

Titulació	Tipus	Curs
Intel·ligència Artificial / Bachelor in Artificial Intelligence	OP	3

Professor/a de contacte

Nom : Xim Cerdà Company

Correu electrònic : joaquin.cerda@uab.cat

Equip docent

Ernest Valveny Llobet

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

There are no official prerequisites, but it is recommended to have completed the subjects of Fundamentals of Programming I and II, Fundamentals of Mathematics I and II, Probability and Statistics, Data Engineering, Fundamentals of Machine Learning, and Fundamentals of Natural Language.

Objectius

This course provides an overview of the Natural Language Processing (NLP) applications, from classical approaches for text processing to advanced methods for person-computer interaction. This course covers both machine learning and deep learning techniques for NLP, considering both text and speech processing.

By the end of this course, students will be able to:

- Understand the fundamental concepts and techniques used in NLP.
- Implement and evaluate various NLP techniques using Python and popular NLP libraries.
- Apply NLP methods to real-world problems and interpret the results.

Resultats d'aprenentatge

- CM14 (Dissenyar solucions a projectes de processament del llenguatge natural en qualsevol àmbit i aplicació.) Dissenyar solucions a projectes de processament del llenguatge natural en qualsevol àmbit i aplicació.
- CM15 (Avaluar l'adequació de sistemes de generació de llenguatge a una determinada aplicació, en funció del seu possible impacte en aspectes ètics, legals i socials com poden ser el biaix, la discriminació, la veracitat o la privacitat i la seguretat de les dades.) Avaluar l'adequació de sistemes de

generació de llenguatge a una determinada aplicació, en funció del seu possible impacte en aspectes ètics, legals i socials com poden ser el biaix, la discriminació, la veracitat o la privacitat i la seguretat de les dades.

- KM32 (Descriure els models, les arquitectures, els procediments d'entrenament, les tècniques i els mètodes d'aprenentatge automàtic que s'utilitzen en el desenvolupament de sistemes de processament de llenguatge natural.) Descriure els models, les arquitectures, els procediments d'entrenament, les tècniques i els mètodes d'aprenentatge automàtic que s'utilitzen en el desenvolupament de sistemes de processament de llenguatge natural.
- SM34 (Utilitzar tècniques i algoritmes de processament de text, modelatge del llenguatge, representació de text i anàlisi de seqüències per a la implementació de solucions de processament de llenguatge natural.) Utilitzar tècniques i algoritmes de processament de text, modelatge del llenguatge, representació de text i anàlisi de seqüències per a la implementació de solucions de processament de llenguatge natural.
- SM35 (Aplicar models de representació i generació de text existents a diferents tipus de problemes de processament del llenguatge.) Aplicar models de representació i generació de text existents a diferents tipus de problemes de processament del llenguatge.
- SM36 (Programar un ús eficient i òptim de les infraestructures i els recursos necessaris per a l'entrenament i l'aplicació de models de processament de llenguatge natural.) Programar un ús eficient i òptim de les infraestructures i els recursos necessaris per a l'entrenament i l'aplicació de models de processament de llenguatge natural.
- SM37 (Aplicar metodologies d'avaluació i anàlisi de sistemes de processament del llenguatge natural que tinguin en compte de manera crítica tant el rendiment com els possibles biaixos i les implicacions ètiques, legals i socials.) Aplicar metodologies d'avaluació i anàlisi de sistemes de processament del llenguatge natural que tinguin en compte de manera crítica tant el rendiment com els possibles biaixos i les implicacions ètiques, legals i socials.

Continguts

1. Fonaments de NLP
2. Anàlisi semàntica i pragmàtica
3. Transformers per a aplicacions de NLP
4. Fonaments dels LLMs
5. Tècniques de millora dels LLMs
6. Extensions dels LLMs
7. Models per al reconeixement de veu i àudio

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Classes de teoria	15	0,6	
Treball en el projecte	50	2	
Estudi individual	25	1	
Sessió de projecte	6	0,24	
Resolució d'exercicis	25	1	
Sessió d'exercicis	25	1	

Sessions will combine three types of teaching activities: theory classes, solving project-based exercises, and a project development. Students will work individually to solve the exercises and in small groups of two or three people for the project.

1. Theory lectures, which will contain the theoretical background needed to solve the exercises and the project

will be presented using a presentation. These presentations will contain theoretical concepts, and mathematical formulation, as well as the corresponding algorithmic solutions.

2. Project-based exercises will be developed using Jupyter notebooks to be able to comment, point by point, the coding solutions. These exercises will be submitted regularly through Campus Virtual, explaining the proposed solution and showing the obtained results. Optionally, these exercises could include a report or a presentation for its evaluation. The exercise's submissions will comprise the portfolio.

3. A project will be carried out during the semester, where students will have to solve a real-world problem. The project will be solved in small groups of two or three students, where each member of the group must contribute a part and put it together with the rest to obtain the final solution. These working groups must be maintained until the end of the semester and must be self-managed in terms of distribution of roles, work planning, assignment of tasks, management of available resources, conflicts, etc. To develop the project, the groups will work autonomously, while the practical sessions will be used (1) for the teacher to present the project theme and discuss possible approaches, (2) for monitoring the status of the project and (3) for the teams to present their final results.

