.

(d) Study whenever possible.

ANSWER: (a) Complete ten revision questions each evening for seven days. The task, quantity and time period can be checked.

Q1 (iii). Before planning revision, a student lists topics she understands and topics she finds difficult. This shows: **[1]**

(a) Delegation

(b) Self-awareness

(c) Data encryption

(d) Avoidance

ANSWER: (b) Self-awareness Recognising strengths and weaknesses supports realistic planning.

Q1 (iv). Why should an important file be backed up to a separate reliable location? **[1]**

(a) To remove the need for access control.

(b) To recover it after accidental loss or device failure.

(c) To prevent all phishing attacks.

(d) To guarantee the file is correct.

ANSWER: (b) To recover it after accidental loss or device failure. A backup provides a recoverable copy; it is not a universal security solution.

Q1 (v). Which statement about entrepreneurship is correct? [1]

(a) Only inherited wealth can create an entrepreneur.

(b) Profit is guaranteed after registration.

(c) An entrepreneur never needs a team.

(d) Skills can be developed through learning and experience.

ANSWER: (d) Skills can be developed through learning and experience. Entrepreneurial abilities are not limited to people born with wealth.

Q1 (vi). Sustainable development aims to meet present needs while: [1]

(a) Using every available resource immediately

(b) Ignoring environmental costs

(c) Protecting future generations' ability to meet theirs

(d) Stopping all economic activity

ANSWER: (c) Protecting future generations' ability to meet theirs Present development should not destroy future options. Q2 - Answer any FIVE of SIX

Q2 (i). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "bright lamps light homes"? [1]

(a) 6

(b) 5

(c) 3

(d) 4

ANSWER: (d) 4 Tokens are bright, lamps, light, homes; repeated occurrences count separately.

Q2 (ii). Which neural-network layer receives the feature values first? [1]

(a) Output layer

(b) Archive layer

(c) Input layer

(d) Footer layer

ANSWER: (c) Input layer The input layer receives the representation supplied to the network.

Q2 (iii). A classifier reports TP=28 and FP=12. What is its precision, rounded to two decimal places? **[1]**

(a) 81.00%

(b) 80.00%

(c) 30.00%

(d) 70.00%

ANSWER: (d) 70.00% Precision = 28/(28+12) x 100 = 70.00%.

Q2 (iv). Why keep a test set separate from training? **[1]**

(a) To ensure all predictions become correct

(b) To make labels unnecessary

(c) To avoid ever collecting data

(d) To estimate performance on unseen examples

ANSWER: (d) To estimate performance on unseen examples Held-out examples give a more realistic evaluation of generalisation.

Q2 (v). In a corpus of 6 documents, the term "solar" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 3

(b) 0.3333

(c) 2

(d) 6

ANSWER: (a) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q2 (vi). A text system resolves what "she" refers to using the previous sentence. This requires: **[1]**

(a) Single-pixel classification

(b) Hardware cleaning

(c) Alphabetical sorting

(d) Discourse integration

ANSWER: (d) Discourse integration Meaning can depend on relationships across sentences. Q3 - Answer any FIVE of SIX

Q3 (i). Use relative error = |actual - predicted|/actual. A numerical prediction is 528 when the actual value is 600. What is the relative error? **[1]**

(a) 0.88

(b) 0.12

(c) 1.12

(d) 0.012

ANSWER: (b) 0.12 |600-528|/600=0.12. This is relative numerical error, not a confusion-matrix count.

Q3 (ii). A high-stakes AI recommendation is reviewed by a responsible person before action. This provides: **[1]**

(a) Anonymity by definition

(b) Data duplication

(c) Human oversight

(d) Automatic infallibility

ANSWER: (c) Human oversight Human review supports responsibility and challenge of outputs.

Q3 (iii). Which sequence best describes a vision pipeline? **[1]**

(a) Interpretation, real scene, sensing device

(b) Real scene, sensing device, interpretation

(c) Stored answer, scene, sensor

(d) Sensor, interpretation, creation of reality

ANSWER: (b) Real scene, sensing device, interpretation The sensor captures visual data that the system processes.

Q3 (iv). Which NLP stage concerns the grammatical structure of a sentence? **[1]**

(a) Pixel normalisation

(b) Model compression

(c) Syntactic analysis

(d) Colour segmentation

ANSWER: (c) Syntactic analysis Syntax concerns how words are structurally combined.

Q3 (v). Avoiding preventable harm from an AI recommendation is the principle of: [1]

(a) Non-maleficence

(b) External motivation

(c) Dimensionality reduction

(d) Profit maximisation alone

ANSWER: (a) Non-maleficence Non-maleficence means avoiding harm.

Q3 (vi). What do weights in an artificial neuron mainly represent? [1]

(a) The size of a password

(b) The influence of input signals on a weighted sum

(c) The number of paper pages

(d) Only the output label's spelling

ANSWER: (b) The influence of input signals on a weighted sum Weights control the contribution of inputs. Q4 - Answer any FIVE of SIX

Q4 (i). A model predicts the majority class for every record and obtains a high overall score. What should be examined next? [1]

(a) Only the file name

(b) Class-specific errors and recall

(c) Only the font used in the report

(d) Only the number of columns

ANSWER: (b) Class-specific errors and recall Class imbalance can conceal failure on a smaller class.

Q4 (ii). Why may a vision model fail on very dark images? [1]

(a) Useful visual detail may be lost or differ from training examples.

(b) Dark images cannot contain pixels.

(c) All labels are automatically deleted.

(d) Resolution must always become zero.

ANSWER: (a) Useful visual detail may be lost or differ from training examples. Visibility and data distribution can affect recognition.

Q4 (iii). An image is 160 pixels wide and 100 pixels high. How many pixels does it contain? [1]

(a) 520

(b) 48000

(c) 260

(d) 16000

ANSWER: (d) 16000 Pixel count = width x height = 160 x 100 = 16000; RGB channels do not triple the number of spatial pixels.

Q4 (iv). Which task is optical character recognition designed to perform? [1]

(a) Cluster music without labels

(b) Extract text characters from an image

(c) Recognise every sound in a recording

(d) Predict future weather from time alone

ANSWER: (b) Extract text characters from an image OCR converts visible characters into machine-readable text.

Q4 (v). Which operation risks data leakage? [1]

(a) Documenting the positive class

(b) Reserving unseen examples

(c) Reporting all confusion-matrix counts

(d) Using the final test results repeatedly to choose the model

ANSWER: (d) Using the final test results repeatedly to choose the model Repeated selection using the final test information contaminates evaluation.

Q4 (vi). Why is an activation function used after a neuron's weighted sum? [1]

(a) To replace all training examples

(b) To rename every training record

(c) To guarantee perfect accuracy

(d) To transform the signal, often introducing nonlinearity

ANSWER: (d) To transform the signal, often introducing nonlinearity Activation enables useful transformations; nonlinear activations support complex patterns. Q5 - Answer any FIVE of SIX

Q5 (i). A neuron receives x1=4, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 6

(b) 4

(c) 3

(d) 5

ANSWER: (b) 4 z = 0.5 x 4 + 1.5 x 2 - 1 = 4.

Q5 (ii). F1 is the harmonic mean of: **[1]**

(a) Accuracy and file size

(b) Precision and recall

(c) TP and TN

(d) Training and test counts

ANSWER: (b) Precision and recall F1 combines precision and recall in one score.

Q5 (iii). A smart assistant interprets "Could you open the window?" as a request. This uses: **[1]**

(a) Folder creation

(b) Pragmatic interpretation

(c) Raw character count only

(d) Image thresholding

ANSWER: (b) Pragmatic interpretation Pragmatics considers intended meaning in context.

Q5 (iv). A framework compares the likely benefits and harms of an AI action. Its emphasis is: **[1]**

(a) File type

(b) Consequences or utility

(c) Alphabetical order

(d) Pixel depth

ANSWER: (b) Consequences or utility Consequence-based reasoning considers the effects of decisions.

Q5 (v). A clinic checks that an AI allocation system does not disadvantage a particular community. Which principle is most directly involved? [1]

(a) Tokenisation

(b) Justice

(c) Encoding

(d) Resolution

ANSWER: (b) Justice Justice concerns fair treatment and distribution.

Q5 (vi). An AI service states what data it uses and explains important limitations. This supports: [1]

(a) Uncontrolled access

(b) Secrecy

(c) Unlimited data retention

(d) Transparency

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (d) Transparency Transparency makes relevant processes and limits understandable. Section B | Two-mark model answers

Q6. Distinguish verbal communication from non-verbal communication. [2]

ANSWER: (1) Verbal communication uses spoken or written words. [1] (2) Non-verbal communication conveys meaning through cues such as gestures, facial expressions and posture. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Explain self-awareness and state how it helps planning. [2]

ANSWER: (1) Self-awareness means recognising personal strengths, weaknesses and emotions. [1] (2) It helps a learner set realistic goals and choose areas needing improvement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. Distinguish copying a file from moving a file. [2]

ANSWER: (1) Copying creates another instance while leaving the original in place. [1] (2) Moving transfers the file to a new location and removes it from the old location. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Distinguish an entrepreneur from an employee. [2]

ANSWER: (1) An entrepreneur organises a venture and bears its business uncertainty. [1] (2) An employee works under agreed terms and receives compensation for assigned responsibilities. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Give two suitable measures for reducing electronic waste. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Repair or reuse suitable working devices to extend their useful life. [1] (2) Send unusable electronics through authorised collection or recycling channels instead of mixing them with ordinary waste. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. State two safeguards against unfair AI decisions. [2]

ANSWER: (1) Use representative data and examine errors across relevant groups. [1] (2) Provide meaningful human review or an appeal process for important decisions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. State the roles of a weight and an activation function in a neuron. [2]

ANSWER: (1) A weight controls an input's contribution to the weighted sum. [1] (2) An activation function transforms that sum, often allowing the network to represent nonlinear relationships. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Distinguish a rule-based system from a learning-based system. [2]

ANSWER: (1) A rule-based system follows explicitly written decision rules. [1] (2) A learning-based system fits patterns from data, so its behaviour depends on training and evaluation. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Distinguish precision from recall using their denominators. [2]

ANSWER: (1) Precision divides TP by TP + FP, the number of predicted positives. [1] (2) Recall divides TP by TP + FN, the number of actual positives. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Define a pixel and image resolution. [2]

ANSWER: (1) A pixel is an individual picture element containing image information. [1] (2) Resolution describes image dimensions, commonly as width and height in pixels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Distinguish stemming from lemmatisation. [2] Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Stemming applies rules that trim words and may create non-dictionary stems. [1] (2) Lemmatisation aims to return an appropriate dictionary base form. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. An automated loan-information service presents estimates as guaranteed decisions and gives no explanation of their limitations. [4] (a) Explain the transparency problem. [1] (b) Give one consequence for users. [1] (c) Suggest a correction to the explanation. [1] (d) Suggest one oversight measure. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Presenting uncertain estimates as guaranteed decisions hides relevant limitations and prevents users from understanding the service. [1]	1
(b)	A person may make an unsuitable financial commitment because the displayed estimate appears to be final approval. [1]	1
(c)	Explain assumptions, uncertainty and data use in understandable language, clearly separating an estimate from an authorised decision. [1]	1
(d)	Provide human review of important outcomes and an accessible route for questions, corrections and complaints. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A robot learns routes by trying actions and receiving a reward for safe arrival and a penalty for collision. [4] (a) Name the learning method. [1] (b) Identify the agent. [1] (c) Explain the reward's role. [1] (d) State one safety limitation. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The robot uses reinforcement learning, improving its action strategy through interaction rather than supplied answers for every situation. [1]	1
(b)	The agent is the robot's decision-making system; the routes, obstacles and responses form its environment. [1]	1
(c)	Reward provides feedback that encourages safer, more successful action sequences and better cumulative outcomes over time. [1]	1
(d)	Unsafe trial and error could damage people or equipment, so exploration needs a controlled environment and appropriate constraints. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A model predicts daily water demand from weather and past consumption. [4] (a) Name a suitable learning category when correct demand values are provided. [1] (b) Identify the task as classification or regression. [1] (c) Give two possible features. [1] (d) Distinguish the target from a feature. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Supervised learning is suitable when past examples include weather and consumption features paired with correct demand values. [1]	1
(b)	The task is regression because the required output is a numerical quantity, not a named category. [1]	1
(c)	Temperature and previous consumption are two relevant input features that may help explain changes in demand. [1]	1

Subpart	Model answer	Marks
(d)	The target is the water demand being predicted; features are the available inputs used to estimate it. [1]	1

Suggested marking: Each subpart is 1 mark. In the features subpart, award 0.5 for each of two relevant features.

Q20. A water-leak alerts model has TP=28, FP=12, FN=7 and TN=53. Calculate accuracy, precision, recall and F1 score. Show formula and substitution for each. Round percentages to two decimal places. [4]

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	(TP + TN)/N x 100 = (28 + 53)/100 x 100 = 81.00%. [1]	1
Precision	TP/(TP + FP) x 100 = 28/(28 + 12) x 100 = 70.00%. [1]	1
Recall	TP/(TP + FN) x 100 = 28/(28 + 7) x 100 = 80.00%. [1]	1
F1	2TP/(2TP + FP + FN) x 100 = 56/(56 + 12 + 7) x 100 = 74.67%. [1]	1

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "solar lamps light lanes" Document 2: "solar panels charge lamps" Document 3: "bright lamps light homes" (a) List the vocabulary in first-appearance order. [1] (b) Construct the count vector for Document 3 in that order. [1] (c) Explain how the vector is useful to a text classifier. [1] (d) State one order-related limitation of this representation. **[1]**

ANSWER: Fixed-vocabulary Bag-of-Words count table

Vocabulary term	Document 1	Document 2	Document 3
solar	1	1	0
lamps	1	1	1
light	1	0	1
lanes	1	0	0
panels	0	1	0
charge	0	1	0
bright	0	0	1
homes	0	0	1

(a)/(b) (a) Vocabulary: [solar, lamps, light, lanes, panels, charge, bright, homes]. [1] (b) Document 3 vector: [0, 1, 1, 0, 0, 0, 1, 1]. [1]

(c) (c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) (d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

SAMPLE PAPER 04 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). Which feedback identifies both a strength and an actionable improvement? [1]

(a) Your chart is clear; label its units to improve interpretation.

(b) Do not try presenting again.

(c) Your presentation is useless.

(d) Everyone agrees you are wrong.

