Minds and Machines (MA) Course Syllabus

Course	Minds and Machines	Faculty	Philosophy
Course Code	NCHAI750	Course Leader	Dr Ioannis Votsis
Credits	15	Author	Dr Ioannis Votsis
FHEQ Level	7	Date Approved	
Compulsory/ Optional	Compulsory for MA in Philosophy and AI	Date Modified	
Pre-requisites	None		

WELCOME

A warm and friendly welcome to Minds and Machines! As we emerge from the most recent AI winter, several concurrent developments in the fields of neuroscience, robotics and machine learning have meant that the question of what makes a mind is more pressing than ever. This course explores the intricate relationships between biological, particularly human, and artificial systems that exhibit intelligent behaviour. If you have any questions, please don't hesitate to raise them either by e-mail (ioannis.votsis@nulondon.ac.uk) or in person.

DESCRIPTION

This course investigates the ontological, epistemological and methodological dimensions of issues that emerge in relation to systems that exhibit intelligent behaviour, whether these are biological or artificial. Students will learn about the main theories of mind and of the way these theories enrich our understanding and design of intelligent machines. Conversely, students will consider how advances in artificial intelligence seek to throw light on the human mind. In more detail, the course will address questions such as: What is the correct theory of mind? Do mental states reduce to brain states? Does the mind extend beyond the confines of our heads? What is computation? Does the mind compute in a similar way to machines? Under what conditions can we say that a physical system computes? What is intelligence? Is thinking more than mere intelligence? Can we standardise intelligence tests for both machines and humans? Are we also machines of some sort? What can we know about the world around us? Is the universe just a massive computer simulation?

AIMS

The course aims to:

- To develop students' skills in understanding and evaluating the main theories of mind and machines and the relations between them.
- To promote students' ability to interpret, analyse and compare key texts in this area.
- To enable students to form, elaborate and defend their own views in this area.

LEARNING OUTCOMES

On successful completion of the course, students should be able to:

KNOWLEDGE AND UNDERSTANDING

K1d Demonstrate wide-ranging knowledge and systematic understanding of key questions, debates, and theories in philosophy, especially those concerned with relation between minds and machines.

K2d Offer detailed critical engagement with the texts and theories of key figures.

K3d Show a fine grasp of logical structure and truth-preserving patterns of inference in this area.

SUBJECT-SPECIFIC SKILLS

S1d Make original use of advanced scholarly techniques to clarify and situate ideas and arguments, especially in relation to computing, data, and artificial intelligence.

S2d Engage with unfamiliar material at the forefront of philosophy and artificial intelligence, selecting and analysing information, questioning assumptions, and critically evaluating competing methodologies, sources of data and arguments.

S3d Identify and employ a range of philosophical devices to articulate, develop and synthesise alternative positions.

TRANSFERABLE AND PROFESSIONAL SKILLS

T1d Take initiative and personal responsibility; work independently, effectively, and to deadlines.

T2d Respond systematically and creatively to complex, wide-ranging, and unpredictable data, theories, and arguments.

T3d Display self-direction to produce original, sophisticated, clear, and persuasive presentations (written and oral).

T4d Consistently apply an excellent level of technical proficiency in written English, using an advanced application of scholarly terminology, that demonstrates the

ability to deal with complex issues both systematically and with sophistication.

LEARNING AND TEACHING PLAN

Teaching and learning strategies for this course will include:

- 15 hours of lectures
- One 1-hour one-to-one tutorial
- Independent study

Course information and supplementary materials are available on the College's Virtual Learning Environment (VLE): Canvas https://nulondon.instructure.com/

Students are required to attend and participate in all timetabled sessions for this course, including their end-of-term Collections meetings (see below). Students are also expected to manage their directed learning and independent study in support of the course.

FEEDBACK

Students will receive written feedback on all submitted essays, normally within two weeks of submission. They will also receive verbal feedback in tutorials.

