Guia docente					
Datos Identificativos					2022/23
Asignatura (*)	IA Explicable y Confiable			Código	614544004
Titulación	Máster Universitario en Intelixencia Artificial				
Descriptores					
Ciclo	Periodo	Curso		Tipo	Créditos
Máster Oficial	2º cuatrimestre	Primero		Obligatoria	3
Idioma	Inglés				
Modalidad docente	Presencial				
Prerrequisitos					
Departamento					
Coordinador/a	Alvarez Estevez, Diego		Correo electrónico	diego.alvareze@udc.es	
Profesorado	Alvarez Estevez, Diego		Correo electrónico	diego.alvareze@udc.es	
Web	www.usc.gal/gl/estudos/masteres/enxenaria-arquitectura/master-universitario-intelixencia-artificial/20222023/ia-explicable-confi				
Descripción general	El objetivo principal de la materia es formar al alumnado en el desarrollo de capacidades para un adecuado tratamiento de la privacidad, confiabilidad, transparencia e interpretabilidad de modelos y resultados asociados a sistemas inteligentes. Se hará especial énfasis en identificar y analizar sesgos y su impacto en el diseño de algoritmos de Inteligencia Artificial. Además de aspectos técnicos, tecnologías disruptivas y herramientas informáticas específicas y generalistas, orientadas a cubrir todas las fases del diseño, análisis y evaluación de sistemas inteligentes, el alumnado aprenderá a conocer y comprender las implicaciones sociales y éticas de la tecnología en general y la Inteligencia Artificial en particular Guia docente centro coordinador (USC): https://www.usc.gal/gl/estudos/masteres/enxenaria-arquitectura/master-universitario-intelixencia-artificial/20222023/ia-explicable-confiable-18828-17979-2-102310				

Competencias del título	
Código	Competencias del título
A6	CE05 - Capacidad para diseñar y desarrollar sistemas inteligentes mediante la aplicación de algoritmos de inferencia, representación del conocimiento y planificación automática
A7	CE06 - Capacidad para reconocer aquellos problemas que necesiten de una arquitectura distribuida que no esté prefijada durante el diseño del sistema, que serán adecuados para la implementación de sistemas multiagente inteligentes
A8	CE07 - Capacidad para entender las implicaciones del desarrollo de un sistema inteligente explicable e interpretable
A9	CE08 - Capacidad para diseñar y desarrollar sistemas inteligentes seguros, en términos de integridad, confidencialidad y robustez
B1	CG01 - Mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo de la Inteligencia Artificial
B2	CG02 - Abordar con éxito todas las etapas de un proyecto de Inteligencia Artificial
B3	CG03 - Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo
B6	CB01 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
B7	CB02 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
B8	CB03 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
B9	CB04 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigu?edades
C2	CT02 - Dominar la expresión y la comprensión de forma oral y escrita de un idioma extranjero

C3	CT03 - Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida
C4	CT04 - Desarrollarse para el ejercicio de una ciudadanía respetuosa con la cultura democrática, los derechos humanos y la perspectiva de género
C5	CT05 - Entender la importancia de la cultura emprendedora y conocer los medios al alcance de las personas emprendedoras
C6	CT06 - Adquirir habilidades para la vida y hábitos, rutinas y estilos de vida saludables
C7	CT07 - Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social
C8	CT08 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad

Resultados de aprendizaje			
Resultados de aprendizaje	Competencias del título		
Desarrollar capacidades para un adecuado tratamiento de la privacidad, confiabilidad, transparencia e interpretabilidad de modelos y resultados	AM5 AM6 AM7 AM8	BM1 BM2 BM3 BM6 BM7 BM8 BM9	CM2 CM3 CM4 CM5 CM6 CM7 CM8
Identificar y analizar sesgos y su impacto en el diseño de algoritmos de Inteligencia Artificial	AM5 AM6 AM7 AM8	BM1 BM2 BM3 BM6 BM7 BM8 BM9	CM2 CM3 CM4 CM5 CM6 CM7 CM8
Conocer y comprender las implicaciones sociales y éticas de la tecnología en general y la Inteligencia Artificial en particular	AM5 AM6 AM7 AM8	BM1 BM2 BM3 BM6 BM7 BM8 BM9	CM2 CM3 CM4 CM5 CM6 CM7 CM8