ANSWER: (a) Your chart is clear; label its units to improve interpretation. Constructive feedback is specific and useful rather than a personal attack.

Q1 (ii). Before planning revision, a student lists topics she understands and topics she finds difficult. This shows: [1]

(a) Delegation

(b) Data encryption

(c) Avoidance

(d) Self-awareness

ANSWER: (d) Self-awareness Recognising strengths and weaknesses supports realistic planning.

Q1 (iii). Which action best shows responsible independent work? [1]

(a) Wait for repeated reminders.

(b) Hide errors until submission.

(c) Refuse all feedback.

(d) Finish the assigned task and seek clarification when needed.

ANSWER: (d) Finish the assigned task and seek clarification when needed. Independence includes responsibility, not refusing necessary help.

Q1 (iv). Which action changes a file's name without intentionally moving its contents? [1]

(a) Format the drive

(b) Compress the folder

(c) Rename

(d) Copy

ANSWER: (c) Rename Renaming changes the file identifier shown to users.

Q1 (v). Revenue from a batch is Rs 2,400 and total cost is Rs 1,800. Profit is: [1]

(a) Rs 600

(b) Rs 4,200

(c) Rs 2,400

(d) Rs 1,800

ANSWER: (a) Rs 600 Profit is revenue minus total cost.

Q1 (vi). Which measure directly reduces avoidable water loss? [1]

(a) Replace all plants with water-intensive species.

(b) Repair leaking pipes.

(c) Ignore meter readings.

(d) Increase tap flow without need.

ANSWER: (b) Repair leaking pipes. Leaks waste water without providing useful service. Q2 - Answer any FIVE of SIX

Q2 (i). Document frequency of a term means: [1]

(a) Number of documents containing it

(b) Only its count in the first document

(c) Number of letters in it

(d) Its total count across all documents

ANSWER: (a) Number of documents containing it Each containing document contributes once to DF.

Q2 (ii). An image is 180 pixels wide and 110 pixels high. How many pixels does it contain? [1]

(a) 580

(b) 19800

(c) 290

(d) 59400

ANSWER: (b) 19800 Pixel count = width x height = 180 x 110 = 19800; RGB channels do not triple the number of spatial pixels.

Q2 (iii). If actual positives are rare and missing them is costly, which metric is especially important? [1]

(a) The vocabulary length

(b) Recall for the positive class

(c) The model's file extension

(d) Only overall accuracy

ANSWER: (b) Recall for the positive class Recall shows how many actual positives are detected.

Q2 (iv). A retailer discovers that two products often occur in the same purchase. This is an example of: [1]

(a) Image segmentation

(b) Linear text order

(c) Association

(d) Manual file naming

ANSWER: (c) Association Association discovers co-occurrence relationships.

Q2 (v). A neural network architecture particularly suited to extracting local visual patterns is a: [1]

(a) Spell-check dictionary

(b) Password manager

(c) Convolutional Neural Network

(d) Folder tree

ANSWER: (c) Convolutional Neural Network CNNs learn patterns through local filters in image-like inputs.

Q2 (vi). Which operation risks data leakage? [1]

(a) Documenting the positive class

(b) Using the final test results repeatedly to choose the model

(c) Reserving unseen examples

(d) Reporting all confusion-matrix counts

ANSWER: (b) Using the final test results repeatedly to choose the model Repeated selection using the final test information contaminates evaluation. Q3 - Answer any FIVE of SIX

Q3 (i). Which domain is most directly used when a system interprets spoken or written human language? [1]

(a) Disk formatting

(b) Computer Vision only

(c) Natural Language Processing

(d) Hardware cleaning

ANSWER: (c) Natural Language Processing NLP concerns computational processing of human language.

Q3 (ii). Which data is semi-structured? [1]

(a) Only a fixed numerical matrix by definition

(b) An unannotated photograph alone

(c) An unrecorded conversation

(d) Records in JSON with keys and nested fields

ANSWER: (d) Records in JSON with keys and nested fields JSON has an organising structure without a single rigid relational table layout.

Q3 (iii). Avoiding preventable harm from an AI recommendation is the principle of: [1]

(a) Profit maximisation alone

(b) Non-maleficence

(c) External motivation

(d) Dimensionality reduction

ANSWER: (b) Non-maleficence Non-maleficence means avoiding harm.

Q3 (iv). Which pair counts all incorrect binary classifications? [1]

(a) TP + FP

(b) TN + FN

(c) FP + FN

(d) TP + TN

ANSWER: (c) FP + FN False positives and false negatives are the errors.

Q3 (v). A neuron receives x1=5, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 6.5

(b) 5.5

(c) 3.5

(d) 4.5

ANSWER: (d) 4.5 z = 0.5 x 5 + 1.5 x 2 - 1 = 4.5.

Q3 (vi). Which representation typically has three colour channels? **[1]**

(a) A word-count vector

(b) A single grayscale channel

(c) A file path

(d) RGB image

ANSWER: (d) RGB image RGB stores red, green and blue components. Q4 - Answer any FIVE of SIX

Q4 (i). Adjusting brightness without deciding what objects are present is mainly: **[1]**

(a) Image processing

(b) Language translation

(c) Object understanding by definition

(d) Sentiment analysis

ANSWER: (a) Image processing Pixel transformation need not include semantic interpretation.

Q4 (ii). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "clean air helps people"? **[1]**

(a) 5

(b) 4

(c) 6

(d) 3

ANSWER: (b) 4 Tokens are clean, air, helps, people; repeated occurrences count separately.

Q4 (iii). A classifier reports TP=81 and FP=9. What is its precision, rounded to two decimal places? **[1]**

(a) 90.00%

(b) 81.82%

(c) 86.50%

(d) 10.00%

ANSWER: (a) 90.00% Precision = 81/(81+9) x 100 = 90.00%.

Q4 (iv). An ethical approach asks whether a decision respects fundamental rights. It is mainly: **[1]**

(a) Pixel-based

(b) Frequency-based

(c) Duty-based (deontological)

(d) Compression-based

ANSWER: (c) Duty-based (deontological) Rights-based reasoning prioritises rights and dignity.

Q4 (v). In a corpus of 6 documents, the term "clean" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 6

(b) 0.3333

(c) 2

(d) 3

ANSWER: (d) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q4 (vi). Which information is needed before comparing two Bag-of-Words vectors? **[1]**

(a) The same ordered vocabulary for both

(b) The colour of the text

(c) Only the number of pages

(d) A different vocabulary for each vector

ANSWER: (a) The same ordered vocabulary for both Vector positions must refer to the same terms. Q5 - Answer any FIVE of SIX

Q5 (i). A framework compares the likely benefits and harms of an AI action. Its emphasis is: [1]

(a) File type

(b) Pixel depth

(c) Alphabetical order

(d) Consequences or utility

ANSWER: (d) Consequences or utility Consequence-based reasoning considers the effects of decisions.

Q5 (ii). A model has no predicted positives. Its precision formula has: [1]

(a) The same denominator as accuracy

(b) No dependence on predictions

(c) A zero denominator

(d) A value of one automatically

ANSWER: (c) A zero denominator TP + FP is zero; a reporting convention is needed for undefined precision.

Q5 (iii). Use relative error = |actual - predicted|/actual. A numerical prediction is 572 when the actual value is 650. What is the relative error? [1]

(a) 0.12

(b) 1.12

(c) 0.88

(d) 0.012

ANSWER: (a) 0.12 |650-572|/650=0.12. This is relative numerical error, not a confusion-matrix count.

Q5 (iv). A grayscale pixel value of 0 usually represents: [1]

(a) Black

(b) A missing file

(c) White

(d) Pure red

ANSWER: (a) Black Under the usual 8-bit convention, zero is black and 255 is white.

Q5 (v). A word appears in all N documents. Under the school convention, its IDF ratio N/DF is: [1]

(a) Undefined

(b) 1

(c) N squared

(d) 0

ANSWER: (b) 1 Here DF equals N, so the ratio is one.

Q5 (vi). A team returns to data collection after evaluation reveals missing examples. This shows that the AI project cycle is: [1]

(a) Iterative

(b) Always one-way

(c) Completed before evaluation

(d) Independent of data

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (a) Iterative Evaluation can reveal a need to revisit earlier stages. Section B | Two-mark model answers

Q6. Name one physical and one semantic communication barrier, with an example of each. [2]

ANSWER: (1) Loud machinery is a physical barrier because it interferes with hearing. [1] (2) An unfamiliar technical word is a semantic barrier because its meaning may be misunderstood. [1]

Suggested marking: For each of the two concepts: 0.5 for the distinction and 0.5 for an appropriate example; total 2 marks.

Q7. What makes a goal measurable and time-bound? [2]

ANSWER: (1) Measurable means progress can be checked using a clear quantity or criterion. [1] (2) Time-bound means a deadline or time period is specified. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. What is a Trojan and why can it mislead a user? [2]

ANSWER: (1) A Trojan is malicious software disguised as a legitimate program. [1] (2) Its useful appearance can persuade a user to run it and permit harmful actions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Explain opportunity identification with an example. [2]

ANSWER: (1) Opportunity identification recognises an unmet need that a venture could address. [1] (2) For example, start a repair service where customers lack nearby repairs. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Explain how education can support sustainable behaviour. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Education helps people understand resource use and environmental consequences. [1] (2) It builds the knowledge and skills needed to adopt and evaluate more sustainable practices. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Define bioethics and name two of its principles. [2]

ANSWER: (1) Bioethics applies ethical reasoning to decisions in healthcare and the life sciences. [1] (2) Autonomy respects informed choices. [0.5] (3) Justice requires fair treatment and access. [0.5]

Suggested marking: Definition: 1 mark; any two correctly named bioethical principles: 0.5 each. Beneficence and non-maleficence are also valid.

Q12. Distinguish labelled and unlabelled examples. [2]

ANSWER: (1) A labelled example includes an input and the expected target. [1] (2) An unlabelled example provides the input without a supplied target

category or value. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. State the roles of the input and output layers of an ANN. [2]

ANSWER: (1) The input layer receives the encoded feature values. [1] (2) The output layer produces the network's task-specific result, such as class scores. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Give two reasons to report more than overall accuracy. [2]

ANSWER: (1) Class imbalance can hide poor detection of a smaller but important class. [1] (2) False positives and false negatives may also have different consequences for users. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Explain the usual 0 to 255 range of an 8-bit grayscale image. [2]

ANSWER: (1) Eight bits allow 2 to the power 8, or 256, values. [1] (2) Under the usual convention, 0 is black and 255 is white. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Distinguish a script bot from a smart bot. [2] Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) A script bot follows predefined response paths. [1] (2) A smart bot can use NLP and learned patterns to handle varied inputs, but may still make errors. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A hospital considers an AI recommendation that could benefit many patients but may expose one patient's confidential record. [4] (a) Explain the autonomy concern. [1] (b) Apply non-maleficence. [1] (c) Apply beneficence. [1] (d) Apply justice. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Autonomy means respecting the patient's informed choices and confidentiality rather than assuming unrestricted permission to use records. [1]	1
(b)	Non-maleficence requires avoiding preventable harm, including exposure or misuse of confidential information during AI processing. [1]	1
(c)	Beneficence means seeking genuine patient benefit through a suitable, carefully assessed and less intrusive data-use approach. [1]	1
(d)	Justice requires fair distribution of benefits and burdens; one patient or group should not be unfairly disadvantaged. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A dataset contains a fixed table of sensor readings, JSON event logs and unlabelled photographs. [4] (a) Identify the structured data. [1] (b) Identify the semi-structured data. [1] (c) Identify the unstructured data. [1] (d) Explain why all three still need preparation. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The fixed sensor-reading table is structured data because rows and columns follow a defined arrangement. [1]	1

Subpart	Model answer	Marks
(b)	JSON event logs are semi-structured: keys organise information without requiring one rigid relational table for every record. [1]	1
(c)	Photographs are unstructured data in this classification because their visual content is not organised as fixed table fields. [1]	1
(d)	All three still need quality checks and preparation; errors, missing information or inconsistent representations can mislead a model. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A vision system is trained to recognise product types using local shapes and edges in images. [4] (a) Name a suitable deep-learning architecture. [1] (b) Explain its relation to deep learning. [1] (c) State one role of learned visual features. [1] (d) State one evaluation requirement. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	A Convolutional Neural Network is suitable for learning local visual patterns in the product photographs. [1]	1
(b)	CNNs are neural-network architectures used within deep learning; multiple learned layers can represent increasingly complex features. [1]	1
(c)	Features such as edges, shapes and their combinations help distinguish product categories without memorising only whole training images. [1]	1
(d)	Evaluate on unseen photographs covering intended lighting, viewpoints and product variations to estimate performance outside training. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. For a crop-leaf screening task, define the condition being detected as positive. Out of 200 examples, 81 actual positives were predicted positive; 18 actual positives were missed; 9 actual negatives were predicted positive; and 92 actual negatives were predicted negative. [4] (a) Construct and label the confusion matrix. [2] (b) Calculate accuracy with working. [1] (c) Find the number of wrong predictions with working. **[1]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	See given computation	1
Precision	See given computation	1
Recall	See given computation	1
F1	See given computation	1

Suggested marking: 1 mark per metric.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "clean water helps plants" Document 2: "plants need clean air" Document 3: "clean air helps people" (a) Find the raw term frequency of each word in Document 1. [1] (b) Construct the document-frequency (DF) table for the corpus. [1] (c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1] (d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. **[1]**

ANSWER: Term frequency, document frequency and IDF ratio

Vocabulary term	TF in D1	DF	IDF ratio (3/DF)
clean	1	3	1
water	1	1	3
helps	1	2	1.5
plants	1	2	1.5
need	0	1	3
air	0	2	1.5
people	0	1	3

(a) Raw TF counts are in column 2. (b) DF counts the documents containing each word. (c) The IDF ratio is N/DF with N=3. (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)).

(c) (c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

SAMPLE PAPER 05 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). A science diagram communicates information through labelled arrows. Its main communication mode is: **[1]**

(a) Oral

(b) Visual

(c) Tactile

(d) Gossip

ANSWER: (b) Visual The diagram conveys meaning through visible symbols and spatial relationships.

Q1 (ii). Which action best shows responsible independent work? **[1]**

(a) Hide errors until submission.

(b) Wait for repeated reminders.

(c) Refuse all feedback.

(d) Finish the assigned task and seek clarification when needed.

ANSWER: (d) Finish the assigned task and seek clarification when needed. Independence includes responsibility, not refusing necessary help.

