CST1
COMPUTER SCIENCE TRIPOS Part IB

Thursday 12 June 2025 13:30 to 16:30

COMPUTER SCIENCE Paper 7

*Answer **five** questions.*

*Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.*

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Artificial Intelligence

(*a*) In the context of algorithms for search in AI, what is a *heuristic* and how is it used? Give a definition of what it means for a heuristic to be *admissible*. [3 marks]

(*b*) You have a collection $h_1, h_2, \ldots, h_n$ of heuristics, each of which is admissible. Prove that the heuristic

$$h(s) = \sum_{i=1}^{n} a_i h_i(s) \text{ where } a_i > 0 \text{ and } \sum_{i=1}^{n} a_i = 1$$

is also admissible. [2 marks]

Your aim is to use supervised learning to attempt to infer a general admissible heuristic for a particular search problem. States are denoted as vectors, and you have managed to gather information on a collection $(s_1, s_2, \ldots, s_m)$ of m states, and the corresponding values $(h'(s_1), h'(s_2), \ldots, h'(s_m))$ for the *exact* distances between states and goals. Your friend has some software called VectorDribble and suggests that you use its built-in function for linear regression to achieve this.

(*c*) How would you address this as a supervised learning problem? Is linear regression likely to be a good method? Explain your answer. [3 marks]

(*d*) You find that VectorDribble's built-in function uses the standard measure of error

$$E(\mathbf{w}) = \sum_{i=1}^{m} \left(h'(\mathbf{s}_i) - f(\mathbf{s}_i, \mathbf{w})\right)^2$$

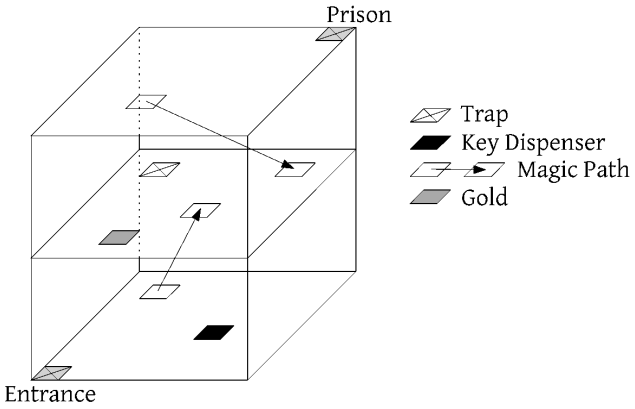
where $\mathbf{w}$ is the learned vector of weights and $f(\mathbf{s}_i, \mathbf{w})$ is the output of the learned function. Your friend claims that this is the error you must use as it is built into the library. Explain why this function should not be used. [2 marks]

(*e*) Suggest a better function to use in place of the built-in $E(\mathbf{w})$. Explain how it improves on the built-in version. [4 marks]

(*f*) As you now can not use VectorDribble to solve the problem, derive a gradient descent-based learning algorithm that is suitable. You may continue to assume that the underlying linearity is appropriate. [6 marks]

2 Artificial Intelligence

This question addresses the *state-variable representation* for *planning problems*. Evil Robot has found himself in a new cave. This cave is 3-dimensional, with locations on a $10 \times 10 \times 10$ grid. An example cave is:



Arbitrary locations can be connected by Magic Paths transporting the user instantly from one place to another, and these are the only means of vertical travel. Magic Paths can only be traversed in one direction. Some locations contain traps that transport the victim to a prison at location $(10, 10, 10)$.

Escape from the prison back to the entance at (1, 1, 1) is possible if a key has been obtained. Some locations contain key dispencers. One location contains some gold, and Evil Robot's aim is to find it.

(a) Describe the elements of the state-variable representation. [5 marks]

(b) Give examples of how the following three elements of Evil Robot's cave problem can be represented in the state-variable representation: the locations of the traps and key dispensers, obtaining a key, and escape from the prison. [5 marks]

(c) Movement is achieved using actions `goLeft`, `goRight`, `goForward` and `goBack`. We wish to use a single action description for each. However the cave is surrounded by walls and not all locations have the same effect when moved into. Explain how *rigid relations* can be used to obtain precisely four action descriptions with the desired behaviour. Make sure you account for the fact that it should not be possible to escape from the prison just by moving. [5 marks]

(d) One of the advantages of the state-variable representation is that planning problems can easily be translated into a form solvable by a *constraint satisfaction solver*. Briefly describe how your `goRight` action would be modelled as part of a constraint satisfaction problem. [5 marks]

3 Economics, Law and Ethics

(*a*) Provide examples to illustrate (i) adverse selection and (ii) moral hazard, defining relevant terms. [5 marks]

