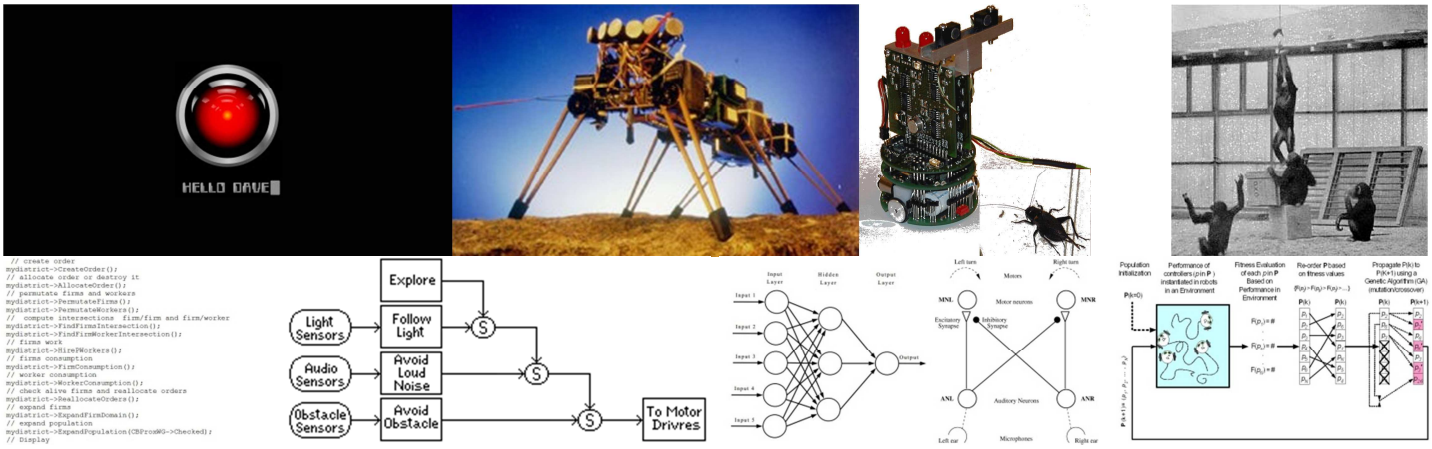

PHIL 447: AI, Robots, and People
TR, 9:40-11:00 am
University of New Hampshire
Professor William Seeley

Office Hours: 321 Nesmith, T/Th 11:00-12:00

william.seeley@unh.edu

www.bates.edu/~wseeley

Can a computer think? If so could it ever learn to think thoughts it hadn't been programmed to think? Can a robot learn to autonomously guide its own behavior in a novel environment, one it wasn't programmed for? What, if anything, can the answers to these and similar questions about teach us about the nature of intelligent behavior or about what we ourselves are like as as conscious thinking beings?

Classical AI models intelligence on the human capacity for rational decision making and general problem solving. On this account thoughts are constructed from abstract symbols that represent aspects of the environment, and minds are treated as general purpose symbol manipulation systems that can be realized in any of a potentially infinite number of natural organisms or artificial systems. Although this computational model of mind has been a powerful research tool in philosophy and cognitive science, it has proven difficult to implement in computer simulations and autonomous robots.

We begin the course by introducing and evaluating Classical AI. We then turn to a range of alternatives models that have emerged from research in autonomous robotics. We conclude the course with a discussion of the potential social consequences of living with smart technology, artificial intelligence, and autonomous robots. Along the way we will encounter some traditional philosophical questions about minds and artificial intelligence. What is the nature of the relation between minds and bodies? Does a system have to be self-conscious to count as an intelligent system? What could it even mean for an autonomous robot to be self-conscious or experience emotions? Or more generally, how does the philosophical concept of a person frame our analysis of the possibility of artificial intelligence.

Topics covered include: the nature of intelligence, the computational theory of mind, classical AI and the symbol system hypothesis, behavior-based robotics, artificial neural networks, genetic algorithms, dynamic systems, embodied cognition, the nature of representation, and the concept of a person.

Requirements:
Students will be asked to write two 3-page papers (worth 15% of the final grade each); a take-home midterm (worth 30% of the final grade each); and a cumulative final exam (worth 40% of the final grade). In addition, there will be five short pop-quizzes on dates to be determined (cumulatively worth 10% of the final grade each). Students will be given a choice between two assigned topics for each paper.

Attendance:
Class participation is a requirement. Attendance is a minimum requirement for class participation. Therefore, excesses absences will affect a student's final grade. Class participation will be factored in as +/- 10% of the final grade

Required Texts:

- Andy Clark, *Being There* (New York: Oxford University Press, 1998).
- *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. Patrick Lin, Keith Abney, and George A. Bekey (Cambridge, MA: MIT Press, 2014).

SYLLABUS: What follows is a provisional syllabus. This schedule may change, as may the readings within it, as a result of class interests and time constraints. In fact it is likely to change in accordance with the pace of our day to day conversations and discussions in class. Any changes will be announced in class prior to the date of the change of assignment. Also, importantly, I will let you know each day how to approach the readings for the next class, whether some readings are more important than others, and, occasionally, whether we will drop a reading from a session (it has happened). You can't get this information if you are not in class.

