Note: This is an **"open book exam"**: You are allowed to access any material, including the course material in online or printed form. However, it is strictly **forbidden to copy answers** (except for brief, clearly marked quotations), **or to interact with other people** to discuss the exam during the exam.

You can answer in Finnish, Swedish, or English.

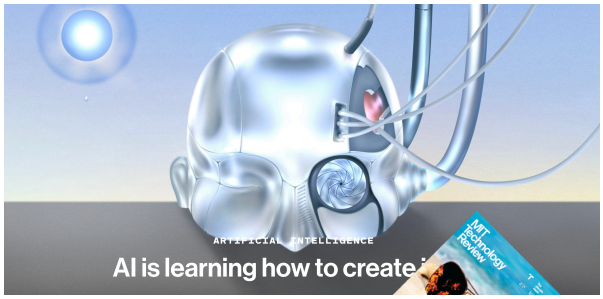
Answer four (4) problems out of five (5). If you answer all problems, only the first four will be taken into account. Each problem is worth max 10 points.

1. **Philosophy, History, and Ethics of AI**

a. (4 p) Essay. A good length is about 2/3 of a page. Topic: "The next generation of AI: Hybrid neuro-symbolic systems?"
It has been suggested that neither the "good-old fashioned AI" (GOFAI) (methods based mostly on symbolic representations) or the modern AI (methods such as machine learning and neural networks) alone can solve all problems, and that what we'll need to get to the next generation of AI is hybrid systems that combine the "best of both worlds".

Briefly discuss how you see the future of AI and to what extent such hybrid "neuro-symbolic" systems might be the missing piece.

b. (3 p) What do you think about the following headline and illustration (see right; MIT Tech Review, May 2021):



"AI is learning how to create itself – Humans have struggled to make truly intelligent machines. Maybe we need to let them get on with it themselves."

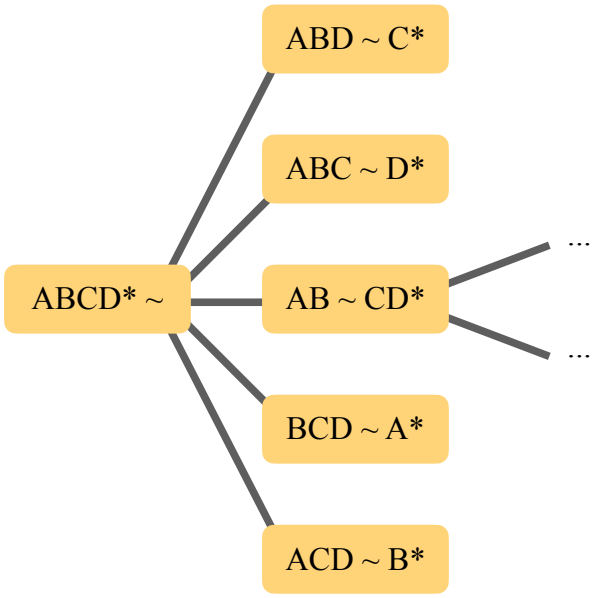
How well does this reflect your understanding of the capabilities of current AI?

c. (3 p) In what ways might AI be a threat to human rights? Mention at least two threats and propose ways to avoid these threats when developing and deploying AI systems.

2. **Problem Solving by Search**

Let's solve a puzzle: Alice (A) and Bob (B) are the proud owners of two remote-controlled robot dogs, named Curveball (C) and Dumbo (D). They want to cross a river with a tiny boat that can only fit only a) one person or b) up to two robot dogs (not, for example, one person and one robot dog). The boat has an electric outboard motor and the robot dogs are capable of operating the boat on their own. How to get all of them across the river dry?

a. (5 p) To get started, we'll draw a state-transition diagram. We've already gotten started with it, see right. The notation shows each of our protagonists as the letters A,B,C,D on either the left or the right side of the river ('~'). The star ('*') indicates on which side of the river the boat is in each state.



For example, in the starting state (the leftmost node in the diagram), everyone is still on the left side with the boat. There are five possible transitions: one for each of the four agents traveling to the right side alone, plus one where Curveball and Dumbo go together (the middle one: "AB ~ CD*").

Draw the rest of the state-transition diagram. Remember to draw each state only once, or otherwise your diagram will become infinitely large. Note that all the transitions are two-way. You shouldn't get more than a couple dozen states in your diagram. If you get more, something's off.

b. (1 p) Using the state-transition diagram, provide a solution to the puzzle in the form of a sequence of transitions from the state "ABCD* ~" to the state "~ ABCD*".

c. (2 p) What search algorithm would be a good choice in this case? What search algorithm would be a poor choice? Why?

d. (2 p) What types of real-world problems can be solved by this type of search techniques? What types of real-world problems can't?

