

CPS 570: Artificial Intelligence

<http://www.cs.duke.edu/courses/fall17/compsci570/>

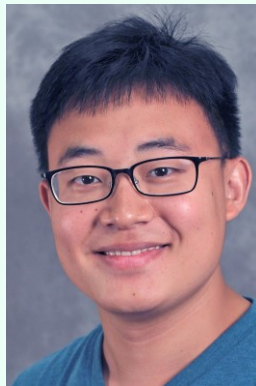
Introduction

Instructor: Vincent Conitzer

Basic information about course

- WF 10:05-11:20am, LSRC D106
- Text: Artificial Intelligence: A Modern Approach
- Instructor: Vincent Conitzer
- TA: Shuzhi Yu (CS Ph.D. student since 2016)
- TA: Rui-Yi Zhang (CS Ph.D. student since 2016)

*Shuzhi
Yu*



*Rui-Yi
Zhang*

Prerequisites

- Comfortable **programming** in general-purpose programming language
- 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%

Some highly visible recent AI successes in games



Watson defeats
Jeopardy champions
(2011)



DeepMind
achieves human-
level performance
on many Atari
games (2015)



AlphaGo defeats Go
champion (2016)



CMU's Libratus defeats
top human poker
players (2017)

Typical picture in news articles



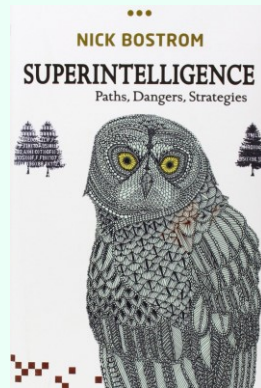
BusinessInsider reporting on the poker match...

Worries about AI - superintelligence



Nick Bostrom
(philosopher at
Oxford)

