

Escuela de Pregrado

PROGRAMA DE ASIGNATURA

Aspectos Generales de la Actividad Curricular

1. Plan de Estudios

Licenciatura en Filosofía

2. Código y Nombre de la Actividad Curricular

FIL06048

Filosofía y Ética de la IA

3. Code and Name of the Curricular Activity

FIL06048

Philosophy and ethics of AI

4. Pre-requisitos

Sin requisitos

5. Número de Créditos SCT – Chile

6

6. Horas Semanales de trabajo

Presenciales: 3

No presenciales: 6

7. Semestre/Año Académico en que se dicta:

Segundo Semestre 2025

8. Línea Formativa

Formación Especializada (FE)

9. Palabras Clave

máquinas programadas; reemplazo humano; ética de la IA; gobernanza de la IA

10. Propósito general del curso

Proporcionar una visión general, los argumentos, teorías y conceptos de la Filosofía y de la Ética de la Inteligencia Artificial

11. General purpose of the course

The course gives an overview of the arguments, theories and concepts of the Philosophy and Ethics of AI

Equipo Docente

12. Nombre Completo del, de la (los/as) Docente(s) Responsable(s)

Rodrigo Alfonso González Fernández

13. Nombre Completo del, de la (los/as) Docente(s) Participante(s)

Guido Vallejos Oportot

14. Unidad Académica / organismo de la unidad académica que lo desarrolla

Departamento de Filosofía

Descripción Curricular

15. Ámbito del conocimiento al que contribuye el curso

Investigativo

16. Competencias a las que contribuye el curso

1. Comprender y analizar críticamente las corrientes fundamentales y los problemas principales de la disciplina., 2. Analizar conceptos, construir y evaluar argumentos, utilizando métodos y estrategias discursivas de la disciplina., 4. Realizar investigaciones en el ámbito de la filosofía., 5. Establecer diálogo con otras disciplinas identificando campos, problemas y contenidos de carácter interdisciplinario.

17. Subcompetencias

1.1. Identificar y comprender los problemas filosóficos según han sido formulados por las distintas tradiciones de la historia de la filosofía., 1.2. Identificar y comprender los problemas

básicos de las diversas áreas que constituyen la disciplina., 1.3. Desarrollar la actitud crítica., 1.4. Reconocer y comprender los aspectos metódicos en las diversas prácticas filosóficas, 2.2. Evaluar críticamente argumentos atendiendo a su propósito discursivo

18. Resultados de Aprendizaje

El curso examina la filosofía de la Inteligencia artificial, para introducir a los estudiantes a esta subdisciplina, y para permitir una comprensión de la Ética de la IA y los problemas que aborda esta en la época contemporánea

19. Saberes / contenidos

1. El origen de la Inteligencia Artificial;
 - 1.1 Descartes: la diferencia entre hombres y máquinas;
 - 1.2 El hombre máquina de La Mettrie;
 - 1.3 La mecanización del pensamiento en el siglo XIX: Babbage y sus máquinas;
 - 1.4 De los engranajes a los algoritmos de la Inteligencia de Máquina;
 - 1.5 De la Inteligencia de Máquina a la Inteligencia Artificial: la revolución de la ciencia de la computación del siglo XX;
 - 1.6 Las Máquinas de Turing y la tesis Church-Turing; El Test de Turing.
2. El paradigma clásico de la Inteligencia Artificial: de las matemáticas a la Filosofía;
 - 2.1 Arquitecturas cognitivas de la ciencia cognitiva
 - 2.2 Reglas y representaciones;
 - 2.3 Block y el modelo computacional de la mente;
 - 2.4 Los procesadores primitivos y los circuitos lógicos;
 - 2.5 El principio de implementación múltiple;
 - 2.6 La hipótesis del sistema universal de símbolos;
 - 2.7 El conexionismo: reconsiderando la biología.
3. ¿Es posible reducir estados mentales a reglas y representaciones?;
 - 3.1 La Pieza China y la distinción entre Inteligencia Artificial Fuerte y Débil;
 - 3.2 Réplicas a la Pieza China;
 - 3.3 Dennett y McCarthy: la atribución de estados mentales intencionales a máquinas;
 - 3.4 Otras críticas a la IA Fuerte;

- 3.5 Lucas y Penrose: Gödel y el halting problem;
- 3.6 El problema de la codificación del conocimiento, el sentido común y la racionalidad;
- 3.7 Cleland y la controvertida causalidad de las máquinas algorítmicas;
- 3.8. Shani y la defensa de la Pieza China con base en la intencionalidad
- 4. La Ética de la IA;
- 3.1 Floridi y la teoría de la información;
- 3.2 Los desiderata de una Ética de la IA;
- 3.3 Los 5 principios de una Ética de la IA: beneficencia, no malevolencia, autonomía, justicia y explicabilidad;
- 3.4 La importancia de la explicabilidad de la IA para que esta sea confiable;
- 3.5 La evolución de la ética de la IA;
- 3.6 Principios para la gobernanza de la IA.