Q1 (iii). A student notices tension, identifies unfinished work as its cause, and creates a timetable. This is: **[1]**

(a) Stress management

(b) Escapism

(c) File management

(d) Uncontrolled risk taking

ANSWER: (a) Stress management The student identifies a stressor and takes a constructive step.

Q1 (iv). A program looks useful but secretly performs harmful actions when run. It is a: **[1]**

(a) Text extension

(b) Trojan

(c) Spreadsheet formula

(d) Backup utility by definition

ANSWER: (b) Trojan A Trojan disguises malicious behaviour as legitimate software.

Q1 (v). A supplier delivers late. Which response best reflects entrepreneurial responsibility? [1]

(a) Hide the delay indefinitely.

(b) Blame customers.

(c) Communicate honestly and arrange a practical alternative.

(d) Stop tracking orders.

ANSWER: (c) Communicate honestly and arrange a practical alternative. Accountability requires a response to the business problem.

Q1 (vi). Separating dry recyclable paper from wet food waste mainly helps: [1]

(a) Contaminating recyclable waste

(b) Increasing litter

(c) Removing the need to collect waste

(d) Material recovery

ANSWER: (d) Material recovery Separation can keep recoverable materials usable. Q2 - Answer any FIVE of SIX

Q2 (i). Use relative error = |actual - predicted|/actual. A numerical prediction is 616 when the actual value is 700. What is the relative error? [1]

(a) 0.012

(b) 0.12

(c) 0.88

(d) 1.12

ANSWER: (b) 0.12 |700-616|/700=0.12. This is relative numerical error, not a confusion-matrix count.

Q2 (ii). A system assigns a class label to each pixel but does not distinguish two cars of the same class. This is: [1]

(a) Speech synthesis

(b) Instance segmentation

(c) File compression

(d) Semantic segmentation

ANSWER: (d) Semantic segmentation Semantic segmentation labels pixel classes without separate instances.

Q2 (iii). A decision-maker values honesty, care and integrity when developing AI. This reflects: **[1]**

(a) Object detection

(b) Random selection

(c) Virtue-based reasoning

(d) Term frequency

ANSWER: (c) Virtue-based reasoning Virtue-based reasoning focuses on responsible character and conduct.

Q2 (iv). A classifier reports TP=24 and FP=16. What is its precision, rounded to two decimal places? **[1]**

(a) 60.00%

(b) 40.00%

(c) 80.00%

(d) 81.67%

ANSWER: (a) 60.00% Precision = 24/(24+16) x 100 = 60.00%.

Q2 (v). With "defective" defined as positive, a faulty item is predicted to be good. The outcome is: **[1]**

(a) False negative

(b) False positive

(c) True negative

(d) True positive

ANSWER: (a) False negative The actual positive was missed.

Q2 (vi). An ethical approach asks whether a decision respects fundamental rights. It is mainly: **[1]**

(a) Pixel-based

(b) Duty-based (deontological)

(c) Compression-based

(d) Frequency-based

ANSWER: (b) Duty-based (deontological) Rights-based reasoning prioritises rights and dignity. Q3 - Answer any FIVE of SIX

Q3 (i). A model has no predicted positives. Its precision formula has: **[1]**

(a) A zero denominator

(b) A value of one automatically

(c) No dependence on predictions

(d) The same denominator as accuracy

ANSWER: (a) A zero denominator TP + FP is zero; a reporting convention is needed for undefined precision.

Q3 (ii). A neuron receives x1=6, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 7

(b) 6

(c) 4

(d) 5

ANSWER: (d) 5 z = 0.5 x 6 + 1.5 x 2 - 1 = 5.

Q3 (iii). Which task predicts one of a set of named categories? **[1]**

(a) Classification

(b) File compression

(c) Regression of a continuous value

(d) Frequency normalisation only

ANSWER: (a) Classification Classification selects a category.

Q3 (iv). A system assigns the label "bicycle" to an entire photograph. The task is: **[1]**

(a) Image classification

(b) Document translation

(c) Instance segmentation

(d) Speech recognition

ANSWER: (a) Image classification A category is assigned to the whole image.

Q3 (v). A term occurs three times in one document. Using raw-count TF, its term frequency there is: **[1]**

(a) 1

(b) 3

(c) 0

(d) The number of documents only

ANSWER: (b) 3 Raw-count TF records repeated occurrences in that document.

Q3 (vi). A team returns to data collection after evaluation reveals missing examples. This shows that the AI project cycle is: **[1]**

(a) Iterative

(b) Always one-way

(c) Independent of data

(d) Completed before evaluation

ANSWER: (a) Iterative Evaluation can reveal a need to revisit earlier stages.

Q4 - Answer any FIVE of SIX

Q4 (i). Why can a basic Bag-of-Words representation confuse opposite statements? **[1]**

(a) It stores only images.

(b) It discards word order and much grammatical structure.

(c) It cannot store zeroes.

(d) It always removes every negative word.

ANSWER: (b) It discards word order and much grammatical structure. Order-sensitive meaning can be lost even if all words are retained.

Q4 (ii). After defining its objective, a team obtains relevant observations from reliable sources. This is: **[1]**

(a) Image segmentation

(b) Reward calculation

(c) Model evaluation

(d) Data acquisition

ANSWER: (d) Data acquisition Acquisition obtains the data needed for the task.

Q4 (iii). In a corpus of 6 documents, the term "smart" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 2

(b) 6

(c) 0.3333

(d) 3

ANSWER: (d) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q4 (iv). An 80:20 split of 250 examples gives how many test examples? **[1]**

(a) 20

(b) 50

(c) 200

(d) 80

ANSWER: (b) 50 The test share is 20 percent of 250.

Q4 (v). Which task selects the important terms that represent a document's subject? **[1]**

(a) Screen locking

(b) Keyword extraction

(c) Instance segmentation

(d) Pixel interpolation

ANSWER: (b) Keyword extraction Keyword extraction identifies useful content terms.

Q4 (vi). A team identifies affected users and the location of a recurring problem before collecting data. The stage is: **[1]**

(a) Problem scoping

(b) Classification

(c) Tokenisation

(d) Model deployment

ANSWER: (a) Problem scoping Problem scoping defines the problem and its context. Q5 - Answer any FIVE of SIX

Q5 (i). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "smart teams clean bins"? **[1]**

(a) 5

(b) 4

(c) 6

(d) 3

ANSWER: (b) 4 Tokens are smart, teams, clean, bins; repeated occurrences count separately.

Q5 (ii). A system draws a separate bounding box around each visible bus. The task is: **[1]**

(a) Object detection

(b) Text generation

(c) Sentiment analysis

(d) Only image storage

ANSWER: (a) Object detection Detection identifies and locates objects.

Q5 (iii). A model flags a sensor reading far outside typical operating behaviour. The task is: **[1]**

(a) Document translation

(b) Alphabetical sorting

(c) Speech synthesis

(d) Anomaly detection

ANSWER: (d) Anomaly detection Anomaly detection looks for unusual observations.

Q5 (iv). An image is 200 pixels wide and 120 pixels high. How many pixels does it contain? **[1]**

(a) 72000

(b) 320

(c) 24000

(d) 640

ANSWER: (c) 24000 Pixel count = width x height = 200 x 120 = 24000; RGB channels do not triple the number of spatial pixels.

Q5 (v). Which report is most transparent? **[1]**

(a) Present training accuracy as a guarantee.

(b) State the dataset, metrics, limitations and group-wise results.

(c) Hide errors in small groups.

(d) Show only the most favourable number.

ANSWER: (b) State the dataset, metrics, limitations and group-wise results. Meaningful evaluation includes context and limitations.

Q5 (vi). In a dataset of houses, floor area is used to predict rent. Floor area is a: **[1]**

(a) Test score

(b) Final decision rule by itself

(c) Predicted target value

(d) Feature

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (d) Feature A feature is an input attribute used by the model. Section B | Two-mark model answers

Q6. Explain constructive feedback with one suitable example. **[2]**

ANSWER: (1) Constructive feedback identifies a specific strength or weakness and suggests an improvement. [1] (2) For example, add a labelled diagram to clarify the process. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Describe two behaviours that demonstrate responsible independent work. [2]

ANSWER: (1) Take initiative and complete agreed tasks without repeated reminders. [1] (2) Monitor progress, accept responsibility and seek clarification when information is genuinely needed. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. Give two safe practices for using a shared computer. [2]

ANSWER: (1) Sign out of personal accounts and lock or end the session. [1] (2) Do not save personal passwords or leave confidential files accessible to others. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. State two reasons to collect customer feedback. [2]

ANSWER: (1) Feedback identifies unmet needs and problems in the current product. [1] (2) It helps test improvements rather than relying only on the owner's assumptions. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. State two ways a community can reduce unnecessary water use. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Detect and repair leaking pipes and taps before water is wasted. [1] (2) Monitor consumption and adopt water-efficient routines to identify and reduce other avoidable losses. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Explain transparency and accountability in an AI service. [2]

ANSWER: (1) Transparency makes relevant data use, limitations and decisions understandable. [1] (2) Accountability means identifiable people or organisations remain responsible for outcomes and corrective action. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Distinguish artificial intelligence from machine learning. [2]

ANSWER: (1) AI is the broader field of systems performing tasks associated with intelligence. [1] (2) Machine learning is an AI approach that learns patterns from examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Explain the relationship between deep learning and machine learning. [2]

ANSWER: (1) Deep learning is a subset of machine learning using multilayer neural networks. [1] (2) It can learn complex representations rather than relying only on manually selected features. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Explain train-test split and its purpose. [2]

ANSWER: (1) Divide data into examples used for training and a separate set used for evaluation. [1] (2) The held-out set estimates performance on unseen data. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Distinguish an RGB image from a grayscale image. [2]

ANSWER: (1) An RGB image usually stores separate red, green and blue channel values. [1] (2) A grayscale image stores one intensity value per pixel. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Distinguish term frequency and document frequency. [2]

Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Term frequency counts occurrences of a word within one document. [1] (2) Document frequency counts the number of documents

containing that word. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A council tests a flood-alert model using examples only from one well-connected town. [4] (a) Identify one scoping issue. [1] (b) Identify one data concern. [1] (c) Propose an evaluation action. [1] (d) Propose a communication safeguard. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	Scoping should identify every intended community and its warning needs, not merely the easiest town to study. [1]	1
(b)	One town's records may fail to represent remote areas, communication conditions or different patterns of flooding. [1]	1
(c)	Evaluate the model on relevant conditions from multiple communities and investigate differences in missed or false warnings. [1]	1
(d)	Use understandable warnings through accessible channels, state uncertainty and ensure a responsible human response remains available. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A machine-monitoring team notices readings far outside normal behaviour, although no fault labels are available. [4] (a) Identify the detection task. [1] (b) Explain the role of typical behaviour. [1] (c) Give one reason an alert may be false. [1] (d) Suggest an appropriate response. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	The task is anomaly detection: identifying observations that differ substantially from the patterns normally seen. [1]	1
(b)	The model or method uses typical behaviour as a reference against which unusual readings are compared. [1]	1
(c)	An unfamiliar but harmless operating condition can also differ from typical behaviour and generate a false alert. [1]	1
(d)	Inspect the equipment and supporting evidence before concluding that a fault exists; retain an appropriate human decision process. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. Two systems estimate delivery time: one uses fixed speed rules and the other learns from past journeys. [4] (a) Identify the approach in each system. [1] (b) Explain one advantage of the fixed rules. [1] (c) Explain one advantage of the learned model. [1] (d) State one limitation of the learned model. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The fixed-speed estimator is rule-based, while the estimator fitted to previous journeys is learning-based. [1]	1
(b)	Explicit rules can be inspected directly and applied consistently when the task and operating assumptions are simple. [1]	1
(c)	Learning from representative examples can capture relationships among several features that simple fixed-speed rules may miss. [1]	1

Subpart	Model answer	Marks
(d)	The learned estimator can still fail when future routes or conditions differ substantially from those represented in training. [1]	1

Suggested marking: Identification of both approaches: 0.5 each. The other three explained subparts: 1 each. Total 4.

Q20. A machine-fault alerts model has TP=24, FP=16, FN=6 and TN=74. Calculate accuracy, precision, recall and F1 score. Show formula and substitution for each. Round percentages to two decimal places. [4]

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	(TP + TN)/N x 100 = (24 + 74)/120 x 100 = 81.67%. [1]	1
Precision	TP/(TP + FP) x 100 = 24/(24 + 16) x 100 = 60.00%. [1]	1
Recall	TP/(TP + FN) x 100 = 24/(24 + 6) x 100 = 80.00%. [1]	1
F1	2TP/(2TP + FP + FN) x 100 = 48/(48 + 16 + 6) x 100 = 68.57%. [1]	1

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "smart bins collect waste" Document 2: "teams sort waste" Document 3: "smart teams clean bins" (a) List the vocabulary in first-appearance order. [1] (b) Construct the count vector for Document 3 in that order. [1] (c) Explain how the vector is useful to a text classifier. [1] (d) State one order-related limitation of this representation. **[1]**

ANSWER: Fixed-vocabulary Bag-of-Words count table

Vocabulary term	Document 1	Document 2	Document 3
smart	1	0	1
bins	1	0	1
collect	1	0	0
waste	1	1	0
teams	0	1	1
sort	0	1	0
clean	0	0	1

(a)/(b) (a) Vocabulary: [smart, bins, collect, waste, teams, sort, clean]. [1] (b) Document 3 vector: [1, 1, 0, 0, 1, 0, 1]. [1]

(c) (c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) (d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

SAMPLE PAPER 06 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). Which sentence uses a courteous request? [1]

(a) Send it now, or else.

(b) Could you please send the revised file by noon?

(c) Only a careless person would delay this.

(d) You never do anything properly.

ANSWER: (b) Could you please send the revised file by noon? Courtesy maintains respect while making a clear request.

Q1 (ii). A student notices tension, identifies unfinished work as its cause, and creates a timetable. This is: [1]

(a) Uncontrolled risk taking

(b) Escapism

(c) File management

(d) Stress management

ANSWER: (d) Stress management The student identifies a stressor and takes a constructive step.

Q1 (iii). What does the M in a SMART goal stand for? [1]

(a) Maximum

(b) Mechanical

(c) Memorable

(d) Measurable

ANSWER: (d) Measurable Progress needs a measure that can be checked.

Q1 (iv). What does an operating system primarily manage? [1]

(a) Hardware resources and program execution

(b) Only the appearance of a website

(c) Only a user's marks

(d) Only printed documents

ANSWER: (a) Hardware resources and program execution The operating system coordinates resources and application processes.