Contenidos	
Tema	Subtema
Explicabilidad e interpretabilidad. Métodos agnósticos al modelo. Explicaciones basadas en ejemplos. FAT-E (imparcialidad, responsabilidad, transparencia y ética). Estudio y tipos de sesgos. Tipos y modelos de explicación. Metodologías de evaluación. Integridad de datos, privacidad, confidencialidad y robustez de modelos. Confiabilidad por diseño	Explicabilidad e interpretabilidad. Métodos agnósticos al modelo. Explicaciones basadas en ejemplos. FAT-E (imparcialidad, responsabilidad, transparencia y ética). Estudio y tipos de sesgos. Tipos y modelos de explicación. Metodologías de evaluación. Integridad de datos, privacidad, confidencialidad y robustez de modelos. Confiabilidad por diseño

Planificación				
Metodologías / pruebas	Competéncias	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales

Prácticas de laboratorio	A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8	11	43	54
Sesión magistral	A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8	10	10	20
Atención personalizada		1	0	1
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	**Descripción**
Prácticas de laboratorio	Las clases interactivas se desarrollarán en el Aula de Informática habilitada para ello en cada Universidad, empleando diversas herramientas software para cada uno de los bloques temáticos, abordando prácticas y proyectos con distintos niveles de complejidad. El alumnado trabajará en puestos individuales con el apoyo constante del profesorado. Los guiones de las prácticas serán auto-explicativos permitiendo la realización de los mismos en horario de trabajo personal. La realización de las prácticas permitirá desarrollar las competencias CG1, CG3, CB6, CB7, CB8, CT3, CT8, CE5, CE6, CE7, CE8, CE9. Estas clases están dedicadas a que el alumnado desarrolle trabajos prácticos que impliquen abordar la resolución de problemas complejos, y el análisis y diseño de soluciones que constituyan un medio para su resolución. Esta actividad puede requerir de los alumnos la presentación oral de los trabajos realizados. Los trabajos realizados por el alumnado se pueden realizar de forma individual o en grupos de trabajo. El alumnado puede trabajar la solución a los problemas planteados de forma individual o en grupos. Esta metodología docente se aplicará a la actividad formativa "Clases prácticas de laboratorio" y se podrá aplicar a la actividad formativa de "Sesiones de aprendizaje basado en problemas, seminarios, estudio de casos y proyectos". Prácticas de laboratorio: el profesorado de la materia plantea al alumnado un problema o problemas de carácter práctico cuya resolución requiere la comprensión y aplicación de los contenidos teórico-prácticos incluidos en los contenidos de la materia. Aprendizaje por proyectos: se plantea al alumnado proyectos prácticos cuyo alcance requiere que se le dedique una parte importante de la dedicación total del alumno a la asignatura. Además, por el alcance de los trabajos a realizar, se requiere que el alumnado aplique competencias de gestión además de competencias de índole técnica. La docencia estará apoyada por la plataforma virtual del máster de la siguiente manera: repositorio de la documentación relacionada con la materia (textos, presentaciones, etc.) y tutorización virtual del alumnado (correo-e y foros). Tutorías: el profesorado atenderá al alumnado en sesiones de tutorías individualizadas dedicadas a la orientación en el estudio y la resolución de dudas sobre los contenidos y trabajos de la asignatura

