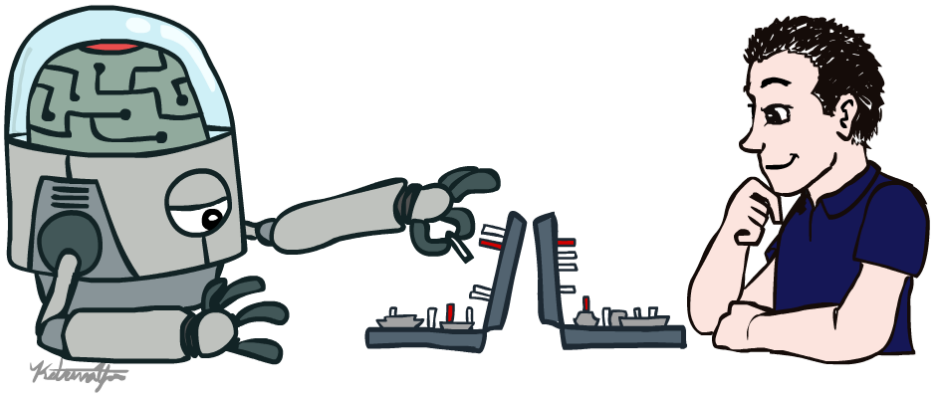
CS 188: Artificial Intelligence

Introduction



Instructors: Sergey Levine and Stuart Russell

Course Staff

Professors

Sergey Levine

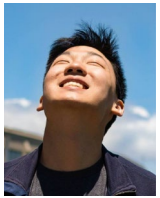
Stuart Russell

GSIs

Aditya Baradwaj

Adam Gleave

Alex Li



Austen Zhu

Avi Singh

Charles Tang

Dennis Lee

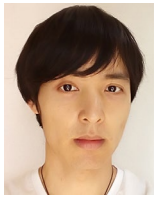
Dequan Wang

Ellen Luo

Fred Ebert

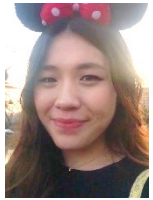
Henry Zhu

Jasmine Deng



Jason Peng

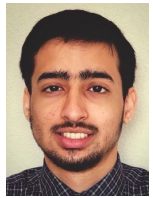
Katie Luo



Laura Smith

Micah Carroll

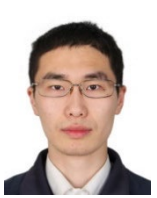
Mike Chang



Murtaza Dalal

Rachel Li

Rishi Veerapaneni



Ronghang Hu

Sid Reddy

Simin Liu

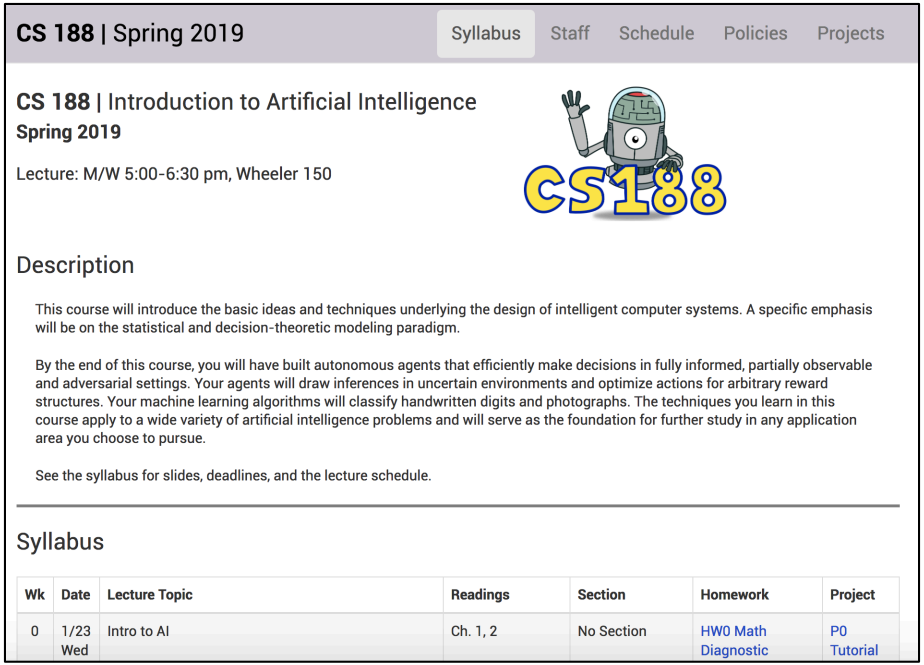
Tony Zhao

Wilson Yan

Xiaocheng (Mesut) Yang

Course Information

http://inst.cs.berkeley.edu/~cs188



CS 188 | Spring 2019 — Syllabus | Staff | Schedule | Policies | Projects

CS 188 | Introduction to Artificial Intelligence
Spring 2019

Lecture: M/W 5:00-6:30 pm, Wheeler 150

Description

This course will introduce the basic ideas and techniques underlying the design of intelligent computer systems. A specific emphasis will be on the statistical and decision-theoretic modeling paradigm.

