

Titulación	Tipo	Curso
Inteligencia Artificial / Artificial Intelligence	FB	2

Contacto

Nombre: María Pilar Dellunde Clave

Correo electrónico: pilar.dellunde@uab.cat

Equipo docente

Nuria Valles Peris

(Externo) Roger Deulofeu Batllori

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

Esta asignatura no tiene prerrequisitos

Objetivos y contextualización

A partir de casos de estudio reales, este curso está diseñado para introducir a los estudiantes las implicaciones éticas y sociales de la inteligencia artificial (IA). La asignatura promoverá la incorporación de estrategias y el uso de herramientas para minimizar los riesgos éticos y desarrollar sistemas de IA en el marco de una IA responsable.

Competencias

- Actuar con responsabilidad ética y con respeto por los derechos y deberes fundamentales, la diversidad y los valores democráticos.
- Actuar en el ámbito de conocimiento propio evaluando las desigualdades por razón de sexo/género.
- Comunicarse de manera efectiva, tanto de forma oral como escrita, utilizando adecuadamente los recursos comunicativos necesarios y adaptándose a las características de la situación y de la audiencia.

- Concebir, diseñar, analizar e implementar agentes y sistemas ciber-físicos autónomos capaces de interactuar con otros agentes y/o personas en entornos abiertos, teniendo en cuenta las demandas y necesidades colectivas.
- Desarrollar pensamiento crítico para analizar de forma fundamentada y argumentada alternativas y propuestas tanto propias como ajenas.
- Identificar, analizar y evaluar el impacto ético y social, el contexto humano y cultural, y las implicaciones legales del desarrollo de aplicaciones de inteligencia artificial y de manipulación de datos en diferentes ámbitos.
- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Trabajar de forma autónoma, con responsabilidad e iniciativa, planificando y gestionando el tiempo y los recursos disponibles, adaptándose a las situaciones imprevistas.

Resultados de aprendizaje

1. Analizar críticamente los principios, valores y procedimientos que rigen el ejercicio de la profesión.
2. Analizar las desigualdades por razón de sexo/género y los sesgos de género en el ámbito de conocimiento propio.
3. Comprender las implicaciones sociales, éticas y legales de la práctica profesional en IA.
4. Comunicarse de manera efectiva, tanto de forma oral como escrita, utilizando adecuadamente los recursos comunicativos necesarios y adaptándose a las características de la situación y de la audiencia.
5. Desarrollar pensamiento crítico para analizar de forma fundamentada y argumentada alternativas y propuestas tanto propias como ajenas.
6. Explicar el código deontológico, explícito o implícito, del ámbito de conocimiento propio.
7. Identificar las principales desigualdades y discriminaciones por razón de sexo/géneros presentes en la sociedad.
8. Identificar los sesgos sociales, culturales y económicos de los algoritmos.
9. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
10. Saber analizar casos de aplicación de la IA desde un punto de vista ético, legal y social.
11. Saber trabajar en equipo en el diseño proyectos interdisciplinares. Ser capaz de colaborar con no profesionales y profesionales de otros sectores.
12. Ser capaz de incorporar los principios de la investigación e innovación responsable en los desarrollos basados en la IA.
13. Ser capaz de incorporar valores adecuados a las necesidades de las personas en el diseño de dispositivos dotados de IA.
14. Trabajar de forma autónoma, con responsabilidad e iniciativa, planificando y gestionando el tiempo y los recursos disponibles, adaptándose a las situaciones imprevistas.
15. Valorar cómo los estereotipos y los roles de género inciden en el ejercicio profesional.
16. Valorar las dificultades, los prejuicios y las discriminaciones que pueden incluir las acciones o proyectos, a corto o largo plazo, en relación con determinadas personas o colectivos.

Contenido

Parte I. Aspectos éticos de la inteligencia artificial

1. Introducción: ¿Por qué los profesionales de la IA deberían estudiar ética?

1.1. Código Ético y de Conducta Profesional de la ACM

1.2. Marcos éticos (consecuencialismo, teoría de la justicia, ética de la virtud...)

- 1.3. Principios éticos (equidad, responsabilidad, justicia, privacidad...)
 - 2. Recogida de datos y privacidad
 - 2.1. La importancia de la privacidad
 - 2.2. Principales técnicas para la privacidad de datos (anonimato, cifrado, privacidad diferencial...)
 - 2.3. Privacidad más allá de los datos (en el contexto, por diseño...)
 - 3. Algoritmos, toma de decisiones y sesgos
 - 3.1. Definiciones técnicas de sesgo en resultados algorítmicos
 - 3.2. Discriminación algorítmica directa e indirecta
 - 3.3. Definición de equidad y métricas de equidad
 - 3.4. Representación del conocimiento normativo y ético en IA
 - 3.5. Directrices éticas para una IA fiable: AI-Fairness Toolkits
 - 4. Explicabilidad
 - 4.1. El impacto sobre la responsabilidad y la rendición de cuentas en los sistemas autónomos, centrándonos en el caso de los vehículos autónomos
 - 4.2. La importancia de las buenas explicaciones en los sistemas de IA
 - 4.3. Herramientas para evaluar la explicabilidad
- Parte II: Aspectos ético-políticos de la inteligencia artificial
- 5. Introducción: ¿Por qué son relevantes los aspectos políticos y sociales de la inteligencia artificial?
 - 5.1. Teoría de la mediación tecnológica
 - 5.2. Narrativa en torno a la inteligencia artificial y determinismo tecnológico
 - 5.3. Innovación e investigación responsable (RRI)
 - 6. Ética y robótica
 - 6.1. Robots y sociedad
 - 6.2. Retos éticos en robótica
 - 6.3. Ejemplos aplicados de robótica en el ámbito cotidiano