(*b*) Ouchbridge hospital handles sensitive medical data from patients. If there is any data breach, Ouchbridge hospital may face large fines. Ouchbridge hospital is considering buying antivirus software. How does adverse selection affect the process of selecting which software to buy? [5 marks]

(*c*) Painbridge hospital has cybersecurity insurance. How does this affect moral hazard? [5 marks]

(*d*) King's Ache & Break hospital has limited financial capacity and must choose to secure either financial data only or medical data only, but not both. A malicious actor will try to get unauthorised access to either financial only or medical data only. Which game does the interaction between the King's Ache & Break hospital and the malicious actor look like? Use relevant game theory terms and discuss what the optimal strategy would be. [5 marks]

4 Economics, Law and Ethics

As AI continues to attract widespread attention, it has become a key legal and policy concern. Drawing on concepts and principles outlined in this course, and using examples where appropriate, respond to the following:

(*a*) There is ongoing debate about the role of law and regulation for digital technologies. Briefly explain different perspectives on this issue and discuss the relationship between the law and technology's development and use. [6 marks]

(*b*) Automated decision-making (ADM) is a purported feature of some AI applications.

(*i*) Discuss two risks that ADM presents. [3 marks]

(*ii*) Explain a regulatory challenge for governing ADM systems. [2 marks]

(*c*) Data used to train models can raise data protection concerns.

(*i*) Explain when the General Data Protection Regulation (GDPR) applies to training data. [1 mark]

(*ii*) Describe two rights given to individuals by the GDPR, discussing how an organisation might implement them in relation to training datasets. [4 marks]

(*d*) The EU AI Act classifies AI systems based on risk, imposing stricter rules on higher-risk uses of AI. Briefly describe two advantages and two concerns of an application-specific, risk-based approach to AI regulation. [4 marks]

5 Formal Models of Language

Consider the following natural language sentences:

- a toast to the queen was raised tonight
- a toast was raised to the queen tonight

(*a*) Provide valid unlabelled dependency parses for both of the sentences. Provide motivation for the dependencies and their direction. [5 marks]

(*b*) Provide a dependency grammar that is consistent with both of your dependency parses. [5 marks]

(*c*) Can you write a context-free grammar that is weakly equivalent to your grammar? Explain your answer. [1 mark]

(*d*) Demonstrate how a modified LR shift-reduce parser could yield your dependencies for the first example sentence, *"a toast to the queen was raised tonight"*. Show the state of the parser, the actions taken and the dependencies recorded. You may finish the demonstration when *"was"* has been shifted onto the stack. [6 marks]

(*e*) Explain why data driven dependency parsing is described as grammarless. [3 marks]

6 Formal Models of Language

(*a*) Provide a definition of a formal language and a natural language. With respect to each, what does it mean for a string to be *grammatical*? [3 marks]

(*b*) With reference to both formal and natural languages, what is meant by *complexity*? [3 marks]

(*c*) Consider the following natural language sentences:

i) A student worked hard succeeds.

ii) A student that is worked hard succeeds.

You are debating with a friend whether sentence i) or ii) is more difficult for a human to process. Which do you think is more difficult and why? [2 marks]

(*d*) You are now considering the following natural language sentences:

iii) The notes the supervisor the students like wrote are useful.

iv) The notes that were written by the supervisor that the students like are useful.

Which sentence, iii) or iv), do you think is more difficult for a human to process and why? [2 marks]

(*e*) Your deliberations on processing language have made you concerned about the language in your upcoming dissertation. You decide to make a tool to: 1) identify sentences that are difficult to process; 2) label the sentences with the reason they are difficult. Describe how your tool works. Explain any theoretical assumptions you are making and describe at least two methods by which you will identify the sentences. [6 marks]

(*f*) Your friend suggests that, rather than building your tool, you could reduce the processing difficulty of your dissertation by writing it in a language that can be defined by a finite-state machine. Do you agree? Explain your answer. [4 marks]

7 Further Graphics

A height-field is defined as a surface in 3D over the xy-plane such that for each x and y coordinate, the z coordinate is given by $h(x, y)$. For a particular height-field, assume the partial derivatives $\frac{\partial h}{\partial x}(0, 0) = \frac{\partial h}{\partial y}(0, 0) = 0$.

