



Universidad de Buenos Aires
FACULTAD DE FILOSOFÍA Y LETRAS

SEMINARIO DEL PROGRAMA DE ACTUALIZACIÓN EN IA DESDE UNA PERSPECTIVA HUMANÍSTICA

El uso de “inteligencia” y otros términos psicológicos en la IA

Docente a cargo: Diana I. Pérez

Carga horaria: 16 hs.

Cuatrimestre, año: Primer cuatrimestre 2024

Fundamentación

La Inteligencia Artificial descansa en la idea de que es posible mecanizar el pensamiento humano, generando artefactos de diseño que sean capaces de emular nuestros pensamientos y conductas, de una manera similar, o aún más eficiente y rápida, que como lo hacemos los humanos. Esta posibilidad descansa en cierta concepción de la mente humana, el comportamiento humano y, en última instancia, en cierta autocomprensión de nuestra naturaleza humana. En este curso se examinarán los supuestos acerca de las mentes y los comportamientos humanos que subyacen a los diversos desarrollos en inteligencia artificial. Para ello se examinarán tanto concepciones filosóficas como científicas de la mente, y sus rasgos fundamentales: la conciencia y la intencionalidad. Asimismo, se analizará la cuestión de nuestras prácticas cotidianas de atribución de estados psicológicos y conductas inteligentes, así como la validez que tiene el Test de Turing como piedra de toque para evaluar el carácter inteligente de la conducta exhibida por una entidad (humana o no).

Objetivos

Que el estudiantado se familiarice con nociones básicas relativas a la naturaleza de la mente humana y las perspectivas para su estudio científico y modelización computacional.

Que el estudiantado se familiarice con los problemas metafísicos, epistemológicos y metodológicos relativos al estudio de la mente y conducta humanas, y su impacto en nuestras concepciones de lo mental y lo humano.

Que el estudiantado evalúe críticamente los supuestos e implicancias del Test de Turing como marca de la inteligencia, distinguiendo posibles tipos de inteligencia involucrados en el desarrollo de la inteligencia artificial.

Que el estudiantado analice los supuestos relativos a la mente y la agencia humanas presupuestos en los diversos programas de desarrollo de la inteligencia artificial.

UNIDAD 1: El Test de Turing: implicaciones y supuestos filosóficos

Contenidos: La formulación original del Test de Turing. Objeciones y respuestas. Requisitos para diseñar una máquina que pueda pasar el Test. ¿Alguna máquina existente ya pasó el Test? ¿Es inteligente toda máquina que pase el Test? Inteligencia humana e inteligencia artificial.

Bibliografía obligatoria:

Oppy, G. y Dowe, D. (2021) "The Turing Test", en Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*.

<https://plato.stanford.edu/archives/win2021/entries/turing-test/>

Turing, A. (1950) "Computing Machinery and Intelligence" *Mind*, vol 59. (Traducción de Garrido y Antón: "¿Puede pensar una máquina?" en Turing, A., Putnam, H. y Davidson, D., *Mentes y Máquinas*, Tecnos, Madrid 1985)

Bibliografía complementaria:

Montemayor, C. (2022) *The Prospect of a Humanitarian Artificial Intelligence*. Bloomsbury, UK. Cap. 1

Neufeld, E. y Finnstad, S. (2020) "Imitation Game: Threshold or Watershed?" *Minds and Machines* <https://doi.org/10.1007/s11023-020-09544-5>

Proudfoot, D. (2020) "Rethinking Turing's Test and the Philosophical Implications" *Minds and Machines* <https://doi.org/10.1007/s11023-020-09534-7>

UNIDAD 2: La IA en el contexto de las ciencias de la mente

Contenido: Las ciencias cognitivas. El funcionalismo. Argumentos a favor y en contra. La IA como modelo de lo mental. Vasos comunicantes entre las ciencias de la mente y los desarrollos en IA. El postcognitivismo.

Bibliografía obligatoria:

Fodor, J. (1994) *Psicosemántica*, Madrid, Técno, cap. 1 y apéndice.

García-Carpintero, M. (1995) "El funcionalismo" en Broncano, F. *La Mente Humana*. Ed. Trotta. Madrid

Gomila, T. y Calvo, P. (2008) "Directions for an embodied cognitive science: toward an integrated approach" en Calvo, P. y Gomila, T. (2008) *Handbook of Cognitive Science. An Embodied Approach*. Elsevier, UK.

Russell, S y Norvig P (2003) *Inteligencia Artificial. Un enfoque Moderno*. (Segunda Edición) Madrid. Pearson Educación. 2004, Cap. 1

Bibliografía complementaria:

Thagard, P. (2023) "Cognitive Science", en Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/spr2023/entries/cognitive-science/>

UNIDAD 3: Un poco de metafísica de la mente

Contenido: La mente cartesiana y el problema mente-cuerpo. Las marcas de lo mental: Conciencia e intencionalidad. ¿Qué es la conciencia? ¿Cómo dar cuenta de la intencionalidad de lo mental? La mente extendida, encarnada, situada, enactiva y afectiva.