20. Metodología de Enseñanza - Aprendizaje

- 1. Clases expositivas y explicativas
- 2. Lectura y análisis de textos
- 3. Discusión dirigida

21. Metodología de Evaluación

Temario: 30%

Ensayo: 70%

22. Requisitos de aprobación

Mínimo de aprobación: 4,0

23. Requisito de asistencia

0,8

Recursos

24. Bibliografía Obligatoria

- 1. Block, N. (1995): "The mind as software of the brain." Extracted and edited in: J. Heil (ed.) Philosophy of Mind: a Guide and Anthology. Oxford: OUP, pp. 267-274;
- 2. _____ (1990): "The computer model of the mind." In: D.N. Osherson and E.E. Smith (eds.) Thinking: An Invitation to Cognitive Science, Vol. 3. Cambridge, Mass.: MIT Press, pp. 247-89;
- 3. Cleland, C. (1993): "Is the Church-Turing thesis true?" Minds and Machines 3, 283-312;
- 4. Copeland, J. (1993): Artificial Intelligence: A Philosophical Introduction. Oxford: Blackwell;
- 5. Churchland P.M. and Churchland P.S. (1990): "Could a machine think?" Scientific American January 1990, pp. 26-31;

6. Descartes, R. (2004): Discourse on Method (Ch. 5). In: S. Shieber (ed.) The Turing Test. Cambridge, Mass.: MIT Press;
7. Dreyfus, H.L. and Dreyfus, S.E. (1990): "Making a mind versus modeling the brain: Artificial Intelligence back at a Branch-point." In: M. Boden (ed.) The Philosophy of Artificial Intelligence. Oxford: OUP;
8. Floridi, L. A unified framework of five principles for AI in society. In: FLORIDI, L. Ethics, governance, and policies in Artificial Intelligence. Cham: Springer, 2021b. p. 5-17.
9. González, R. (2007): "El Test de Turing: dos mitos, un dogma". Revista de Filosofía Universidad de Chile, Vol. 63, 37-53;
10. _____ (2012): "La pieza china: un experimento mental con sesgo cartesiano". Revista Chilena de Neuropsicología, Vol. 7, edición especial, 1-6.
11. _____ (2024) "Can AI be trustworthy?" Aufklärung, João Pessoa, v.11, n. Especial, Nov., 2024, p.11-24. DOI: <https://doi.org/10.18012/arf.v11iEspecial.69910>
12. McCarthy, J. (1983): "The little thoughts of thinking machines." At: <http://www-formal.stanford.edu/jmc/>;
13. McCulloch, W.S. and Pitts, W.H. (1943): "A logical calculus of the ideas immanent in nervous activity." Bulletin of Mathematical Biophysics 5, 115-33;
14. Penrose, R. (1993): "Setting the scene: The claim and the issues." In: D. Broadbent (ed.) The Simulation of Human Intelligence. Oxford: Blackwell, pp. 1-32;
15. Searle, J. (1980): "Minds, brains and programs." Behavioral and Brain Sciences 3, 417-24. Reprinted in: M. Boden (ed.) The Philosophy of Artificial Intelligence. Oxford: OUP, pp. 67-88;
16. _____ (1990): "Is the brain's mind a computer program?" Scientific American, January 1990, 20-25;
17. _____ (2004): Mind: A Brief Introduction. New York: OUP;
18. Shani, I. (2005): "Computation and Intentionality: A recipe for an epistemic impasse." Minds and Machines, Vol. 15, 2, 207-228;
19. Shieber, S. (2004): The Turing Test: Verbal Behavior as the Hallmark of Intelligence. Cambridge, Mass.: MIT Pres.
20. Swade, D. (2000): The Difference Engine: Charles Babbage and the Quest to build the First Computer. London: Penguin;
21. Turing, A.M. (1950): "Computing intelligence and machinery." Mind LIX, no. 2236, (Oct. 1950), 433-60. Reprinted in: M.A. Boden (ed.) The Philosophy of Artificial Intelligence. Oxford: OUP, pp. 40-66. Alianza Editorial, diversas reimpresiones