TEACHING SCHEDULE

Hilary Term		
Week	**Topic**	**Reading**
1	General Theories of Mind	Funkhouser, E. (2007) 'Multiple Realizability', *Philosophy Compass*, 2(2), 303-315. Polger, T. W. (2009) 'Identity Theories', *Philosophy Compass*, 4(5), 822-834.
2	Extending the Mind	Adams, F. and K. Aizawa (2001) 'The Bounds of Cognition', *Philosophical Psychology*, 14(1): 43–64. Clark, A. and D. Chalmers (1998) 'The Extended Mind', *Analysis*, *58*(1), 7-19.
3	Human and Machine Intelligence	Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: 417–57. Turing, A. (1950) 'Computing Machinery and Intelligence', *Mind*, 59: 433–460.
4	Computationalism	Ritchie, J. B., & Piccinini, G. (2018) 'Computational implementation' in M. Sprevak

Hilary Term		
Week	**Topic**	**Reading**
		& M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind*, New York: Routledge, pp. 192-204. Sprevak, M. (2010) 'Computation, Individuation, and the Received View on Representation', *Studies in History and Philosophy of Science*, 41: 260–270.
5	Science and AI	Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, *346*(6206), 171-172. Gillies, D. (1996) *Artificial Intelligence and Scientific Method*, Oxford: OUP, ch. 2.
6	Creativity and AI	Boden, M. A. (2014). *Creativity and Artificial Intelligence: A Contradiction in Terms?* in E. S. Paul and S. B. Kaufman (Eds.), *The Philosophy of Creativity: New Essays*. Oxford University Press, New York, NY, pp. 224-244. Elkins, K., & Chun, J. (2020). 'Can GPT-3 Pass a Writer's Turing Test?', *Journal of Cultural Analytics*, vol. 5(2): 1-16.
7	**READING WEEK**	
8	Emotion and AI	DeLancey, C. (2002) *Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence*, Oxford: OUP, Chs. 12-13. Scarantino, A. (2004). Book Review of Craig DeLancey's Passionate Engines: What Emotions Reveal about the Mind and Artificial Intelligence, *Philosophy of Science*, 71(2), 227-230.
9	Superintelligence	Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*, Oxford, UK: Oxford University Press, Ch. 2 and 9.

Hilary Term		
Week	**Topic**	**Reading**
		Russell, S. (2017). 'Artificial intelligence: The future is superintelligent'. *Nature*, 548(7669), 520-521.
10	Scepticism and Computer Simulations	Bostrom, N. (2003) 'Are we Living in a Computer Simulation?', *The Philosophical Quarterly*, 53(211), 243-255. Moskowitz, C. (2016) 'Are we living in a Computer Simulation?', *Scientific American*, April 7th, 2016, https://www.scientificamerican.com/article/are-we-living-in-a-computer-simulation/
11	**Consolidation Lecture**	
12	**Collections**	

Please refer to your CELCAT timetable for exact dates and times of lectures, seminars and tutorials.

ASSESSMENT

FORMATIVE

Please refer to the 'Formative and Summative Assessment Planner' for **SUBMISSION DATES** of your formative and summative assignments.

Students will be given a mark for all formative work, not a precise numerical mark (unless appropriate, as e.g. for a maths exercise), but a qualitative indication of where in the range Distinction (72-100); Merit (62-68); Pass (52-58) a piece of work lies. The mark will be communicated to students at the time feedback is given, for it constitutes an element of that feedback.

No.	Assignment Type	Details	Length
FA1/ AE1	*Essay/Presentation/Question set/Other*	*Title/Other*	*xx words/xx minutes*
FA1	Formative Essay	See below	2000-2500 words

FA1 ESSAY TOPICS (Choose one of the following):

1. Is multiple realisability a knockdown argument against identity theories of mind?

2. Critically evaluate two central objections to the extended mind hypothesis.
3. 'Symbol manipulation does not entail understanding'. Discuss.
4. Assess the pancomputationalist objection to computationalist views of the mind.
5. What does the use of artificial intelligence in science teach us about the scientific method?
6. Discuss Boden's view of creativity in relation to the prospects of creative artificial agents.
7. Compare the evaluative tradition to the feeling tradition about emotion.
8. What is the control problem in the context of discussions about superintelligence?

SUMMATIVE

Summative assignment briefs can be found on the course page on the College's VLE: Canvas https://nchlondon.instructure.com/.

No.	Assignment Type	Details	Length
FA1/ AE1	*Essay/Presentation/Question set/Other*	*Title/Other*	*xx words/xx minutes*
AE1	Summative Essay	Choose one of the summative essay topics listed in the Summative Assessment Brief. **Note**: Students are free to choose the same topic for both a formative and the summative essay provided: (a) it appears on the summative essay topic list and (b) the summative essay is sufficiently more developed and does not merely copy material from the formative.	4000 words

READING

Please note that due to copyright restrictions I may not be able to upload all the required readings on Canvas.

Disclaimer: Both the required and further readings can be changed at any time in order to respond to the needs of course participants. It is important that students check for updates on Canvas every week.