By the end of this course, you will have built autonomous agents that efficiently make decisions in fully informed, partially observable and adversarial settings. Your agents will draw inferences in uncertain environments and optimize actions for arbitrary reward structures. Your machine learning algorithms will classify handwritten digits and photographs. The techniques you learn in this course apply to a wide variety of artificial intelligence problems and will serve as the foundation for further study in any application area you choose to pursue.

See the syllabus for slides, deadlines, and the lecture schedule.

Syllabus

Wk	Date	Lecture Topic	Readings	Section	Homework	Project
0	1/23 Wed	Intro to AI	Ch. 1, 2	No Section	HW0 Math Diagnostic	P0 Tutorial

- Communication:
  - Announcements, questions on Piazza
  - Staff email: cs188@berkeley.edu
  - Office hours in 730 Sutardja Dai Hall
    - Sergey: Monday 9-10, after lectures
    - Stuart Tuesday 9-11 (not next week)
  - Sections, tutoring signup, videos
- Course technology:
  - Website
  - Piazza
  - Gradescope
  - This course is webcast

Course Information

- Prerequisites:
  - (CS 61A or CS 61B) and (CS 70 or Math 55)
    - Recommended: CS 61A and CS 61B and CS 70
  - There will be some math and some programming
- Work and Grading:
  - 5 programming projects (25%): Python, groups of 1 or 2
    - 5 late days for semester, maximum 2 per project
  - 11 homework assignments (15%):
    - Electronic component: Online, interactive, solve alone/together, submit alone
    - Written component: On paper, solve alone/together, submit alone, self-assess
  - One midterm (20%), one final (40%)
  - Fixed grading scale (85% A, 80% A-, etc.)
  - Participation (class, section, Piazza, contests) can help on margins
  - Academic integrity policy

Exam Dates

- Midterm: March 20th, 7:00pm-9:00pm
- Final: May 16th, 7.00pm-10.00pm

- There will be no alternate exams
  - Conflict with other class final exam: see web site form

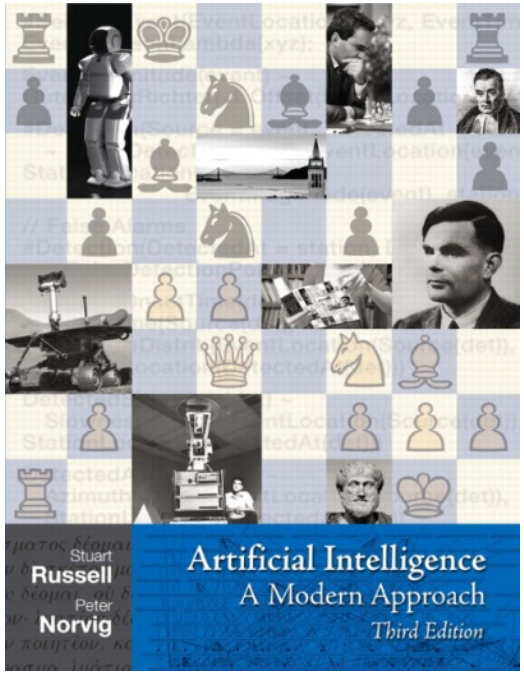
Discussion Section

- Topic: review / warm-up exercises / questions not handled in class
  - There will also be recorded videos of how to think through the solution process
- Currently, none of you are assigned to sections
- You are welcome to attend any section of your preference
- Piazza survey later this week to help keep sections balanced
- From past semesters' experience we know sections will be (over)crowded the first two weeks of section, but then onwards section attendance will be lower and things will sort themselves out
- Sections begin next week (1/28).

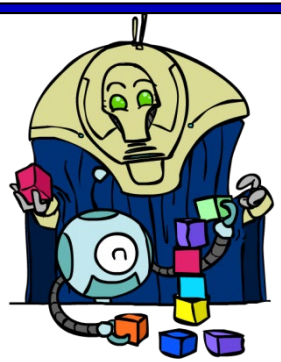
Textbook

Russell & Norvig, AI: A Modern Approach, 3rd Ed.

(sorry!)