The above activities will be complemented by a system of tutoring and consultations outside class hours.

All the information of the subject and the related documents that the students need will be available at the virtual campus (cv.uab.cat).

Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Note: Fifteen minutes of one class session, within the schedule established by the school/degree programme, will be reserved for students to complete the surveys evaluating both the teaching performance of the instructor and the course/module

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, perquè l'alumnat completi les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Projecte	30%	2	0,08	CM14, CM15, KM32, SM34, SM35, SM36, SM37
Portfoli	49%	0	0	CM14, CM15, KM32, SM34, SM35, SM36, SM37
Test	21%	2	0,08	CM14, CM15, KM32, SM34, SM35, SM36, SM37

To assess the level of student learning, a formula is established that combines knowledge acquisition, the ability to solve problems and the ability to work as a team, as well as the presentation of the results obtained.

Final grade

The final grade is calculated in the following way and according to the different activities that are carried out:

Final grade = 0.7 * Exercises Grade + 0.3 * Project Grade

This formula will be applied as long as the Exercises and Project grades, are higher than 4. If the formula yields ≥ 5 , but the student does not reach the minimum required in any of the evaluation activities, then a final grade of 4.5 will be assigned.

Exercises Grade

The aim of the exercises is to become familiar with the use of the theoretical concepts, and to apply them to a real-world problem. The regular submission of problem solutions will be used as evidence of this work.

In order to obtain a grade for exercises, it is necessary that more than 50% of the exercises are submitted during the semester. In the contrary, the portfolio grade will be 0.

A test about the exercises will be performed individually at the end of the semester. The final problems grade will be the combination of the exercise's portfolio and this test.

Problems Grade = $0.7 * \text{Portfolio evaluation} + 0.3 * \text{Test}$

The formula will be applied as long as the Test grade is higher than 4.

At the end of the semester, students will have the opportunity to re-submit two different deliveries for retaking, and they will also be able to retake the Test. After the retakes, the maximum grade which can be obtained is 8.

Project Grade

The project carries an essential weight in the overall mark of the subject. Developing the project requires that the students work in groups and design an integral solution to the defined challenge. In addition, the students must demonstrate their teamwork skills and present the results to the class.

The project is evaluated through its report, an oral presentation that students will present in class, and an individual-evaluation process. The participation of students in all three activities (preparing the report, presentation and individual evaluation) is necessary in order to obtain a projects grade. The grade is calculated as follows:

Project Grade = $0.6 * \text{Report} + 0.2 * \text{Presentation} + 0.2 * \text{Individual evaluation}$

If performing the above calculation yields ≥ 5 but the student did not participate in any of the activities (report, presentation, individual evaluation), then a final grade of 4.5 will be given to the corresponding project.

There will be a retake of the project in case the final project grade does not reach the minimum of 4. In case of copy, there will be no recovery and the subject will be considered failed. The maximum project grade that can be obtained in case of retake is 7.

In this subject, the use of generative Artificial Intelligence technologies is allowed in a controlled manner. The student must clearly identify which parts of his/her work have been carried out with the support of generative AI tools and, in any case, must be able to understand, explain and justify the work carried out. The lack of transparency in the use of generative AI will be considered a lack of academic honesty and may lead to a partial or total penalty in the grade of the activity.

Bibliografia

- D. Jurafsky, J.H. Martin. *Speech and Language Processing*. Third Edition. 2021 < <https://web.stanford.edu/~jurafsky/slp3/>>
- R.S.T. Lee. *Natural Language Processing*. 2024. Springer
- G. Paab. *Foundation Models for Natural Language Processing*. 2023. Springer
- J. Eisenstein. *Natural Language Processing*. 2018. MIT Press
- H. Lane, C. Howard, H. M. Hapke. *Natural Language Processing in Action*. 2019. Manning Publications
- Kenny, Dorothy, ed. *Machine translation for everyone*. Lanugage Science Press, 2022. < <https://langsci-press.org/catalog/book/342>>
- Rowe, Bruce M., and Diane P. Levine. *A concise introduction to linguistics*. Routledge, 2018.

Programari

For the problems and projects of the course we will use Python, along with some Python libraries for NLP that will be specified during the course.

Grups i idiomes de l'assignatura

La informació proporcionada és provisional fins al 30 de novembre.

A partir d'aquesta data, podreu consultar l'idioma de cada grup a través d'aquest [enllaç](#). Per accedir a la informació, caldrà introduir el CODI de l'assignatura

Tipus de docència	Grup	Idioma	Semestre	Torn
(TE) Teoria	71	Anglès	segon quadrimestre	tarda
(PAUL) Pràctiques d'aula	711	Anglès	segon quadrimestre	tarda
(PLAB) Pràctiques de laboratori	711	Anglès	segon quadrimestre	tarda