Sesión magistral	La metodología didáctica se basará en el trabajo individual del alumnado, en la discusión con el profesorado en clase y en las tutorías individuales. En las Clases de teoría (expositivas), la Exposición oral será complementada con el uso de medios audiovisuales y la introducción de algunas preguntas dirigidas a los estudiantes, con la finalidad de transmitir conocimientos y facilitar el aprendizaje. Además del tiempo de exposición oral por parte del profesor, esta actividad formativa requiere del alumno la dedicación de un tiempo para preparar y revisar por cuenta propia los materiales objeto de la clase. Para cada tema o bloque temático de las clases expositivas, el profesorado preparará los contenidos, explicará los objetivos del tema al alumnado en clase, presentará cada tema con el objetivo de facilitar un conjunto de información con alcance concreto, sugerirá bibliografía, proporcionará material de trabajo adicional, etc. Esta metodología docente se aplicará a la actividad formativa "Clases de teoría". En estas clases expositivas se trabajarán las competencias CG1, CG3, CB6, CB7, CB8, CB9, CE5, CE6, CE7, CE8, CE9. Además, el profesorado propondrá al alumnado un conjunto de actividades a realizar, de forma individual o en grupo (estudio de casos, trabajos, presentaciones, lecturas, etc.). El alumnado deberá entregar obligatoriamente una selección de ellas para su evaluación. Estas actividades permitirán desarrollar las competencias CG3, CB7, CB8, CB9, CT2, CT3, CT4, CT6, CT8, CE7, CE8

Atención personalizada	
Metodologías	**Descripción**
Prácticas de laboratorio	

Evaluación			
Metodologías	**Competéncias**	**Descripción**	**Calificación**
Sesión magistral	A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8	examen de la parte teórica (45%)	45
Prácticas de laboratorio	A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8	evaluación de las entregas asociadas a las sesiones interactivas (35%), la entrega de un trabajo personal y la presentación oral del mismo (15%) y la evaluación continua de cada estudiante a lo largo del curso (5%)	55

Observaciones evaluación

La evaluación del aprendizaje considera tanto un examen de la parte teórica (45%) como la evaluación de las entregas asociadas a las sesiones interactivas (35%), la entrega de un trabajo personal y la presentación oral del mismo (15%) y la evaluación continua de cada estudiante a lo largo del curso (5%).

Será requisito indispensable aprobar todas las partes (expositiva, interactiva, trabajo, evaluación continua), considerando los siguientes criterios:

1. Examen (45%): la parte teórica de la asignatura se evaluará en un único examen a realizar en la fecha oficial, que constará de preguntas relacionadas con todos los temas del programa. El examen estará orientado especialmente a evaluar la comprensión de los conocimientos expuestos en las clases de teoría. La calificación del examen será la media ponderada de los módulos de la asignatura, que sólo se calculará en el caso de tener calificación igual o superior a 4 en cada módulo.

2. Entregas interactivas (35%): habrá entregas obligatorias asociadas a las sesiones interactivas relacionadas con cada módulo teórico. Se evaluarán las soluciones propuestas por el alumnado a las prácticas planteadas. La evaluación de prácticas puede llevarse a cabo mediante una corrección por parte del profesor, una defensa de la solución aportada por parte del alumno ante el profesor o una presentación oral de la solución desarrollada. (Aplicable a los resultados de las actividades formativas "Clases prácticas de laboratorio", "Aprendizaje basado en problemas, seminarios, estudio de casos y proyectos" y "Realización de trabajos tutelados"). La nota media sólo se calculará en el caso de tener calificación superior o igual a 4/10 en todas las entregas. Además, es obligatoria la asistencia presencial al menos al 60% de las clases interactivas.

3. Trabajo (15%): el alumnado deberá entregar un trabajo personal y hacer la presentación oral del mismo según el calendario que se establezca al inicio del cuatrimestre. La evaluación del trabajo tutelado se llevará a cabo mediante una defensa en la que el alumnado explica su propuesta y conclusiones ante el profesorado, o mediante una presentación oral de la solución ante el aula. La calificación obtenida será la media de la evaluación del trabajo escrito y su presentación oral. Sólo se realizará la media si se obtiene una nota igual o superior a 4 en cada parte.

