TEXTS:

- Haugeland, John; *Artificial Intelligence: The Very Idea.* (Hereafter referred to as AI:TVI)
- Boden, Margaret, ed.; *The Philosophy of Artificial Intelligence.* (Hereafter PAI)
- Dreyfus, Hubert; *What Computers Still Can’t Do.* (Hereafter CCD)
- Johnson, Deborah; *Computer Ethics.* (Hereafter CE)
- Seech, Z.; *Writing Philosophical Papers*

Hand-out readings will also be distributed occasionally in class.

(Reading assignments will ordinarily range from 50 to 100 pages per week.)

GENERAL COURSE DESCRIPTION:

This course covers major issues in what has come to be called *Cognitive Science* -- the study of the conditions for the possibility of artificially producing "intelligent" computers (or any intelligence whatsoever). In addition, social and ethical issues relating to computer science will be separately considered.

COURSE REQUIREMENTS:

- Grades will be assigned on the basis of class participation, including discussion, homework, examinations and papers. Since some material will be presented only in class, class attendance is essential.

- There will be two examinations (a mid-term and a final), there will be two papers due (3-6 pages each), several small homework assignments, including reading questions. Completing the required reading assignments is essential to passing this course.

Following is a tentative schedule of topics (weeks and topics are approximate):

BACKGROUND

24 August (first day):

Introduction, course requirements, grading practices, texts, course topics, purpose of the course, etc. Overview of topics.
26 August:


**Readings:** (PAI: *Introduction*)

Weeks 2 & 3:


**Readings:** (Dreyfus, CCD, *Introduction* (pp. 67-87); Introduction to the MIT Press Edition (pp. ix-lii), AI:TVI, Intro & Ch 1; Seeth, Intro, Ch1, Ch2 (through pp.23).

Week 4:


**Readings:** (AI:TVI, Ch. 2)

Meanings and interpretations of Formal Systems: Syntax vs. Semantics, where do meanings come from? What are they, anyway? Internal vs external meanings. Interpretations and truth. Assumptions of AI: intelligence can be wholly captured by rule governed systems, such systems are media independent, etc. Classical AI corollaries.

**Readings:** (AI:TVI, Ch 3).

Week 5:

The philosophical importance of Turing Machines, what Turing machines are, equivalence notions, Universal Turing Machines, Computability, Church/Turing thesis. How do we know the Church/Turing thesis applies to human beings (pro/con)?

**Readings:** (AI:TVI, Ch 4)

Does computer architecture matter? Symbolic Computing, Von Neumann Architecture; Neural Nets, Connectionist models.

Landmark Attempts at AI: Heuristic Search (Logic Theorist, General Problem Solver, Chess programs), Turing tests and the "Doctor" programs (ELIZA), Micro-worlds (SHRDLU), Schank's "Frames."

**Readings:** (AI:TVI, Ch.'s 4 & 5).
Week 6:

Examining Strong AI assumptions: What are people really like?? Rule governed action vs. holistic action; knowing how and knowing that (Can you know how and not that?); learning and expertise; the intricate and unisolatable network of human knowledge and meaning; finite provinces of meaning; common sense (everyday life) as foundation ("paramount reality") of all meaning; self-awareness, imagination, feelings, moods, strivings, purposes, will.

**Readings:** (Dreyfus; CCD, *Introduction to the Revised Edition* pp. 1-65)

**THE CHALLENGE OF HUMAN CAPABILITIES**

(Weeks 7 - 12)

The Turing Test (PAI: Turing, "Computing Machinery and Intelligence"; Handout: "Lessons from a Restricted Turing Test."). Searle's Criticism of AI (PAI: "Minds, Brains, and Programs")

The Knowledge Representation Problem: Frames and Scripts. **Readings:** PAI: Dennett; "Cognitive Wheels: The Frame Problem of AI"

*Connectionism*: A new paradigm for AI: Neural Nets, how they work; What they explain. NN shortcomings: Fodor/Smolensky debate. **Readings:** (Copland, Jack; AI's Fresh Start, Parallel Distributed Processing (handout); Dreyfus and Dreyfus, "Making a Mind vs. Modeling a Brain", Ch. 13 in PAI)


**INFORMATION ETHICS**

Weeks 13-15


*Ethics and Computers:*

**Piracy:** copying software: why is it wrong? The swimming pool analogy; loss of autonomy. What sort of intellectual property is software? Legal protections: trade secrets, copyrights, patents. The Lotus "look and feel" problem.

**Privacy:** The Dangerous Dossier -- The risks of anonymous data accumulation: miscollection, misinterpretation, faulty distribution. Existing dangers: True Name Theft. Denial of service, employment.

**Misuse:** What do hackers ethically do wrong? The no-harm, no foul fallacy. How secure must a computer system be?
Professional Ethics: Risks and Liabilities – what are the ethical responsibilities of the computer scientist? Who is responsible when something goes wrong? Diffuse responsibility, following orders. Contractual responsibilities.

Social Implications of Information Technology: Computers as tools, as extensions of one’s intellect. Dangers of dependence and social over-reliance on information technology. Computation as a substitute for thought. The costs of technology - cultural and fiscal. Computers as an agent of good or evil.

Readings: CE.

Course Requirements:

In addition to requisite reading and class attendance, this course requires two substantial papers, a mid-term and a final, reading assignments and homework assignments, including questions drawn from readings to be turned in from each reading assignment, to be weighted as follows:

- Paper I: 20%
- Paper II: 25%
- Mid-term: 20%
- Reading Questions: 5%
- Homework and discussion: 5%
- Final: 25%

Papers will be 3-6 typed, double-spaced, pages in length on topics approved in advance. The first paper will be due the end of the 6th week of class. Paper II will be due about the 12th week of class. Papers will be graded on philosophical acumen, argumentation, coherence, English usage, and style. Late papers will be severely penalized.