Bibliografía obligatoria:

Burdman, F. (2015) "El post-cognitivismo en cuestión: extensión, corporización y enactivismo". *Principia. An International Journal of Epistemology*, vol. 19, nro. 3, pp. 475-495

Chalmers, D. (1999) *La mente consciente*, Gedisa, Barcelona, cap 1 y Cap 3, secc 1.

Dennett, D. (1991) "Evolución, error e intencionalidad" en Dennett (1991) *La actitud intencional*. Buenos Aires. Gedisa.

Pérez, D (2002) "Naturalismo y mente humana. De las intrincadas relaciones entre la filosofía, la ciencia, el sentido común, el alma humana, y la madre naturaleza", en Pérez, D. (comp.) *Los Caminos del Naturalismo*, Eudeba, Buenos Aires, pp. 107-130

Searle, J. (1980) "Mentes, cerebros y Programas", en D. Hofstadter & D. Dennett (eds.) (1981) *El ojo de la mente*, Buenos Aires, Sudamericana, 1983

Bibliografía complementaria:

Noe, A. (2010) *Fuera de la cabeza. Por qué no somos el cerebro, y otras lecciones de la biología de la conciencia*. Ed. Kairós. Barcelona, Caps. 1 y 2.

Pérez, D. (2007) "¿La conciencia? ¿Qué es eso?" *Revista Estudios de Psicología*. 28 (2), pp. 127-140. <https://doi.org/10.1174/021093907780962845>

UNIDAD 4: Las atribuciones psicológicas

Contenido: Lenguaje y conceptos psicológicos. El acceso a las mentes: la primera, la segunda y la tercera persona. Las prácticas de atribución psicológica y sus supuestos: racionalidad, teorización, simulación, empatía, expresión emocional, corporalidad. Atribuciones psicológicas a seres no-humanos.

Bibliografía obligatoria:

Dennett, D. (1991) "Verdaderos creyentes" en Dennett (1991) *La actitud intencional*. Buenos Aires. Gedisa.

Gomila, A. (2002) La perspectiva de segunda persona de la atribución mental. *Azafea* 1: 123-138.

Pérez, D. (2022) "La segunda persona y la teoría de la mente" *Revista Ideação* v. 1 n. 45 DOI: <https://doi.org/10.13102/ideac.v1i45>
<http://periodicos.uefs.br/index.php/revistaideacao/issue/view/336>

Russell, S. y Norvig P. (2003) *Inteligencia Artificial. Un enfoque Moderno*. (Segunda Edición) Madrid. Pearson Educación. 2004, Cap. 2.

Bibliografía complementaria:

Pérez, D. (2013) *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*. Prometeo, Buenos Aires, caps 3, 4 y 5.

Pérez, D. "Inteligencia artificial, actitud intencional y segunda persona" en Pérez Jiménez, M.A., Suarez, C. y Mantilla, A. *Intersubjetividad y segunda persona* (en

prensa)

Bibliografía general

Balmaceda, T. y Pedace, K (2018) *Temas de Filosofía de la mente. Atribución psicológica*. Ed. SADAFA. Buenos Aires

Calvo, P. y Gomila, T. (2008) *Handbook of Cognitive Science. An Embodied Approach*. Elsevier, UK.

Clark, A. y Chalmers, D. (2011) *La mente extendida*. KRK, Oviedo.

Damasio, A. (1994) *El error de Descartes*. Ed- Andrés Bello. Santiago de Chile.

Davidson, D. (1982) “Animales racionales” en Davidson, D. (2003) *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, pp. 141-155.

Frankish, K. & Ramsey, W. (eds.) (2012). *The Cambridge Handbook of Cognitive Science*. Cambridge: Cambridge University Press.

Hurley, S. (1998) *Consciousness in Action*. Harvard University Press

Lakoff, G. y Johnson, M. (1999) *Philosophy in the flesh*. Basic Books, NY.

Montemayor, C. (2022) *The Prospect of a Humanitarian Artificial Intelligence*. Bloomsbury, UK.

Pérez, D. (2013) *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*. Prometeo, Buenos Aires.

Pérez, D. (2018) “Filosofía, Ciencias Cognitivas y Sentido Común: el caso de la segunda persona de la atribución mental”, *Revista de Filosofía Moderna e Contemporânea*, Brasilia, v.6, n.1, jul., p. 49-72 49 ISSN: 2317-9570
<https://periodicos.unb.br/index.php/fmc/article/view/18654>

Pérez, D. & Gomila, A. (2022). *Social cognition and the second person in human interaction*. London: Routledge, 182 páginas. (ISBN 9780367678579)

Pérez, D. y Melamed, A. (2020) “La razón en las pasiones y las pasiones en la razón. Un entretejido desatendido”, en Vidal, J. y Muñoz, C. (eds.) *Perspectivas en filosofía de la subjetividad*. Editorial de la Universidad de Concepción, Concepción, Chile. pp. 213-230.

Russell, S y Norvig P [2003] *Inteligencia Artificial. Un enfoque Moderno*. (Segunda Edición) Madrid. Pearson Educación. 2004

Modalidad de cursada: a distancia

Formas de evaluación

Los seminarios se aprueban con una evaluación (puede ser realizado 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 6 (seis) o superior.