(a) Write an implicit function for the height field. [2 marks]

(b) What is the surface normal at $x = 0, y = 0$? [2 marks]

(c) Write an implicit function for the tangent plane at $x = 0, y = 0$. [2 marks]

(d) Recall that normal curvature at a surface point for a direction is defined as the curvature of the curve formed by intersecting the surface with a particular plane. The plane intersects the surface at $x = 0, y = 0$, and is given in the direction $\mathbf{d} = [d_x, d_y, d_z]$, where $\mathbf{d}$ lies on the plane tangent to the surface. Write an explicit equation for the points on the plane used to compute the curvature. [6 marks]

(e) Write an expression for the curve given by the intersection of the above plane with the surface. [4 marks]

(f) For $\mathbf{d} = [1, 0, 0]$, derive the condition on the height-field such that we can use the expression $\frac{\partial^2 \mathbf{c}(t)}{\partial t^2} = \kappa(t)\mathbf{n}(t)$ to derive the curvature $\kappa(t)$ of the curve $\mathbf{c}$ in the previous question. [4 marks]

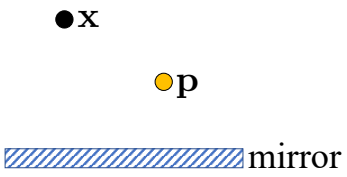
8 Further Graphics

(*a*) (*i*) Show that quaternions that represent rotation around the same axis commute, that is $\mathbf{q}_1\mathbf{q}_2 = \mathbf{q}_2\mathbf{q}_1$. [2 marks]

(*ii*) Derive the general condition on the rotation axes of two quaternions representing rotations in order for them to commute. [3 marks]

(*iii*) Prove that linear blending of quaternions with normalisation satisfies coordinate invariance. (Hints: recall that this means if we first transform all quaternions with the same rotation and then blend the resulting quaternions, we get the same result as first blending them and then applying the rotation, and $(\mathbf{q}_1\mathbf{q}_2)^* = \mathbf{q}_2^*\mathbf{q}_1^*$ for quaternions $\mathbf{q}_1$ and $\mathbf{q}_2$.) [5 marks]

(*b*) You are given a point light source at $\mathbf{p}$ that emits light with constant radiance L in all directions. There is an ideal planar mirror that reflects all light received if the angle between incoming light direction and surface normal is equal to that between reflected light direction and surface normal, and reflects no light otherwise.



(*i*) Write the expression for the BRDF of the mirror, show that it satisfies the perfect mirror property by deriving the reflected light radiance. [4 marks]

(*ii*) How much light is received at $\mathbf{x}$ in the figure? [2 marks]

(*iii*) Assume an imperfect mirror where the BRDF f_r of a perfect mirror is convolved with a Gaussian kernel. Explain the deformation of the image we will see as our reflection in the resulting mirror. [2 marks]

(*iv*) For some mirrors, the image we see in the mirror might have slightly different colours, explain why. [2 marks]

9 Further Human–Computer Interaction

Consider the design of a future integrated software development environment (IDE) that supports three different modes of interaction: (1) direct manipulation via a touch screen, (2) text and command entry via a keyboard, and (3) voice audio interaction.

(*a*) For *each of these three* interaction modes, describe a specific software development task that will be especially appropriate for completion using that mode. [3 marks]

(*b*) The three tasks that you have described, together with the three interaction modes, define a matrix of nine different task-mode combinations, some of which will be appropriate and effective, and some of which will be inappropriate or ineffective. Choose *five* of these possible task-mode combinations, and for *each of the five*, use one of the Cognitive Dimensions of Notations to justify why you expect this combination to be a good or bad design choice. [10 marks]

(*c*) Describe how you would set up a study to test these predictions, considering choice of participants, instructions you would give them, and data you would collect. [2 marks]

(*d*) Sketch an original design for an IDE suitable for interactive devices where use of an alphanumeric keyboard is not practical (for example, a smart watch), so that only modes 1 and 3 are available. Include an explanation of how this would be used for a simple programming task. [5 marks]

10 Further Human–Computer Interaction

Imagine that you have been engaged as an HCI consultant, by the manufacturer of this dual-function remote control car / MP3 player.

To save costs, they propose to stop shipping the remote control, remove the MP3 functionality, and make software modifications so that the six control panel buttons can be used to program a sequence of car movements. To further save costs, only the internal programming will be changed — the labels on the buttons (as in the inset) will not be changed.

(*a*) Describe the general approach that you propose, to repurpose the six button control panel as a simple programming interface. [2 marks]

(*b*) Explain the implications of changing direct manipulation controls into programming commands, from the perspective of the Attention Investment model of abstraction use. [2 marks]

(*c*) Make an initial evaluation of the usability of your programming interface, by applying the Cognitive Dimensions of Notations to the resulting system. Choose three relevant cognitive dimensions, and explain the design issues and/or opportunities arising from each of them. [6 marks]

(*d*) Describe the new interpretations that you propose for each of the symbols on the control panel, and explain why these are appropriate in terms of principles of visual correspondence. Do this for each of the six symbols from left to right. [6 marks]

(*e*) Describe a qualitative and a quantitative approach to summative evaluation of this redesign. Which approach would be most useful? [4 marks]

END OF PAPER