3. **Reasoning under Uncertainty**

Suppose we are building a probabilistic model of the following scenario: The risk of catching a contagious disease[1] depends on the age of a person and whether they have been exposed to an infected person. The symptoms of the disease include fever which occurs more often with elderly patients than with the young. There is a test that can identify the disease but the test isn't perfect.

a. (3 p) Draw a Bayesian network to describe the above scenario.

   The network should contain five nodes: Age (young/old), Exposed (yes/no), Disease (yes/no), Fever (yes/no), and Test (pos/neg). The parents of each node should include the variables that influence the node in question: for example, the parents of the node Disease, should be Age and Exposed. (You can assume, for the sake of simplicity, that Age and Exposed are independent of each other, and that the test accuracy is the same for everyone.)

b. (3 p) Come up with imaginary, but reasonable, conditional probability tables for the network.

c. (2 p) Pretend to draw a sample of size $n = 10$ from the network. List the values of the five variables for each of the ten cases (or tuples) like this:

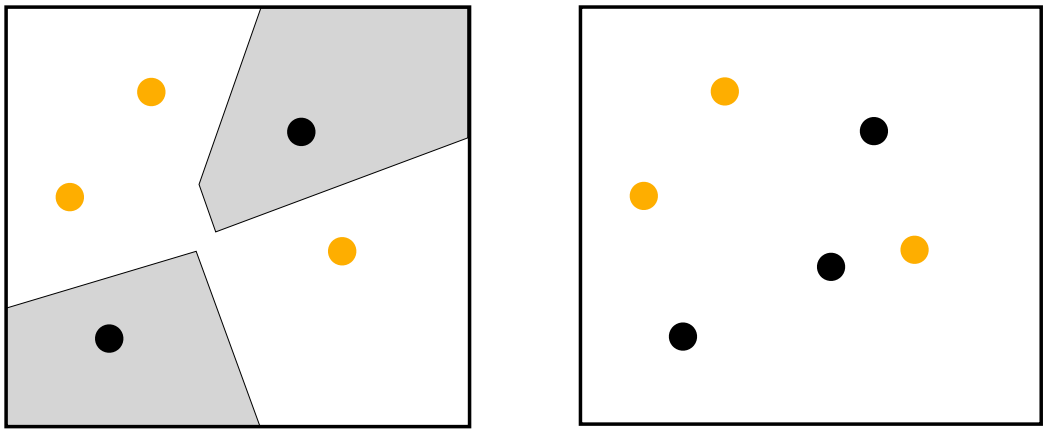
AGE	EXPOSED	DISEASE	FEVER	TEST
young	yes	no	no	neg
old	yes	yes	yes	pos
...				

d. (2 p) What could such a network be useful for? How would it be used?

[1] Any similarity to actual persons or events is purely coincidental.

4. **Machine Learning**

Consider a simple binary classification task where we classify two-dimensional data points into two classes: orange vs black. Given labeled training data, a nearest neighbour classifier splits the (x,y)-plane into regions based on the class label of the nearest training data point. The left figure below shows an example of the resulting classification boundaries: any test data point in the shaded regions would be classified as black.



a. (3 p) Draw a similar diagram showing the classification boundaries for the training data in the figure on the right, where there is one additional black data point.

b. (2 p) Do the same with a 3-NN classifier (using the data on the right).

c. (3 p) Now suppose we weren't dealing with simple 2D data but, say, images or audio samples. In what kinds of image or audio classification tasks would the *k*-NN classifier be likely to perform well and in what kinds not so well?

d. (2 p) What type of an approach would you try instead in the cases from the previous item where you expect the *k*-NN classifier to perform poorly?

5. Natural Language Processing

a. (3 p) Consider a context-free grammar with the following rules:

$S \longrightarrow S\ S$ $\quad B \longrightarrow S\ B$ $\quad S \longrightarrow$ bada

$S \longrightarrow A\ B$ $\quad A \longrightarrow$ big $\quad B \longrightarrow$ boom

S is the starting symbol. From S, we can generate, for example, $S \longrightarrow AB \longrightarrow ASB \longrightarrow ... \longrightarrow$ big bada boom.

Does the language contain the following three sentences (or strings): "boom boom", "bada bada bada bada", "big boom big boom"? Justify your answer.

b. (5 p) Apply the CYK algorithm to fill in the following table in order to determine whether the sentence "big boom bada bada big boom" belongs to the language. *Hint:* Cells (5,5), (6,6), and (5,6) have already been filled in. Recall that, e.g., cell (5,6) covers the words 5–6: "big boom".

(1,6)					
(1,5)	(2,6)				
(1,4)	(2,5)	(3,6)			
(1,3)	(2,4)	(3,5)	(4,6)		
(1,2)	(2,3)	(3,4)	(4,5)	S (5,6)	
(1,1)	(2,2)	(3,3)	(4,4)	A (5,5)	B (6,6)
big	*boom*	*bada*	*bada*	*big*	*boom*

c. (2 p) What is the corresponding parse tree (or parse trees), if any? Does the sentence belong to the language or not?