

IA i/o ÈTICA. Dilemes d'un futur present

Humberto Ruiz Gil

Introducció

El curs explora la intersecció entre la Intel·ligència Artificial (IA) i l'Ètica. Al llarg del curs, els participants obtindran una comprensió profunda de l'ètica com a fonamentació de les morals, que possibilitarà un apropament crític als dilemes que ens planten les situacions o trobades amb la IA. S'abordaran temes com l'ètica en la recopilació de dades, el desenvolupament de models d'IA, la presa de decisions automatitzades, l'automatització laboral, la salut, la seguretat, la justícia, l'educació, la creativitat, la política i el futur de la IA. A més, s'analitzaran casos concrets i es fomentarà la discussió per desenvolupar una perspectiva informada i reflexiva sobre els desafiaments ètics al camp de la IA i el seu impacte a la societat. El curs està dissenyat per fomentar la reflexió crítica i promoure la presa de decisions ètiques a l'ús de la IA.

Objectius

- Familiaritzar-se amb els conceptes bàsics de la intel·ligència artificial i la seva aplicació a diversos camps.
- Reconèixer i comprendre els desafiaments morals i ètics associats amb el desenvolupament, implementació i ús de la intel·ligència artificial.
- Avaluar l'impacte social i ètic de la intel·ligència artificial en àrees com l'ocupació, la salut, l'educació o la justícia entre d'altres.
- Estudiar casos d'estudi reals que il·lustren dilemes ètics en el camp de la intel·ligència artificial i participar-hi en discussions crítiques.

Programa

Sessió 1: Introducció a l'IA i l'Ètica (2 hores)

- Definició d'IA i la seva rellevància ètica.
- Ètica, moral i altres conceptes de filosofia pràctica.

Sessió 2: Ètica a la Recopilació de Dades (2 hores)

- Privadesa i consentiment informat.
- Biaixos en les dades i les seves implicacions ètiques.

Sessió 3: Ètica al Desenvolupament de Models d'IA (2 hores)

- Biaixos algorítmics.
- *Fairness* a la IA.

Sessió 4: Ètica a la Presa de Decisions d'IA (2 hores)

- Ètica a la presa de decisions autònomes.
- Responsabilitat i rendició de comptes a la IA.

Sessió 5: Ètica a l'Automatització de Treball (2 hores)

- Impacte laboral de la IA.
- Consideracions ètiques a l'automatització.

Sessió 6: Ètica a la IA i la Salut (2 hores)

- Ètica a l'atenció mèdica assistida per IA.
- Privadesa de les dades de salut.

Sessió 7: Ètica a la IA i la Seguretat (2 hores)

- Ciberseguretat i ètica.
- Armes autònomes i ètica.

Sessió 8: Ètica a la IA i la Justícia (2 hores)

- Justícia algorítmica.
- Biaixos a la justícia i l'aplicació de la llei.

Sessió 9: Ètica a la IA i l'Educació (2 hores)

- Ètica a l'educació assistida per IA.
- Biaixos en l'aprenentatge automatitzat.

Sessió 10: Ètica a la IA i la Creativitat (2 hores)

- Generació de contingut per IA i propietat intel·lectual.
- Implicacions ètiques a l'art i la creativitat.

Sessió 11: Ètica a la IA i la Política (2 hores)

- Influència de la IA a la política i la democràcia.
- Manipulació de l'opinió pública.

Sessió 12: Ètica a la IA i els Assistents personals (2 hores)

- La convivència domèstica amb les IA.
- Drets i afectes

Sessió 13: Ètica i Regulació de la IA (2 hores)

- Normatives i marcs legals.
- Desafiaments en la regulació de la IA.

Sessió 14: Casos d'estudi i discussió (2 hores)

- Anàlisi de casos reals de dilemes ètics a la IA.
- Debat i discussió en grup.

Sessió 15: La IA com a singularitat tecnocientífica (2 hores)

- Superintel·ligència i control ètic.
- Ètica a l'exploració espacial i la colonització.

Metodologia

La metodologia de les sessions de l'assignatura combina:

- Part expositiva: Cada sessió començarà amb una part expositiva, en la qual es presentaran els conceptes clau i les teories fonamentals relacionades amb el tema de la sessió. S'utilitzarà un llenguatge formal i es proporcionaran exemples clars i rellevants per a facilitar la comprensió dels estudiants.
- Materials multimèdia: Per a complementar la part expositiva i enriquir l'experiència d'aprenentatge, s'utilitzaran materials multimèdia, com a imatges, gràfics, vídeos i animacions. Aquests recursos visuals ajudaran a visualitzar els conceptes abstractes i a proporcionar exemples pràctics relacionats amb els temes tractats.
- Debat i discussió: Després de la part expositiva, es fomentarà la participació dels estudiants a través de debats i discussions. Es plantejaran preguntes i es donarà espai perquè els estudiants comparteixin les seves opinions, plantegin dubtes i generin un diàleg enriquidor. Això permetrà aprofundir en els temes, explorar

diferents perspectives i promoure el pensament crític.

- Anàlisi de casos i estudis de recerca: En algunes sessions, es presentaran casos d'estudi i estudis de recerca rellevants per a il·lustrar l'aplicació pràctica dels conceptes teòrics. S'analitzaran els resultats i les implicacions d'aquests casos, estimulant així la capacitat dels estudiants per a connectar la teoria amb la pràctica.
- Lectures complementàries i referències bibliogràfiques: Es proporcionaran lectures complementàries i referències bibliogràfiques perquè els estudiants aprofundeixin en els temes tractats en cada sessió. Aquestes lectures els permetran explorar diferents enfocaments, perspectives i estudis rellevants, i ampliar la seva comprensió del tema.

La combinació d'aquestes metodologies, que inclouen la part expositiva, l'ús de materials multimèdia, el debat i la discussió, així com l'anàlisi de casos i estudis de recerca, garantirà una experiència d'aprenentatge dinàmica, interactiva i enriquidora per als estudiants.

Bibliografia

- "Ética de la inteligencia artificial" Adela Cortina. Editorial Dykinson. 2019
- "Artificial Intelligence: A Modern Approach" Peter Norvig y Stuart Russell. Athenaeum Uitgeverij. 2021
- Ética en la Inteligencia Artificial y en los Sistemas Basados en Datos.Fuente: Grupo de Expertos de la Comisión Europea sobre Ética en la Inteligencia Artificial (AI HLEG). Fecha: Publicado en abril de 2019.
- Reglamento de la Comisión Europea sobre la Inteligencia Artificial. Fuente: Comisión Europea. Fecha: Publicado en abril de 2021.
- Ética y Derechos Fundamentales en la Inteligencia Artificial: Implicaciones Jurídicas. Fuente: Parlamento Europeo. Fecha: Aprobado en febrero de 2021.