



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

105001054 - Inteligencia Artificial Y Ciencia Abierta En Ingeniería De Software Científico

### PLAN DE ESTUDIOS

10CD - Grado En Ciencia De Datos E Inteligencia Artificial

### CURSO ACADÉMICO Y SEMESTRE

2023/24 - Segundo semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	2
5. Descripción de la asignatura y temario.....	3
6. Cronograma.....	6
7. Actividades y criterios de evaluación.....	8
8. Recursos didácticos.....	10
9. Otra información.....	11

## 1. Datos descriptivos

---

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	105001054 - Inteligencia Artificial y Ciencia Abierta en Ingeniería de Software Científico
<b>No de créditos</b>	3 ECTS
<b>Carácter</b>	Optativa
<b>Curso</b>	Cuarto curso
<b>Semestre</b>	Octavo semestre
<b>Período de impartición</b>	Febrero-Junio
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10CD - Grado en Ciencia de Datos e Inteligencia Artificial
<b>Centro responsable de la titulación</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Curso académico</b>	2023-24

## 2. Profesorado

---

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías</b> *
Oscar Corcho Garcia (Coordinador/a)		oscar.corcho@upm.es	- -

\* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

## 3. Conocimientos previos recomendados

---

### 3.1. Asignaturas previas que se recomienda haber cursado

El plan de estudios Grado en Ciencia de Datos e Inteligencia Artificial no tiene definidas asignaturas previas recomendadas para esta asignatura.

### 3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Web Semántica y Grafos de Conocimientos
- Inteligencia Artificial
- Ingeniería del software

## 4. Competencias y resultados de aprendizaje

---

### 4.1. Competencias

CB05 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE06 - Capacidad para describir los fundamentos de las infraestructuras de gestión e intercambio de datos: hardware, sistemas operativos, bases de datos, redes de computadores.

CE12 - Capacidad de comunicar de forma efectiva el proceso de análisis a partir de los datos y la interpretación de los resultados del mismo, seleccionando y utilizando para ello las técnicas y herramientas de visualización de datos más adecuadas.

CG01 - Capacidad de trabajo en equipo, en entornos interdisciplinarios y complejos, negociando y resolviendo conflictos, diseñando soluciones eficientes, fiables, robustas y responsables.

## 4.2. Resultados del aprendizaje

RA145 - Ability to identify and address real world problems where AI techniques applied to research software engineers can help

RA143 - Basic knowledge of parallelization techniques for efficient execution

RA149 - Ability to clean, integrate and exploit data from knowledge graphs

RA152 - Ability to read, understand and implement research publications

RA147 - Ability to create an abstract sketch of a research method

RA151 - Ability to create knowledge graphs of research software and data

RA144 - Ability to apply cutting edge machine learning techniques to help classify, recommend and organize research software and data

RA150 - Ability to create software containers that can run software components in different computational infrastructures

RA148 - Ability to describe metadata and provenance for research software and data in machine-readable formats

RA146 - Ability to read, understand and implement standard recommendations and guidelines (e.g., international committees, World Wide Web Consortium (W3C), etc.)

## 5. Descripción de la asignatura y temario

---

### 5.1. Descripción de la asignatura

El objetivo de este curso es que los estudiantes aprendan los fundamentos de la IA en la Ingeniería de Software Científico, con un enfoque especial en aplicaciones del mundo real. El curso incluirá conocimientos teóricos y metodológicos sobre técnicas avanzadas de IA diseñadas para ayudar a la reproducibilidad y reutilización de software, datos y sus metadatos en el ámbito de investigación, con aplicaciones en industria. Específicamente, el curso abordará la importancia de la reproducibilidad en ciencia, enfoques existentes para administrar, empaquetar y entregar un producto de investigación, la planificación y ejecución en paralelo de experimentos computacionales complejos, la creación de grafos de conocimiento para facilitar la búsqueda y la configuración de software científico; así como técnicas de aprendizaje automático para identificar similitudes entre software. La asignatura presentará aplicaciones que combinan todo lo anterior para facilitar la integración y comprensión de datos y

software científicos.

The objective of this course is for students to learn the foundations of AI in Research Software Engineering, with a special focus on real-world applications. The course will include theoretical and methodological knowledge on cutting edge AI techniques designed to aid the reproducibility, repurpose and reuse of research software, data and their metadata. More specifically, the course will address the importance of reproducibility in Science, approaches for managing, containerizing and delivering a research product, the role of planning and parallelization in computational-heavy experiments, the creation of knowledge graphs to ease research software findability and set up; the role of machine learning to help identify similarities between different code bases and applications that combine all of the above to ease research data and software integration and understanding.