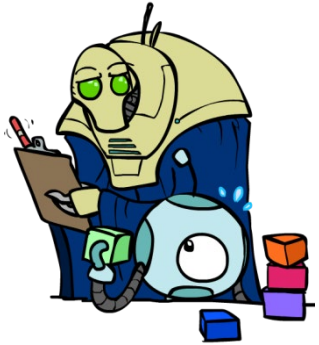


Instruction vs. Assessment



Instruction

Grow knowledge, collaborate, work until success



Assessment

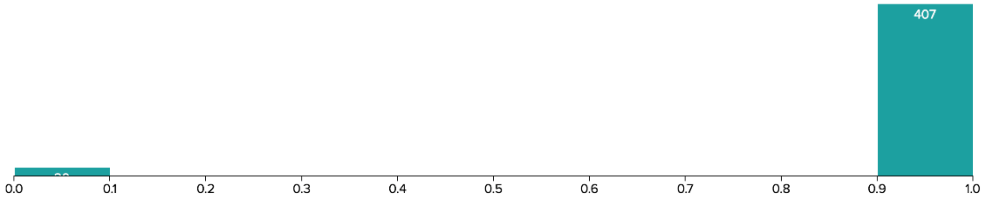
Measure knowledge, each student on their own, stopped before success

Our experience: these two goals don't mix

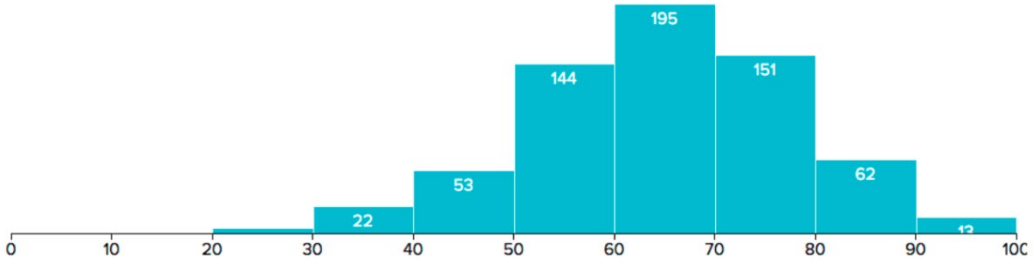
- Lecture / Section / OH / Piazza / Homework / Projects are instruction
  - collaborative, work until success (but please no spoilers, no cheating)
- Exams are assessment
  - on your own

Some Historical Statistics

- Homework and projects: work alone/together, iterate/learn till you nailed it



- Exams: assessment



Announcements This Week

- Important this week:
  - **Check out website:** https://inst.eecs.berkeley.edu/~cs188 (has links to homework, projects)
  - **Register** on **Gradescope and Piazza** (check your email for links)
  - **HW0: Math self-diagnostic** is online now (due on Monday 1/28 at 11:59pm)
  - **P0: Python tutorial** is online now (due on Monday 1/28 at 11:59pm)
  - **One-time (optional) P0 lab hours** (Thursday 7-8.30pm, Friday 6-7.30pm, 330 Soda Hall)
  - **Instructional accounts:** if you want one, go to https://inst.eecs.berkeley.edu/webacct
- Also important:
  - **Waitlist**: See https://eecs.berkeley.edu/resources/undergrads/cs/degree-reqs/enrollment-policy or google "Berkeley EECS enrollment"
  - **Concurrent enrollment** (with certain administrative exceptions) occurs when waitlist is empty

Laptops in Lecture

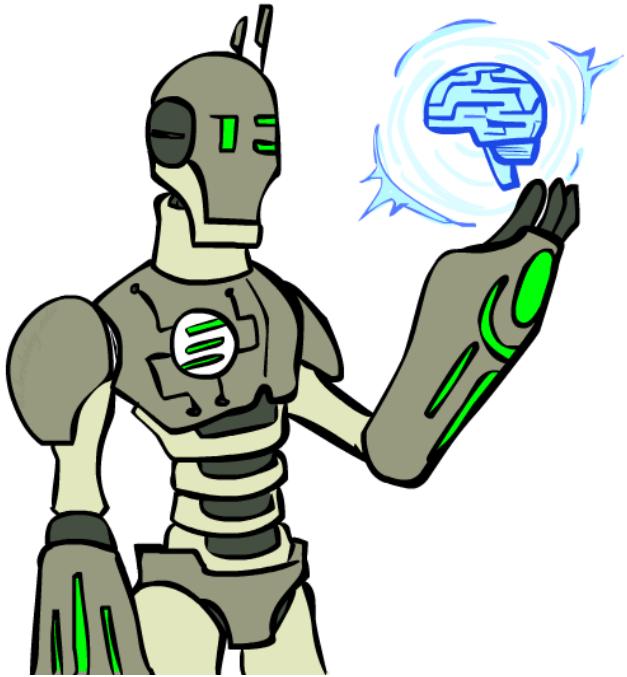
- Laptops can easily distract students behind you

