Welcome to CIS 521

- **Professor:** Mitch Marcus, mitch@<cis standard>
- **Levine 503**
- **TAs:**
  - Head TA: Toma Pigli
  - TAs: Gedas Bertasius, Dean Fulgoni, David Liao, Daniel Moroz, Eddie Smith
- Email addresses and office hours will be announced on Piazza
- **Course Administrator:** Cheryl Hickey, cherylh@<seas standard>
  Levine 502, 215-898-3538

---

Grading & Homework

**Grading:**
- 50% Homeworks
- 25% Midterm 1
- 25% Midterm 2

**Homework:**
- Homework will be due at 11:59 on specified dates.
- You can submit up to two homeworks up to 2 days late, but extensions after that will be granted only for true emergencies.
- Your lowest homework grade will be dropped.

**ALL HOMEWORKS MUST BE YOUR OWN INDEPENDENT WORK**
- Violations of Penn's Code of Academic Integrity and in particular academic dishonesty as defined in the Code of Integrity will not be tolerated. PENALTIES WILL BE SUBSTANTIAL

---

On to the Real Stuff:

**WHAT THE COURSE IS AND ISN’T ABOUT**

"I want to design a machine that will be proud of me – Danny Hillis"
“I want to design a machine that will be proud of me – Danny Hillis”

“Startup Funded $143M to Create Sentient Computing” – EETimes 12/2014

- Now a startup with $143 million in funding [is] describing a sentient distributed artificial intelligence that sounds like a nice-guy version of Skynet from the cinema flick Terminator. According to the technology gurus at Sentient Technologies Holdings Ltd. of San Francisco, the software for sentient computers, which they are already installing at key customer sites, goes beyond natural language recognition, unstructured searching, machine learning, and deep knowledge.

- "Reasoning and logic are one thing, but beyond that is true intelligence -- what we call sentience," Babak Hojat, cofounder and chief scientist tells EE Times.

- "Sentience is being aware, having perceptions, being mindful, and has implications of autonomy," chief technology officer Nigel Duffy said.

Recent Significant Advances In NLP

- IBM’s Watson
  - Web-scale information extraction & question answering

- Apple’s Siri
  - Interactive Dialogue Systems

- Google Translate
  - Automatic Machine Translation

Broadcast Monitoring
BBN MAPS & Language Weaver MT
Older Real Accomplishments of AI

- 1991: AI Logistics Planning for Gulf War
- 1997: Deep Blue defeated the reigning world chess champion Garry Kasparov
- 1998: Deep Space 1 (launched) – Remote Agent Experiment

“Invisible” AI
- Computer Algebra Systems (Maple, Mathematica)
- Machine Learning
  - Credit Evaluation, Fraud Detection
  - Internet Search, Spam Filtering
  - Handwritten character recognition (checks, US mail)

What is AI?
Views of AI fall into four categories:

<table>
<thead>
<tr>
<th>Thinking humanly</th>
<th>Thinking rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acting humanly</td>
<td>Acting rationally</td>
</tr>
</tbody>
</table>

"We will focus on "acting rationally"

Acting humanly: Turing Test

- Turing (1950) "Computing machinery and intelligence": “Can machines think?” “Can machines behave intelligently?”
- Operational test for intelligent behavior: the Imitation Game

- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated most major arguments against AI
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Acting humanly: Social robots

- Cynthia Breazeal: MIT
Thinking humanly: Cognitive modeling

- 1960s "cognitive revolution": Information-processing psychology, a.k.a. cognitive psychology
- Requires scientific theories of internal activities of the brain
- How to validate? Requires
  1) Predicting and testing behavior of human subjects
  2) Direct identification from neurological data (bottom-up)
- Both approaches are now distinct from AI
- Caveat: ACT-R & SOAR communities do computational modeling of high level mental functions

Thinking rationally: "Laws of thought" aka Logic

- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of logic: notation and rules of derivation for thoughts; may or may not have proceeded to the idea of mechanization
- Direct line through mathematics and philosophy to modern AI
- Problems:
  1. Not all intelligent behavior is mediated by logical deliberation
  2. What is the purpose of thinking? What thoughts should I have?
  3. Ignores the hard problem of perception
  4. All attempts to encode what we know in logic have failed
  5. Most logical inference is intractable

Acting rationally: rational agents

- Rational behavior: doing the right thing
- The right thing: that which is expected to maximize goal achievement, given the available information
- Doesn't necessarily involve thinking – e.g., blinking reflex – but thinking should be in the service of rational action

Rational agents

Rational agent: An agent is an entity that perceives and acts

This course is about effective programming techniques for designing rational agents

Course Overview – First Half

Module 0: Introduction
- Intelligent Agents
- Python Programming

Module 1: Search Strategies
- Uninformed & Informed Search (Homeworks: Puzzle Solvers)
- Constraint Satisfaction (Homework: Sudoku Solver)
- Adversarial Search

Course Overview – Second Half

Module 2: Machine Learning and Natural Language Processing
- Review of Probability
- Naive Bayes (Spam Filtering) & Bayesian Networks (Homework: Build a spam filter)
- Perceptrons and Support Vector Machines
- Hidden Markov Models & Part of Speech Tagging (Homework: Generate fake Frankenstein text, Build Part of Speech Tagger)

Module 3: Knowledge Representation and Logic
- Logical Agents (Homework: Logic Puzzle Solver)
- The Singularity: A critique
Two Approaches to AI

- **Logical representations (Modules 1&3)**
  - Dominant BEFORE 1995
  - Relations between entities
    - "Mitch's bicycle is red"
    - (isa B3241 bicycle) (color B3231 red) (owns B3241 P119)
    - (isa P119 person) (name P119 'Mitch')
  - Explicit logical models
  - Logical inference, Search
  - Chess, Sudoku, computer games, ...

- **Statistical models (Module 2)**
  - Dominant SINCE 2000
  - Prediction by look-up or by weighted combinations
    - P(y=bicycle) = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 + …
  - Machine Learning, Machine vision, speech recognition, …

The last lecture: Kurweil's singularity vision

"With artificial intelligence we're summoning the demon"
- Elon Musk

"Full artificial intelligence could spell the end of the human race"
- Steven Hawking

Intelligent Agents

AIMA, Chapter 2.1-2.2

Agents and environments

- An agent is specified by an agent function f: P → a that maps a sequence of percept vectors P to an action a from a set A:
  - \[ P = [p_0, p_1, ... , p_t] \]
  - \[ A = [a_0, a_1, ... , a_k] \]

Agents

- An agent is anything that can be viewed as
  - perceiving its environment through sensors and
  - acting upon that environment through actuators

- **Human agent:**
  - Sensors: eyes, ears, ...
  - Actuators: hands, legs, mouth, ...

- **Robotic agent:**
  - Sensors: cameras and infrared range finders
  - Actuators: various motors

- **Agents include humans, robots, softbots, thermostats, …**
Agent function & program

- The *agent program* runs on the physical *architecture* to produce *f*
  - agent = architecture + program
- “Easy” solution: table that maps every possible sequence $Y$ to an action $a$
  - One small problem: exponential in length of $Y$

Rational agents II

- **Rational Agent**: For each possible percept sequence, a rational agent should select an action that is expected to *maximize* its *performance measure*.
- **Performance measure**: An objective criterion for success of an agent’s behavior, given the evidence provided by the percept sequence.
- A performance measure for a vacuum-cleaner agent might include one or more of:
  - +1 point for each clean square in time $T$
  - +1 point for clean square, -1 for each move
  - -1000 for more than $k$ dirty squares

Rationality is not omniscience

- **Ideal agent**: maximizes *actual* performance, but needs to be *omniscient*.
  - Usually impossible…..
    - But consider tic-tac-toe agent…
  - Rationality ≠ Success
- Agents can perform actions in order to modify future percepts so as to obtain useful information (*information gathering, exploration*)
- **Caveat**: computational limitations make *perfect rationality* unachievable
  - design best *program* for given machine resources