25. Bibliografía Complementaria

1. Block, N. (1980): "What is Functionalism?" In: J. Heil (ed.) *Philosophy of Mind: A Guide and Anthology*. Oxford: OUP, pp. 183-99;
2. Copeland, J. (2000): "The Turing test." In: J.H. Moor (ed.) *The Turing test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer Academic Publishers, pp. 1-21;
3. Dennett, D. (1988): "When Philosophers encounter Artificial Intelligence". In: S.R. Graubard (ed.) *The Artificial Intelligence Debate False Starts Real Foundations*. Cambridge, Mass.: MIT Press;
4. Dubber, M., Pasquale, F. y Das, S. (2020): *The Oxford Handbook of Ethics of AI*. Oxford: OUP.
5. González, R. (2007): *The Chinese Room Revisited: Artificial Intelligence and the Nature of Mind*. Dissertation presented to fulfill the requirements for the degree of Doctor (Ph.D.) in Philosophy. Centre for Logic and Analytic Philosophy, Institute of Philosophy, Katholieke Universiteit Leuven;
6. _____ (2012): "El humanismo de Humberto Sábato: ¿Visionario del último engranaje?" *Revista Mapocho*, No. 72, segundo semestre, 13-26.
7. Heil, J. (2004): "Functionalism." In: *Philosophy of Mind: A Guide and Anthology*. Oxford: OUP, pp. 139-49;
8. Kuhn, T.S. (1964): "A function for thought experiments." Reprinted in: *The Essential Tension*. Chicago: University of Chicago Press, pp. 240-65;
9. Moor, J.H. (1987): "Turing test." In: S.C. Shapiro (ed.) *Encyclopedia of Artificial Intelligence*, Vol. 2. New York: Wiley, pp. 1126-30;
10. Paupert, S. (1988): "One AI or many?" In: S.R. Graubard (ed.) *The Artificial Intelligence Debate False Starts Real Foundations*. Cambridge, Mass.: MIT Press;
11. Rucker, R. (1982): *Software*. New York: HarperCollins;
12. Saygin, A.P., Cicekli, I. and Akman, V. (2000): "Turing test: 50 years later." In: J.H. Moor (ed.) *The Turing test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer Academic Publishers, pp. 23-78;
13. Schank, R.C. and Abelson, R.P. (1977): *Scripts, Plans, Goals, and Understanding*. Hillsdale, N.J.: Erlbaum;
14. Searle, J. (2002): "Twenty-one years in the Chinese Room." In: J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford: OUP, pp. 51-69;
15. Smith, B.C. (2019): *The Promise of Artificial Intelligence: Reckoning and Judgement*. Cambridge, Mass.: MIT Press.

16. Schwartz, J. (1988): The New Connectionism: Developing Relationships Between Neuroscience and Artificial Intelligence. In: S.R. Graubard (ed.) The Artificial Intelligence Debate False Starts Real Foundations. Cambridge, Mass.: MIT Press;
17. Sokolowski, R. (1988): "Natural and Artificial Intelligence" In: S.R. Graubard (ed.) The Artificial Intelligence Debate False Starts Real Foundations. Cambridge, Mass.: MIT Press;
18. Weizenbaum, J. (1984): Computer Power and Human Reason: From Judgement to Calculation. Harmondsworth: Pelican

26. Recursos web

STANFORD ENCYCLOPEDIA OF PHILOSOPHY:

<http://plato.stanford.edu/>

Por una Facultad comprometida con una educación no-sexista y el respeto por los DDHH, te invitamos a conocer los instrumentos de Equidad que rigen en nuestra Comunidad Universitaria:

Política de corresponsabilidad en cuidados: En conformidad con la Política de Igualdad de Género de nuestra Universidad los y las estudiantes padres y madres cuidadores de menores de 6 años pueden solicitar apoyos económicos, pre y postnatal y medidas de flexibilidad académica para compatibilizar sus responsabilidades estudiantiles y de cuidados. Para más información sobre beneficios y procedimientos, revisa: Kit corresponsabilidad y [Link WEB DiGenDiFil](#)

Uso de Nombre Social: Gracias al instructivo Mara Rita cuentas con la posibilidad de establecer oficialmente dentro del espacio universitario el nombre y los pronombres por los que quieres ser llamado, según tu identidad sexo genérica. Para saber más sobre el procedimiento, revisa: KIT MARA RITA [Link WEB DiGenDiFil](#) y si quieres editar tu firma de correo electrónico con tus pronombres, participa de la campaña [#MiPronombre](#)

Protocolo de actuación ante denuncias sobre acoso sexual, violencia de género y discriminación arbitraria. Porque #NosCansamos del Abuso, #LaChileDiceNo al acoso sexual. Si vives alguna de estas situaciones, puedes dirigirte a DAEC o DiGenDiFil, para buscar apoyos y orientación en tus procesos personales y de denuncias. Para contactarnos escribe al daec@uchile.cl o digenfil@uchile.cl y para más información sobre procedimientos, revisa [DIGEN UCHILE](#)