Q1 (v). Which is primarily an entrepreneurial function? [1]

(a) Avoiding every decision

(b) Waiting for all needs to disappear

(c) Guaranteeing that no competitor exists

(d) Organising resources to provide a useful product or service

ANSWER: (d) Organising resources to provide a useful product or service Entrepreneurs combine resources and take decisions to create value.

Q1 (vi). Which action reduces energy consumption without changing the intended task? [1]

(a) Leave all equipment on overnight.

(b) Ignore idle equipment.

(c) Switch off unused lights and devices.

(d) Use more devices for the same task.

ANSWER: (c) Switch off unused lights and devices. Avoiding unnecessary operation saves energy. Q2 - Answer any FIVE of SIX

Q2 (i). Which task gives different pixel masks to two separate people in one photograph? [1]

(a) Instance segmentation

(b) Colour inversion only

(c) Tokenisation

(d) Whole-image classification

ANSWER: (a) Instance segmentation Instance segmentation separates individual objects at pixel level.

Q2 (ii). A Bag-of-Words vector normally preserves: [1]

(a) The complete original sentence order

(b) Every grammatical relationship

(c) Term counts in a fixed vocabulary order

(d) Tone of voice automatically

ANSWER: (c) Term counts in a fixed vocabulary order The positions refer to vocabulary entries, not original word order.

Q2 (iii). An image is 220 pixels wide and 130 pixels high. How many pixels does it contain? **[1]**

(a) 85800

(b) 700

(c) 350

(d) 28600

ANSWER: (d) 28600 Pixel count = width x height = 220 x 130 = 28600; RGB channels do not triple the number of spatial pixels.

Q2 (iv). After defining its objective, a team obtains relevant observations from reliable sources. This is: **[1]**

(a) Image segmentation

(b) Model evaluation

(c) Data acquisition

(d) Reward calculation

ANSWER: (c) Data acquisition Acquisition obtains the data needed for the task.

Q2 (v). With "defective" defined as positive, a faulty item is predicted to be good. The outcome is: **[1]**

(a) False positive

(b) True positive

(c) False negative

(d) True negative

ANSWER: (c) False negative The actual positive was missed.

Q2 (vi). Which question belongs to the WHY part of a 4Ws canvas? **[1]**

(a) What exactly is the problem?

(b) Who experiences the problem?

(c) Where is the problem observed?

(d) Why is solving this problem valuable?

ANSWER: (d) Why is solving this problem valuable? WHY explains the value or need for solving the problem. Q3 - Answer any FIVE of SIX

Q3 (i). A model estimates the number of minutes needed for a journey. This is usually: **[1]**

(a) Tokenisation

(b) Clustering

(c) Classification into fixed names

(d) Regression

ANSWER: (d) Regression The target is a numerical quantity rather than a category label.

Q3 (ii). A neuron receives x1=7, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 6.5

(b) 4.5

(c) 7.5

(d) 5.5

ANSWER: (d) 5.5 z = 0.5 x 7 + 1.5 x 2 - 1 = 5.5.

Q3 (iii). A team plots observations to detect outliers and patterns. This belongs mainly to: **[1]**

(a) Password management

(b) Profit calculation

(c) Final publication only

(d) Data exploration

ANSWER: (d) Data exploration Exploration helps understand the data before modeling.

Q3 (iv). In a corpus of 6 documents, the term "local" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 6

(b) 2

(c) 3

(d) 0.3333

ANSWER: (c) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q3 (v). A classifier reports TP=48 and FP=8. What is its precision, rounded to two decimal places? [1]

(a) 90.00%

(b) 85.71%

(c) 14.29%

(d) 80.00%

ANSWER: (b) 85.71% Precision = 48/(48+8) x 100 = 85.71%.

Q3 (vi). A digital image's smallest addressable picture element is a: [1]

(a) Paragraph

(b) Token

(c) Record label

(d) Pixel

ANSWER: (d) Pixel A pixel stores image information at a spatial position. Q4 - Answer any FIVE of SIX

Q4 (i). Which denominator belongs to precision? [1]

(a) TP + FP

(b) TP + FN

(c) All cases in every situation

(d) TN + FN

ANSWER: (a) TP + FP Precision conditions on predicted positives.

Q4 (ii). With "spam" defined as positive, a legitimate message is blocked as spam. The outcome is: [1]

(a) False negative

(b) True positive

(c) True negative

(d) False positive

ANSWER: (d) False positive A negative item was incorrectly predicted positive.

Q4 (iii). Examples contain measurements together with the correct category for each measurement. The learning setting is: **[1]**

(a) Data encryption

(b) Unsupervised learning

(c) Random sorting

(d) Supervised learning

ANSWER: (d) Supervised learning Known target labels guide supervised learning.

Q4 (iv). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "trains connect towns"? **[1]**

(a) 3

(b) 2

(c) 4

(d) 5

ANSWER: (a) 3 Tokens are trains, connect, towns; repeated occurrences count separately.

Q4 (v). A team identifies affected users and the location of a recurring problem before collecting data. The stage is: **[1]**

(a) Classification

(b) Problem scoping

(c) Model deployment

(d) Tokenisation

ANSWER: (b) Problem scoping Problem scoping defines the problem and its context.

Q4 (vi). What is removed when a process deletes a specified list of common words? **[1]**

(a) All proper nouns automatically

(b) Every meaningful term

(c) All punctuation only

(d) Stop words

ANSWER: (d) Stop words Stop-word removal uses a chosen list of common terms. Q5 - Answer any FIVE of SIX

Q5 (i). Which denominator belongs to recall? [1]

(a) TN only

(b) TP + FP

(c) TN + FP

(d) TP + FN

ANSWER: (d) TP + FN Recall conditions on actual positives.

Q5 (ii). Use relative error = |actual - predicted|/actual. A numerical prediction is 660 when the actual value is 750. What is the relative error? [1]

(a) 1.12

(b) 0.012

(c) 0.88

(d) 0.12

ANSWER: (d) 0.12 |750-660|/750=0.12. This is relative numerical error, not a confusion-matrix count.

Q5 (iii). A model works badly for a community absent from the training examples. Which concern is central? [1]

(a) Word order only

(b) A larger font size

(c) A file extension error by definition

(d) Representational bias

ANSWER: (d) Representational bias Unrepresentative data can create unequal performance.

Q5 (iv). Unlabelled records are grouped by similarity. Which task is being performed? [1]

(a) Label transcription

(b) Decryption

(c) Clustering

(d) Supervised regression

ANSWER: (c) Clustering Clustering seeks groups without supplied group labels.

Q5 (v). A system assigns a complaint to "delivery", "payment" or "product". This is: [1]

(a) Text classification

(b) Image segmentation

(c) Face detection

(d) Regression of pixel values

ANSWER: (a) Text classification The system assigns a text category.

Q5 (vi). An 8-bit grayscale channel has how many possible values? [1]

(a) 256

(b) 8

(c) 255

(d) 16

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (a) 256 Eight bits represent 2 to the power 8 values, from 0 to 255. Section B | Two-mark model answers

Q6. State the purpose of a clear subject line and a polite closing in an email. [2]

ANSWER: (1) A clear subject line tells the reader the purpose of an email. [1] (2) A polite closing states the requested next step without sounding demanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Suggest two improvements to an unrealistic revision timetable. [2]

ANSWER: (1) Set achievable priorities rather than assigning excessive work each day. [1] (2) Allocate time for rest and review progress so the timetable can be adjusted. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. State two functions of an operating system. [2]

ANSWER: (1) It manages hardware resources such as memory and storage. [1] (2) It provides services and an interface for running applications and handling files. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. A batch earns Rs 3,600 revenue and costs Rs 2,850 in total. Find profit and distinguish it from revenue. [2]

ANSWER: (1) Profit = revenue - total cost = 3,600 - 2,850 = Rs 750. [1] (2) Revenue is total sales income before deducting costs. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. State two factors, apart from purchase price, for judging the sustainability of a device. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Consider the energy and resources a device consumes throughout everyday use. [1] (2) Also assess its durability, repairability and options for responsible collection or recycling at end of life. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Why is the AI project cycle described as iterative? [2]

ANSWER: (1) Evaluation can reveal weaknesses in the data, model or original problem definition. [1] (2) The team may revisit earlier stages and improve the solution using that feedback. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Distinguish structured and unstructured data with examples. [2]

ANSWER: (1) Structured data follows a defined arrangement, such as a table of sales records. [1] (2) Unstructured data lacks that table-like schema, such as photographs or free text. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Explain semi-structured data with one example. [2]

ANSWER: (1) Semi-structured data contains organising tags or keys without requiring one fixed table schema. [1] (2) A JSON event log with named fields is one example. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Explain underfitting and contrast it with overfitting. [2]

ANSWER: (1) Underfitting occurs when a model misses important patterns and performs poorly even in training. [1] (2) Overfitting fits training details but generalises poorly. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. State two reasons a visual model may fail in a new environment. [2]

ANSWER: (1) Lighting, viewing angle or obstruction may hide important features. [1] (2) The new images may differ from those represented in its training data. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain why sentiment analysis must consider context. [2]

Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Positive words may occur in sarcastic or negative statements. [1] (2) Context and negation can change the intended sentiment beyond individual word meanings. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. An AI assistant recommends jobs using historical records that contain very few examples from some communities. [4] (a) Identify the fairness risk. [1] (b) State a rights-based concern. [1] (c) Give one technical response. [1] (d) Give one organisational response. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Historical underrepresentation can make suitable applicants from some communities less likely to receive appropriate job recommendations. [1]	1
(b)	A rights-based approach requires fair consideration rather than exclusion through irrelevant characteristics or proxies for protected groups. [1]	1
(c)	Improve representation and test relevant group-wise outcomes to identify and investigate avoidable disparities in recommendations. [1]	1
(d)	Publish meaningful criteria, retain human review and provide an accessible way for applicants to challenge unsuitable outcomes. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. An ANN receives numerical weather features and predicts whether outdoor work should be delayed. [4] (a) Explain the purpose of input features. [1] (b) Explain weights. [1] (c) Explain bias. [1] (d) Explain activation. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Input features numerically encode relevant weather information so that the network can process the observed conditions. [1]	1

Subpart	Model answer	Marks
(b)	Weights scale each input's contribution, determining how strongly that signal influences a neuron's combined input. [1]	1
(c)	Bias shifts the combined input independently of the feature values, allowing a more flexible position for a decision boundary. [1]	1
(d)	An activation function transforms the combined signal before it is passed onward; nonlinear activations enable more complex relationships. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A service needs to assign messages to named departments and estimate the time each message will take to resolve. [4] (a) Identify the first supervised task. [1] (b) Identify the second supervised task. [1] (c) State what labels are needed for the first task. [1] (d) State what targets are needed for the second task. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Assigning a message to a named department is classification because the output belongs to a set of categories. [1]	1
(b)	Estimating time to resolve the message is regression because the output is a numerical duration. [1]	1
(c)	The classification training examples should pair message information with the correct department category determined reliably. [1]	1

Subpart	Model answer	Marks
(d)	The regression examples should pair available message features with actual recorded resolution times measured in a consistent unit. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. For a message-spam detection task, define the condition being detected as positive. Out of 200 examples, 48 actual positives were predicted positive; 12 actual positives were missed; 8 actual negatives were predicted positive; and 132 actual negatives were predicted negative. [4] (a) Construct and label the confusion matrix. [2] (b) Calculate accuracy with working. [1] (c) Find the number of wrong predictions with working. **[1]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	See given computation	1
Precision	See given computation	1
Recall	See given computation	1
F1	See given computation	1

Suggested marking: 1 mark per metric.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "local trains save time" Document 2: "local buses save fuel" Document 3: "trains connect towns" (a) Find the raw term frequency of each word in Document 1. [1] (b) Construct the document-frequency (DF) table for the corpus. [1] (c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1] (d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. **[1]**

ANSWER: Term frequency, document frequency and IDF ratio

Vocabulary term	TF in D1	DF	IDF ratio (3/DF)
local	1	2	1.5
trains	1	2	1.5
save	1	2	1.5
time	1	1	3
buses	0	1	3
fuel	0	1	3
connect	0	1	3
towns	0	1	3

(a) Raw TF counts are in column 2. (b) DF counts the documents containing each word. (c) The IDF ratio is N/DF with N=3. (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)).

(c) (c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

SAMPLE PAPER 07 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). A receiver misunderstands a correct spoken message because of loud machinery. The obstacle is mainly: **[1]**

(a) Feedback

(b) Physical noise

(c) Non-verbal encouragement

(d) A grammatical article

ANSWER: (b) Physical noise Environmental sound interferes with the channel.

Q1 (ii). What does the M in a SMART goal stand for? **[1]**

(a) Mechanical

(b) Memorable

(c) Measurable

(d) Maximum

ANSWER: (c) Measurable Progress needs a measure that can be checked.

Q1 (iii). A revision plan includes no rest and several impossible daily targets. Which correction is best? **[1]**

(a) Add more tasks each day.

(b) Avoid reviewing progress.

(c) Set achievable targets and allow breaks.

(d) Remove all deadlines.

ANSWER: (c) Set achievable targets and allow breaks. A workable plan should be realistic and sustainable.

Q1 (iv). Which practice is safest on a shared computer? **[1]**

(a) Share passwords with the next user.

(b) Sign out after work and do not save personal passwords.

(c) Disable the screen lock.

(d) Leave the session unlocked.

ANSWER: (b) Sign out after work and do not save personal passwords. Ending the session reduces unauthorised access.

Q1 (v). Customer feedback shows demand for a smaller package. Revising the package shows: [1]

(a) Denial

(b) Unemployment

(c) Monopoly by definition

(d) Adaptability

ANSWER: (d) Adaptability The business adapts its offering to evidence about customer needs.

Q1 (vi). A product designed to be repaired and used longer supports: [1]

(a) Resource efficiency

(b) Uncontrolled disposal

(c) Planned waste generation

(d) Unlimited extraction

ANSWER: (a) Resource efficiency Extending useful life can reduce replacement demand. Q2 - Answer any FIVE of SIX

Q2 (i). An AI service states what data it uses and explains important limitations. This supports: [1]

(a) Transparency

(b) Uncontrolled access

(c) Secrecy

(d) Unlimited data retention

ANSWER: (a) Transparency Transparency makes relevant processes and limits understandable.