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Asistencia a clase y participación activa	30	1,2	3, 4, 5, 9, 10, 12, 13, 14
Casos de estudios	50	2	1, 3, 5, 8, 7, 12, 15

La orientación del curso es predominantemente práctica. En general, cada clase comenzará con la presentación de un caso de estudio real, que dará lugar a una discusión grupal. A continuación, se introducirán y explicarán los conceptos, métodos o sistemas de IA relacionados con los retos éticos planteados por el caso de estudio. Por último, el alumnado realizará prácticas individuales o grupales para reforzar el aprendizaje del contenido de la clase. En algunas sesiones se reserva tiempo para repasar y corregir estas prácticas. Algunas clases consistirán en visitas a centros de investigación en IA.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Práctica evaluativa 1	34%	7	0,28	1, 2, 3, 4, 5, 6, 8, 7, 9, 10, 12, 13, 14, 15, 16
Práctica evaluativa 2	33%	7	0,28	1, 3, 4, 5, 6, 8, 7, 9, 10, 14, 16
Práctica evaluativa 3	33%	6	0,24	4, 5, 9, 11, 12, 13, 14

La evaluación de la asignatura se puede llevar a cabo mediante dos modalidades: evaluación continua o evaluación única.

Evaluación continua

El estudiante deberá completar individualmente tres actividades de evaluación: dos correspondientes a la Parte I y una a la Parte II.

Actividad 1: prueba escrita en clase con preguntas cortas sobre la Parte I y análisis de un caso real de aplicación de un sistema de IA.

Actividad 2: uso de kits de herramientas de IA para evaluar y discutir métricas éticas de un sistema de IA.

Actividad 3: prueba escrita en clase sobre la Parte II

Para poder ser evaluado mediante esta modalidad, es necesario haber realizado las tres actividades. Para superar la asignatura, se debe obtener una nota mínima de 5 sobre 10 en cada una de ellas. En caso contrario, el estudiante deberá presentarse a la recuperación. La nota final será la media aritmética de las tres actividades de evaluación.

Evaluación única

El estudiante realizará un examen final individual en enero, estructurado en tres partes equivalentes a las actividades de la evaluación continua, tanto por el tipo de prueba como por su peso. Para aprobar, se debe obtener una nota mínima de 5 sobre 10 en cada una de las tres partes. La nota final será la media aritmética de las tres calificaciones.

Recuperación

Podrán acceder a la recuperación:

- Estudiantes que hayan completado las tres actividades de evaluación (evaluación continua)
- Estudiantes que se hayan presentado al examen de enero (evaluación única)

La recuperación consistirá en un examen final individual. Para aprobar, se debe obtener una nota mínima de 5. Esta será la nota final de la asignatura.

Revisión de calificaciones

Después de cada actividad de evaluación, el profesorado informará a través de Moodle sobre las calificaciones obtenidas y el procedimiento y la fecha para la revisión.

Matrícula de honor

Se otorgarán a estudiantes con una nota final de 10. Si hay más estudiantes con esta nota que matrículas de honor asignadas a esta asignatura, se realizará una prueba adicional para determinar a quién se otorgan.

No evaluable

El estudiante recibirá la calificación de "No evaluable" si no se presenta a más de una actividad de evaluación (evaluación continua) o si no se presenta al examen de enero (evaluación única).

Estudiantes repetidores

No se prevé ningún tratamiento diferenciado para estudiantes repetidores.

Uso de la Inteligencia Artificial (IA)

En esta asignatura, no se permite el uso de tecnologías de IA en ninguna de sus fases. Cualquier trabajo que incluya fragmentos generados con IA será considerado una falta de honestidad académica y podrá conllevar una penalización parcial o total en la nota de la actividad, o sanciones mayores en casos graves.

Irregularidades

Cualquier irregularidad que pueda alterar significativamente la calificación de una actividad conllevará una nota de cero en dicha actividad. En caso de múltiples irregularidades, la nota final de la asignatura será cero, independientemente de cualquier proceso disciplinario.

Adaptación a formato en línea

En caso de que las pruebas o exámenes no puedan realizarse de forma presencial, se adaptarán a un formato en línea disponible a través de las herramientas virtuales de la UAB (se mantendrá la ponderación original). Los deberes, actividades y la participación en clase se realizarán mediante foros, wikis y/o discusiones en Teams, etc. El profesorado velará por que los estudiantes puedan acceder a estas herramientas virtuales o les ofrecerá alternativas viables.

Bibliografía

1. Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
2. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center for Internet & Society.
3. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
4. van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, 30(3), 385-409.
5. van Wynsberghe, A. (2013). Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics*, 19(2), 407-433.
6. Vallès-Peris, N., & Domènech, M. (2020). Roboticians' imaginaries of robots for care: The radical imaginary as a tool for an ethical discussion. *Engineering Studies*, 12(3), 156-176.
7. Vallès-Peris, N., & Domènech, M. (2021). Caring in the in-between: A proposal to introduce responsible AI and robotics to healthcare. *AI & Society*.

8. Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology & Human Values*, 31(3), 361-380.

Software

Por determinar (Parte I).

Grupos e idiomas de la asignatura

La información proporcionada es provisional hasta el 30 de noviembre de 2025. A partir de esta fecha, podrá consultar el idioma de cada grupo a través de este [enlace](#). Para acceder a la información, será necesario introducir el CÓDIGO de la asignatura

Nombre	Grupo	Idioma	Semestre	Turno
(PAUL) Prácticas de aula	711	Inglés	primer cuatrimestre	tarde
(TE) Teoría	71	Inglés	primer cuatrimestre	tarde