  Please consider sitting towards the back if using your laptop in lecture

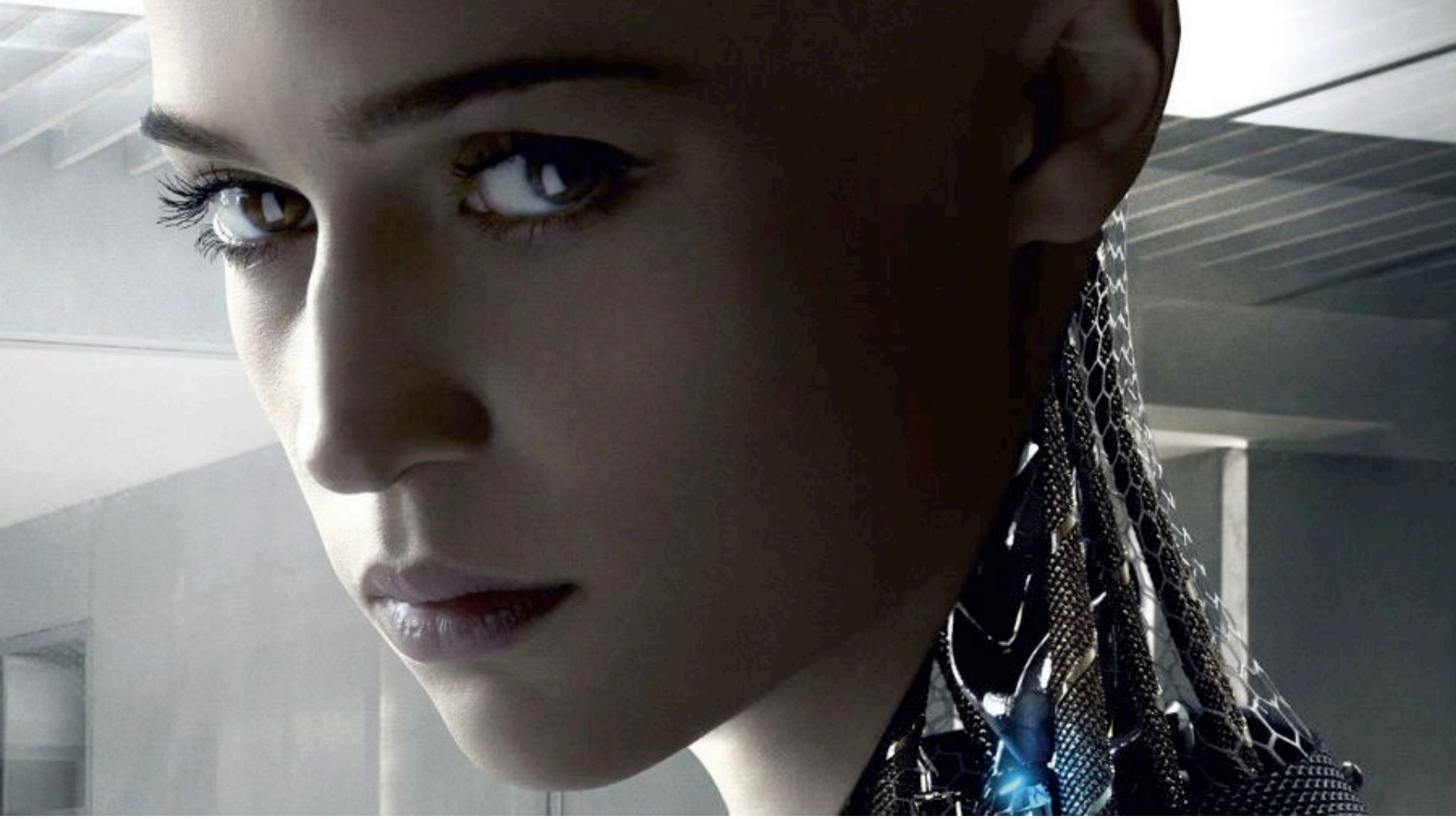
Today

- What is artificial intelligence?

- Past: how did the ideas in AI come about?
- Present: what is the state of the art?
- Future: will robots take over the world?



Movie AI




YESTERDAY DR. WILL CASTER WAS ONLY HUMAN
JOHNNY DEPP
REBECCA HALL
PAUL BETTANY
KATE MARA
CILLIAN MURPHY
AND MORGAN FREEMAN
TRANSCENDENCE
entertainmentfilms.co.uk

Stuart Russell Become a fan

Computer science professor at Berkeley; Co-author, 'Artificial Intelligence: a Modern Approach'

