Universidad de Chile
Facultad de Filosofía y Humanidades
Departamento de Filosofía

PROGRAMA

I. IDENTIFICACIÓN

Seminario:	"Filosofía de la Mente y de la Inteligencia Artificial"
Año académico:	Segundo semestre de 2011
Profesor:	Rodrigo González, Manuel Rodríguez y Guido Vallejos.
Horario:	Martes de 18.00 a 21.00
Sala:	P1, 4º piso, Fac. de Filosofía y Humanidades
Programas en que se imparte:	Magíster en Estudios Cognitivos

II. DESCRIPCIÓN

Aunque el desarrollo de la Inteligencia Artificial parece totalmente ajeno a la Filosofía, es posible sostener que existe una importante relación entre ambas, incluso desde los orígenes de la primera. En los albores de lo que entonces se denominaba "inteligencia de máquina" Charles Babbage descubrió el problema filosófico fundamental que atañe a la Inteligencia Artificial, esto es, la supuesta vida mental de máquinas programadas. Dicho problema no resultaba ajeno a un proyecto nacido en el contexto de la época industrial decimonónica, a saber, la mecanización del pensamiento. Diversas necesidades, incluso de orden práctico, como por ejemplo la erradicación de los errores de los cálculos matemáticos, o la puesta en escena de máquinas que pudiesen ahorrar esfuerzo mental al hombre, explican la aparición de dicho proyecto.

Frente al problema de si las máquinas programadas poseen estados mentales de diverso orden, Babbage y otros adoptaron la posición que posteriormente Searle bautizó como Inteligencia Artificial Débil, o la tesis de que máquinas implementando algoritmos no poseen vida mental de manera literal, sino que su funcionamiento puede ser descrito instrumentalmente de tal forma. Por otra parte, el lógico matemático Alan Turing, además de sentar las bases de la Inteligencia Artificial con la Máquina de Turing y con el controvertido Test de Turing, criticó la posición cauta de pensadores como Babbage, al sostener que las máquinas programadas poseen estados mentales como cuestión de hecho. Tanto esta última tesis, llamada por Searle Inteligencia Artificial Fuerte, como la de Babbage, dieron origen a uno de los debates contemporáneos más prolíficos de la Filosofía Analítica, a tal punto que diversas teorías, experimentos mentales y argumentos han sido esbozados en pro y en contra de cada posición.

Teniendo en consideración el mencionado debate, este curso examinará los problemas filosóficos de mayor relevancia asociados a la Inteligencia Artificial, con énfasis en la disputa sobre la atribución de estados mentales a máquinas programadas. A pesar de que se expondrá de manera breve y sucinta el origen de la Inteligencia Artificial, se explorará la conexión e influencia recíproca entre la Filosofía de la Mente y esta última disciplina, especialmente en

relación con los tópicos centrales de la última. Así, con el fin de investigar la controversia acerca de la supuesta vida mental de máquinas de naturaleza algorítmica, este curso explorará de manera detallada los conceptos fundamentales de la Inteligencia Artificial.

III. 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
1.7 El Test de Turing

2. El paradigma clásico de la Inteligencia Artificial: de las matemáticas a la Filosofía
2.1 Reglas y representaciones
2.2 Block y el modelo computacional de la mente
2.3 Los circuitos lógicos
2.4 El principio de implementación múltiple
2.5 La hipótesis del sistema universal de símbolos
2.6 El conexionismo: reconsiderando la biología

3. ¿Es posible reducir estados metales 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 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

III. EVALUACIÓN

30% Temario
20% Proyecto de *paper**
50% *Paper*

*Sin proyecto de *paper* aprobado no se recibirá el *paper*

Bibliografía básica

Block, N. (1980): "What is Functionalism?" In: J. Heil (ed.) *Philosophy of Mind: A Guide and Anthology*. Oxford, OUP, pp. 183-99.

_____ (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.

_____ (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.

Cleland, C. (1993): “Is the Church-Turing thesis true?” *Minds and Machines* 3, 283-312.

Copeland, B.J. (1993): *Artificial Intelligence: A Philosophical Introduction*. Oxford, Blackwell.

_____ (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.

Churchland P.M. and Churchland P.S. (1990): “Could a machine think?” *Scientific American* January 1990, pp. 26-31.

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.

Descartes, R. (2004): *Discourse on Method (Ch. 5)*. In: S. Shieber (ed.) *The Turing Test*. Cambridge, Mass.: MIT Press.

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.

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.

_____ (2007) “El Test de Turing: dos mitos, un dogma”. *Revista de Filosofía Universidad de Chile*, Vol. 63, 37-53.

Heil, J. (2004): “Functionalism.” In: *Philosophy of Mind: A Guide and Anthology*. Oxford, OUP, pp. 139-49.

Kuhn, T.S. (1964): “A function for thought experiments.” Reprinted in: *The Essential Tension*. Chicago, University of Chicago Press, pp. 240-65.

McCarthy, J. (1983): “The little thoughts of thinking machines.” At: http://www-formal.stanford.edu/jmc/

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.

Moor, J.H. (1987): “Turing test.” In: S.C. Shapiro (ed.) *Encyclopedia of Artificial Intelligence*, Vol. 2. New York, Wiley, pp. 1126-30.

Paupert, S. (1988): “One AI or many?” In: S.R. Graubard (ed.) *The Artificial Intelligence Debate False Starts Real Foundations*. Cambridge, Mass.: MIT Press.

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.

Rucker, R. (1982): *Software*. New York, HarperCollins.

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.

Schank, R.C. and Abelson, R.P. (1977): *Scripts, Plans, Goals, and Understanding. Hillsdale, N.J., Erlbaum.*

_____ (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.

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.

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.

_____ (1990): “Is the brain’s mind a computer program?” *Scientific American*, January 1990, 20-25.

_____ (2004): *Mind: A Brief Introduction*. New York, OUP.

Shani, I. (2005): “Computation and Intentionality: A recipe for an epistemic impasse.” *Minds and Machines*, Vol. 15, 2, 207-228.

Sokolowski, R. (1988): “Natural and Artificial Intelligence” In: S.R. Graubard (ed.) *The Artificial Intelligence Debate False Starts Real Foundations*. Cambridge, Mass.: MIT Press.

Swade, D. (2000): *The Difference Engine: Charles Babbage and the Quest to build the First Computer*. London, Penguin.

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.

Weizenbaum, J. (1984): *Computer Power and Human Reason: From Judgement to Calculation*. Harmondsworth, Pelican.