## 5.2. Temario de la asignatura

1. Introducción a la Ingeniería del Software de Investigación / Introduction to Research Software Engineering
  - 1.1. Motivación: Reproducibilidad y los principios FAIR para datos y software científicos / Motivation: Reproducibility and the FAIR principles for research data and software
  - 1.2. Software y preservación de datos: repositorios de software y registros de metadatos/ Software and data conservation: code repositories and metadata registries
  - 1.3. Iniciativas abiertas para gestionar datos y software / Open initiatives for managing Data and Software
  - 1.4. Descripción de datos y software para reproducibilidad y reutilización / Describing data and software for reproducibility and reuse
2. Métodos computacionales / Computational scientific methods
  - 2.1. Notebooks / Computational notebooks
  - 2.2. Infraestructura computacional y contenedores de software / Computational infrastructure and software containers
  - 2.3. Flujos de trabajo científicos / Scientific workflows
  - 2.4. Composición de flujos de trabajo y razonamiento / Workflow composition and reasoning
3. Provenance en Ingeniería de Software de Investigación / Provenance in Research Software Engineering
  - 3.1. Introducción a provenance / Introduction to provenance
  - 3.2. El estándar W3C PROV / The W3C PROV standard
  - 3.3. Repositorios abiertos para recolectar provenance / Open repositories for collecting provenance
  - 3.4. Aplicaciones para capturar y usar provenance en la Web / Applications for capturing and exploiting provenance in the Web

#### 4. Grafos de Conocimiento Científicos / Scientific Knowledge Graphs

4.1. Representación de Datos y software procesable por máquinas / Machine-readable data and software representation

4.2. Capturando el contexto de un experimento: Objetos de Investigación / Capturing the context of research: Research Objects

4.3. Aplicaciones y ejemplos sobre grafos de Conocimiento Científicos / Applications for Scientific Knowledge Graphs

#### 5. Aprendizaje automático en Ingeniería del Software de Investigación / Machine Learning for Research Software Engineering

5.1. Clustering y clasificación de software científico y sus metadatos / Clustering and classification of research software and its metadata

5.2. Reconocimiento de Entidades en software científico / Named entity recognition in research software

5.3. Aplicaciones de aprendizaje automático en Ingeniería del Software de Investigación / Applications of machine learning models in Research Software Engineering

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>Themes 1.1 to 1.4</b> Duración: 01:15 LM: Actividad del tipo Lección Magistral	<b>Themes 1.1 to 1.4</b> Duración: 00:45 PL: Actividad del tipo Prácticas de Laboratorio		
2	<b>Themes 2.1 and 2.2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Themes 2.1 and 2.2</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
3	<b>Theme 2.2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 2.2</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
4	<b>Theme 2.2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 2.2</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	<b>Theme 2.3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 2.3</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>Create a computational experiment, containerize it, and describe its provenance</b> TI: Técnica del tipo Trabajo Individual Evaluación continua No presencial Duración: 15:00
6	<b>Theme 2.4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 2.4</b> Duración: 01:30 PL: Actividad del tipo Prácticas de Laboratorio		
7	<b>Theme 3.1 to 3.4</b> Duración: 01:15 LM: Actividad del tipo Lección Magistral	<b>Theme 3.1 to 3.4</b> Duración: 00:45 PL: Actividad del tipo Prácticas de Laboratorio		
8	<b>Theme 4.1</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 4.1</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
9	<b>Theme 4.2</b> Duración: 00:30 LM: Actividad del tipo Lección Magistral	<b>Theme 4.2</b> Duración: 01:30 PL: Actividad del tipo Prácticas de Laboratorio		<b>Design and implement an initial version of a scientific knowledge graph using and relating the computational experiments and integrating data and software from at least two open code repositories</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 15:00



10	<b>Theme 4.3</b> Duración: 00:30 LM: Actividad del tipo Lección Magistral	<b>Theme 4.3</b> Duración: 01:30 PL: Actividad del tipo Prácticas de Laboratorio		
11		<b>Homework discussion and questions</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas		
12	<b>Theme 5.1</b> Duración: 00:30 LM: Actividad del tipo Lección Magistral	<b>Theme 5.1</b> Duración: 01:30 PL: Actividad del tipo Prácticas de Laboratorio		
13	<b>Theme 5.2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 5.2</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>Enhance a scientific knowledge graph with AI techniques and explain them.</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 20:00
14	<b>Theme 5.3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Theme 5.3</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
15		<b>Homework discussion and questions</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas		
16	<b>Final group presentations</b> Duración: 02:00 OT: Otras actividades formativas			<b>Final group presentation</b> PG: Técnica del tipo Presentación en Grupo Evaluación continua Presencial Duración: 02:00
17	<b>Written test</b> Duración: 02:00 OT: Otras actividades formativas			<b>Final exam (written test)</b> EX: Técnica del tipo Examen Escrito Evaluación continua y sólo prueba final Presencial Duración: 02:00

