



| Guía Docente | | | | |
|-----------------------|---|--------------------|-----------------------|-----------|
| Datos Identificativos | | | | 2022/23 |
| Asignatura (*) | IA Explicable e Confiable | | Código | 614544004 |
| Titulación | Máster Universitario en Intelixencia Artificial | | | |
| Descritores | | | | |
| Ciclo | Período | Curso | Tipo | Créditos |
| Mestrado Oficial | 2º cuatrimestre | Primeiro | Obrigatoria | 3 |
| Idioma | Inglés | | | |
| Modalidade docente | Presencial | | | |
| Prerrequisitos | | | | |
| Departamento | | | | |
| Coordinación | 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 | | | |
| Descrición xeral | <p>O obxectivo principal da materia é formar ao alumnado no desenvolvemento de habilidades para un tratamento adecuado da privacidade, fiabilidade, transparencia e interpretabilidade dos modelos e resultados asociados a sistemas intelixentes. Farase especial fincapé na identificación e análise de sesgos e o seu impacto no deseño de algoritmos de Intelixencia Artificial. Ademais dos aspectos técnicos, tecnoloxías disruptivas e ferramentas informáticas específicas e xerais, dirixidas a cubrir todas as fases do deseño, análise e avaliación de sistemas intelixentes, o alumnado aprenderá a coñecer e comprender as implicacións sociais e éticas da tecnoloxía en xeral e da Intelixencia Artificial en particular</p> <p>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</p> | | | |

| Competencias do título | |
|------------------------|--|
| Código | Competencias do título |
| A6 | CE05 - capacidade para deseñar e desenvolver sistemas intelixentes mediante a aplicación de algoritmos de inferencia, representación do coñecemento e planificación automática |
| A7 | CE06 - capacidade para recoñecer aqueles problemas que necesiten dunha arquitectura distribuída que non estea prefixada durante o deseño do sistema, que serán axeitados para a implementación de sistemas multiaxe intelixentes |
| A8 | CE07 - capacidade para entender as implicacións do desenrolo dun sistema intelixente explicable e interpretable |
| A9 | CE08 - capacidade para deseñar e desenvolver sistemas intelixentes seguros, en termos de integridade, confidencialidade e robustez |
| B1 | CG01 - Manter e estender os plantexamentos teóricos fundados para permitir a introducción e explotación de tecnoloxías novas e avanzadas no eido da Intelixencia Artificial |
| B2 | CG02 - Abordar con éxito todas as etapas dun proxecto de Intelixencia Artificial |
| B3 | CG03 - Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo |
| B6 | CB01 - Poseer e comprender coñecementos que aporten unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación |
| B7 | CB02 - Que os estudantes saiban aplicar os coñecementos adquiridos e posúan capacidade de resolución de problemas en entornos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa su área de estudo |
| B8 | CB03 - Que os estudantes sexan capaces de integrar coñecementos e enfrentarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos |
| B9 | CB04 - Que os estudantes saiban comunicar as súas conclusións e os coñecementos e razóns últimas que as sustentan a públicos especializados e non especializados dun xeito claro e sen ambigüidades |
| C2 | CT02 - Dominar a expresión e comprensión, de xeito oral e escrito, dun idioma estranxeiro |



| | |
|----|---|
| C3 | CT03 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida |
| C4 | CT04 - Desenvolverse para o exercicio dunha cidadanía respetuosa coa cultura democrática, os dereitos humanos e a perspectiva de xénero |
| C5 | CT05 - Entender a importancia da cultura emprendedora e coñecer os medios ao alcance das persoas emprendedoras |
| C6 | CT06 - Adquirir habilidades para a vida e hábitos, rutinas e estilos de vida saudables |
| C7 | CT07 - Desenvolver a capacidade de traballar en equipos interdisciplinares ou transdisciplinares, para ofrecer propostas que contribúan a un desenvolvemento sostible ambiental, económico, político e social |
| C8 | CT08 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade |

| Resultados da aprendizaxe | | | | |
|--|--|------------------------|-----|-----|
| Resultados de aprendizaxe | | Competencias do título | | |
| Desenvolver capacidades para un adecuado tratamento da privacidade, confiabilidade, transparencia e interpretabilidade de modelos e resultados | AM5 | BM1 | CM2 | |
| | AM6 | BM2 | CM3 | |
| | AM7 | BM3 | CM4 | |
| | AM8 | BM6 | CM5 | |
| | | BM7 | CM6 | |
| | | BM8 | CM7 | |
| | Identificar e analizar rumbos e o seu impacto no deseño de algoritmos de Intelixencia Artificial | AM5 | BM1 | CM2 |
| | | AM6 | BM2 | CM3 |
| | | AM7 | BM3 | CM4 |
| AM8 | | BM6 | CM5 | |
| | | BM7 | CM6 | |
| | | BM8 | CM7 | |
| Coñecer e comprender as implicacións sociais e éticas da tecnoloxía en xeral e a Intelixencia Artificial en particular | | AM5 | BM1 | CM2 |
| | | AM6 | BM2 | CM3 |
| | | AM7 | BM3 | CM4 |
| | AM8 | BM6 | CM5 | |
| | | BM7 | CM6 | |
| | | BM8 | CM7 | |
| | AM8 | BM9 | CM8 | |

| Contidos | |
|--|--|
| Temas | Subtemas |
| Explicabilidade e interpretabilidade. Métodos agnósticos de modelos. Explicacións a partir de exemplos. FAT-E (xusto, rendición de contas, transparencia e ética). Estudo e tipos de prexuízos. Tipos e modelos de explicación. Metodoloxías de avaliación. Integridade dos datos, privacidade, confidencialidade e robustez de modelos. Fiabilidade polo deseño | Explicabilidade e interpretabilidade. Métodos agnósticos de modelos. Explicacións a partir de exemplos. FAT-E (xusto, rendición de contas, transparencia e ética). Estudo e tipos de prexuízos. Tipos e modelos de explicación. Metodoloxías de avaliación. Integridade dos datos, privacidade, confidencialidade e robustez de modelos. Fiabilidade polo deseño |

| Planificación | | | | |
|-----------------------|--------------|-------------------|---|--------------|
| Metodoloxías / probas | Competencias | Horas presenciais | Horas non presenciais / traballo autónomo | Horas totais |
| | | | | |



| | | | | |
|--|---|----|----|----|
| 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 maxistral | 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 |
| *Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado | | | | |

| Metodoloxías | |
|--------------------------|--|
| Metodoloxías | Descrición |
| Prácticas de laboratorio | <p>As clases interactivas desenvolveranse na Aula de Informática habilitada para iso en cada Universidade, utilizando diversas ferramentas informáticas para cada un dos bloques temáticos, abordando prácticas e proxectos con diferentes niveis de complexidade.</p> <p>Estas clases están dedicadas a que o alumnado desenvolva traballos prácticos que impliquen abordar a resolución de problemas complexos, así como a análise e deseño de solucións que constitúan un medio para a súa resolución. Esta actividade pode requirir que os alumnos expoñan o seu traballo oralmente. O traballo realizado polo alumnado pode realizarse individualmente ou en grupos de traballo.</p> <p>O alumnado traballará en postos individuais co apoio constante do profesorado. Os guións das prácticas serán autoexplicativos, permitindo a súa realización en horario persoal. A realización das prácticas permitirá o desenvolvemento de competencias CG1, CG3, CB6, CB7, CB8, CT3, CT8, CE5, CE6, CE7, CE8, CE9.</p> <p>O alumnado pode traballar a solución dos problemas plantexados individualmente ou en grupo. Esta metodoloxía docente aplicarase á actividade formativa "Clases prácticas de laboratorio" e poderá aplicarse á actividade formativa de "Sesións de aprendizaxe baseada en problemas, seminarios, estudos de casos e proxectos".</p> <p>Prácticas de laboratorio: o profesorado da materia plantexa ao alumnado un problema ou problemas de carácter práctico cuxa resolución esixe a comprensión e aplicación dos contidos teórico-prácticos incluídos na materia.</p> <p>Aprendizaxe por proxectos: preséntanse ao alumnado proxectos prácticos cuxo alcance require unha parte importante da dedicación total do estudante á materia. Ademais, debido ao alcance do traballo que se vai realizar, requírese que o alumnado aplique habilidades de xestión, así como habilidades técnicas.</p> <p>A docencia apoiarase na plataforma virtual do máster da seguinte forma: repositorio de documentación relacionada coa materia (textos, presentacións, etc.) e titoría virtual do alumnado (correo electrónico e foros).</p> <p>Titorías: o profesorado asistirá ao alumnado en titorías individualizadas dedicadas á orientación no estudo e á resolución de dúbidas sobre os contidos e traballos da materia</p> |



