

Inteligência Artificial Centrada no Humano

2024/25 – 1º Semestre

Exame de Época Normal

03/janeiro/2025 – 90 min

Mestrado em Engenharia e Ciência de Dados

Departamento de Engenharia Informática

Question:	1	2	3	4	Total
Points:	50	20	15	15	100
Score:					

Pay attention to:

- **Important:** Fraud denotes a serious lack of ethics and constitutes unacceptable behavior for a higher education student and future professional with a master's degree. Any attempt at fraud will result in failure in the course, for both the facilitator and the perpetrator.
- This test is individual and open-book. You may not use electronic devices, including mobile phones. Only handwritten notes on paper are allowed. You may not exchange notes with colleagues.
- Answer the questions in the spaces provided for this purpose. If you need additional paper, contact the supervising instructor in your room.
- For **True/False questions requiring justification**, answers without justification will receive zero points.
- For **True/False questions without a justification requirement**, incorrect answers will incur a 30% penalty of the question's score; if you do not answer, you will receive zero points for the question.

1. (50%) A hospital has a team of robots to transport medications, laboratory samples, and medical equipment between different locations within the hospital. Each robot is equipped with sensors to navigate the hospital environment, avoiding people and obstacles, and is connected to the hospital management system, which optimally distributes and monitors the robots' tasks. Each robot has the ability to generate its own plans to carry out its tasks.

- (a) Indicate, with justification, whether there is cooperation or collaboration among the robots.

.....

.....

.....

.....

.....

.....

.....

.....

- (b) From the perspective of the AI used in the system, what are the risks associated with this system?

.....

.....

.....

.....

.....

.....

.....

.....

- (c) Complete the scenario so that the system operates in a Human-in-the-Loop context.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

- (d) Complete the scenario so that the system operates in an AI-in-the-Loop context.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

- (e) Indicate, with justification, which Learning by Demonstration techniques you would incorporate into the system and how you would implement them.

[illegible]

2. (20%) Consider a collection with the following textual documents and answer the questions:

stunning masterpiece is Positive
 stunning romance is Positive
 iconic masterpiece is Positive
 iconic comedy is Positive
 lacks depth is Negative
 overrated comedy is Negative

- (a) The collection was tokenized based on spaces only, and used as the only data for training two probabilistic language models, one based on bigrams, another based on trigrams, without any smoothing. Compute the probability of the sequence “*iconic comedy is Negative*” given by each model:

1. Bigram model:

2. Trigram model:

- (b) Both for the bigram and for the trigram model, the most probable word after “*crazy masterpiece is*” is *Positive*. Explain why.

- (c) Now consider a causal left-to-right LLM, such as GPT or Llama. The sequences in the collection can be used as training examples for a task in the scope of Sentiment Analysis. What is the name of such task?

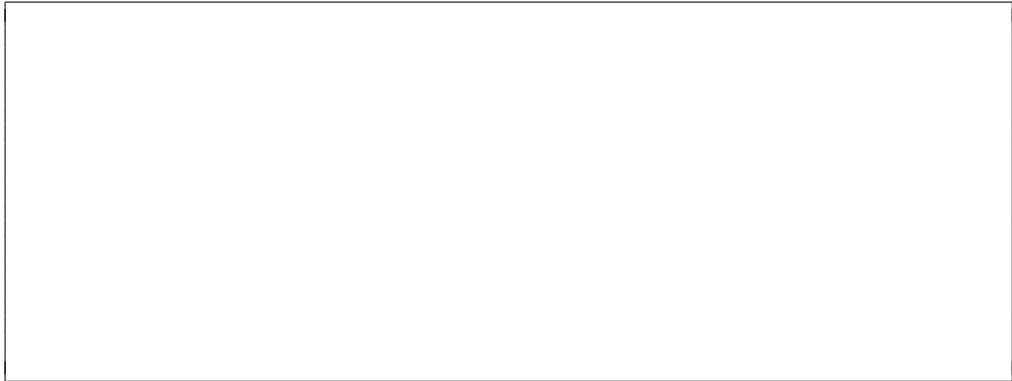
- (d) Think of a crowdsourcing process where, instead of human annotators, we have access to five different LLMs. In this scenario, describe, step-by-step, how the sequence "*crazy movie*" could be annotating, explicitly indicating the adopted *crowdsourcing* model.