Q2 (ii). Which ethical principle concerns a person's informed control over a decision affecting them? [1]

(a) Compression

(b) Autonomy

(c) Clustering

(d) Regression

ANSWER: (b) Autonomy Autonomy concerns informed choice and agency.

Q2 (iii). A classifier reports TP=35 and FP=5. What is its precision, rounded to two decimal places? **[1]**

(a) 80.00%

(b) 87.50%

(c) 12.50%

(d) 70.00%

ANSWER: (b) 87.50% Precision = 35/(35+5) x 100 = 87.50%.

Q2 (iv). Which task is optical character recognition designed to perform? **[1]**

(a) Recognise every sound in a recording

(b) Predict future weather from time alone

(c) Cluster music without labels

(d) Extract text characters from an image

ANSWER: (d) Extract text characters from an image OCR converts visible characters into machine-readable text.

Q2 (v). An image is 240 pixels wide and 140 pixels high. How many pixels does it contain? **[1]**

(a) 100800

(b) 760

(c) 380

(d) 33600

ANSWER: (d) 33600 Pixel count = width x height = 240 x 140 = 33600; RGB channels do not triple the number of spatial pixels.

Q2 (vi). A model performs poorly on both training and test data because it is too simple. This suggests: **[1]**

(a) Perfect calibration

(b) No model error

(c) Ideal complexity

(d) Underfitting

ANSWER: (d) Underfitting The model fails to capture important relationships.
Q3 - Answer any FIVE of SIX

Q3 (i). A high-stakes AI recommendation is reviewed by a responsible person before action. This provides: **[1]**

(a) Data duplication

(b) Automatic infallibility

(c) Anonymity by definition

(d) Human oversight

ANSWER: (d) Human oversight Human review supports responsibility and challenge of outputs.

Q3 (ii). A model works badly for a community absent from the training examples. Which concern is central? **[1]**

(a) Word order only

(b) Representational bias

(c) A larger font size

(d) A file extension error by definition

ANSWER: (b) Representational bias Unrepresentative data can create unequal performance.

Q3 (iii). Which choice describes lemmatisation? **[1]**

(a) Assigning words random colours

(b) Mapping an inflected form to an appropriate dictionary base form

(c) Counting image pixels

(d) Deleting random letters

ANSWER: (b) Mapping an inflected form to an appropriate dictionary base form Lemmatisation aims at meaningful base forms using linguistic information.

Q3 (iv). A rule says, "If stock is below ten units, send an alert." No examples are used to learn it. The approach is: **[1]**

(a) Learned regression by definition

(b) Unsupervised clustering

(c) Reinforcement learning

(d) Rule-based

ANSWER: (d) Rule-based The decision rule was explicitly specified.

Q3 (v). Use relative error = |actual - predicted|/actual. A numerical prediction is 704 when the actual value is 800. What is the relative error? **[1]**

(a) 1.12

(b) 0.012

(c) 0.12

(d) 0.88

ANSWER: (c) 0.12 |800-704|/800=0.12. This is relative numerical error, not a confusion-matrix count.

Q3 (vi). Which sequence best describes a vision pipeline? **[1]**

(a) Interpretation, real scene, sensing device

(b) Sensor, interpretation, creation of reality

(c) Real scene, sensing device, interpretation

(d) Stored answer, scene, sensor

ANSWER: (c) Real scene, sensing device, interpretation The sensor captures visual data that the system processes. Q4 - Answer any FIVE of SIX

Q4 (i). Which question belongs to the WHY part of a 4Ws canvas? **[1]**

(a) Why is solving this problem valuable?

(b) What exactly is the problem?

(c) Who experiences the problem?

(d) Where is the problem observed?

ANSWER: (a) Why is solving this problem valuable? WHY explains the value or need for solving the problem.

Q4 (ii). A menu-driven bot responds through fixed paths. It is mainly a: **[1]**

(a) Script bot

(b) Learning-based bot by definition

(c) Vision sensor

(d) Regression dataset

ANSWER: (a) Script bot The response paths are explicitly designed.

Q4 (iii). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "readers discuss ideas"? **[1]**

(a) 3

(b) 5

(c) 2

(d) 4

ANSWER: (a) 3 Tokens are readers, discuss, ideas; repeated occurrences count separately.

Q4 (iv). A sentence "Wonderful, the train is late again" expresses frustration despite a positive word. The difficulty is: **[1]**

(a) Disk capacity

(b) Image size

(c) Sarcasm and context

(d) A missing file extension

ANSWER: (c) Sarcasm and context The intended sentiment differs from a literal word-level reading.

Q4 (v). A neuron receives x1=8, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 5

(b) 7

(c) 8

(d) 6

ANSWER: (d) 6 z = 0.5 x 8 + 1.5 x 2 - 1 = 6.

Q4 (vi). Which statement correctly relates AI, ML and DL? **[1]**

(a) Machine learning contains every form of intelligence.

(b) Deep learning is a part of machine learning, which is a part of AI.

(c) Deep learning and AI have no connection.

(d) All AI must use deep learning.

ANSWER: (b) Deep learning is a part of machine learning, which is a part of AI. The fields form a nested relationship, but not all AI is learning-based. Q5 - Answer any FIVE of SIX

Q5 (i). Why keep a test set separate from training? [1]

(a) To make labels unnecessary

(b) To ensure all predictions become correct

(c) To avoid ever collecting data

(d) To estimate performance on unseen examples

ANSWER: (d) To estimate performance on unseen examples Held-out examples give a more realistic evaluation of generalisation.

Q5 (ii). An agent changes its actions after receiving rewards from an environment. This is: [1]

(a) Image compression

(b) Rule copying

(c) File sorting

(d) Reinforcement learning

ANSWER: (d) Reinforcement learning The agent learns an action strategy from feedback on its actions.

Q5 (iii). Which denominator belongs to recall? [1]

(a) TP + FP

(b) TN only

(c) TP + FN

(d) TN + FP

ANSWER: (c) TP + FN Recall conditions on actual positives.

Q5 (iv). A model memorises training details but performs poorly on new examples. This suggests: [1]

(a) Ideal generalisation

(b) Overfitting

(c) A guaranteed fair model

(d) Underfitting only

ANSWER: (b) Overfitting The learned behaviour does not generalise well.

Q5 (v). In a corpus of 6 documents, the term "readers" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 3

(b) 0.3333

(c) 2

(d) 6

ANSWER: (a) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q5 (vi). Why may a vision model fail on very dark images? **[1]**

(a) All labels are automatically deleted.

(b) Dark images cannot contain pixels.

(c) Resolution must always become zero.

(d) Useful visual detail may be lost or differ from training examples.

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (d) Useful visual detail may be lost or differ from training examples. Visibility and data distribution can affect recognition. Section B | Two-mark model answers

Q6. Define a pronoun and illustrate it in a sentence. **[2]**

ANSWER: (1) A pronoun replaces a noun to avoid unnecessary repetition. [1] (2) In "Meera checked the file; she corrected it", "she" replaces Meera and "it" replaces the file. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. List four steps for building self-motivation. **[2]**

ANSWER: (1) Identify personal strengths and areas for improvement. [0.5] (2) Set clear and achievable goals. [0.5] (3) Prepare a practical action plan with manageable steps. [0.5] (4) Continue working and reviewing progress despite difficulties. [0.5]

Suggested marking: 0.5 mark for each of four distinct relevant steps; total 2 marks.

Q8. Explain how a Trojan can threaten a computer. [2]

ANSWER: (1) A Trojan disguises itself as useful or legitimate software. [1] (2) When executed, it can perform harmful actions, such as stealing information or permitting unauthorised access. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Explain innovation with a simple business example. [2]

ANSWER: (1) Innovation applies a useful new or improved idea to meet a need. [1] (2) For example, a repair business could introduce convenient scheduled collection for customers. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Explain responsible consumption with one example. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Responsible consumption uses resources carefully and avoids unnecessary waste. [1] (2) For example, buying a durable, repairable product can reduce repeated replacement and disposal. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Explain the Who and Why questions in problem scoping. [2]

ANSWER: (1) Who identifies the people affected by the problem and intended solution. [1] (2) Why explains the reason for solving it and the benefit the solution should provide. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Define clustering and state why labels are not required. [2]

ANSWER: (1) Clustering groups observations that are similar under a chosen representation. [1] (2) It discovers groups from the data rather than learning supplied correct group labels. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Identify the agent and the environment in a learning robot game. [2]

ANSWER: (1) The agent is the robot or program selecting actions. [1] (2) The environment is the game world that responds to those actions with observations and feedback. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Explain why false negatives matter in an early-warning system. [2]

ANSWER: (1) A false negative misses an actual positive event that needed attention. [1] (2) Missing a real warning condition may delay a response, so recall deserves particular attention. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Distinguish image processing from computer vision. [2]

ANSWER: (1) Image processing changes or improves pixel data, for example adjusting brightness. [1] (2) Computer vision interprets visual content, for example identifying and locating a vehicle. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain semantic analysis with one example. [2] Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Semantic analysis examines meaning in words and sentences. [1] (2) For example, it distinguishes a financial bank from a river bank using relevant information from the text. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A team releases a model after reporting its best test result while omitting several weak results. [4] (a) Identify the transparency failure. [1] (b) Identify the accountability requirement. [1] (c) Suggest an honest evaluation practice. [1] (d) Explain why this matters to users. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	Reporting only the strongest test result conceals weak performance and creates a misleading impression of reliability. [1]	1
(b)	Accountability requires identifiable people or teams to explain the evaluation and correct unsupported performance claims. [1]	1
(c)	Report relevant test conditions, representative metrics, variation and failures instead of selectively choosing the most favourable result. [1]	1
(d)	Users need balanced evidence to judge suitability and avoid relying on a system in circumstances it cannot handle. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A rule-based safety alert triggers whenever a temperature crosses a fixed threshold. [4] (a) Explain why it is rule-based. [1] (b) State one strength. [1] (c) State one limitation. [1] (d) Suggest a situation where learning from data could help. [1]

ANSWER:

Subpart	Model answer	Marks
(a)	The alert is rule-based because a person specifies the threshold condition rather than learning it from examples. [1]	1
(b)	Its operation is straightforward to inspect: the same temperature condition triggers the same stated response. [1]	1
(c)	One fixed threshold may miss complex situations involving other features or generate unnecessary alerts in harmless conditions. [1]	1

Subpart	Model answer	Marks
(d)	Learning may help when representative operating histories link several relevant measurements to reliably labelled safe and unsafe situations. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A team uses labelled photographs to train a classifier but accidentally includes the same images in its test set. (a) Identify the learning setting. [1] (b) Identify the data-splitting mistake. [1] (c) Explain the likely effect on reported performance. [1] (d) Suggest a correction. **[1]** [4]

ANSWER:

Subpart	Model answer	Marks
(a)	This is supervised learning because photographs are paired with target categories that guide the classifier's training. [1]	1
(b)	Repeating training images in the test set creates overlap and leaks information into the supposed unseen evaluation. [1]	1
(c)	Reported accuracy may be unrealistically favourable because familiar images do not provide an independent generalisation check. [1]	1
(d)	Create appropriate non-overlapping training and test sets, prevent similar-record leakage where relevant, and repeat the performance evaluation. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. A road-sign recognition model has TP=35, FP=5, FN=15 and TN=45. Calculate accuracy, precision, recall and F1 score. Show formula and substitution for each. Round percentages to two decimal places. **[4]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	(TP + TN)/N x 100 = (35 + 45)/100 x 100 = 80.00%. [1]	1
Precision	TP/(TP + FP) x 100 = 35/(35 + 5) x 100 = 87.50%. [1]	1
Recall	TP/(TP + FN) x 100 = 35/(35 + 15) x 100 = 70.00%. [1]	1
F1	2TP/(2TP + FP + FN) x 100 = 70/(70 + 5 + 15) x 100 = 77.78%. [1]	1

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "readers share useful books" Document 2: "useful books teach ideas" Document 3: "readers discuss ideas" (a) List the vocabulary in first-appearance order. [1] (b) Construct the count vector for Document 3 in that order. [1] (c) Explain how the vector is useful to a text classifier. [1] (d) State one order-related limitation of this representation. **[1]**

ANSWER: Fixed-vocabulary Bag-of-Words count table

Vocabulary term	Document 1	Document 2	Document 3
readers	1	0	1
share	1	0	0
useful	1	1	0
books	1	1	0
teach	0	1	0

Vocabulary term	Document 1	Document 2	Document 3
ideas	0	1	1
discuss	0	0	1

(a)/(b) (a) Vocabulary: [readers, share, useful, books, teach, ideas, discuss]. [1]
(b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1]. [1]

(c) (c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) (d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

SAMPLE PAPER 08 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). Which statement is a complete sentence? [1]

(a) Because of the labels.

(b) Checking every.

(c) After the volunteers.

(d) The volunteers checked every label.

ANSWER: (d) The volunteers checked every label. A complete sentence expresses a complete thought with a suitable subject and predicate.

Q1 (ii). A revision plan includes no rest and several impossible daily targets. Which correction is best? [1]

(a) Remove all deadlines.

(b) Set achievable targets and allow breaks.

(c) Avoid reviewing progress.

(d) Add more tasks each day.

ANSWER: (b) Set achievable targets and allow breaks. A workable plan should be realistic and sustainable.

Q1 (iii). A learner misses a target and reviews the cause before changing the plan. This mainly demonstrates: [1]

(a) Denial

(b) Self-regulation

(c) Carelessness

(d) Dependence

ANSWER: (b) Self-regulation Monitoring and adjusting behaviour supports progress.

Q1 (iv). A folder hierarchy is mainly useful for: [1]

(a) Organising related files for retrieval

(b) Making every file executable

(c) Eliminating malware automatically

(d) Increasing the physical size of a disk

ANSWER: (a) Organising related files for retrieval Folders group files in a structured way.

Q1 (v). Which distinction is generally valid? [1]

(a) Employees never exercise initiative.

(b) Entrepreneurs never receive income.

(c) Every employee owns the employer's business.

(d) An entrepreneur bears business uncertainty; an employee works under agreed employment terms.

ANSWER: (d) An entrepreneur bears business uncertainty; an employee works under agreed employment terms. Ownership and business uncertainty differ from an employment role.

Q1 (vi). Which statement about renewable resources is most accurate? [1]

(a) They can never be depleted locally.

(b) They are identical to all fossil fuels.

(c) They always have zero environmental impact.

(d) They can replenish naturally, but still require responsible use.