| | |
|------------------|---|
| Sesión maxistral | <p>A metodoloxía didáctica basearase no traballo individual do alumnado, na discusión co profesorado na clase e nas titorías individuais.</p> <p>Clases teóricas (expositivas): Exposición oral complementada co uso de medios audiovisuais e a introdución dalgunhas preguntas dirixidas ao alumnado, co fin de transmitir coñecementos e facilitar a aprendizaxe. Ademais do tempo de exposición oral por parte do profesor, esta actividade formativa require que o alumno dedique un tempo para preparar e revisar os materiais da clase por si mesmo.</p> <p>Para cada tema ou bloque temático das clases expositivas, o profesorado elaborará os contidos, explicará os obxectivos da temática ao alumnado na clase, exporá cada tema co obxectivo de achegar un conxunto de información cun alcance específico, suxerirá unha bibliografía, proporcionará material de traballo adicional, etc. Esta metodoloxía docente aplicarase á actividade formativa ?Clases teóricas?.</p> <p>Nestas clases expositivas, o alumnado desenvolverá as habilidades CG1, CG3, CB6, CB7, CB8, CB9, CE5, CE6, CE7, CE8, CE9. Ademais, o profesorado proporalle ao alumnado un conxunto de actividades para realizar, individualmente ou en grupo (casos prácticos, traballos, exposicións, lecturas, etc.). O alumnado deberá presentar unha selección deles para a súa avaliación. Estas actividades axudarán a desenvolver habilidades CG3, CB7, CB8, CB9, CT2, CT3, CT4, CT6, CT8, CE7, CE8</p> |
|------------------|---|

Atención personalizada

| Metodoloxías | Descrición |
|--------------------------|------------|
| Prácticas de laboratorio | |

Avaliación

| Metodoloxías | Competencias | Descrición | Cualificación |
|--------------------------|---|---|---------------|
| Sesión maxistral | A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8 | Examen teórico | 45 |
| Prácticas de laboratorio | A6 A7 A8 A9 B1 B2 B3 B6 B7 B8 B9 C2 C3 C4 C5 C6 C7 C8 | Sesións interactivas (35%), entrega de traballo persoal e exposición do mesmo (15%), e avaliación continua ao longo do curso (5%) | 55 |

Observacións avaliación



A avaliación da aprendizaxe contempla tanto un exame da parte teórica (45%) como a avaliación das entregas asociadas ás sesións interactivas (35%), a entrega dun traballo persoal e a exposición oral do mesmo (15%), e a avaliación continua do alumnado ao longo do curso (5%).