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

\* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

## 7. Actividades y criterios de evaluación

### 7.1. Actividades de evaluación de la asignatura

#### 7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Create a computational experiment, containerize it, and describe its provenance	TI: Técnica del tipo Trabajo Individual	No Presencial	15:00	20%	/ 10	CB05 CE06
9	Design and implement an initial version of a scientific knowledge graph using and relating the computational experiments and integrating data and software from at least two open code repositories	TG: Técnica del tipo Trabajo en Grupo	No Presencial	15:00	12.5%	5 / 10	CG01 CE06 CE12
13	Enhance a scientific knowledge graph with AI techniques and explain them.	TG: Técnica del tipo Trabajo en Grupo	No Presencial	20:00	12.5%	5 / 10	CG01 CE06 CE12
16	Final group presentation	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	10%	5 / 10	CE12
17	Final exam (written test)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	100%	5 / 10	CB05 CE06

#### 7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Final exam (written test)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	100%	5 / 10	CB05 CE06

### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Final exam (written test)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	100%	5 / 10	CE06 CE12 CB05

## 7.2. Criterios de evaluación

Se evaluarán los siguientes elementos, con los pesos que se especifican a continuación:

1. El material proporcionado por el alumno y la interacción con el repositorio de la asignatura sobre el trabajo propuesto en los temas 2 y 3 (peso: 20 %)
2. El material proporcionado por el alumno y la interacción con el repositorio de la asignatura sobre el trabajo propuesto en el tema 4 (peso: 12.5 %)
3. El material proporcionado por el alumno y la interacción con el repositorio de la asignatura sobre el trabajo propuesto en el tema 5 (peso: 12.5 %)
4. El papel del alumno en la presentación y su capacidad para responder preguntas sobre el trabajo del proyecto (peso: 10 %)
5. Examen escrito final (peso: 45%)

Los estudiantes han de aprobar las tareas asignadas para poder hacer el examen escrito (con peso 45%).

\*\*\*\*\* The following items will be evaluated, with the weights that are specified next:

1. The material provided by the student and the interaction in the course code and materials repository on the work proposed in section 2 and 3 (weight: 20%)
2. The material provided by the student and the interaction in the course code and materials repository on the work proposed in section 4 (weight: 12.5%)
3. The material provided by the student and the interaction in the course code and materials repository on the work

proposed in section 5 (weight: 12.5%)

4. The role of the student on the presentation and the ability to answer questions about the project work (weight: 10%)

5. Final written exam (weight: 45%)

Note that all assignments must be passed with a minimum score in order to attend the final written test.

## 8. Recursos didácticos

### 8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
The Scientific Paper of the future	Bibliografía	The Scientific Paper of the future:  <a href="https://www.scientificpaperofthefuture.org/">https://www.scientificpaperofthefuture.org/</a>
Jim Gray on eScience: A Transformed Scientific Method	Bibliografía	Jim Gray on eScience: A Transformed Scientific Method: <a href="https://www.microsoft.com/en-us/research/wpcontent/uploads/2009/10/Fourth_Paradigm.pdf">https://www.microsoft.com/en-us/research/wpcontent/uploads/2009/10/Fourth_Paradigm.pdf</a>
Introduction to containers	Recursos web	<a href="https://biocontainersedu.biocontainers.pro/en/latest/introduction.html">https://biocontainersedu.biocontainers.pro/en/latest/introduction.html</a>
Linked Data: Evolving the Web into a Global Data Space (1st edition).	Bibliografía	Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). . Morgan & Claypool. Available from  <a href="http://linkeddatabook.com/editions/1.0/">http://linkeddatabook.com/editions/1.0/</a>
PROV Model primer	Bibliografía	PROV Model primer: <a href="https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/">https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/</a>
An introduction to Scikit Learn	Recursos web	An introduction to Scikit Learn: <a href="https://scikit-learn.org/stable/tutorial/basic/tutorial.html">https://scikit-learn.org/stable/tutorial/basic/tutorial.html</a>

## 9. Otra información

---

### 9.1. Otra información sobre la asignatura

The course will be taught in English, following a hands-on style of learning. Given the practical nature of the course, students are required to bring their laptops to work during most of the lectures.

All the tools described and taught in this course are open source.