ANSWER: (d) They can replenish naturally, but still require responsible use. Renewability does not remove the need for careful use. Q2 - Answer any FIVE of SIX

Q2 (i). A clinic checks that an AI allocation system does not disadvantage a particular community. Which principle is most directly involved? [1]

(a) Encoding

(b) Resolution

(c) Tokenisation

(d) Justice

ANSWER: (d) Justice Justice concerns fair treatment and distribution.

Q2 (ii). A text system resolves what "she" refers to using the previous sentence. This requires: [1]

(a) Hardware cleaning

(b) Alphabetical sorting

(c) Single-pixel classification

(d) Discourse integration

ANSWER: (d) Discourse integration Meaning can depend on relationships across sentences.

Q2 (iii). Use relative error = |actual - predicted|/actual. A numerical prediction is 748 when the actual value is 850. What is the relative error? **[1]**

(a) 0.12

(b) 1.12

(c) 0.012

(d) 0.88

ANSWER: (a) 0.12 |850-748|/850=0.12. This is relative numerical error, not a confusion-matrix count.

Q2 (iv). Avoiding preventable harm from an AI recommendation is the principle of: **[1]**

(a) Non-maleficence

(b) External motivation

(c) Dimensionality reduction

(d) Profit maximisation alone

ANSWER: (a) Non-maleficence Non-maleficence means avoiding harm.

Q2 (v). Why is an activation function used after a neuron's weighted sum? **[1]**

(a) To replace all training examples

(b) To transform the signal, often introducing nonlinearity

(c) To rename every training record

(d) To guarantee perfect accuracy

ANSWER: (b) To transform the signal, often introducing nonlinearity Activation enables useful transformations; nonlinear activations support complex patterns.

Q2 (vi). An AI service states what data it uses and explains important limitations. This supports: **[1]**

(a) Transparency

(b) Secrecy

(c) Uncontrolled access

(d) Unlimited data retention

ANSWER: (a) Transparency Transparency makes relevant processes and limits understandable. Q3 - Answer any FIVE of SIX

Q3 (i). Which representation typically has three colour channels? **[1]**

(a) RGB image

(b) A word-count vector

(c) A single grayscale channel

(d) A file path

ANSWER: (a) RGB image RGB stores red, green and blue components.

Q3 (ii). Why keep a test set separate from training? **[1]**

(a) To make labels unnecessary

(b) To ensure all predictions become correct

(c) To estimate performance on unseen examples

(d) To avoid ever collecting data

ANSWER: (c) To estimate performance on unseen examples Held-out examples give a more realistic evaluation of generalisation.

Q3 (iii). A framework compares the likely benefits and harms of an AI action. Its emphasis is: **[1]**

(a) File type

(b) Alphabetical order

(c) Pixel depth

(d) Consequences or utility

ANSWER: (d) Consequences or utility Consequence-based reasoning considers the effects of decisions.

Q3 (iv). An image is 260 pixels wide and 150 pixels high. How many pixels does it contain? **[1]**

(a) 820

(b) 117000

(c) 410

(d) 39000

ANSWER: (d) 39000 Pixel count = width x height = 260 x 150 = 39000; RGB channels do not triple the number of spatial pixels.

Q3 (v). A model predicts the majority class for every record and obtains a high overall score. What should be examined next? **[1]**

(a) Class-specific errors and recall

(b) Only the number of columns

(c) Only the font used in the report

(d) Only the file name

ANSWER: (a) Class-specific errors and recall Class imbalance can conceal failure on a smaller class.

Q3 (vi). A high-stakes AI recommendation is reviewed by a responsible person before action. This provides: **[1]**

(a) Automatic infallibility

(b) Data duplication

(c) Human oversight

(d) Anonymity by definition

ANSWER: (c) Human oversight Human review supports responsibility and challenge of outputs. Q4 - Answer any FIVE of SIX

Q4 (i). Which operation risks data leakage? **[1]**

(a) Reserving unseen examples

(b) Using the final test results repeatedly to choose the model

(c) Reporting all confusion-matrix counts

(d) Documenting the positive class

ANSWER: (b) Using the final test results repeatedly to choose the model Repeated selection using the final test information contaminates evaluation.

Q4 (ii). F1 is the harmonic mean of: **[1]**

(a) Training and test counts

(b) TP and TN

(c) Precision and recall

(d) Accuracy and file size

ANSWER: (c) Precision and recall F1 combines precision and recall in one score.

Q4 (iii). A grayscale pixel value of 0 usually represents: **[1]**

(a) Black

(b) White

(c) Pure red

(d) A missing file

ANSWER: (a) Black Under the usual 8-bit convention, zero is black and 255 is white.

Q4 (iv). A neuron receives x1=9, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 7.5

(b) 8.5

(c) 6.5

(d) 5.5

ANSWER: (c) 6.5 z = 0.5 x 9 + 1.5 x 2 - 1 = 6.5.

Q4 (v). Adjusting brightness without deciding what objects are present is mainly: **[1]**

(a) Sentiment analysis

(b) Image processing

(c) Object understanding by definition

(d) Language translation

ANSWER: (b) Image processing Pixel transformation need not include semantic interpretation.

Q4 (vi). What do weights in an artificial neuron mainly represent? [1]

(a) Only the output label's spelling

(b) The size of a password

(c) The influence of input signals on a weighted sum

(d) The number of paper pages

ANSWER: (c) The influence of input signals on a weighted sum Weights control the contribution of inputs. Q5 - Answer any FIVE of SIX

Q5 (i). In a corpus of 6 documents, the term "sensors" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? [1]

(a) 6

(b) 0.3333

(c) 2

(d) 3

ANSWER: (d) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q5 (ii). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "fans cool warm rooms"? [1]

(a) 4

(b) 6

(c) 3

(d) 5

ANSWER: (a) 4 Tokens are fans, cool, warm, rooms; repeated occurrences count separately.

Q5 (iii). A smart assistant interprets "Could you open the window?" as a request. This uses: [1]

(a) Folder creation

(b) Raw character count only

(c) Pragmatic interpretation

(d) Image thresholding

ANSWER: (c) Pragmatic interpretation Pragmatics considers intended meaning in context.

Q5 (iv). A classifier reports TP=90 and FP=15. What is its precision, rounded to two decimal places? **[1]**

(a) 87.50%

(b) 90.00%

(c) 85.71%

(d) 14.29%

ANSWER: (c) 85.71% Precision = 90/(90+15) x 100 = 85.71%.

Q5 (v). Which NLP stage concerns the grammatical structure of a sentence? **[1]**

(a) Model compression

(b) Colour segmentation

(c) Syntactic analysis

(d) Pixel normalisation

ANSWER: (c) Syntactic analysis Syntax concerns how words are structurally combined.

Q5 (vi). Which neural-network layer receives the feature values first? **[1]**

(a) Output layer

(b) Footer layer

(c) Input layer

(d) Archive layer

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (c) Input layer The input layer receives the representation supplied to the network. Section B | Two-mark model answers

Q6. Give one advantage and one limitation of written communication. [2]

ANSWER: (1) Written communication provides a record that can be checked later. [1] (2) However, immediate clarification may be unavailable, so unclear wording can cause misunderstanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. What does SMART stand for? State why time-bound goals help. [2]

ANSWER: (1) SMART means Specific, Measurable, Achievable, Realistic and Time-bound. [1] (2) A deadline helps a learner plan tasks and check whether progress is timely. [1]

Suggested marking: Complete SMART expansion: 1 mark; a relevant explanation of the deadline: 1 mark.

Q8. Distinguish copying a file from moving a file. [2]

ANSWER: (1) Copying creates another instance while retaining the file at its original location. [1] (2) Moving transfers the file to another location and removes the original location entry. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. State two management functions of an entrepreneur. [2]

ANSWER: (1) An entrepreneur coordinates resources and people to deliver the product or service. [1] (2) The entrepreneur also makes decisions and monitors costs, quality and customer requirements. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. State two benefits of repairing a usable product. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Repair can extend the product's useful life and delay disposal. [1] (2) It can also reduce demand for replacement materials and the resources needed to manufacture a new product. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Distinguish a sector-based ethical framework from a value-based framework. [2]

ANSWER: (1) A sector-based framework addresses needs within a field, such as healthcare. [1] (2) A value-based framework examines principles such as rights, beneficial outcomes or virtuous conduct. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. State one difference between a feature and a target. [2]

ANSWER: (1) A feature is an input attribute, such as parcel weight. [1] (2) The target is the output to predict, such as the parcel's delivery time. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Explain why an ANN may need an activation function. [2]

ANSWER: (1) An activation function transforms a neuron's combined input before passing the signal onward. [1] (2) Nonlinear activations allow networks to represent relationships that are not merely linear. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. State two practices that make an evaluation report transparent. [2]

ANSWER: (1) Identify the test dataset, metrics and conditions under which results were obtained. [1] (2) Report relevant limitations and error patterns rather than showing only the most favourable score. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Define object localisation and explain how it differs from classification. [2]

ANSWER: (1) Localisation identifies where an object is, often through coordinates or a bounding box. [1] (2) Classification identifies its category without necessarily supplying its position. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain discourse integration with one example. Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks) [2]

ANSWER: (1) Discourse integration connects meaning across sentences rather than treating each independently. [1] (2) For example, it identifies whom "she" refers to from the preceding sentence. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A transport application proposes retaining every journey record forever for unspecified future use. [4] (a) Explain one purpose-limitation concern. [1] (b) Suggest a minimisation measure. [1] (c) Suggest a retention measure. [1] (d) Suggest an access-control measure. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Unspecified future use prevents users from understanding why journey records are collected and whether collection is necessary. [1]	1
(b)	Limit collection to the fields required for a clearly described function; avoid unnecessary identifiable journey detail. [1]	1
(c)	Set a justified retention period, apply it consistently and delete information when its documented purpose ends. [1]	1
(d)	Restrict access to authorised roles, protect stored records and maintain responsibility for investigating any unauthorised disclosure. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A clustering tool groups neighbourhoods by resource use without being given correct group names. [4] (a) Identify the learning category. [1] (b) Explain what a cluster represents. [1] (c) Give two possible features. [1] (d) Explain why human interpretation is still needed. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The learning category is unsupervised because no correct group names are supplied for the neighbourhoods. [1]	1
(b)	Each cluster contains examples that the chosen representation and method judge similar, not necessarily examples with identical causes. [1]	1
(c)	Water use and electricity consumption are two possible numerical features, measured consistently for fair comparison. [1]	1
(d)	Human interpretation is necessary because a cluster alone does not explain behaviour or justify policy decisions affecting its members. [1]	1

Suggested marking: Each subpart is 1 mark. In the features subpart, award 0.5 for each of two relevant features.

Q19. An agent chooses actions in a game rather than predicting a supplied correct label for each situation. [4] (a) Name the learning category. [1] (b) Identify what an action is. [1] (c) Identify what reward information provides. [1] (d) Distinguish it from supervised learning. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The agent uses reinforcement learning, improving a strategy through actions and feedback from the game environment. [1]	1

Subpart	Model answer	Marks
(b)	An action is a choice available to the agent, such as moving, waiting or selecting a route. [1]	1
(c)	Reward indicates how desirable an outcome is and helps compare strategies in terms of cumulative returns. [1]	1
(d)	Supervised learning instead uses examples with supplied target outputs; it does not rely on the same action-environment feedback structure. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. For a positive-review classification task, define the condition being detected as positive. Out of 200 examples, 90 actual positives were predicted positive; 10 actual positives were missed; 15 actual negatives were predicted positive; and 85 actual negatives were predicted negative. [4] (a) Construct and label the confusion matrix. [2] (b) Calculate accuracy with working. [1] (c) Find the number of wrong predictions with working. **[1]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	See given computation	1
Precision	See given computation	1
Recall	See given computation	1
F1	See given computation	1

Suggested marking: 1 mark per metric.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "sensors track warm rooms" Document 2: "sensors track cool rooms" Document 3: "fans cool warm rooms" (a) Find the raw term frequency of each word in Document 1. [1] (b) Construct the document-frequency (DF) table for the corpus. [1] (c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1] (d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. **[1]**

ANSWER: Term frequency, document frequency and IDF ratio

Vocabulary term	TF in D1	DF	IDF ratio (3/DF)
sensors	1	2	1.5
track	1	2	1.5
warm	1	2	1.5
rooms	1	3	1
cool	0	2	1.5
fans	0	1	3

(a) Raw TF counts are in column 2. (b) DF counts the documents containing each word. (c) The IDF ratio is N/DF with N=3. (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)).

(c) (c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.

SAMPLE PAPER 09 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). During a discussion, which action best demonstrates active listening? [1]

(a) Prepare a reply without attending to the speaker.

(b) Ask a relevant clarifying question after the speaker finishes.

(c) Interrupt at every pause.

(d) Change the topic immediately.

ANSWER: (b) Ask a relevant clarifying question after the speaker finishes. Clarification after attentive listening checks understanding.

Q1 (ii). A learner misses a target and reviews the cause before changing the plan. This mainly demonstrates: [1]

(a) Denial

(b) Self-regulation

(c) Dependence

(d) Carelessness

ANSWER: (b) Self-regulation Monitoring and adjusting behaviour supports progress.

Q1 (iii). Which practice directly supports effective time management? [1]

(a) Leave difficult work until the last minute.

(b) Prioritise important tasks and allocate time to them.

(c) Switch constantly between unrelated tasks.

(d) Start every activity without a deadline.

ANSWER: (b) Prioritise important tasks and allocate time to them. Prioritising and scheduling make use of limited time.

Q1 (iv). Which extension conventionally identifies a plain text file? [1]

(a) .mp3

(b) .txt

(c) .exe

(d) .jpg

ANSWER: (b) .txt A plain text file commonly uses the .txt extension.

Q1 (v). Why conduct market research before launching a service? [1]

(a) To understand needs, demand and competition

(b) To remove the need for planning

(c) To ensure every person will buy it

(d) To copy a competitor's protected branding

ANSWER: (a) To understand needs, demand and competition Research informs decisions; it cannot guarantee sales.

Q1 (vi). A community reduces waste, sorts recyclables and composts suitable organic waste. These actions support: [1]

(a) A ban on every form of production

(b) Maximum landfill expansion

(c) Responsible consumption and production

(d) Disposal without segregation

ANSWER: (c) Responsible consumption and production The actions reduce waste and recover useful material. Q2 - Answer any FIVE of SIX

Q2 (i). Which data is semi-structured? [1]

(a) Only a fixed numerical matrix by definition

(b) An unrecorded conversation

(c) An unannotated photograph alone

(d) Records in JSON with keys and nested fields

ANSWER: (d) Records in JSON with keys and nested fields JSON has an organising structure without a single rigid relational table layout.