Week 1: 01/26 & 01/28
Course introduction. The goal of these sessions is to introduce students to a definition of intelligence and some fundamental concepts in the philosophical study of artificial intelligence.

Week 2: 02/02 & 02/04
The goal of class readings and discussion this week is introduce students to *classical AI*, or what philosophers call g*ood old fashioned artificial intelligence* (*GOFAI)*. We will discuss *the symbol system hypothesis*, then nature of *algorithms*, and the idea of *functional equivalence*. The central claim of this view in AI is that the procedures governing the behaviors of computing machines are functionally equivalent to those governing natural intelligence.

Week 3: 02/09 & 02/11
The goal of class readings and discussion this week is to introduce students to the Chinese Room and symbol grounding problems for classical AI. These puzzles raise a question about whether robots and AI programs could understand what they are doing.

Week 4: 02/16 & 02/18
The goal of class readings and discussion this week is introduce students to the frame problem in AI research. The frame problem is a puzzle about how to best represent knowledge for easy, efficient retrieval on the fly that would enable machines to recognize only what is relevant in the environment for their current actions, to track change in what is relevant in the environment, and to understand the consequences of their actions. Common sense enables us to do this easily, but it turns out to be difficult for symbol systems.

Week 5: 02/23 & 02/25
The goal of class readings and discussion this week is to introduce *behavior-based robotics* as an alternative to classical AI and evaluate whether this approach resolves the symbol grounding and frame problems. Along the way we will discuss some foundational concepts in *embodied cognition* to help understand how research with these autonomous robots relates to natural intelligence.

Week 6: 03/01 & 03/03
The goal of class readings and discussion this week is to continue the discussion of behavior-based robotics and intelligence. We will look at Braitenberg Vehicles and phonotaxis in crickets (a technique female crickets use to locate their mates in a noisy environment). Braitenberg vehicles will give us a better sense of how these robots work and cricket behavior will give us a sense of these robots can be used to model intelligent behavior.

Week 7: 03/08 & 03/10
The goal of class readings and discussion this week is to discuss how research in embodied cognition and behavior-based robotics scales up to human intelligence. We start this week by discussion research that pairs dynamic systems theory and developmental psychology.

Week 8: 03/22 & 03/24
The goal of class readings and discussion this week is to continue the discussion of autonomous robots and embodied cognition with an eye to the nature of representation and the role of language in intelligent behaviors.

Week 9: 03/29 & 03/31
The goal of class readings and discussion this week is to discuss the role played by the body in information processing in embodied systems. This is often called *morphological computation*. The general idea is that the shapes of our bodies have evolved to facilitate locating and encoding information, and in many cases act as cognitive shortcuts to simplify cognition.

Week 10: 04/05 & 04/07
The goal of class readings and discussion this week is to introduce students to foundational concepts in artificial neural network modeling as an alternative to behavior-based robotics and discuss the nature of learning and representation in these systems.

Week 11: 04/12 & 04/14
The goal of class readings and discussion this week is to introduce students to *evolutionary robotics* as an alternative to Hebbian and supervised learning algorithms in artificial neural networks. Evolutionary robotics uses *genetic algorithms* to optimize the behaviors of autonomous robots. These algorithms search the total range of possible weights an artificial neural network could adopt to settle on the most efficient configuration for the behaviors they encode.

Week 12: 04/19 & 04/21
We turn to robot ethics this week. The goal of class discussion and readings is to develop a sense of what it means to ask if ethics is computable and what it would take to program an autonomous robot as an *Autonomous Moral Agent*.

Week 13: 04/21 & 04/26

The goal of class discussion and readings is to discuss some ethical problems that have emerged in discussions of autonomous military drones, self-driving cars, and robot care givers.

Week 13: 05/03 & 05/05

The goal of class discussion and readings is to introduce the concept of moral personhood, the concept of a person more generally, and to discuss how these concepts influence our evaluation of intelligent machines.

ASSIGNMENTS: All assignments must be handed in both in hard copy and electronically via the dropbox for that assignment on CANVAS

Writing Assignment 1: The Chinese Room Problem

Please write a 3-4 page (double-spaced, 12 point font, 1" margins) paper (1000-1200 words) on the following topic: Does the robot reply add anything to the Outdoor$_1$ version of the Chinese Room argument as we defined it in class. In composing your answer make sure to discuss any relevant logical problems with the Chinese Room argument, identify Searle's real point in the Chinese Room argument, explain how the Robot Reply is supposed to answer Searle, and explain why you think the robot reply either succeeds or doesn't.

Please upload your paper to the assignment dropbox on Canvas by the due date.
Due Date: Sunday, February 21 @ 11:55pm

Mid Term Take Home Exam

Please upload your exam to the assignment dropbox on Canvas by the due date.
Due date: Thursday April 14th at 5pm on CANVAS

Writing Assignment 2: Robot Ethics

Please write a 3 page (double-spaced, 12 point font, 1" margins) paper (1000 words) on the following topics: Does the possibility of functionally moral robots run afoul of the symbol grounding problem…or is their a Moral Chinese Problem standing in the way of the possibility of moral robots? In constructing your answer be sure to consider Searle's discussion of whether or not robots could understand the reasons for their actions.