Transcendence: An AI Researcher Enjoys Watching His Own Execution

THE BLOG 04/19/2014 09:14 am ET | **Updated** Jun 19, 2014

Transcending Complacency on Superintelligent Machines

By Stephen Hawking, Max Tegmark, Stuart Russell, and Frank Wilczek

News AI

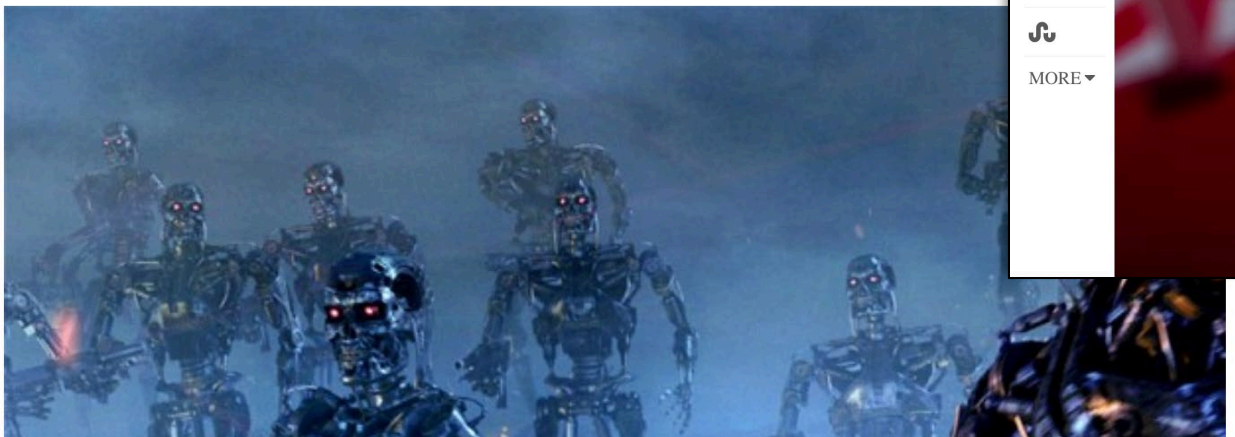
AI is the biggest risk we face as a civilisation, Elon Musk says

Billionaire burn: Musk says Zuckerberg's understanding of AI threat 'is limited'

HOME » FINANCE » FINANCE TOPICS » DAVOS

'Sociopathic' robots could overrun the human ra a generation

Computers should be trained to serve humans to reduce their th human race, says a leading expert on artificial intelligence

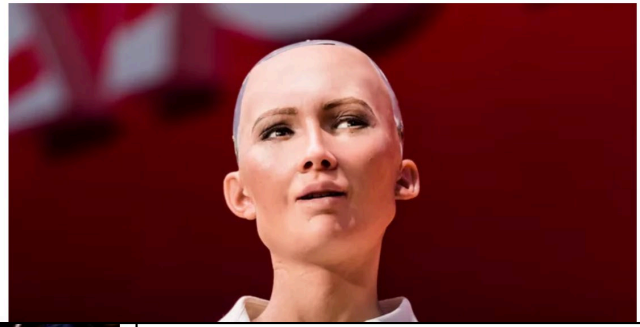


LIVESCIENCE NEWS TECH HEALTH PLANET EAR

Live Science › Tech

Lifelike 'Sophia' Robot Granted Citizenship to Saudi Arabia

By Mindy Weisberger, Senior Writer | October 30, 2017 03:39pm ET



News AI

TECH • ARTIFICIAL INTELLIGENCE

United Kingdom Plans $1.3 Billion Intelligence Push

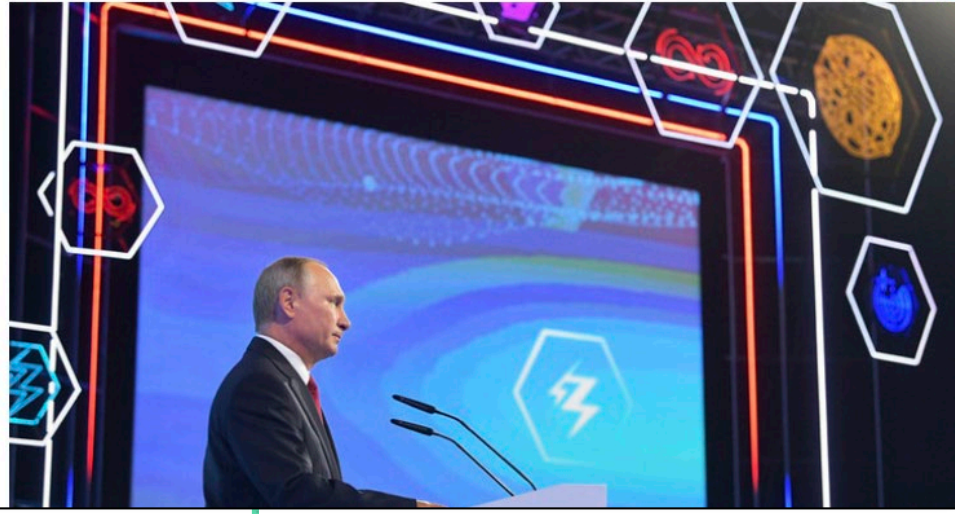
France to spend $1.8 billion on compete with U.S., China

EU wants to invest £18b development

China's Got a Huge Art Intelligence Plan

'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day

Published time: 1 Sep, 2017 14:08
Edited time: 1 Sep, 2017 14:40



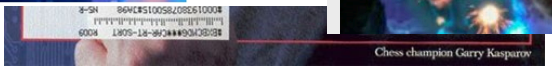
News AI

IBM's Watson Jeopardy Computer Shuts Down Humans in Final Game

Silicon prevails ir

DAILY NEWS 9 March 2016

'I'm in shock! world's best



Blizzard will show off Google's Deepmind AI in StarCraft 2 later this week

By Andy Chalk 4 hours ago

Google and Blizzard launched the artificial intelligence project in 2016.

