Last Name: First Name:

Matriculation Number:

Exam
Artificial Intelligence 2

September 24, 2025

Please ignore the QR codes; do not write on them, they are for grading support

	To be used for grading, do not write here												
prob.	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2	4.3	5.1	5.2	Sum	grade
total	9	8	10	10	9	8	9	9	8	6	4	90	
reached													

Organizational Information

Please read the following directions carefully and acknowledge them with your signature.

1. Please place your student ID card and a photo ID on the table for checking.
2. You can reach 90 points if you fully solve all problems. You will only need 85 points for a perfect score, i.e. 5 points are bonus points.
3. No resources or tools are allowed except for a pen.
4. You have 90 min (sharp) for the exam.
5. Write the solutions directly on the sheets, no other paper will be graded.
6. If you have to abort the exam for health reasons, your inability to sit the exam must be certified by an examination at the University Hospital. Please notify the exam proctors and have them give you the respective form.
7. Please make sure that your copy of the exam is complete (15 pages excluding cover sheet and organizational information pages) and has a clear print. **Do not forget to add your personal information on the cover sheet and to sign this declaration.**

Declaration: With my signature I certify having received the full exam document and having read the organizational information above.

Erlangen, September 24, 2025 .....
(signature)

Please consider the following guidelines to avoid losing points:

- If you continue an answer on another page, clearly give the problem number on the new page and a page reference on the old page.
- You can always ask for the translation or explanation of a non-technical word.
- If you do not want something to be graded, clearly cross it out. Adding a wrong statement to a correct solution may lead to deductions.
- The instructions "Give X", "List X" or similar mean that only X is needed. If you additionally justify your answer, we may or may not give you partial credit.
- The instruction "Assume X" means that X is information that your answer may use.
- The instruction "Model X as a Y" means that you have to describe X formally and exactly as an instance of Y using the definition of Y from the lecture.
- If you are uncertain how long or complex an answer should be, use the number of points as an indication: 1 point roughly corresponds to 1 minute.
- In all calculation questions, you have to simplify as much as reasonably possible without a calculator. For example, $\log 2$ or 3^7 should not be calculated, but $0.4 \cdot 0.3 \cdot 0.5 = 0.06$ should be.

1 Probabilities

Problem 1.1 (Python)

1. Consider the Python program below where J holds the joint probability distribution of two random variables such that $J[x][y] = P(X = x, Y = y)$. Which probability-related operation does the function `foo` compute? 4 Points

```
def foo(J):
  res = 0
  for x in range(len(J)):
    for y in range(len(J[x])):
      res += (x+y)*J[x][y]
  return res
```

2. Consider a state space $\{0, \ldots, n-1\}$, a transition model such that $T[i][j]$ is the probability of state i transitioning into state j, and a probability distribution of the current state, i.e., $S[i]$ is the probability of currently being in state i. 5 Points
Complete the function definition below to return the probability distribution of the state after one transition.

```
def next(T,S):
```

Problem 1.2 (Working with Distributions)
Assume Boolean random variables X, Y, Z. Their joint probability distribution is given as follows:

x	y	z	$P(X = x, Y = y, Z = z)$
0	0	0	a
0	0	1	b
0	1	0	c
0	1	1	d
1	0	0	e
1	0	1	f
1	1	0	g
1	1	1	h

1. What is the sample space Ω here?

2 Points

2. Give all subsets of the probabilities $\{a, \ldots, h\}$ that must sum to 1.

2 Points

3. In terms of $a, \ldots, h$, give $P(X = 0|Y = Z)$.

2 Points

4. If $e = f = g = h = 0$, under what conditions are X and Y independent?

2 Points

This page was intentionally left blank for extra space

2 Bayesian Reasoning

Problem 2.1 (Basic Rules)
Let A and B be Boolean random variables (using values 0 and 1) expressing that *Alice* and *Bob*, respectively, will pass a test. Let H be a Boolean random variable expressing that the test will be hard.