REQUIRED READINGS

If these texts have not been uploaded - see copyright comment above - then the students must get them through Senate House Library or through some other arrangement (e.g. from another library or buy them).

FURTHER READINGS

These are useful for deepening the student's knowledge on the particular topic but are not necessary.

****

Brief Surveys:

Boden, M. (2018) *Artificial Intelligence*: *A Very Short Introduction*, Oxford: Oxford University Press.

HT Week 1: GENERAL THEORIES OF MIND

Required Reading:

Funkhouser, E. (2007). Multiple realizability. *Philosophy Compass*, 2(2), 303-315.
Polger, T. W. (2009). Identity theories. *Philosophy Compass*, 4(5), 822-834.

Further Reading:

Block, N. (1980) 'Troubles with Functionalism' in N. Block (ed.), *Readings in the Philosophy of Psychology, Volumes 1 and 2*, Cambridge, MA: Harvard University Press.
Chalmers, D.M., 1996, *The Conscious Mind*, New York: Oxford University Press.
Dennett, D. (1981) *Brainstorms*, Cambridge, MA: MIT Press.
Fodor, Jerry A. (1975) *The Language of Thought*, New York: Thomas Crowell.
Kim, J., (1989) 'Mechanism, Purpose, and Explanatory Exclusion', in J. Kim, Supervenience and Mind, Cambridge, Cambridge University Press.
Jackson, F. (1986) 'What Mary Didn't Know', *Journal of Philosophy*, 83: 291–295.
Peacocke, C. (1999) *Being Known*, Oxford: Oxford University Press.
Nagel, T. (1974) 'What is it Like to be a Bat?', *Philosophical Review*, 83: 435–450.
Putnam, H. (1960) 'Minds and Machines', in S. Hook (ed.), *Dimensions of Mind*, New York: New York University Press.
Ryle, G. (1949) *The Concept of Mind*, London: Hutcheson.
Smart, J. J. C. (2017) 'The Mind/Brain Identity Theory', *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/mind-identity/>.

HT Week 2: EXTENDING THE MIND

Required Reading:

Adams, F. and K. Aizawa (2001) 'The Bounds of Cognition', *Philosophical Psychology*, 14(1): 43–64.
Clark, A. and D. Chalmers (1998) 'The Extended Mind', *Analysis*, *58*(1), 7-19.

Further Reading:

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617-645.
Burge, T. (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4: 73–121.
Drayson, Z. (2010). Extended cognition and the metaphysics of mind. *Cognitive Systems Research*, *11*(4), 367-377.
Farkas, K. (2008) *The Subject's Point of View*, Oxford: Oxford University Press.
Hurley, S. L. (1998) *Consciousness in Action*, Cambridge, MA: Harvard University Press.
Rowlands, M. (2009) 'Extended Cognition and the Mark of the Cognitive', *Philosophical Psychology*, 22(1): 1–19.
Shapiro, L. (2007). The embodied cognition research programme. *Philosophy compass*, *2*(2), 338-346.
Sprevak, M., (2009). 'Extended cognition and functionalism', *Journal of Philosophy*, 106(9): 503–527.
Varela, F. J., E. Thompson, and E. Rosch (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

HT Week 3: HUMAN AND MACHINE INTELLIGENCE

Required Reading:

Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: 417–57.
Turing, A. (1950) 'Computing Machinery and Intelligence', *Mind*, 59: 433–460.

Further Reading:

Boden, M. (1988) *Computer Models of the Mind*, Cambridge: Cambridge University Press.
Copeland, J. (2002) 'The Chinese Room from a Logical Point of View', in Preston and Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, New York: Oxford University Press, pp. 104–122.
Simon, H. and Eisenstadt, S., 2002, 'A Chinese Room that Understands', in Preston and Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, New York: Oxford University Press, pp. 95–108.
Sprevak, M. (2007) 'Chinese Rooms and Program Portability', *British Journal for the Philosophy of Science*, 58(4): 755–776.
Thagard, P. (1986) 'The Emergence of Meaning: An Escape from Searle's Chinese Room', *Behaviorism*, 14: 139–46.

HT Week 4: COMPUTATIONALISM

Required Reading:

Ritchie, J. B., & Piccinini, G. (2018) 'Computational implementation' in M. Sprevak & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind*, New York: Routledge, pp. 192-204.
Sprevak, M. (2010) 'Computation, Individuation, and the Received View on Representation', *Studies in History and Philosophy of Science*, 41: 260–270.