Será requisito indispensable a superación de todas as partes (exame, interactivas, traballo, avaliación continua), tendo en conta os seguintes criterios:

1. Exame (45%): A parte teórica da materia avaliarase nun único exame que se realizará na data oficial, que constará de preguntas relacionadas con todos os temas do programa. O exame estará especialmente orientado a avaliar a comprensión dos coñecementos expostos nas clases teóricas. A nota do exame será a media ponderada dos módulos da materia, que só se computará no caso de ter unha nota igual ou superior a 4 en cada módulo.

2. Interactivas (35%): Haberá entregas obrigatorias asociadas ás sesións interactivas relacionadas con cada módulo teórico. Avaliaranse as solucións propostas polo alumnado ás prácticas propostas. A avaliación das prácticas pode realizarse mediante unha corrección por parte do profesor, unha defensa da solución aportada polo alumno ante o profesor ou unha exposición oral da solución desenvolvida. (Aplicable aos resultados das actividades formativas "Clases prácticas de laboratorio", "Aprendizaxe por problemas, seminarios, estudos de casos e proxectos" e "Realización de traballos tutelados"). A nota media só se computará no caso de ter unha nota superior ou igual a 4/10 en todas as entregas. Ademais, é obrigatoria a asistencia presencial a polo menos o 60% das clases interactivas.

3. Traballo (15%): O alumnado deberá presentar un traballo persoal e realizar unha exposición oral do mesmo segundo o calendario establecido ao comezo do cuadrimestre. A avaliación do traballo tutelado realizarase mediante unha defensa na que o alumnado expoña ao profesor a súa proposta e conclusións, ou mediante unha exposición oral da solución diante da aula. A cualificación obtida será a media da avaliación do traballo escrito e da súa exposición oral. A media só se realizará se en cada parte se obtén unha nota igual ou superior a 4.

4. Avaliación continua (5%): Terase en conta a asistencia e participación activa do alumnado tanto nas clases expositivas como na presentación de traballos, debates, seminarios, como nas sesións interactivas que se realicen ao longo do curso. É obrigatoria a asistencia a polo menos o 60% das sesións de presentación e seminarios.



A cualificación final da materia será a suma das catro cualificacións parciais, agás nos supostos sinalados anteriormente. Cando non se supere algunha parte, a cualificación final da oportunidade será a mínima das cualificacións parciais.

O alumnado que non participase en ningunha das actividades de avaliación obterá a cualificación de non presentado.

O alumnado que teña exención oficial de asistencia a clase deberá realizar, en todo caso, o exame final escrito, así como todas as entregas de prácticas e traballos que se establezan como obrigatorias ao longo do curso e, no seu caso, realizar a exposición oralmente a partir delas. Nesta modalidade, a titoría e as entregas serán virtuais e as presentacións poderán facerse en telepresencia.

Na segunda oportunidade, o alumnado deberá superar as actividades de avaliación pendentes da primeira oportunidade, de acordo cos criterios anteriores.

Para os casos de realización fraudulenta de exercicios ou probas será de aplicación o disposto na Normativa de avaliación do rendemento académico do alumnado e revisión das cualificacións. A copia total ou parcial de calquera exercicio práctico ou teórico suporá automaticamente unha nota de 0,0 na materia e oportunidade



Fontes de información

| | |
|------------------------------------|--|
| Bibliografía 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 Develop and 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 responsible AI, 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 from the 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 evaluation approaches 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, ACM Computing 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 fuzzy systems to explainable AI systems. Springer International Publishing, 2021, ISBN: 978-3-030-71098-9, https://doi.org/10.1007/978-3-030-71098-9 |
| Bibliografía complementaria | |

Recomendacións

Materias que se recomenda ter cursado previamente

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións

Recoméndase

levar a materia ao día e o uso de titorías para aclarar dúbidas e asesorar no seu desenvolvemento. Ademais, recoméndase que o alumnado resolva, verifique e valide todos os exercicios e prácticas propostos ao longo do curso (non só os avaluables)

(*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente de acordo coa normativa vixente que establece o proceso de elaboración de guías