

1 2 9 0



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE  
**COIMBRA**

## Inteligência Artificial Centrada no Humano

2022/23 – 1º Semestre

**Exame de Época Normal**  
**04/janeiro/2023 – 90 min**

*Mestrado em Engenharia e Ciéncia de Dados  
Departamento de Engenharia Informática*

Question:	1	2	3	4	Total
Points:	25	25	25	25	100
Score:					

### Read carefully:

- **Important note:** Fraud denotes a serious lack of ethics and constitutes a conduct that is not admissible in a student and future licensed professional. Any attempt at fraud leads to cancellation of the exam of both the facilitator as the harasser.
- Each page must clearly include the student's name and number.
- Only paper materials can be used for consultation.
- Answers must be written by the student, individually, with no support of a third party. It is strictly forbidden to communicate with anyone else but the watchers. Electronic devices are not allowed.
- In **True/False questions** with justification, unjustified answers will be scored with zero.
- For **True vs False questions**, you should consider that if you provide an incorrect answer, you get a penalty of 30%, if you don't provide an answer you get 0%, and if you provide a correct answer you get 100%.

1. (25%) Classify the following sentences as True or False (important note, for each sentence: if you give the correct answer, you get 100%; if you don't give an answer, you receive 0%; if you give an incorrect answer, you receive a penalty of 30% of the value of the question):
  - (a) Active Learning maybe classified as a kind of unsupervised learning because it assumes that a non-empty subset of the data set is unlabelled.  
(a) \_\_\_\_\_
  - (b) Human-AI collaboration can happen only when the AI agent needs help for learning.  
(b) \_\_\_\_\_
  - (c) Imitation Learning is a form of Interactive Learning.  
(c) \_\_\_\_\_
  - (d) Active Learning agents can also be called curiosity-based agents.  
(d) \_\_\_\_\_
  - (e) Linear Models assume the existence of only three kinds of features: numerical, binary, and categorical features.  
(e) \_\_\_\_\_
  - (f) The interpretation for a categorical feature in Linear Models is: Changing the feature from the reference category to another category changes the estimated outcome by the feature's weight.  
(f) \_\_\_\_\_
  - (g) Bayesian networks are compact specifications of full joint distributions that cannot be learned from data.  
(g) \_\_\_\_\_
  - (h) For Explainable AI (XAI), the beliefs and goals of an AI system are irrelevant, while decisions are of primary importance; XAI should focus only on explaining decisions.  
(h) \_\_\_\_\_
  - (i) We need XAI to debug wrong predictions, to avoid bias, not to correct models.  
(i) \_\_\_\_\_
  - (j) Random Forest is a very accurate model that is interpretable and thus it does not need explanations.  
(j) \_\_\_\_\_
  - (k) Linear Models and Graphical Models are interpretable and do not demand explanations at all.  
(k) \_\_\_\_\_
  - (l) An explainable Machine Learning model goes beyond interpretability in that it helps humans in understanding in an human-readable fashion how and why a model came up with a particular answer given a particular input.  
(l) \_\_\_\_\_

2. (25%) Consider the utility matrix below for the rating of five products (P1, ..., P5) by four users (A, B, C, and D) in which a User Based, Nearest Neighbor Collaborative Recommender System relies on. The ratings belong to [1, 5]. What is the prediction for the rating of product P3 by user A, using the typical prediction function ( $pred(u, p)$ ; see Equation 1 below) and considering a user neighbor set of size 2, computed with the Jaccard Index defined below?

$$JaccardIndex = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$JaccardDistance = 1 - JaccardIndex$$

$$pred(u, p) = r_{u,p} = \bar{r}_u + \frac{\sum_{i \in users} sim(u, i) * (r_{i,p} - \bar{r}_i)}{\sum_{i \in users} sim(u, i)} \quad (1)$$

$$\text{, where } \bar{r}_u = \frac{\sum_p r_{u,p}}{\sum_p}$$

	<b>P1</b>	<b>P2</b>	<b>P3</b>	<b>P4</b>	<b>P5</b>
<b>A</b>	2	3	?		4
<b>B</b>	3	2	4	3	
<b>C</b>	1		3	2	4
<b>D</b>		5	3	1	4

3. (25%) Great changes are expected in the business model of a retail company. In order to assess the impact of some, a second company was hired for developing a system for the automatic classification of the sentiment expressed by the clients, in the Mastodon social network.