Further Reading:

Chalmers, D. (2011). A computational foundation for the study of cognition. *Journal of Cognitive Science*, 12(4), 323–57.
Churchland, P.S. (1986) *Neurophilosophy*, Cambridge, MA: MIT Press.
Copeland, J. (1996) 'What is Computation?', *Synthese*, 108: 335–359.
Dennett, D. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.
Egan, F. (1992) 'Individualism, computation, and perceptual content', *Mind*, *101*(403), 443-459.
Piccinini, G. (2015). *Physical computation*: *A mechanistic account*. Oxford: Oxford University Press.
Putnam, H. (1960). Minds and machines, In S. Hook (Ed.), *Dimensions of mind: A symposium* (pp. 138–164). New York: Collier.
Rescorla, M. (2014). A theory of computational implementation, *Synthese*, 191, 1277–1307.
Searle, J. (1990). Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association*, 64, 21-37.
Shagrir, O. (2018). The brain as an input–output model of the world. *Minds and Machines*, 28(1), 53-75.
Sprevak, M. (2018). Triviality arguments about computational implementation. In M. Sprevak & M. Colombo (Eds.), *Routledge handbook of the computational mind*. London: Routledge

HT Week 5: SCIENCE AND AI

Required Reading:

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, *346*(6206), 171-172.
Gillies, D. (1996) *Artificial Intelligence and Scientific Method*, Oxford: OUP, ch. 2.

Further Reading:

Buckner, C. (2019). 'Deep learning: A philosophical introduction', *Philosophy Compass*, *14*(10), e12625.
Bridewell, W. and Langley, P. (2010) 'Two Kinds of Knowledge in Scientific Discovery', *Topics in Cognitive Science*, vol. 2: 36–52.
Darden, L. (1997). Recent work in computational scientific discovery. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 161-166). Mahwah, New Jersey: Lawrence Erlbaum.
Glymour, C., Scheines, R., & Spirtes, P. ([1986]2014). *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. Academic Press.
Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987) *Scientific Discovery: Computational Explorations of the Creative Processes*. MIT press.
Langley, P. and Arvay, A. (2019) 'Scientific Discovery, Process Models, and the Social Sciences', In *Scientific discovery in the social sciences* (pp. 173-190). Springer, Cham.
Rissanen, J. (1983) 'Universal prior for integers and estimation by minimum description length', *The Annals of Statistics*, 11(2), 416–431.
Shanahan, M. (1997), *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press.
Thagard, P. (1993). *Computational philosophy of science*. MIT press.
Votsis, I. (2016) 'Ad hoc Hypotheses and the Monsters within', in V. C. Müller (ed.), *Fundamental Issues of Artificial Intelligence (Synthese Library)*, 2016, Berlin: Springer, pp. 299-313.

HT Week 6: CREATIVITY AND AI

Required Reading:

Boden, M. A. (2014). *Creativity and Artificial Intelligence: A Contradiction in Terms?* in E. S. Paul and S. B. Kaufman (Eds.), *The Philosophy of Creativity: New Essays*. Oxford University Press, New York, NY, pp. 224-244.
Elkins, K., & Chun, J. (2020). 'Can GPT-3 Pass a Writer's Turing Test?', *Journal of Cultural Analytics*, vol. 5(2): 1-16.

Further Reading:

Carnovalini, F., & Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3(14): 1-20.
Carruthers, P. (2011). Creative action in mind. *Philosophical Psychology*, *24*(4), 437-461.
Dartnall, T. (Ed.). (2013). *Artificial intelligence and creativity: An interdisciplinary approach* (Vol. 17). Springer Science & Business Media.
Meheus and T. Nickles (eds), *Models of Discovery and Creativity*, Dordrecht: Springer, 127–66.
Nersessian, N. (2009) 'Conceptual Change: Creativity, Cognition, and Culture', *Hypatia*: 226–37.
Thagard, P. (2012). Creative combination of representations: Scientific discovery and technological invention. *Psychology of science: Implicit and explicit processes*, 389-405.
Stokes, D., (2011). 'Minimally Creative Thought', *Metaphilosophy*, 42: 658–81.
Van Der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). 'Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory', *Music & Science*, vol. 1: 1-18.

HT Week 7: READING WEEK

HT Week 8: EMOTION AND AI

Required Reading:

DeLancey, C. (2002) *Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence*, Oxford: OUP, Chs. 12-13.
Scarantino, A. (2004). Book Review of Craig DeLancey's Passionate Engines: What Emotions Reveal about the Mind and Artificial Intelligence, *Philosophy of Science*, 71(2), 227-230.