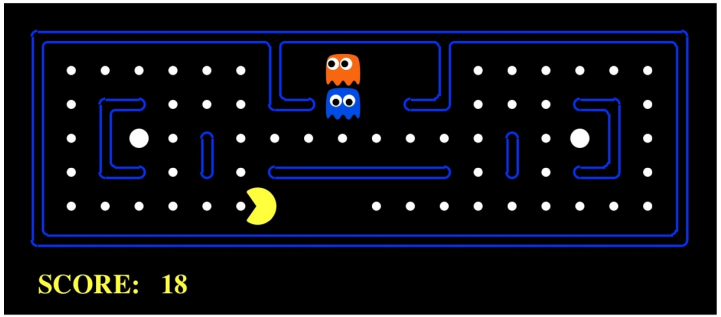
COMMENTS

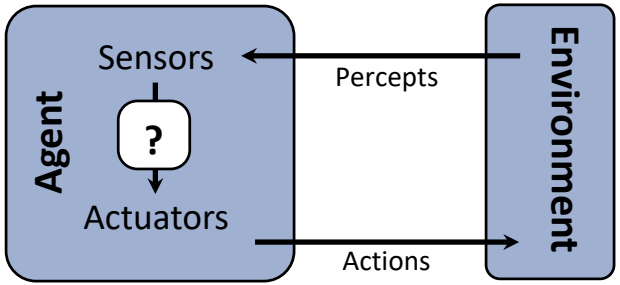
AI as computational rationality

- Humans are intelligent to the extent that our actions can be expected to achieve our objectives
- Machines are intelligent to the extent that their actions can be expected to achieve their objectives
  - Control theory: minimize cost function
  - Economics: maximize expected utility
  - Operations research: maximize sum of rewards
  - Statistics: minimize loss function
  - AI: all of the above, plus logically defined goals
- AI ≈ computational rational agents

Designing Rational Agents

- An **agent** is an entity that *perceives* and *acts*.
- A **rational agent** selects actions that maximize its (expected) **utility**.
- Characteristics of the **percepts, environment,** and **action space** dictate techniques for selecting rational actions
- **This course** is about:
  - General AI techniques for many problem types
  - Learning to choose and apply the technique appropriate for each problem

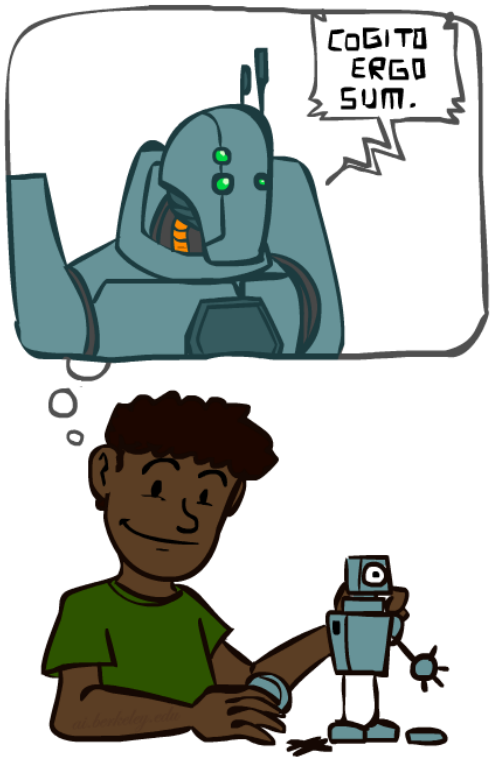




What About the Brain?

- Brains (human minds) are very good at making rational decisions, but far from perfect; they result from accretion over evolutionary timescales
- We don’t know how they work
- “Brains are to intelligence as wings are to flight”
- Lessons learned from human minds: memory, knowledge, feature learning, procedure formation, and simulation are key to decision making

A (Short) History of AI

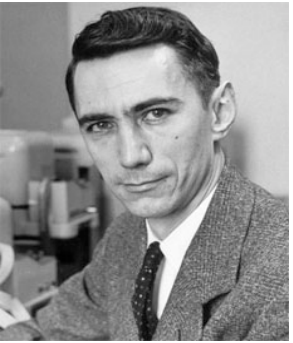


Demo: HISTORY – MT1950.wmv

A short prehistory of AI

- Prehistory:
  - **Philosophy** from Aristotle onwards
  - **Mathematics** (logic, probability, optimization)
  - **Neuroscience** (neurons, adaptation)
  - **Economics** (rationality, game theory)
  - **Control theory** (feedback)
  - **Psychology** (learning, cognitive models)
  - **Linguistics** (grammars, formal representation of meaning)
- Near miss (1842):
  - Babbage design for universal machine
  - Lovelace: “a thinking machine” for “all subjects in the universe.”