- (e) Write a possible *prompt* that takes advantage of the collection for applying the aforementioned task to the sequence "*crazy movie*", in a *2-shot learning* mode.

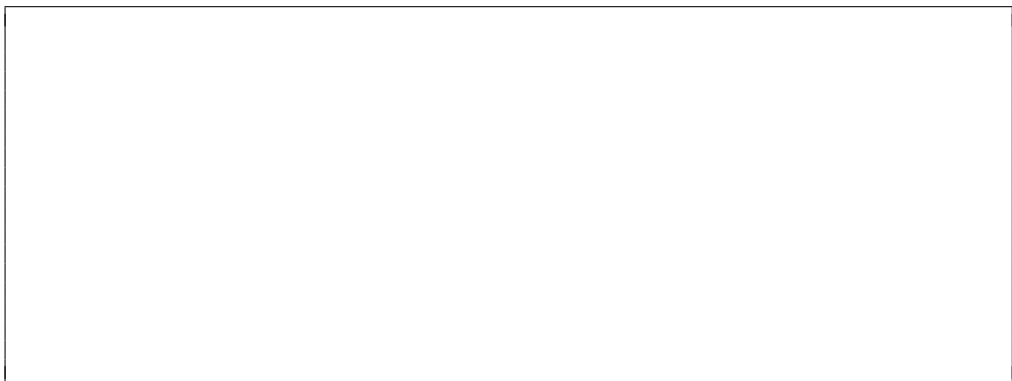
3. (15%) In order to better understand how the task of the previous question could be solved, a larger group of data was used to train a supervised model that learns how to split data in branches, based on feature values. The result is a tree-like structure where each leaf represents a class, and the path between the root and each leaf represents a series of decisions based on the features.

In the scope of Explainable AI, answer the following questions and explain why.

- (a) Is the method intrinsic or post-hoc?



- (b) Is the method model-specific or model-agnostic?



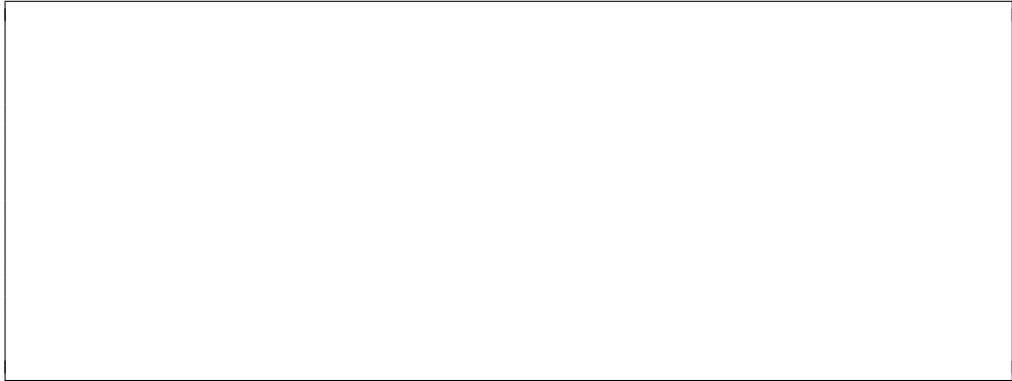
- (c) Are the provided explanations local or global?



4. (15%) Consider the following description of an AI-based tool:

ChatBotHelper is designed to assist users with customer service inquiries by providing automated responses to common questions. It uses natural language processing to understand and respond to user queries, helping to streamline customer-support operations.

- (a) According to the *EU AI Act*, what is the risk level of the described tool? Explain why.



- (b) According to the European Union, Trustworthy AI is based on three components. Identify which one is compromised if the following features are not part of the tool.

1. Provide correct responses.

A. Legal / Lawful B. Ética / Ethical C. Robusto / Robust D. Nenhuma

2. Inform the users that they are interacting with a computational system.

A. Legal / Lawful B. Ética / Ethical C. Robusto / Robust D. Nenhuma

3. Avoid biases in the given responses.

A. Legal / Lawful B. Ética / Ethical C. Robusto / Robust D. Nenhuma

- (c) Modern chatbots are based on LLMs. Comment the following statement, while clarifying the difference between a faithful and a plausible explanation.

Besides being black-box models, if prompted for that, LLMs may generate natural language explanations for their responses.