You estimate that Alice has a 60% chance to pass in general but only a 30% chance if the test is hard. 8 of the 10 most recent tests were hard, and you expect this to accurately predict the difficulty of the next test.

1. Give the following probabilities: 2 Points

$$P(H = 1) = \square$$

$$P(A = 1|H = 1) = \square$$

2. Calculate the updated probability that the test was hard after finding out that Alice has passed. 3 Points

3. Now assume you know nothing about Alice, but you know that $P(B = 1, H = 1) = 30\%$ and $P(B = 0, H = 0) = 5\%$.
Calculate the probability that Bob passes. 3 Points

4. Assume Alice and Bob have only studied together all semester. Explain in about 2 sentences, how that knowledge affects the probability analysis in this situation. 2 Points

Problem 2.2 (Bayesian Networks)
Consider the following situation about a car:

- Your car is unusable if it is out of gas or if it is broken. These two are the only causes.
- You might be late for work if your car does not work or if you oversleep. These two are the only causes.

You want to model this situation as a Bayesian network using Boolean random variables.

1. Give an appropriate set of random variables and their meaning. Give a good variable ordering and draw the resulting Bayesian network. 3 Points

2. Give the probability of the car being unusable in terms of the entries of the conditional probability table of your network. 2 Points

3. Now you decide to make the car-unusable node deterministic. Explain (in about 2 sentences) why that choice is justified based on the description above, and how it affects the conditional probability table of that node. 2 Points

4. Now you decide to make the late-for-work node a noisy disjunction node. Explain (in about 2 sentences) which two properties must hold about its probability distribution for this decision to 3 Points

be justified. Judge if these are backed by the description.

3 Markovian Reasoning

Problem 3.1 (Hidden Markov Models)
Consider the following situation:

- You make daily observations about your business B. Each day business is either good (b_1) or bad (b_2).
- You know this is caused by the general economic situation G, which you cannot easily observe, and which can be getting worse (g_1), be stable (g_2), or getting better (g_3).
- You have previously obtained the following information:
  - when the economy gets worse, your business is good 36% of the time,
  - when the economy is stable, your business is good 84% of the time,
  - when the economy gets better, your business is good 90% of the time,
  - half the time, the economy is the same as on the previous day,
  - when the economy changes from one day to the next, each change is equally likely.

You want to model this situation as a hidden Markov model with two families of random variables indexed by day number d.

1. Give the state and evidence variables and their domains.

2 Points

2. How can you tell that the sensor model is stationary here?

1 Points

3. What order does the model have?

1 Points

4. Complete the following sentence: The transition model T is given by the matrix

$$T = \begin{pmatrix} & \\ & \end{pmatrix} \quad \text{where} \quad T_{ij} = P(G_{d+1} = g_j \mid G_d = g_i).$$

2 Points

5. Complete the following sentence: The sensor model S is given by the matrix

$$S = \begin{pmatrix} & \\ & \end{pmatrix} \quad \text{where} \quad S_{ij} = P(B_d = b_j \mid G_d = g_i).$$

2 Points

6. Assume you want to apply filtering after observing good business at $t = 1$. Give the diagonal sensor matrix O_1 to use in this case. 1 Points

Problem 3.2 (Decisions and Utility)
You want to calculate the utilities of state sequence $\vec{s} = s_0, s_1, \ldots$ experienced by an agent. You want to use a reward function R from states to real numbers.

1. In about 1 sentence, explain the problem of naively computing the total reward as $R(\vec{s}) = \Sigma_i R(s_i)$. 2 Points

2. Finite horizon and reward discounting are two possible solutions for this problem. Explain each in about 1 sentence or 1 formula. 2 Points

3. Fill in the gaps in the following text: 4 Points

Both the value iteration and the policy evaluation algorithms use update rules derived from the [ ] equation to iteratively recompute the utility of all states. In each iteration, the utility of each state s is recomputed by applying one action and adding the [ ] of s and the expected utility of the next state (computed using the current utility values) multiplied by [ ] .