Further Reading:

Barrett, L. F. (2017) *How Emotions are Made: The Secret Life of The Brain*, New York, NY: Houghton-Mifflin-Harcourt.
DeLancey, C. and A. Kovach (2005) 'On Emotions and the Explanation of Behavior', *Noûs*, 39 (1): 106-122.
Goldie, P. (2007) 'Emotion', *Philosophy Compass*, *2*(6), 928-938.
Evans, D. (2002) *Emotion: The science of sentiment*. Oxford University Press, USA.
Hurley, S. L. (1998) Consciousness in Action, Cambridge, MA: Harvard University Press.
Hutto, D. D. and E. Myin (2013) *Radicalizing Enactivism: Basic Minds without Content*, Cambridge, MA: MIT Press.
Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster.
Todd, C. (2014). Emotion and value. *Philosophy Compass*, 9(10), 702-712.

HT Week 9: SUPERINTELLIGENCE

Required Reading:

Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*, Oxford, UK: Oxford University Press, Ch. 2 and 9.
Russell, S. (2017). 'Artificial intelligence: The future is superintelligent'. *Nature*, 548(7669), 520-521.

Further Reading:

Batin, M., Turchin, A., Sergey, M., Zhila, A., & Denkenberger, D. (2017). *Artificial intelligence in life extension: from deep learning to superintelligence*. Informatica, 41(4).
Chalmers, D. (2009). The singularity: A philosophical analysis. *Science fiction and philosophy: From time travel to superintelligence*, 171-224.
Everitt, T., G. Lea, and M. Hutter (2018) 'AGI Safety Literature Review', in *International Joint Conference on Artificial Intelligence (IJCAI)*, arXiv: 1805.01109.
Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, *62*, 729-754.
Hardcastle, V.G. (ed.) (2012) Symposium on David Chalmer's The Singularity, *Journal of Consciousness Studies*, 19(7-8).
Kurzweil, R. (1999) *The Age of Spiritual Machines: When computers exceed human intelligence*, New York, Viking Press.
Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental issues of artificial intelligence* (pp. 555-572). Springer, Cham.

HT Week 10: SCEPTICISM AND COMPUTER SIMULATIONS

Required Reading:

Bostrom, N. (2003) 'Are we Living in a Computer Simulation?', *The Philosophical Quarterly*, 53(211), 243-255.
Moskowitz, C. (2016) 'Are we living in a Computer Simulation?', *Scientific American,* April 7th, 2016, https://www.scientificamerican.com/article/are-we-living-in-a-computer-simulation/

Further Reading:

Beisbart, C. (2014) 'Are we sims? how computer simulations represent and what this means for the simulation argument.' *The Monist* 97.3: 399-417.
Birch, J. (2013) 'On the "Simulation Argument" and Selective Scepticism.' *Erkenntnis* (1975-), vol. 78, no. 1, pp. 95–107.
Greene, P. (2020) 'The Termination Risks of Simulation Science', *Erkenntnis*, *85*(2), 489-509.
Hanson, R. (2001) 'How to live in a simulation.' *Journal of Evolution and Technology* 7.1 (2001): 3-13.
Kipping, D. (2020) 'A Bayesian Approach to the Simulation Argument', *Universe* 6.8 (2020): 109.
Lewis, P. J. (2013) 'The doomsday argument and the simulation argument', *Synthese* 190.18: 4009-4022.
Mizrahi, M. (2017). The fine-tuning argument and the simulation hypothesis. *Think*, 16(46), 93-102.
Richmond, A. M. (2017) 'Why doomsday arguments are better than simulation arguments', *Ratio* 30.3 (2017): 221-238.
Steinhart, E. (2010) 'Theological implications of the simulation argument', *Ars Disputandi* 10.1: 23-37.
Weatherson, B. (2003) 'Are You a Sim?', *The Philosophical Quarterly*, vol. 53, no. 212, pp. 425–431.

HT Week 11: CONSOLIDATION LECTURE

HT Week 12: COLLECTIONS

APPENDIX A COMMON ASSESSMENT MARKING SCHEME

Distinction		**Merit**		**Pass**		**Fail**	
100	Highest possible distinction	68	High merit	58	High pass	48	High fail
90	Extremely high distinction	65	Mid merit	55	Mid pass	42	Mid Fail
85	Very high distinction	62	Low merit	52	Low pass	35	Clear Fail
80	High distinction					20	Fail
75	Mid distinction					5	Almost no attempt
72	Low distinction					0	No attempt Late Submission