Q2 (ii). A model has no predicted positives. Its precision formula has: [1]

(a) The same denominator as accuracy

(b) A value of one automatically

(c) No dependence on predictions

(d) A zero denominator

ANSWER: (d) A zero denominator TP + FP is zero; a reporting convention is needed for undefined precision.

Q2 (iii). A retailer discovers that two products often occur in the same purchase. This is an example of: [1]

(a) Association

(b) Manual file naming

(c) Image segmentation

(d) Linear text order

ANSWER: (a) Association Association discovers co-occurrence relationships.

Q2 (iv). If actual positives are rare and missing them is costly, which metric is especially important? [1]

(a) Recall for the positive class

(b) The vocabulary length

(c) The model's file extension

(d) Only overall accuracy

ANSWER: (a) Recall for the positive class Recall shows how many actual positives are detected.

Q2 (v). Which operation risks data leakage? [1]

(a) Reserving unseen examples

(b) Documenting the positive class

(c) Using the final test results repeatedly to choose the model

(d) Reporting all confusion-matrix counts

ANSWER: (c) Using the final test results repeatedly to choose the model Repeated selection using the final test information contaminates evaluation.

Q2 (vi). Avoiding preventable harm from an AI recommendation is the principle of: [1]

(a) Non-maleficence

(b) Profit maximisation alone

(c) External motivation

(d) Dimensionality reduction

ANSWER: (a) Non-maleficence Non-maleficence means avoiding harm. Q3 - Answer any FIVE of SIX

Q3 (i). A neural network architecture particularly suited to extracting local visual patterns is a: [1]

(a) Folder tree

(b) Spell-check dictionary

(c) Convolutional Neural Network

(d) Password manager

ANSWER: (c) Convolutional Neural Network CNNs learn patterns through local filters in image-like inputs.

Q3 (ii). Which information is needed before comparing two Bag-of-Words vectors? [1]

(a) The same ordered vocabulary for both

(b) A different vocabulary for each vector

(c) Only the number of pages

(d) The colour of the text

ANSWER: (a) The same ordered vocabulary for both Vector positions must refer to the same terms.

Q3 (iii). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "careful checks prevent faults"? [1]

(a) 4

(b) 3

(c) 6

(d) 5

ANSWER: (a) 4 Tokens are careful, checks, prevent, faults; repeated occurrences count separately.

Q3 (iv). An ethical approach asks whether a decision respects fundamental rights. It is mainly: [1]

(a) Duty-based (deontological)

(b) Frequency-based

(c) Compression-based

(d) Pixel-based

ANSWER: (a) Duty-based (deontological) Rights-based reasoning prioritises rights and dignity.

Q3 (v). A system draws a separate bounding box around each visible bus. The task is: **[1]**

(a) Text generation

(b) Only image storage

(c) Object detection

(d) Sentiment analysis

ANSWER: (c) Object detection Detection identifies and locates objects.

Q3 (vi). A framework compares the likely benefits and harms of an AI action. Its emphasis is: **[1]**

(a) Pixel depth

(b) Consequences or utility

(c) File type

(d) Alphabetical order

ANSWER: (b) Consequences or utility Consequence-based reasoning considers the effects of decisions. Q4 - Answer any FIVE of SIX

Q4 (i). A system assigns a class label to each pixel but does not distinguish two cars of the same class. This is: **[1]**

(a) Speech synthesis

(b) File compression

(c) Semantic segmentation

(d) Instance segmentation

ANSWER: (c) Semantic segmentation Semantic segmentation labels pixel classes without separate instances.

Q4 (ii). A system assigns the label "bicycle" to an entire photograph. The task is: **[1]**

(a) Instance segmentation

(b) Speech recognition

(c) Image classification

(d) Document translation

ANSWER: (c) Image classification A category is assigned to the whole image.

Q4 (iii). A classifier reports TP=56 and FP=14. What is its precision, rounded to two decimal places? **[1]**

(a) 80.00%

(b) 70.00%

(c) 81.00%

(d) 20.00%

ANSWER: (a) 80.00% Precision = 56/(56+14) x 100 = 80.00%.

Q4 (iv). Which pair counts all incorrect binary classifications? **[1]**

(a) TP + TN

(b) TP + FP

(c) FP + FN

(d) TN + FN

ANSWER: (c) FP + FN False positives and false negatives are the errors.

Q4 (v). A team returns to data collection after evaluation reveals missing examples. This shows that the AI project cycle is: **[1]**

(a) Independent of data

(b) Completed before evaluation

(c) Always one-way

(d) Iterative

ANSWER: (d) Iterative Evaluation can reveal a need to revisit earlier stages.

Q4 (vi). Use relative error = |actual - predicted|/actual. A numerical prediction is 792 when the actual value is 900. What is the relative error? **[1]**

(a) 0.88

(b) 0.012

(c) 1.12

(d) 0.12

ANSWER: (d) 0.12 |900-792|/900=0.12. This is relative numerical error, not a confusion-matrix count. Q5 - Answer any FIVE of SIX

Q5 (i). Document frequency of a term means: **[1]**

(a) Number of documents containing it

(b) Only its count in the first document

(c) Its total count across all documents

(d) Number of letters in it

ANSWER: (a) Number of documents containing it Each containing document contributes once to DF.

Q5 (ii). A word appears in all N documents. Under the school convention, its IDF ratio N/DF is: **[1]**

(a) Undefined

(b) 1

(c) 0

(d) N squared

ANSWER: (b) 1 Here DF equals N, so the ratio is one.

Q5 (iii). In a corpus of 6 documents, the term "careful" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 6

(b) 0.3333

(c) 2

(d) 3

ANSWER: (d) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q5 (iv). A neuron receives x1=10, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 8

(b) 6

(c) 7

(d) 9

ANSWER: (c) 7 z = 0.5 x 10 + 1.5 x 2 - 1 = 7.

Q5 (v). Which domain is most directly used when a system interprets spoken or written human language? [1]

(a) Disk formatting

(b) Natural Language Processing

(c) Computer Vision only

(d) Hardware cleaning

ANSWER: (b) Natural Language Processing NLP concerns computational processing of human language.

Q5 (vi). An image is 280 pixels wide and 160 pixels high. How many pixels does it contain? [1]

(a) 880

(b) 44800

(c) 440

(d) 134400

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (b) 44800 Pixel count = width x height = 280 x 160 = 44800; RGB channels do not triple the number of spatial pixels. Section B | Two-mark model answers

Q6. Distinguish a complete sentence from a sentence fragment. [2]

ANSWER: (1) A complete sentence expresses a complete thought with an appropriate subject and predicate. [1] (2) A fragment, such as "because the printer stopped", leaves the thought unfinished. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. Explain one difference between self-awareness and self-confidence. [2]

ANSWER: (1) Self-awareness is understanding personal strengths, limitations and emotions. [1] (2) Self-confidence is trust in the ability to act or learn, while still recognising areas for improvement. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. State two precautions when installing software. [2]

ANSWER: (1) Obtain software from a trusted source and verify that it is appropriate. [1] (2) Review requested permissions and keep security protection active during installation and later use. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Explain why entrepreneurship does not guarantee a profit. [2]

ANSWER: (1) Sales, costs and customer preferences are uncertain and may change. [1] (2) An entrepreneur can reduce risk through planning and testing, but cannot ensure every venture earns profit. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Define renewable energy and give one example. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Renewable energy comes from sources replenished naturally on useful human timescales. [1] (2) For example, solar panels convert incoming sunlight into electricity for homes or other buildings. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. What is deployment? Why does monitoring continue afterwards? [2]

ANSWER: (1) Deployment makes a tested AI solution available in its intended setting. [1] (2) Monitoring checks changing conditions and unexpected errors so that the solution can be improved. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. State two reasons why a rule-based system may be suitable for a small task. [2]

ANSWER: (1) Explicit rules can implement a small set of clear, stable requirements. [1] (2) Their decisions can also be inspected directly without collecting a large labelled training dataset. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Explain why incorrect training labels can reduce model quality. [2]

ANSWER: (1) Labels tell a supervised model which outputs its examples should produce. [1] (2) Incorrect labels supply misleading learning signals and can produce unreliable decisions on new examples. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. What does F1 measure, and when can it be useful? [2]

ANSWER: (1) F1 is the harmonic mean of precision and recall. [1] (2) It is useful when both false alarms and missed positives matter and neither precision nor recall should be ignored. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. Explain optical character recognition with a use case. [2]

ANSWER: (1) Optical character recognition identifies text characters in an image and converts them into machine-readable text. [1] (2) It can extract printed invoice details for later processing. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain pragmatic analysis with one example. [2] Answer any THREE of Q17-Q21. For prose, aim for 50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) Pragmatic analysis interprets intended meaning using the situation and purpose. [1] (2) For example, "Could you close the door?" is understood as a request, not merely an ability question. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A developer says that high overall accuracy proves a model is fair to every group. [4] (a) Explain why the claim is insufficient. [1] (b) Name one further check. [1] (c) Suggest one correction if a disparity is found. [1] (d) State one human safeguard. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Overall accuracy averages performance and may conceal poor results for a smaller, differently represented user group. [1]	1
(b)	Compare relevant subgroup errors, including false-negative rates where missed positives can produce significant harm for users. [1]	1
(c)	Investigate data representation and design choices, make appropriate corrections and repeat the group-wise evaluation after changes. [1]	1
(d)	Keep human review and an accessible appeal process for important decisions instead of treating every prediction as final. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A camera application must recognise handwritten digits from photographs. [4] (a) Identify a suitable supervised subcategory. [1] (b) Describe appropriate labelled training examples. [1] (c) Name a suitable neural architecture. [1] (d) State one data-quality issue. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Classification is suitable because each input image must be assigned one of the digit categories. [1]	1
(b)	Training examples should pair varied digit photographs with accurate labels indicating the digit actually present. [1]	1
(c)	A CNN is a suitable neural architecture because learned local filters can capture relevant visual patterns. [1]	1
(d)	Blurred images, unrepresentative handwriting or incorrect labels can weaken training and evaluation, so examples require careful quality checks. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A technician wants to explain the path from a feature vector to an ANN prediction. [4] (a) State what reaches the input layer. [1] (b) State what hidden units compute. [1] (c) State what training changes. [1] (d) State what the output represents. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The input layer receives the encoded feature vector; every position should have a consistent meaning across examples. [1]	1
(b)	Hidden units combine weighted inputs and bias, then apply activation functions to form useful intermediate representations. [1]	1
(c)	During training, the learning procedure adjusts weights and biases using feedback about the required task. [1]	1

Subpart	Model answer	Marks
(d)	The output provides task-specific predictions or scores; it must be evaluated and is not a guarantee that the result is correct. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. A wildlife-presence detection model has TP=56, FP=14, FN=24 and TN=106. Calculate accuracy, precision, recall and F1 score. Show formula and substitution for each. Round percentages to two decimal places. [4]

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	(TP + TN)/N x 100 = (56 + 106)/200 x 100 = 81.00%. [1]	1
Precision	TP/(TP + FP) x 100 = 56/(56 + 14) x 100 = 80.00%. [1]	1
Recall	TP/(TP + FN) x 100 = 56/(56 + 24) x 100 = 70.00%. [1]	1
F1	2TP/(2TP + FP + FN) x 100 = 112/(112 + 14 + 24) x 100 = 74.67%. [1]	1

Suggested marking: each metric earns 0.5 for the correct formula with relevant substitution and 0.5 for the result. Equivalent correct fraction/decimal or percentage is acceptable when labelled. Do not round P and R before finding F1.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "careful drivers check brakes" Document 2: "drivers check lights" Document 3: "careful checks prevent faults" (a) List the vocabulary in first-appearance order. [1] (b) Construct the count vector for Document 3 in that order. [1] (c) Explain how the vector is useful to a text classifier. [1] (d) State one order-related limitation of this representation. **[1]**

ANSWER: Fixed-vocabulary Bag-of-Words count table

Vocabulary term	Document 1	Document 2	Document 3
careful	1	0	1
drivers	1	1	0
check	1	1	0
brakes	1	0	0
lights	0	1	0
checks	0	0	1
prevent	0	0	1
faults	0	0	1

(a)/(b) (a) Vocabulary: [careful, drivers, check, brakes, lights, checks, prevent, faults]. [1] (b) Document 3 vector: [1, 0, 0, 0, 0, 1, 1, 1]. [1]

(c) (c) The fixed-length numerical vector supplies word-count features that a text model can use to learn a classification pattern. [1]

(d) (d) Standard BoW discards the original word order, so different sentences with the same word counts can receive the same representation. [1]

Suggested marking: vocabulary 1, requested document vector 1, feature-extraction explanation 1 and the word-order limitation 1. In these exercises the specified order must be followed.

SAMPLE PAPER 10 - SOLVED EDITION

Class X AI (417) | Theory | 50 marks | 2 hours. Answers follow each question.

Q1 (i). A sender converts an idea into spoken words. This is: [1]

(a) Feedback

(b) Archiving

(c) Decoding

(d) Encoding

ANSWER: (d) Encoding Encoding creates the message; decoding interprets it.

Q1 (ii). Which practice directly supports effective time management? [1]

(a) Prioritise important tasks and allocate time to them.

(b) Leave difficult work until the last minute.

(c) Switch constantly between unrelated tasks.

(d) Start every activity without a deadline.

ANSWER: (a) Prioritise important tasks and allocate time to them. Prioritising and scheduling make use of limited time.

Q1 (iii). Riya studies robotics because solving problems interests her, even without prizes. This is: [1]

(a) Financial reward

(b) External punishment

(c) Intrinsic motivation

(d) Peer pressure

ANSWER: (c) Intrinsic motivation Her motivation comes from interest in the activity itself.

Q1 (iv). Which maintenance action is appropriate before physically cleaning a desktop computer? [1]

(a) Shut it down and disconnect power safely.

(b) Open all unknown software first.

(c) Pour water onto the keyboard.

(d) Keep it running while removing components.

ANSWER: (a) Shut it down and disconnect power safely. Power should be disconnected before physical cleaning.

Q1 (v). A failed pilot leads an entrepreneur to improve the design and test again. This shows: [1]

(a) Persistence with learning

(b) Refusal to use evidence

(c) Complete risk avoidance

(d) A guarantee of future profit

ANSWER: (a) Persistence with learning Persistence is most useful when accompanied by learning.