They differ in which actions are chosen:

The latter considers only one action per [ ] , which is given by a fixed [ ] .

The former picks the action that [ ] .

4 Learning

Problem 4.1 (Learning for Data with Attributes)
Consider the following data observed for some patients during a medical trial of a new drug:

Duration of Treatment 1, 2, or 3 weeks	Dosage 5 or 10 mg/week	Success yes/no
1	5	no
1	10	yes
2	5	no
2	10	yes
3	5	yes
3	10	yes

1. Give the decision tree for the attribute order "duration, dosage" that determines the patient's condition given the duration and dosage of the treatment. 3 Points

2. Give the entropy of the attribute "dosage". 1 Points

3. Give the entropy of the attribute "duration". 1 Points

To give a logical formulation, we use binary predicate symbols *Dur* and *Dos* and a unary predicate symbol *Suc*. For example, $Dur(p, 2) \wedge Suc(p)$ represents that patient p was successfully treated for 2 weeks.

4. Give the shortest first-order formula that correctly represents the treatment results for all patients p. 2 Points

5. Which clause(s) would inductive logic programming learn to represent the predicate *Suc*? 2 Points

Problem 4.2 (Loss)
Our goal is to find a linear approximation $h_a(x) = ax$ for the series of squares $0, 1, 4, 9, 16$ of the numbers $0, 1, 2, 3, 4$.

1. Model this situation as an inductive learning problem.

2 Points

2. Assuming all 5 possible examples are equally probable, compute the generalized loss using the squared error loss function.

3 Points

3. Explain in about 2 sentences how you would continue to eventually determine that $h^*(x) = 10x/3$.

2 Points

4. What is the error rate of h^*?

2 Points

Problem 4.3 (Bayesian Learning)
Consider an experiment with 2 different results 0 and 1. You repeat the experiment 3 times obtaining results $d = (d_1, d_2, d_3) = (1, 1, 0)$.
Your hypothesis space contains the functions h_β given by $h_\beta(i) = \beta^i$ for $0 \leq \beta \leq 1$ where $h_\beta(i)$ is the probability that the i-th repetition (starting at $i = 1$) will yield 1.

1. Calculate $P(d|h_{1/2})$, i.e., the likelihood of the data d under hypothesis $h_{1/2}$. 2 Points

2. Why does it make sense to exclude hypotheses h_β for $\beta > 1$? 2 Points

3. Before performing the experiment, you suspected β to be large and judged the probability of h_β to be $1 - \beta$. 2 Points
   Now that you've obtained the data d, state the formula obtained from Bayes' rule that updates the probability of h_β.

4. Let us call the probability obtained in the previous problem $P(\beta)$. Explain in about 2 sentences how you can choose a hypothesis using the Maximum a Posteriori approximation. 2 Points

5 Natural Language Processing

Problem 5.1 (Language Models)
Consider a language L over the alphabet with characters x, y, and z and a corpus for it consisting of the four words

$$x\,x\,x\,y\,z \qquad x\,x\,x\,y\,y \qquad x\,x\,x\,y\,z\,y \qquad x\,x\,x\,z\,y\,z$$

You want to build a 3-gram model using random variables $c_1, ...$

1. Seen as a stationary Markov process, what is the meaning of the c_i and the order of such a model? 2 Points

2. Give the value of $P(c_i = z | c_{i-2} = x, c_{i-1} = y)$ in the resulting model. 2 Points

3. Consider the word prefix xxx. By applying your model twice, which two characters would be predicted to follow? 2 Points

Problem 5.2 (Information Retrieval)
Consider a corpus of n documents. You have already computed the `tfidf` vector for each document.

1. Informally, explain in about 2 sentences how can you use the `tfidf` vectors to choose the 3 most relevant document for the query q. 2 Points

2. Let TP, TN, FP, FN be the number of true/false positive/negative results to a query. In terms of those, give the definitions of 2 Points
   1. precision:
   2. recall: