

# **CPS 570: Artificial Intelligence**

<http://www.cs.duke.edu/courses/fall14/compsci570/>

## **Introduction**

Instructor: Vincent Conitzer

# Basic information about course

- WF 10:05-11:20am, LSRC D106
- Text: Artificial Intelligence: A Modern Approach
- Instructor: Vincent Conitzer
  - Research on computational aspects of (micro)economics, game theory, systems with multiple intelligent agents
- TA: Andrew Kephart
  - 2<sup>nd</sup>-year CS Ph.D. student at Duke working with Vince

# Prerequisites

- Comfortable **programming** in language such as C (or C++), Java, ...
- Some knowledge of **algorithmic concepts** such as running times of algorithms; having some rough idea of what NP-hard means
- Some familiarity with **probability** (we will go over this from the beginning but we will cover the basics only briefly)
- Not scared of **mathematics**, some background in discrete mathematics, able to do simple mathematical proofs
- If you do not have a **standard undergraduate computer science background**, talk to me first.
- Well-prepared **undergraduates** are certainly welcome
- You do **not** need to have taken an **undergraduate AI course** (though of course it will help if you have)

# Grading

- Assignments: 35%
  - May discuss with another person; writeup and code must be your own
- Midterm exams: 30%
- Final exam: 30%
- Participation: 5%

# What is artificial intelligence?

- Popular conception driven by science fiction
  - Robots good at everything except emotions, empathy, appreciation of art, culture, ...
    - ... until later in the movie.
  - Perhaps more representative of human autism than of (current) real robotics/AI
    - “It is my belief that the existence of autism has contributed to [the theme of the intelligent but soulless automaton] in no small way.” [Uta Frith, “Autism”]
    - Current AI is also bad at lots of simpler stuff!
    - There **is** a lot of AI work on thinking about what other agents are thinking



# Real AI

- A serious science.
- **General-purpose AI** like the robots of science fiction is incredibly hard
  - Human brain appears to have lots of special and general functions, integrated in some amazing way that we really do not understand at all (yet)
- **Special-purpose AI** is more doable (nontrivial)
  - E.g., chess/poker playing programs, logistics planning, automated translation, voice recognition, web search, data mining, medical diagnosis, keeping a car on the road, ... ..

# Definitions of AI

if our system can be more rational than humans in some cases, why not?

focus on **action**

Systems that **think**  
**like humans**

Systems that **think**  
**rationally**

sidesteps philosophical issues such as “is the system conscious” etc.

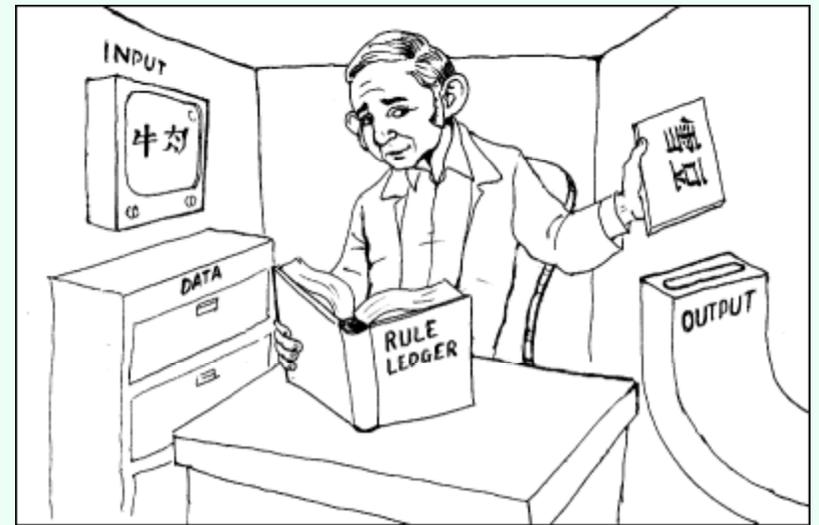
Systems that **act**  
**like humans**

Systems that **act**  
**rationally**

- We will mostly follow “**act rationally**” approach
  - Distinction may not be that important
    - acting rationally/like a human presumably requires (some sort of) thinking rationally/like a human,
    - humans much more rational anyway in complex domains

# “Chinese room” argument [Searle 1980]

image from <http://www.unc.edu/~prinz/pictures/c-room.gif>



- **Person** who knows English but not Chinese sits in room
- Receives notes in Chinese
- Has systematic English **rule book** for how to write new Chinese characters based on input Chinese characters, returns his notes
  - Person=CPU, rule book=AI program, really also need lots of paper (storage)
  - Has no understanding of what they mean
  - But from the outside, the room gives perfectly reasonable answers in Chinese!
- Searle’s argument: the room has no intelligence in it!

# Turing Test

- (Human) judge communicates with a human and a machine over text-only channel,
- Both human and machine try to act like a human,
- Judge tries to tell which is which.
- Numerous variants
- Loebner prize
  - 2013 winner: <http://mitsuku.com/>

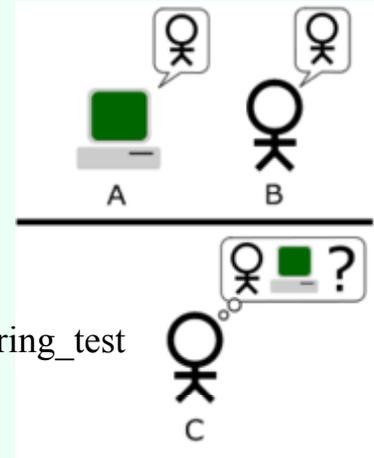


image from [http://en.wikipedia.org/wiki/Turing\\_test](http://en.wikipedia.org/wiki/Turing_test)

# Turing Test on unsuspecting judges

- It is possible to (temporarily) fool humans who do not realize they may be talking to a bot
- ELIZA program [Weizenbaum 66] rephrases partner's statements and questions (~psychotherapist)

# Modern variant: AOLiza

[fragment from <http://www.macworld.com/article/1014418/25aoliza.html>]

**five:** just asking if richard can tell me about some good lobster diving spots that  
it if this is a problem then forget it i dont know whats going on here

**AOLiza:** that is interesting. please continue.

**five:** continue what just ask richard when he gets home about some lobster  
diving up hat way thats all if you got a problem with it then forget it, this isnt worth  
it i dont know whats been going on or what bev has told you but i havent done  
anything to anybody iam the one that got used and dumped

**AOLiza:** tell me more about that.

**five:** about what

**AOLiza:** what do you think?

**five:** theres nothing to tell she used me then got tired of me said i wasnt her  
type anymore i was possessive and controlling

**AOLiza:** perhaps i already know you were possessive and controlling.

**five:** i was huh you nuts i gave her all the room she wanted never told her what  
to do or where to go

# Is Turing Test the right goal?

- “Aeronautical engineering texts do not define the goal of their field as making ‘machines that fly so exactly like pigeons that they can fool even other pigeons.’” [Russell and Norvig]

# Lessons from AI research

- **Clearly-defined** tasks that we think require intelligence and education from humans tend to be doable for AI techniques
  - Playing chess, drawing logical inferences from clearly-stated facts, performing probability calculations in well-defined environments, ...
  - Although, **scalability** can be a significant issue
- **Complex, messy, ambiguous** tasks that come natural to humans (in some cases other animals) are much harder
  - Recognizing your grandmother in a crowd, drawing the right conclusion from an ungrammatical or ambiguous sentence, driving around the city, ...
- Humans better at coming up with **reasonably good** solutions in complex environments
- Humans better at **adapting/self-evaluation/creativity** (“My usual strategy for chess is getting me into trouble against this person... Why? What else can I do?”)

# Early history of AI

- 50s/60s: Early successes! AI can draw logical conclusions, prove some theorems, create simple plans... Some initial work on neural networks...
- Led to **overhyping**: researchers promised funding agencies spectacular progress, but started running into difficulties:
  - **Ambiguity**: highly funded translation programs (Russian to English) were good at syntactic manipulation but bad at disambiguation
    - “The spirit is willing but the flesh is weak” becomes “The vodka is good but the meat is rotten”
  - **Scalability/complexity**: early examples were very small, programs could not scale to bigger instances
  - Limitations of **representations** used

# History of AI...

- 70s, 80s: Creation of **expert systems** (systems specialized for one particular task based on experts' knowledge), wide industry adoption
- Again, overpromising...
- ... led to **AI winter(s)**
  - Funding cutbacks, bad reputation

# Modern AI

- More **rigorous, scientific, formal/mathematical**
- Fewer grandiose promises
- Divided into many **subareas** interested in particular aspects
- More directly connected to “neighboring” disciplines
  - Theoretical computer science, statistics, economics, operations research, biology, psychology/neuroscience, ...
  - Often leads to question “Is this really AI”?
- Some senior AI researchers are calling for **re-integration** of all these topics, return to more grandiose goals of AI
  - Somewhat risky proposition for graduate students and junior faculty...

# Some AI videos

- Note: there is a lot of AI that is not quite this “sexy” but still very valuable!
  - E.g. logistics planning – DARPA claims that savings from a single AI planning application during 1991 Persian Gulf crisis more than paid back for all of DARPA’s investment in AI, ever. [Russell and Norvig]
- <https://www.youtube.com/watch?v=1JJsBFiXGI0>
- <https://www.youtube.com/watch?v=s6VIWDUHTa4>
- [http://www.aaaivideos.org/2007/aibo\\_ingenuity/](http://www.aaaivideos.org/2007/aibo_ingenuity/)
- [http://www.aaaivideos.org/2012/ai\\_vs\\_ai\\_chatbots/](http://www.aaaivideos.org/2012/ai_vs_ai_chatbots/)
- <https://www.youtube.com/watch?v=yJptrlCVDHI>
- <https://www.youtube.com/watch?v=ScXX2bndGJc>

# This course

- Focus on **general AI techniques** that have been useful in many applications
- Will try to **avoid application-specific techniques** (still interesting and worthwhile!)

# Topics

- Search
- Constraint satisfaction problems
- Game playing
- Logic, knowledge representation
- Planning
- Probability, decision theory, game theory, reasoning under uncertainty
- Machine learning, reinforcement learning (briefly, if time allows)

# Nonexhaustive list of AI publications

- General AI conferences: IJCAI, AAAI, ECAI
- Reasoning under uncertainty: UAI
- Machine learning: ICML, NIPS
- Multiagent systems: AAMAS
- Vision: ICCV, CVPR
  
- Some journals: Artificial Intelligence, Journal of AI Research, Machine Learning, Journal of ML Research, Journal of Autonomous Agents and Multi Agent Systems
- AI Magazine

# AI at Duke



- Vince Conitzer
  - Systems with multiple self-interested agents, game theory, economics



- George Konidaris
  - Robotics, planning, reinforcement learning,



- Ron Parr
  - Reasoning under uncertainty, reinforcement learning, robotics



- Carlo Tomasi
  - Computer vision, medical imaging



- Alex Hartemink
  - Computational biology, machine learning, reasoning under uncertainty



- Bruce Donald
  - Computational biology & chemistry



- Sayan Mukherjee
  - Statistics, machine learning
- Duke Robotics, Intelligence, and Vision (DRIV) seminar (=AI seminar)
  - Website: <http://driv.cs.duke.edu/>
  - Mailing list: <https://lists.duke.edu/sympa/info/drive>