→
writes



→
influences



Elon Musk

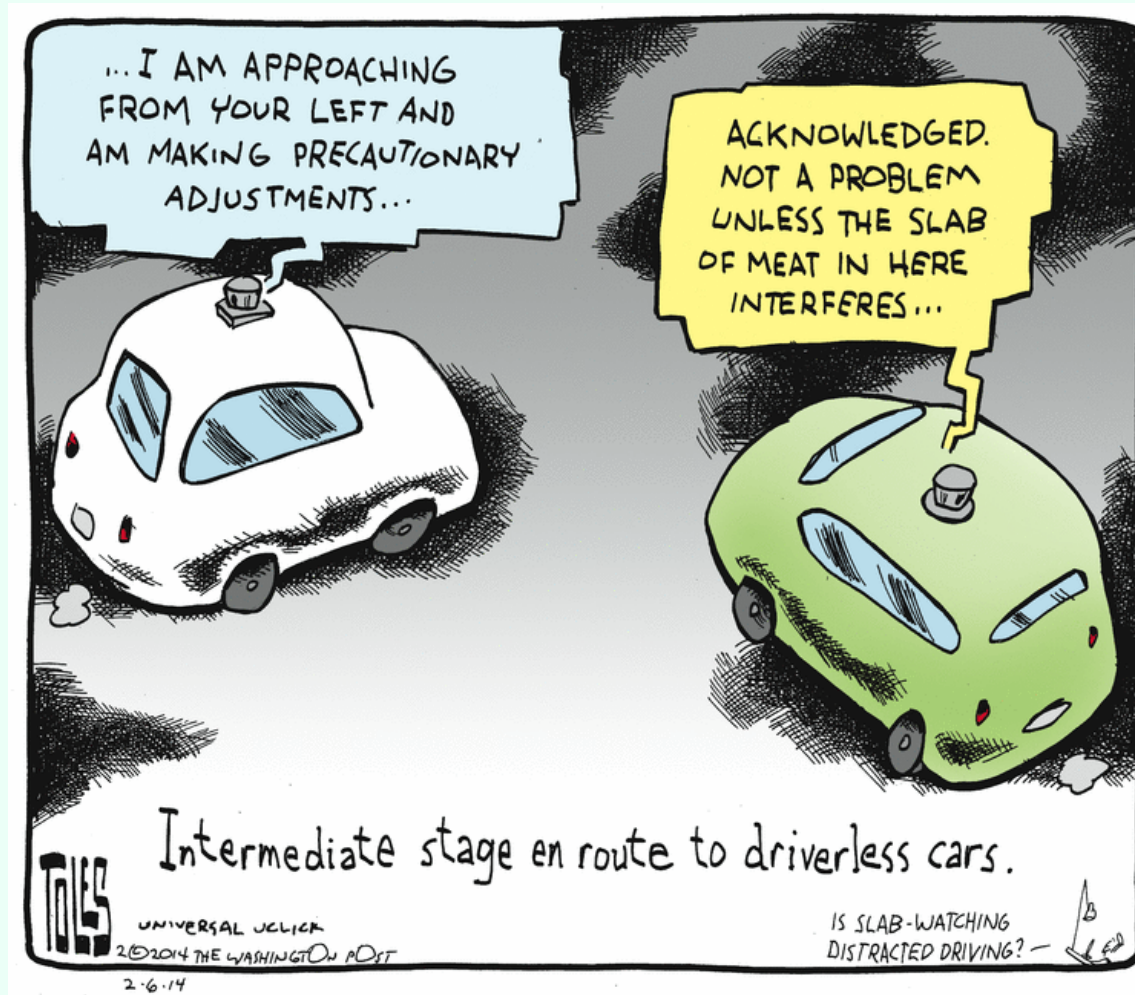
→
donates to



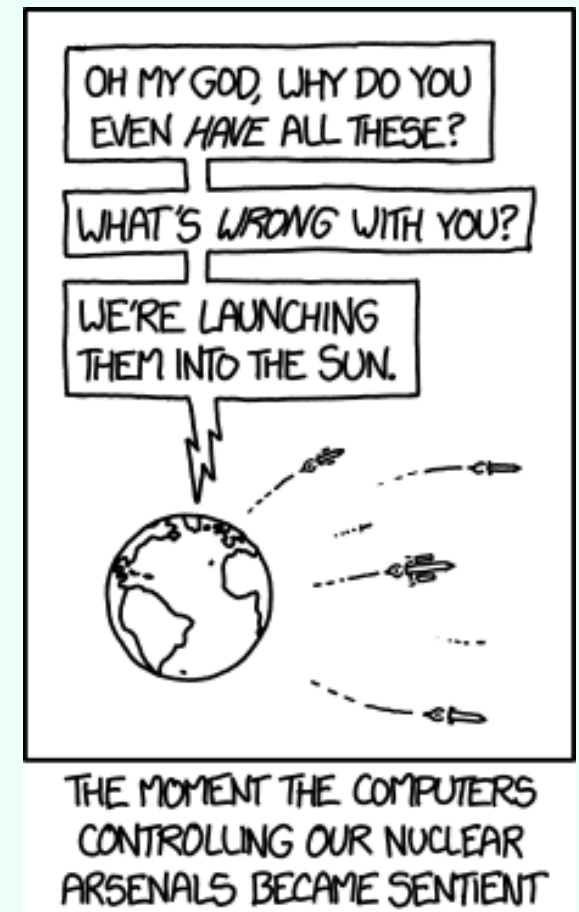


technological unemployment

Worries about AI - near term



autonomous vehicles – legal and other issues



autonomous
weapon systems

...

(Some of my own thoughts)

Prospect
The leading magazine of ideas

HOME BLOGS POLITICS ECONOMICS & FINANCE WORLD ARTS & BOOKS LIFE

HOME > SCIENCE & TECHNOLOGY

Artificial intelligence: where's the philosophical scrutiny?

AI research raises profound questions—but answers are lacking

by Vincent Conitzer / May 4, 2016 / [Leave a comment](#)



A humanoid robot, equipped with an artificial intelligence, helps a teacher with a science class at Kelo University Kindergarten in Shibuya Ward, Tokyo on 25th January, 2016 ©Miho Ikeya/AP/Press Association Images

The idea of Artificial Intelligence has captured our collective imagination for decades. Can behaviour that we think of as intelligent be replicated in a machine? If so, what consequences could this have for society? And what does it tell us about ourselves as

MIT
Technology
Review

Topics+ Top Stories

A View from **Vincent Conitzer**

Today's Artificial Intelligence Does Not Justify Basic Income

Even the simplest jobs require skills—like creative problem solving—that AI systems cannot yet perform competently.

October 31, 2016

Not a day goes by when we do not hear about the threat of AI taking over the jobs of everyone from **truck drivers** to **accountants** to **radiologists**. An **analysis coming out of McKinsey** suggested that “currently demonstrated technologies could automate 45 percent of the activities people are paid to perform.” There are even **online tools** based on research from the University of Oxford to estimate the probability that various jobs will be automated.

Prospect
The leading magazine of ideas

HOME BLOGS POLITICS ECONOMICS & FINANCE WORLD ARTS & BOOKS LIFE

HOME > BRITISH ACADEMY

The AI debate must stay grounded in reality

Sponsored feature

Research works best when it takes account of multiple views

by Vincent Conitzer / March 6, 2017 / [Leave a comment](#)



Are driverless cars the future © Fabio De Paola/PA Wire/PA Images

Progress in artificial intelligence has been rapid in recent years. Computer programs are dethroning humans in games ranging from leopardv to Go to poker. Self-driving cars are

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.
- 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 (yet)
- **Special-purpose AI** is more doable (nontrivial)
 - E.g., chess/poker/Go playing programs, logistics planning, automated translation, speech and image 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.

