Universidad de Buenos Aires
FACULTAD DE FILOSOFÍA Y LETRAS

SEMINARIO del Programa de Actualización de Posgrado "Inteligencia artificial desde una perspectiva humanística"

Desafíos éticos de la IA

Docente a cargo: Dr. Tomás Balmaceda

Carga horaria: 16hs.

Cuatrimestre, año: 1ero, 2024

Fundamentación

La tecnología acompaña a hombres y mujeres desde los inicios de los tiempos, incluso desde antes que existieran los primeros indicios de lo que hoy llamaríamos civilización. Es motivo de debate entender cuándo fue que se creó el primer artefacto tecnológico pero podríamos pensar que todo comenzó en el momento mismo en que un homínido le imprimió su intención a una rama para volverla un bastón que lo ayudara a caminar o cuando afiló una roca para usarla para cazar a una presa. Desde ese instante, el destino de la especie y el de los productos que fue creando o utilizando a partir de elementos naturales quedó sellado en una unión inseparable.

Esto generó, desde ese momento, responsabilidades. Sin embargo, la ubicuidad y potencia de las tecnologías que conocemos como Inteligencia Artificial (IA) a veces parecen escapar de esas reflexiones. O al menos se cree que son desarrollos cualitativamente diferentes a otros. Pero nunca antes en la historia de la humanidad una tecnología reciente tuvo la capacidad de afectar a tantas personas y de manera tan profunda,

por lo que se impone un espacio de reflexión crítica. Este seminario busca ahondar en algunos de los desafíos éticos de la IA, buscando primero un posible marco filosófico para pensar acerca del diseño de artefactos y de sistemas tecnológicos para, luego, enfocarnos en la responsabilidad, la privacidad, los sesgos y otros debates éticos porque, sin la reflexión filosófica, muchos aspectos relevantes permanecen ocultos en la "caja negra" del sistema tecnológico, afectando el pasado y el presente de la educación.

Objetivos

- Brindar una presentación actualizada y crítica del mapa actual de debates éticos alrededor de los sistemas que involucran IA desde una perspectiva filosófica.
- Generar una actitud reflexiva en las y los estudiantes en relación con promover decisiones informadas acerca del impacto de herramientas que involucran IA.
- Propiciar que las y los estudiantes puedan abordar críticamente los temas tratados, respetando cánones de claridad expositiva y argumentativa.
- Vincular los temas de ética con los contenidos vistos en seminarios anteriores y preparar el terreno para los debates en seminarios posteriores

UNIDAD 1: RESPONSABILIDAD

Contenidos:

Tecnología y valores. “Máquinas arresponsables” y decisiones inexplicables. Agencialidad. Transparencia y explicabilidad. Seguridad. El futuro del trabajo.

Bibliografía obligatoria:

Coeckelbergh, M. (2022). Robot ethics. MIT Press. (selección)
Danesi, C y Balmaceda, T. “Inteligencia artificial a la luz de la filosofía y el derecho” en Inteligencia artificial, tecnologías emergentes y derecho
Lin, Patrick, Keith Abney, and Ryan Jenkins, eds. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. New York: Oxford University Press, 2017 (Selección)

Bibliografía complementaria:

Brynjolfsson, Erik, and Andrew McAfee. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W. W. Norton and Company, 2014

Bryson J (2010) Robots should be slaves. In: Wilks Y (ed) Close engagement with artificial companions. John Benjamins, Amsterdam, pp 63–74

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, «Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems»

Matthias A (2004) The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics andInformation Technology 6(3): 175–183

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. y Floridi, L., The ethics of algorithms: Mapping the debate, en Big Data & Society, 2016.

UNIDAD 2 - PRIVACIDAD

Contenidos:

Privacidad de los datos. Enfoque de privacidad por defecto y desde el diseño. Medidas especiales de protección para niños, niñas y adolescentes. Manipulación, explotación y usuarios vulnerables. Impacto en las relaciones personales. Clasificación de los sistemas de IA con un enfoque basado en el riesgo

Bibliografía obligatoria:

OECD (2019), Artificial Intelligence in Society, OECD Publishing.

Comisión Europea. 2020. White Paper on Artificial Intelligence. A European approach to excellence and trust. European Commission. COM(2020) 65 final. (Brussels, 19.2.2020)

Bibliografía complementaria:

Andreu, G. R. (2021). Libro Blanco de la Comisión Europea sobre Inteligencia Artificial. Un enfoque europeo hacia la excelencia y la confianza. Revista Ius et Praxis, 27(1), 264-270.

Glenn, T. y Monteith, S., Privacy in the Digital World: Medical and Health Data Outside of HIPAA Protections, en Current psychiatry reports, 16, 494, 10.1007/s11920-014-0494-4, 2014.

Gómez Mont, C., Del Pozo, C. M., Martínez Pinto, C., & Martín del Campo Alcocer, A. V. (2020). La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países. *BID*.

Schermer, B., The limits of privacy in automated profiling and data mining, en Computer Law and Security Review, 27, 2011, pp. 45-52.

UNIDAD 3 - SESGOS

Contenidos:

Definiciones de sesgos. IA & vínculos: chatbots y robots sexuales. La IA desde la perspectiva de los derechos humanos.