AI's official birth: Dartmouth, 1956

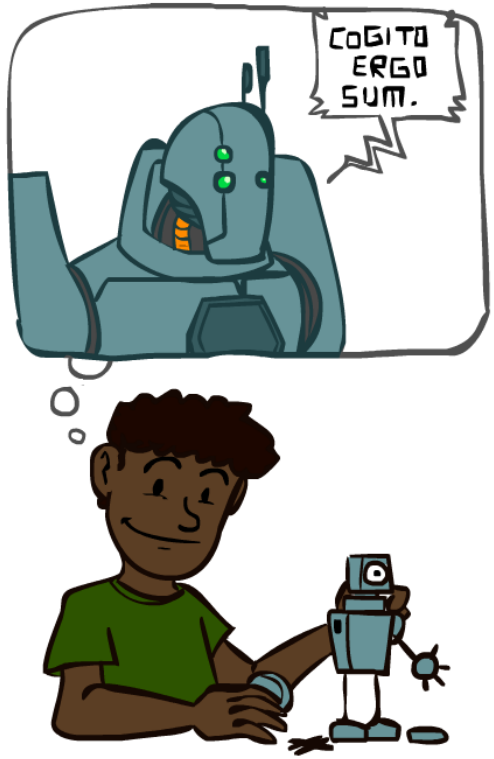


"An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. ***We think that a significant advance can be made if we work on it together for a summer."***

John McCarthy and Claude Shannon
Dartmouth Workshop Proposal

A (Short) History of AI

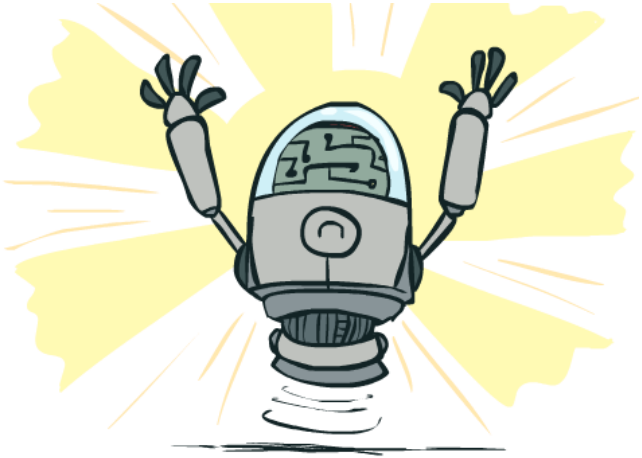
- 1940-1950: Early days
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence”
- 1950—70: Excitement: Look, Ma, no hands!
  - 1950s: Early AI programs: chess, checkers program, theorem proving
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1965: Robinson's complete algorithm for logical reasoning
- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: “AI Winter”
- 1990— 2012: Statistical approaches + subfield expertise
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems… “AI Spring”?
- 2012— ___: Excitement: Look, Ma, no hands again?
  - Big data, big compute, neural networks
  - Some re-unification of sub-fields
  - AI used in many industries



What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Play a decent game of table tennis?
- ✓ Play a decent game of Jeopardy?
- ✓ Drive safely along a curving mountain road?
- ? Drive safely along Telegraph Avenue?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley Bowl?
- ? Discover and prove a new mathematical theorem?
- ✗ Converse successfully with another person for an hour?
- ? Perform a surgical operation?
- ✓ Translate spoken Chinese into spoken English in real time?
- ? Fold the laundry and put away the dishes?
- ✗ Write an intentionally funny story?



Unintentionally Funny Stories

Janelle Shane
@JanelleCShane

Follow

Tried retraining the neural net on just "what do you get when you cross a X with a X?" jokes. Results did not improve. And for some reason, bungees are its favorite thing.

What do you get when you cross a dog and a vampire? A bungee

What do you get when you cross a gorilla and a bull? A spider.

What do you get when you cross a cow with a rhino? A bungee with a dog.

What do you get when you cross a mountain and a bungee and a cow? A cow with a rhinocero.

