UAB
Universitat Autònoma de Barcelona

Ética

Código: 106559
Créditos ECTS: 6

2024/2025

Titulación	Tipo	Curso
2504392 Inteligencia Artificial / Artificial Intelligence	FB	2

Contacto

Nombre: Vicente Costa Bueno

Correo electrónico: vicente.costa@uab.cat

Equipo docente

Nuria Valles Peris

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 ético-políticos de la inteligencia artificial

1. Introducción: ¿Por qué son relevantes los aspectos políticos y sociales de la inteligencia artificial?

1.1.Teoría de la mediación tecnológica

1.2. Narrativa en torno a la inteligencia artificial y determinismo tecnológico

1.3. Innovación e investigación responsable (RRI)

2. Ética y robótica

2.1. Robots y sociedad

2.2. Retos éticos en robótica

2.3. Ejemplos aplicados de robótica en el ámbito cotidiano

Parte II: Aspectos éticos de la inteligencia artificial

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

3.1. Código Ético y de Conducta Profesional de la ACM

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

3.3. Principios éticos (equidad, responsabilidad, justicia, privacidad...)

4. Recogida de datos y privacidad

4.1. La importancia de la privacidad

4.2. Principales técnicas para la privacidad de datos (anonimato, cifrado, privacidad diferencial...)

4.3. Privacidad más allá de los datos (en el contexto, por diseño...)

5. Algoritmos, toma de decisiones y sesgos

5.1. Definiciones técnicas de sesgo en resultados algorítmicos

5.2. Discriminación algorítmica directa e indirecta

5.3. Definición de equidad y métricas de equidad

5.4. Representación del conocimiento normativo y ético en IA

5.5. Directrices éticas para una IA fiable: AI-Fairness Toolkits

6. Explicabilidad

6.1. El impacto sobre la responsabilidad y la rendición de cuentas en los sistemas autónomos, centrándonos en el caso de los vehículos autónomos

6.2. La importancia de las buenas explicaciones en los sistemas de IA

6.3. Herramientas para evaluar la explicabilidad

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Asistencia a clase y participación activa	30	1,2	10, 4, 5, 12, 13, 9, 3, 14
Casos de estudios	50	2	1, 5, 15, 7, 8, 12, 3
Prácticas y ejercicios	50	2	10, 2, 5, 16, 6, 7, 13, 9, 14

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	10, 2, 4, 1, 5, 15, 16, 6, 7, 8, 12, 13, 9, 3, 14
Práctica evaluativa 2	33%	7	0,28	10, 4, 1, 5, 16, 6, 7, 8, 9, 3, 14
Práctica evaluativa 3	33%	6	0,24	4, 5, 12, 13, 9, 11, 14

La evaluación se puede realizar de las dos maneras descritas a continuación.

Evaluación continua.
Los estudiantes deben realizar tres tareas evaluativas: una correspondiente a la Parte I y dos relacionadas con la Parte II. La tarea correspondiente a la Parte I será un examen para hacer en clase (presumiblemente en octubre). La segunda tarea será una prueba evaluativa para hacer en clase, que incluirá preguntas cortas sobre la Parte II y un análisis de un caso real de la aplicación de un sistema de IA. La tercera tarea consistirá en el uso de herramientas de IA para evaluar y discutir diferentes métricas relacionadas con la ética de un sistema de IA.
Para ser elegible para la evaluación continua, el estudiante debe haber completado las tres tareas evaluativas. Además, para aprobar la asignatura, es necesario obtener al menos una nota de 5 en las tres tareas evaluativas. De lo contrario, el estudiante tendrá que hacer la recuperación (ver la sección de Recuperación). La nota final de la asignatura en esta modalidad se determinará como la media de las tres tareas evaluativas.

Evaluación única.
El estudiante realizará un examen final en enero. El examen contendrá tres partes y, para aprobar la asignatura, el estudiante deberá obtener al menos una nota de 5 (sobre 10) en cada parte. La nota final de la asignatura en esta modalidad será la media de las notas obtenidas en cada parte del examen.

Recuperación.
Para ser elegible para la recuperación, los estudiantes deben haber completado las tres tareas evaluativas (evaluación continua) o haber hecho el examen en enero (evaluación única).
Solo se realizará un examen final de recuperación individual. Para aprobar la asignatura en esta modalidad, la notadel examen final de recuperación debe ser igual o superior a 5. La nota final será la nota del examen final de recuperación.
Si el estudiante está matriculado en la evaluación continua, solo necesitará recuperar las partes en el examen correspondientes a sus tareas evaluativas suspendidas.
Al realizar cada actividad de evaluación, los profesores informarán a los estudiantes (en Moodle) de los procedimientos que se deben seguir para revisar todas las notas otorgadas y la fecha en la que se realizará dicha revisión.
En el caso de que un estudiante cometa alguna irregularidad que pueda llevar a una variación significativa en

la nota otorgada a una actividad evaluativa, el estudiante recibirá un cero en dicha actividad, independientemente de cualquier proceso disciplinario que se pueda llevar a cabo. En caso de varias irregularidades en actividades evaluativas de la misma asignatura, el estudiante recibirá un cero como nota final de esa asignatura.
En el caso de que las pruebas o exámenes no se puedan realizar presencialmente, se adaptarán a un formato en línea disponible a través de las herramientas virtuales de la UAB (se mantendrá la ponderación original). Las tareas, actividades y la participación en clase se realizarán mediante foros, wikis y/o discusiones en Teams, etc. Los profesores se asegurarán de que los estudiantes puedan acceder a estas herramientas virtuales o les ofrecerán alternativas viables.

Bibliografía

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI." Berkman Klein Center for Internet & Society, 2020.

Mehrabi N., Morstatter F., Saxena N-, Lerman K., Galstyan A. *A Survey on Bias and Fairness in Machine Learning.* Association for Computing Machinery Surveys, (2021), 54(6)

Vallès-Peris N and Domènech M (2020) Roboticists' Imaginaries of Robots for Care: The Radical Imaginary as a Tool for an Ethical Discussion. Engineering Studies, 12 (3): 156-176.

Vallès-Peris, N., Domènech, M. (2021) Caring in the in-between: a proposal to introduce responsible AI and robotics to healthcare. AI & Society.

van de Poel, I. (2020) 'Embedding Values in Artificial Intelligence (AI) Systems', Minds and Machines, 30(3), pp. 385-409.

van Wynsberghe, A. (2013) 'Designing Robots for Care: Care Centered Value-Sensitive Design', Science and Engineering Ethics, 19(2), pp. 407-433.

Verbeek, P.-P. (2006) 'Materializing Morality: Design Ethics and Technological Mediation', Science, Technology & Human Values, 31(3), pp. 361-380.

Software

Por determinar (Parte II).

Lista de idiomas

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