Bibliografía obligatoria:

Coeckelbergh, M. (2020). AI ethics. Mit Press (selección)
Balmaceda, T., Pedace, K., Lawler D., Pérez D., Zeller M. "Pensar la tecnología digital con perspectiva de género" (selección)

Bibliografía complementaria:

Dautenhahn K (2007) Socially intelligent robots: dimensions of human-robot interaction. Philos Trans R Soc Lond B Biol Sci 362:679–704
Duffy BR (2003) Anthropomorphism and the social robot. Robot Auton Syst 42:177–190
Hao, K. (2019a). Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo. MIT Technology Review, febrero 8, 2019.
Ishiguro H (2006) Interactive humanoids and androids as ideal interfaces for humans. In Proceedings of the 11th international conference on Intelligent user interfaces (IUI '06). Association for Computing Machinery, New York, NY, USA, 2–9
Wallach, Wendell y Allen, Colin, Moral Machines: Teaching Robots Right from Wrong, Oxford, Oxford University Press, 2009.
Turkle, Sherry, Alone Together: Why We Expect More from Technology and Less from Each Other, Nueva York, Basic Books, 2011.

UNIDAD 4: ¿ÉTICA PROACTIVA?

Contenidos:

¿Es posible programar la ética en un sistema? Ética proactiva: innovación responsable y valores integrados en el diseño

Bibliografía obligatoria:

Floridi, Luciano; Cowls, Josh; Beltrametti, Monica; Chatila, Raja; Chazerand, Patrice; Dignum, Virginia; Luetge, Christoph; Madelin, Robert; Pagallo, Ugo; Rossi, Francesca; Schafer, Burkhard; Valcke, Peggy Y Vayena, Effy, «AI4People— An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations», Minds and Machines 28, núm. 4: 2018, 689-707.
Fry, Hannah, Hola mundo: cómo seguir siendo humanos en la era de los algoritmos, Barcelona, Blackie Books, 2019.

Bibliografía complementaria:

Carpenter, Julie. "Deus Sex Machina: Loving Robot Sex Workers and the Allure of an Insincere Kiss." In Robot Sex: Social and Ethical Implications, edited by John Danaher and Neil McArthur, 261–287. Cambridge, MA: MIT Press, 2017.
Wallach, Wendell, and Colin Allen. Moral Machines: Teaching Robots Right from Wrong. Oxford: Oxford University Press, 2009.
Noble, La religión de la tecnología, Barcelona, Paidós, 1999.

Bibliografía general

Bostrom, Nick, Superinteligencia, Zaragoza, TEELL, 2016.
Burrell, J., How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society. Disponible en: https://doi.org/10.1177/2053951715622512, 2016.
Caro, M. (2019). Algoritmos machistas: los datos (escondidos) que no quieren a las mujeres. El País, abril 9, 2019.
Coeckelbergh, Mark, Growing Moral Relations: Critique of Moral Status Ascription, Nueva York, Palgrave Macmillan, 2012.
Coeckelbergh M (2015) The tragedy of the master: automation, vulnerability, and distance. Ethics Inf Technol 173:219–229
Domingos, P., A few useful things to know about machine learning, en Communications of the ACM, v.55 n.10, octubre de 2012.
Dreyfus, Hubert L., What Computers Can't Do, Nueva York, Harper & Row, 1972.
Eubanks, V. (2021). La automatización de la desigualdad. Herramientas de tecnología avanzada para supervisar y castigar a los pobres. Capitán Swing.
Floridi, L. (2021). Establishing the rules for building trustworthy AI. Ethics, Governance, and Policies in Artificial Intelligence, 41-45.
Harari, Yuval Noah, Homo deus: breve historia del mañana, Madrid, Debate, 2016.
O'Neil, C. (2018). Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia. Capitán Swing.
Tene, O. y Polonetsky, J., Judged by the Tin Man: Individual Rights in the Age of Big Data, en Journal of Telecommunications and High Technology Law, agosto de 2013.
Turilli, M. y Floridi, L. The ethics of information transparency, en Ethics Inf Technol, 2009, 11: 105. https://doi.org/10.1007/s10676-009-9187-9
Zoller, Y., The costs of overprotecting the young - iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy -and completely unprepared for adulthood- and what that means for the rest of us by Jean M. Twenge, en The American Journal of Psychology, 2019,132, pp. 115-119.

Modalidad de cursada:

Las clases se desarrollarán con la modalidad de alternancia entre actividades y referencias a recursos de manera asincrónica y una instancia sincrónico-virtual de desarrollo de conceptos generales, su instanciación, consultas y puesta en común de ideas.

Formas de evaluación:

El seminario se aprueba con un trabajo escrito que unirá los contenidos de este seminario con el de regulación de la IA (puede ser realizada de manera individual o en equipos de hasta 2 estudiantes) cuyo carácter se especificará oportunamente.
El término de presentación del trabajo de cada seminario será de tres meses con derecho a una prórroga por otros tres meses

Requisitos para la aprobación del seminario:

Aprobar el trabajo final del curso con una nota de 4 (cuatro) o superior.