4. Evaluación continua (5%): Se tendrá en cuenta la asistencia y participación activa del alumnado tanto en las clases expositivas como en la presentación de trabajos, discusiones, seminarios, y en las sesiones interactivas que se celebren a lo largo del curso. Es obligatoria

la asistencia al menos al 60% de las sesiones de presentación de trabajos y seminarios.

La calificación final de la materia será la suma de las cuatro calificaciones parciales, excepto en aquellas situaciones indicadas anteriormente. Cuando no se supere alguna de las partes, la calificación final de la oportunidad será el mínimo de las calificaciones parciales.

Obtendrá la calificación de no presentado el alumnado que no haya participado en ninguna de las actividades de evaluación.

El alumnado que tenga exención oficial de asistencia a clase deberá realizar, en todo caso, el examen final escrito, así como todas las entregas de prácticas y trabajos que se establezcan como obligatorios a lo largo del curso y, en su caso, realizar la presentación oral de los mismos. En esta modalidad, la tutorización y las entregas serán virtuales y las presentaciones podrán realizarse de forma telepresencial.

En la segunda oportunidad, el alumnado deberá superar las actividades de evaluación pendientes de la primera oportunidad, de acuerdo con los criterios anteriores.

Para los casos de realización fraudulenta de ejercicios o pruebas será de aplicación lo recogido en la Normativa de evaluación del rendimiento académico del alumnado y de revisión de calificaciones. La copia total o parcial de algún ejercicio de prácticas o teoría supondrá automáticamente una calificación de 0.0 en la asignatura y oportunidad

Fuentes de información	
Básica	Aportaranse notas ou material específico na aula virtual para seguir a materia. Dada a heteroxeneidade dos temas a tratar na materia, con cada un dos temas achegaranse referencias a recursos bibliográficos e outro tipo de contidos (titoriais, multimedia, etc.) para os aspectos máis específicos da materia. As seguintes referencias son de tipo complementario, tratan aspectos xerais relacionados coa IA explicable e fiable. 1. V. Dignum. Responsible Artificial Intelligence. How to Developand Use AI in a Responsible Way. Springer Nature Switzerland AG, 2019, ISBN: 978-3-030-30370-9 , https://doi.org/10.1007/978-3-030-30371-6 2. A. Barredo Arrieta et al., Explainable Artificial Intelligence(XAI): Concepts, taxonomies, opportunities and challenges toward responsibleAI, Information Fusion, 58:82-115, Elsevier 2020, https://doi.org/10.1016/j.inffus.2019.12.012 3. T. Miller, Explanation in artificial intelligence: Insights fromthe social sciences. Artificial Intelligence, 267:1-38, Elsevier 2019, https://doi.org/10.1016/j.artint.2018.07.007 4. G. Vilone, L. Longo, Notions of explainability and evaluationapproaches for Explainable Artificial Intelligence, Information Fusion,76:89-106, Elsevier 2021, https://doi.org/10.1016/j.inffus.2021.05.009 5. R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti,D. Pedreschi, A Survey of Methods for Explaining Black Box Models, ACMComputing Surveys, 51(5):1?42, 2019, https://dl.acm.org/doi/10.1145/3236009 6. J.M. Alonso, C. Castiello, L. Magdalena, C.Mencar, Explainable Fuzzy Systems. Paving the way from interpretable fuzzysystems to explainable AI systems. SpringerInternational Publishing, 2021, ISBN: 978-3-030-71098-9, https://doi.org/10.1007/978-3-030-71098-9
Complementária	

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Asignaturas que se recomienda cursar simultáneamente
Asignaturas que continúan el temario
Otros comentarios
Se recomienda llevar la asignatura al día y el uso de tutorías para aclarar dudas y asesorar en su desarrollo. Además, se recomienda que el alumnado resuelva, verifique y valide todos los ejercicios y prácticas propuestos a lo largo del curso (no solamente los evaluables)