Q1 (vi). Which decision considers both environmental and economic effects? [1]

(a) Compare operating cost, durability and resource use before purchase.

(b) Choose only by colour.

(c) Ignore repairability.

(d) Treat waste disposal as irrelevant.

ANSWER: (a) Compare operating cost, durability and resource use before purchase. Sustainable choices consider several consequences. Q2 - Answer any FIVE of SIX

Q2 (i). Which report is most transparent? [1]

(a) Present training accuracy as a guarantee.

(b) Show only the most favourable number.

(c) Hide errors in small groups.

(d) State the dataset, metrics, limitations and group-wise results.

ANSWER: (d) State the dataset, metrics, limitations and group-wise results. Meaningful evaluation includes context and limitations.

Q2 (ii). After defining its objective, a team obtains relevant observations from reliable sources. This is: [1]

(a) Image segmentation

(b) Data acquisition

(c) Reward calculation

(d) Model evaluation

ANSWER: (b) Data acquisition Acquisition obtains the data needed for the task.

Q2 (iii). A term occurs three times in one document. Using raw-count TF, its term frequency there is: **[1]**

(a) 3

(b) 1

(c) The number of documents only

(d) 0

ANSWER: (a) 3 Raw-count TF records repeated occurrences in that document.

Q2 (iv). An ethical approach asks whether a decision respects fundamental rights. It is mainly: **[1]**

(a) Compression-based

(b) Duty-based (deontological)

(c) Frequency-based

(d) Pixel-based

ANSWER: (b) Duty-based (deontological) Rights-based reasoning prioritises rights and dignity.

Q2 (v). Which task predicts one of a set of named categories? **[1]**

(a) Regression of a continuous value

(b) Frequency normalisation only

(c) File compression

(d) Classification

ANSWER: (d) Classification Classification selects a category.

Q2 (vi). A model has no predicted positives. Its precision formula has: **[1]**

(a) A zero denominator

(b) The same denominator as accuracy

(c) A value of one automatically

(d) No dependence on predictions

ANSWER: (a) A zero denominator TP + FP is zero; a reporting convention is needed for undefined precision. Q3 - Answer any FIVE of SIX

Q3 (i). A team returns to data collection after evaluation reveals missing examples. This shows that the AI project cycle is: **[1]**

(a) Always one-way

(b) Iterative

(c) Independent of data

(d) Completed before evaluation

ANSWER: (b) Iterative Evaluation can reveal a need to revisit earlier stages.

Q3 (ii). Use relative error = |actual - predicted|/actual. A numerical prediction is 836 when the actual value is 950. What is the relative error? **[1]**

(a) 0.12

(b) 0.88

(c) 0.012

(d) 1.12

ANSWER: (a) 0.12 |950-836|/950=0.12. This is relative numerical error, not a confusion-matrix count.

Q3 (iii). A classifier reports TP=42 and FP=18. What is its precision, rounded to two decimal places? **[1]**

(a) 70.00%

(b) 30.00%

(c) 87.00%

(d) 84.00%

ANSWER: (a) 70.00% Precision = 42/(42+18) x 100 = 70.00%.

Q3 (iv). A model flags a sensor reading far outside typical operating behaviour. The task is: **[1]**

(a) Speech synthesis

(b) Anomaly detection

(c) Alphabetical sorting

(d) Document translation

ANSWER: (b) Anomaly detection Anomaly detection looks for unusual observations.

Q3 (v). Which task selects the important terms that represent a document's subject? [1]

(a) Instance segmentation

(b) Keyword extraction

(c) Pixel interpolation

(d) Screen locking

ANSWER: (b) Keyword extraction Keyword extraction identifies useful content terms.

Q3 (vi). Why can a basic Bag-of-Words representation confuse opposite statements? [1]

(a) It cannot store zeroes.

(b) It discards word order and much grammatical structure.

(c) It stores only images.

(d) It always removes every negative word.

ANSWER: (b) It discards word order and much grammatical structure. Order-sensitive meaning can be lost even if all words are retained. Q4 - Answer any FIVE of SIX

Q4 (i). An 80:20 split of 250 examples gives how many test examples? [1]

(a) 80

(b) 20

(c) 50

(d) 200

ANSWER: (c) 50 The test share is 20 percent of 250.

Q4 (ii). Use whitespace word tokenisation, with punctuation excluded. How many tokens are in "artists paint paint signs"? [1]

(a) 4

(b) 3

(c) 5

(d) 6

ANSWER: (a) 4 Tokens are artists, paint, paint, signs; repeated occurrences count separately.

Q4 (iii). A neuron receives x1=11, x2=2, weights w1=0.5, w2=1.5 and bias b=-1. Find the weighted sum z = w1*x1 + w2*x2 + b before activation. **[1]**

(a) 9.5

(b) 6.5

(c) 7.5

(d) 8.5

ANSWER: (c) 7.5 z = 0.5 x 11 + 1.5 x 2 - 1 = 7.5.

Q4 (iv). In a corpus of 6 documents, the term "artists" occurs in 2 different documents. Under the convention IDF ratio = N/DF, what is the ratio? **[1]**

(a) 0.3333

(b) 2

(c) 3

(d) 6

ANSWER: (c) 3 IDF ratio = 6/2 = 3. Take log10 of this ratio only when calculating TF-IDF.

Q4 (v). In a dataset of houses, floor area is used to predict rent. Floor area is a: **[1]**

(a) Test score

(b) Predicted target value

(c) Feature

(d) Final decision rule by itself

ANSWER: (c) Feature A feature is an input attribute used by the model.

Q4 (vi). An 8-bit grayscale channel has how many possible values? **[1]**

(a) 256

(b) 8

(c) 16

(d) 255

ANSWER: (a) 256 Eight bits represent 2 to the power 8 values, from 0 to 255.
Q5 - Answer any FIVE of SIX

Q5 (i). A decision-maker values honesty, care and integrity when developing AI. This reflects: **[1]**

(a) Random selection

(b) Virtue-based reasoning

(c) Term frequency

(d) Object detection

ANSWER: (b) Virtue-based reasoning Virtue-based reasoning focuses on responsible character and conduct.

Q5 (ii). With "defective" defined as positive, a faulty item is predicted to be good. The outcome is: **[1]**

(a) False negative

(b) True negative

(c) False positive

(d) True positive

ANSWER: (a) False negative The actual positive was missed.

Q5 (iii). Which task gives different pixel masks to two separate people in one photograph? **[1]**

(a) Whole-image classification

(b) Tokenisation

(c) Colour inversion only

(d) Instance segmentation

ANSWER: (d) Instance segmentation Instance segmentation separates individual objects at pixel level.

Q5 (iv). A team identifies affected users and the location of a recurring problem before collecting data. The stage is: **[1]**

(a) Tokenisation

(b) Model deployment

(c) Problem scoping

(d) Classification

ANSWER: (c) Problem scoping Problem scoping defines the problem and its context.

Q5 (v). An image is 300 pixels wide and 170 pixels high. How many pixels does it contain? **[1]**

(a) 153000

(b) 470

(c) 51000

(d) 940

ANSWER: (c) 51000 Pixel count = width x height = 300 x 170 = 51000; RGB channels do not triple the number of spatial pixels.

Q5 (vi). A digital image's smallest addressable picture element is a: **[1]**

(a) Paragraph

(b) Pixel

(c) Token

(d) Record label

SECTION B | SUBJECTIVE QUESTIONS | 26 MARKS

Employability Skills: answer any THREE of Q6-Q10 in 20-30

words each. (2 x 3 = 6 marks)

ANSWER: (b) Pixel A pixel stores image information at a spatial position. Section B | Two-mark model answers

Q6. Explain how cultural differences can create a communication barrier. **[2]**

ANSWER: (1) A gesture or expression can have different meanings across cultures. [1] (2) Using respectful language and checking understanding helps

prevent assumptions from turning into misunderstanding. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q7. State two responsible ways of responding to a difficult task. [2]

ANSWER: (1) Break the task into manageable steps and begin with a realistic plan. [1] (2) Seek specific guidance when necessary, then take responsibility for completing and checking the work. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q8. Explain the use of a file extension with an example. [2]

ANSWER: (1) A file extension helps identify the file type and suitable applications. [1] (2) For example, .txt commonly identifies a plain-text file that a text editor can open. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q9. Explain how market research supports a business decision. [2]

ANSWER: (1) Market research gathers evidence about customers, alternatives and unmet needs. [1] (2) It helps a business assess demand before committing resources to a proposed product or service. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q10. Explain why sustainability involves more than environmental protection. [2] AI Subject Skills: answer any FOUR of Q11-Q16 in 20-30 words each. (2 x 4 = 8 marks)

ANSWER: (1) Sustainability also concerns social wellbeing and economically workable choices. [1] (2) For example, an environmental measure should consider fair access and the livelihoods of affected people. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q11. Explain autonomy and non-maleficence in healthcare AI. [2]

ANSWER: (1) Autonomy respects a patient's informed choices and control over relevant decisions. [1] (2) Non-maleficence requires avoiding preventable harm when developing or using a healthcare AI system. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q12. Define a CNN and identify a suitable kind of input. [2]

ANSWER: (1) A Convolutional Neural Network uses learned filters to extract local patterns. [1] (2) Images are suitable inputs because neighbouring pixels often contain related visual information. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q13. Explain why a learned association does not prove causation. [2]

ANSWER: (1) An association reports that items or events occur together in the observed data. [1] (2) Other factors may explain their co-occurrence, so causation requires additional evidence. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q14. Explain one reason for checking performance across user groups. [2]

ANSWER: (1) A high overall score can hide poor performance for a smaller group. [1] (2) Group-wise evaluation helps reveal unequal errors that may require better data or design. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q15. State two data precautions for a face-recognition service. [2]

ANSWER: (1) Collect only necessary face data for a clear, justified purpose. [1] (2) Protect access and retention, and evaluate errors across representative conditions and intended user groups. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording.

Q16. Explain why a common word can receive a low TF-IDF weight. [2] Answer any THREE of Q17-Q21. For prose, aim for

50-80 words; include required tables and numerical working. (4 x 3 = 12 marks)

ANSWER: (1) A word appearing in many documents has a high document frequency. [1] (2) Its inverse-frequency contribution is therefore small, reducing its ability to distinguish one document from others. [1]

Suggested marking: 1 mark for each required, correct element; accept equivalent accurate wording. Four-mark model answers

Q17. A community app could reduce waiting time but excludes people who cannot read its only supported language. [4] (a) Identify an affected stakeholder group. [1] (b) Explain the access concern. [1] (c) Suggest an inclusive design change. [1] (d) Explain how to evaluate the change. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	People unable to read the application's supported language are affected stakeholders even if they rarely use it. [1]	1
(b)	A single-language interface can deny equal access to a public benefit and hide unmet needs in usage statistics. [1]	1
(c)	Offer understandable local-language interfaces or assisted alternatives appropriate to the community rather than assuming everyone reads one language. [1]	1
(d)	Test completion rates, errors and comprehension with representative users of each alternative, and improve the design using their feedback. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q18. A business finds an association between buying a toolkit and buying protective gloves. [4] (a) Identify the technique. [1] (b) State what the association shows. [1] (c) Explain why it does not establish causation. [1] (d) Suggest one sensible business use. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	The technique is association learning, which examines recurring relationships among items in recorded transactions. [1]	1
(b)	The discovered pattern means toolkits and protective gloves occur together relatively often in the observed purchases. [1]	1
(c)	This does not prove one purchase causes the other; a shared activity or another factor may explain both. [1]	1
(d)	Use the finding to test a relevant recommendation or inventory adjustment, then evaluate whether it genuinely helps customers. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q19. A team is deciding whether every intelligent application must use deep learning. [4] (a) State the relation of AI to ML. [1] (b) State the relation of ML to DL. [1] (c) Give one non-learning AI approach. [1] (d) Explain why task suitability matters. **[1]**

ANSWER:

Subpart	Model answer	Marks
(a)	Artificial intelligence is the broader field; machine learning is one approach within it that learns from data. [1]	1
(b)	Deep learning is a subset of machine learning using neural networks with multiple learned processing layers. [1]	1
(c)	Explicit rule-based reasoning is an AI approach that need not learn its decision rules from examples. [1]	1

Subpart	Model answer	Marks
(d)	Choose methods for the task and available evidence; a complex model is not automatically better than a clear, suitable simpler approach. [1]	1

Suggested marking: 1 mark for each correctly answered subpart; total 4. Equivalent relevant explanations are acceptable.

Q20. For a recyclable-item sorting task, define the condition being detected as positive. Out of 200 examples, 42 actual positives were predicted positive; 8 actual positives were missed; 18 actual negatives were predicted positive; and 132 actual negatives were predicted negative. [4] (a) Construct and label the confusion matrix. [2] (b) Calculate accuracy with working. [1] (c) Find the number of wrong predictions with working. **[1]**

ANSWER: Evaluation metrics

Metric	Formula / substitution and result	Marks
Accuracy	See given computation	1
Precision	See given computation	1
Recall	See given computation	1
F1	See given computation	1

Suggested marking: 1 mark per metric.

Q21. Use the following three-document corpus. Lowercase the text. Retain every word; do not remove stop words or apply stemming. Use the first-appearance vocabulary order throughout. [4] Document 1: "artists mix blue paint" Document 2: "blue paint covers walls" Document 3: "artists paint paint signs" (a) Find the raw term frequency of each word in Document 1. [1] (b) Construct the document-frequency (DF) table for the corpus. [1] (c) Construct the IDF-ratio table using N/DF, where N is the number of documents. [1] (d) Write TF-IDF(w,d) in terms of TF and the base-10 logarithm of the IDF ratio. **[1]**

ANSWER: Term frequency, document frequency and IDF ratio

Vocabulary term	TF in D1	DF	IDF ratio (3/DF)
artists	1	2	1.5
mix	1	1	3
blue	1	2	1.5
paint	1	3	1
covers	0	1	3
walls	0	1	3
signs	0	1	3

(a) Raw TF counts are in column 2. (b) DF counts the documents containing each word. (c) The IDF ratio is N/DF with N=3. (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)).

(c) (c) The last column gives the inverse-document-frequency ratio N/DF for every term. [1]

(d) (d) TF-IDF(w,d) = TF(w,d) x log10(N/DF(w)). Here N/DF is the IDF ratio; the logarithm is applied once. [1]

Suggested marking: one mark for each requested table/column or formula. Partial table credit, where justified, must remain within that part's one-mark maximum. This is not a published official 2026-27 marking scheme.