Please upload your paper to the assignment dropbox on Canvas by the due date.
Due Date: Friday, May 6 @ 11:55pm

Final Exam (Cumulative) at the scheduled exam time for the course.

Schedule of Readings:

Date	Readings	Assignments
01/26	Introduction: a short primer on the nature of intelligence ***(no reading)***	
01/28	The Mind Project: Can a machine think? http://www.mind.ilstu.edu/curriculum/ai_can_a_machine_think/ai_machine_think_1.php	
02/02	Haugeland, What is Mind Design?: 8-21 ***(e-reserve)***	
02/04	Haugeland, What is Mind Design?: 8-21 ***(e-reserve)***	
02/09	Searle, Minds, Brains, Programs: 183-196. ***(e-reserve)***	
02/11	Copeland, The Chinese room from a Logical…: 109-122. ***(e-reserve)***	1st Paper topics distributed
02/16	Pfeifer & Scheier, The Frame Problem: 63-69. ***(e-reserve)*** Boden, Some philosophical problems: 769-775. ***(e-reserve)*** Copeland, Knowledge Representation: 91-95. ***(e-reserve)***	
02/18	Pfeifer & Scheier, The Frame Problem: 63-69. ***(e-reserve)*** Boden, Some philosophical problems: 769-775. ***(e-reserve)*** Copeland, Knowledge Representation: 91-95. ***(e-reserve)***	1st paper due: Sunday Feb 21 @ midnight
02/23	Brooks, Intelligence without Representation: 395-420. ***(e-reserve)*** * Brooks, Elephants Don't Play Chess?: 3-15. ***(e-reserve)*** * Pfeifer & Scheir, The Subsumption Architecture: 199-225. ***(e-reserve)***	
02/25	Brooks, Intelligence without Representation: 395-420. ***(e-reserve)*** * Brooks, Elephants Don't Play Chess?: 3-15. ***(e-reserve)*** * Pfeifer & Scheir, The Subsumption Architecture: 199-225. ***(e-reserve)***	
03/01	Braitenberg, Vehicles 1-4: 1-19. ***(online)*** https://drive.google.com/a/bates.edu/file/d/0B8DzH-_rb9JvQWUtNXdjM2VCaTQ/edit Webb & Harrison, Phonotaxis in crickets and robots: 533-552. ***(online)***	
03/03	Webb & Harrison, Phonotaxis in crickets and robots: 533-552. ***(online)*** * Webb, A Cricket Robot: 94-99. ***(online)*** http://users.sussex.ac.uk/~ezequiel/iam/webb1996.pdf *Huber & Thorson, Cricket Auditory Communications: 60-68. ***(online)***	
03/08	Clark, The Situated Infant: 35-51. ***(C)***	
03/10	Clark, Mind and World: 53-69. ***(C)***	
03/15	***SPRING BREAK***	
03/17	***SPRING BREAK***	
03/22	Clark, Being, Computing, Representing: 143-175. ***(C)***	Mid-term distributed
03/24	Clark, Language: The Ultimate Artifact: 193-218. ***(C)***	
03/29	Pfeifer & Bongard, Ecological Balance: 123-137. ***(e-reserve)***	
03/31	Pfeifer & Bongard, From Locomotion to Cognition: 145-164 (***e-reserve)***	Mid-term's Due Sunday April 3 @ midnight

04/05	The Mind Project, Connectionism: An Introduction http://www.mind.ilstu.edu/curriculum/connectionism_intro/connectionism_1.php Pfeifer & Scheier, Neural Networks for Adaptive Behavior ***(e-reserve)***	
04/07	The Mind Project, Connectionism: An Introduction http://www.mind.ilstu.edu/curriculum/connectionism_intro/connectionism_1.php Pfeifer & Scheier, Neural Networks for Adaptive Behavior ***(e-reserve)***	
04/12	Pfeifer & Schier: Artificial Evolution & Artificial Life (excerpt): 227-241. (***e-reserve)***	
04/14	Pfeifer & Schier: Artificial Evolution & Artificial Life (excerpt): 227-241. (***e-reserve)***	
04/19	Moor, Is Ethics Computable? ***(e-reserve)***	
04/21	Allen & Wallach, Moral Machines: 55-68. ***(DAB)***	
04/26	Sharkey, Killing made Easy, Joysticks to Politics: 111-128. ***(DAB)*** Self-driving cars and the trolley problem https://www.youtube.com/watch?v=FypPSJfCRFk	2nd paper topics distributed
04/28	Borenstein & Pearson, Robot Caregivers: Ethical Issues: 251-166. ***(DAB)***	
05/03	Sparrow, Can Machines be People: 301-316. ***(DAB)***	
05/05	***Discussion: intelligence and moral behavior in AI & Robots***	2nd Paper Due Friday May 6th @ midnight
	FINAL EXAM	