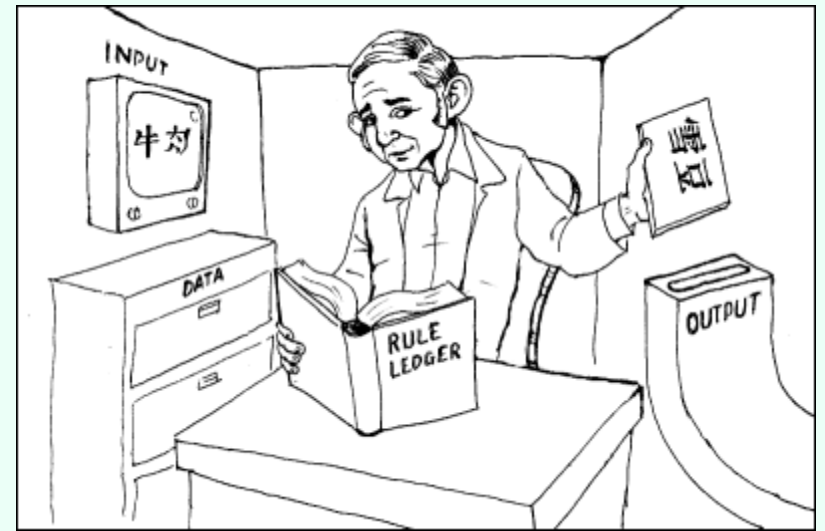
Systems that act
like humans

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
 - 2016 winner: <http://mitsuku.com/>

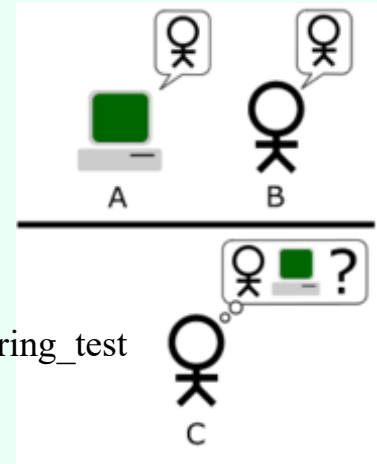


image from http://en.wikipedia.org/wiki/Turing_test

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 thats it if this is a problem then forget iot 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 worht 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: therers nothing to tell she us3ed me then got tired of me said i wasnt her type anymore i was possesive and controlling

AOLiza: perhaps i already know you were possesive and controlling.

five: i was huh you nuts i gave her all the room she wanted never told nhewre 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 naturally to humans (in some cases other animals) are much harder...
- ... though recent years have seen remarkable progress, especially in machine learning for narrow domains
 - Image recognition, speech recognition, reinforcement learning in computer games, self-driving cars
- AI systems still lack: broad understanding of the world, common sense, ability to learn from very few examples, truly out-of-the-box creativity...
- We don't understand consciousness. (Does it matter for AI?)

Some areas where humans shine

- Coming up with **reasonably good** solutions in complex messy environments
- **Adapting/self-evaluation/creativity** (“My usual approach to chess is getting me into trouble against this person... Why? Is there something entirely different I can do?”)
- **Analogical reasoning, transfer learning** (applying insights from one domain to another)
- **Explaining** our reasoning
- Tasks that require a **broad understanding** of the (human) world
- Knowing **what it's like to be human**
- **Humor**
- ...

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]
- <http://aivideocompetition.org/videos/>
- <https://www.youtube.com/watch?v=1JJsBFiXGI0>
- <https://www.youtube.com/watch?v=s6VIWUDUHTa4>
- <https://www.youtube.com/watch?v=C5Xnxjq63Zg>
- <https://www.youtube.com/watch?v=ScXX2bndGJc>
- <https://www.youtube.com/watch?v=V1eYniJ0Rnk>

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; ML is a separate course)

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

Lots of AI at Duke!



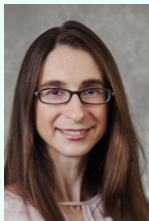
- [Vince Conitzer](#) (multiagent systems, game theory, economics)



- [Ron Parr](#) (reasoning under uncertainty, reinforcement learning, robotics)



- [Carlo Tomasi](#) (computer vision, medical imaging)



- [Cynthia Rudin](#) (machine learning (especially interpretable ML), data mining, knowledge discovery)



- [Alex Hartemink](#) (computational biology, machine learning, reasoning under uncertainty)



- [Bruce Donald](#) (computational biology & chemistry)



- [Kris Hauser](#) (robot motion planning and control, integrating perception and planning)



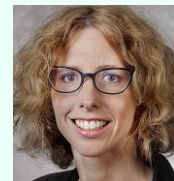
- [Missy Cummings](#) (humans+AI, unmanned vehicles)



- [Guillermo Sapiro](#) (computer vision, image analysis, machine learning)



- [Larry Carin](#) (machine learning, neural networks)



- [Katherine Heller](#) (statistics, machine learning)



- [Sayan Mukherjee](#) (statistics, machine learning)

• ...

Getting more involved

- Duke Robotics, Intelligence, and Vision (DRIV) seminar (=AI seminar)
 - Website: <http://driv.cs.duke.edu/>
 - Mailing list: <https://lists.duke.edu/sympa/subscribe/drive>
- Machine Learning at Duke
 - Website: <http://machinelearning.duke.edu/>
 - Mailing list: <https://lists.duke.edu/sympa/subscribe/machine-learning>
- “Moral AI”
 - https://lists.duke.edu/sympa/subscribe/moral_ai