(a) Describe a task in the scope of Sentiment Analysis that is adequate to the problem. Then, think of a model for this task and, when applied for the problem, specify what would be: (i) the inputs; (ii) possible outputs.

(b) Through the social network's API, it is possible to automatically collect all the publications with mentions to the company, their publication data, and their author. What is lacking from these data for making them ready for training and evaluating supervised models with the desired goal? Describe, in detail, a possible approach for obtaining it, while supporting its suitability and considering, as far as possible, the answer given in a.

- (c) Describe two transformations that these data may suffer at the pre-processing level, explain why and when they can be useful for the previous problem and, finally, illustrate the result of these transformations when applied to the following publication:

*Did they get that we are unhappier now?*

- (d) How would a neural language model based on transformers be useful to this problem? Explain what could be its role, as well as an advantage and a disadvantage of its utilization.

4. (25%) DALL-E is an image generation system based on computational learning, created by the company OpenAI. It is trained to generate images from textual descriptions, using a dataset of text-image pairs. The system works online and is capable of generating a wide variety of images, from photorealistic to highly stylised, from abstract to purely descriptive, and can even generate images of objects and scenes that do not exist in the real world.

Suppose the company wants DALL-E to follow the Ethical Guidelines for Trustworthy AI defined by the High-Level Expert Group on Artificial Intelligence set up on the initiative of the European Commission.

- (a) Analyse carefully the following sentences that expose concerns about the operation of systems such as DALL-E. Tick, for each one, which of the three components of Trustworthy AI will be compromised if the situation arises.

- (i) *The system will be only as good as the data it is trained with. If the data is biased, the system may also produce biased results, which can lead to discrimination against certain groups of people.*

Concern regarding (tick one option):

Legality: \_\_\_\_\_ Ethics: \_\_\_\_\_ Robustness: \_\_\_\_\_

- (ii) *The system uses large amounts of data to generate results, which may raise privacy concerns for individuals whose personal data is being used.*

Concern regarding (tick one option):

Legality: \_\_\_\_\_ Ethics: \_\_\_\_\_ Robustness: \_\_\_\_\_

- (iii) *Systems like DALL-E can replace human jobs, particularly in the arts, and can lead to unemployment for artists and other related professions.*

Concern regarding (tick one option):

Legality: \_\_\_\_\_ Ethics: \_\_\_\_\_ Robustness: \_\_\_\_\_

- (iv) *DALL-E has the potential to generate images and objects that do not exist in reality, which can lead to the misuse of the technology for malevolent purposes, such as creating fake news and offensive content, or manipulating public opinion.*

Concern regarding (tick one option):

Legality: \_\_\_\_\_ Ethics: \_\_\_\_\_ Robustness: \_\_\_\_\_

- (b) Classify the following statements as True or False (considering the Ethical Guidelines mentioned above), justifying your answer (important note, for each sentence: if you give the correct answer, you get 100%; if you do not give an answer, you get 0%; if you give an incorrect answer, you get a penalty of 30% of the question value):

- (i) Compliance with the Data Protection Regulations in force is an ethical imperative for a system like DALL-E.

*True/False:* \_\_\_\_\_

Justify:

---

---

---

---

---

---

- (ii) The ability to explain its results is an legal imperative for this system.

*True/False:* \_\_\_\_\_

Justify:

---

---

---

---

---

---

- (iii) The protection of the dignity, mental integrity and physical integrity of the human being, the avoidance of situations where the system may cause or aggravate negative impacts due to power or information asymmetries, are all ethical imperatives for systems such as DALL-E.

*True/False:* \_\_\_\_\_

*Justify:*

---

---

---

---

---