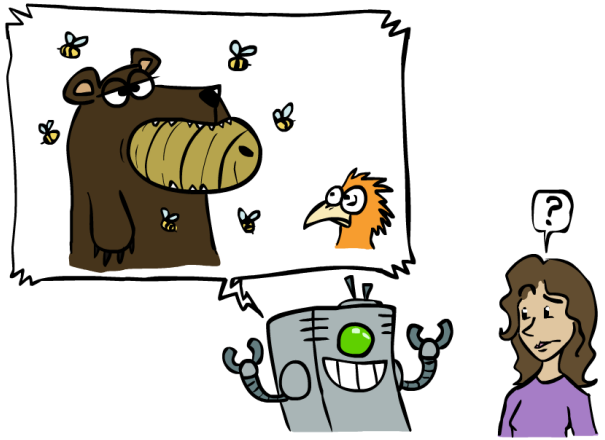
What do you get when you cross a street and a bungee with a cow? A bungee and a pig with a cow.

What do you get when you cross a street and a cow? A bungee with a bungee and a rhino.

What do you get when you cross a pig with a cow with a party? Because the engineers with a dog.

iend
l him
ed to

the
sitting.
owned.

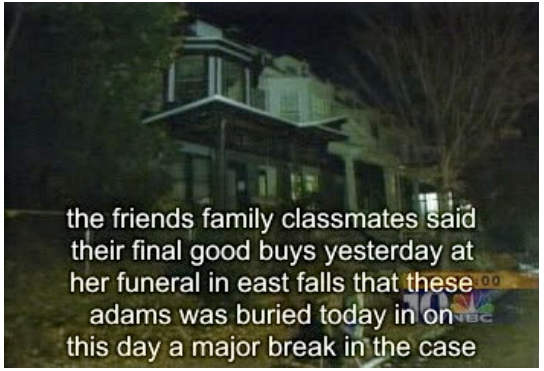


nd a vain crow. One day the
cheese in his mouth. He noticed
ecame hungry, and swallowed
The End.

[Shank, Tale-Spin System, 1984]

Natural Language

- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems

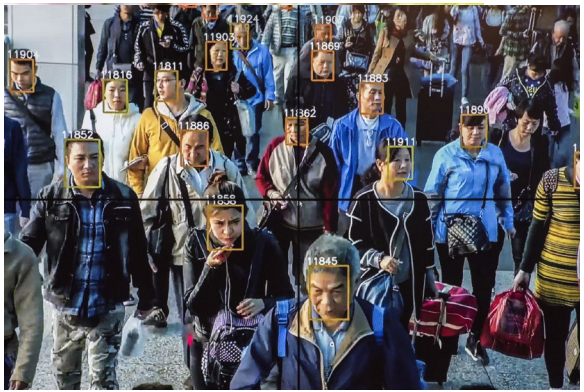


- Language processing technologies
  - Question answering
  - Machine translation

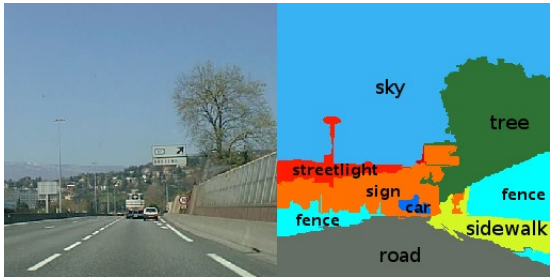
  - Web search
  - Text classification, spam filtering, etc…

Vision (Perception)

Face detection and recognition

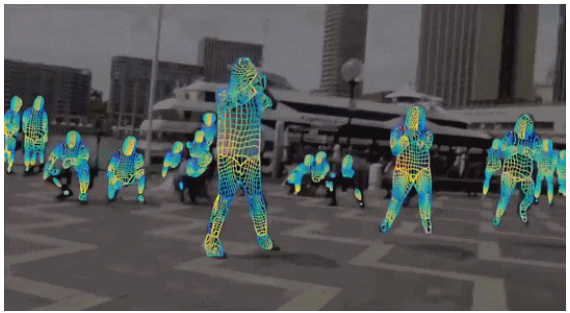


Semantic Scene Segmentation



[Caesar et al, ECCV 2017]

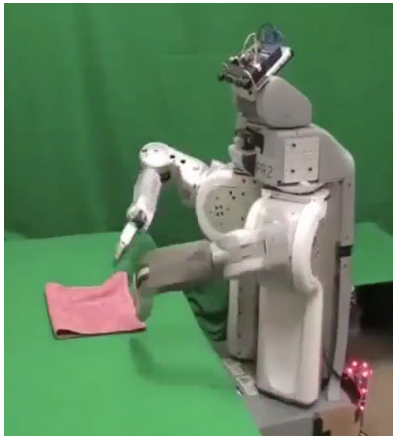
3-D Understanding



[DensePose]

Robotics

- Robotics
  - Part mech. eng.
  - Part AI
  - Reality much harder than simulations!
- In this class:
  - We ignore mechanics
  - Methods for planning
  - Methods for control

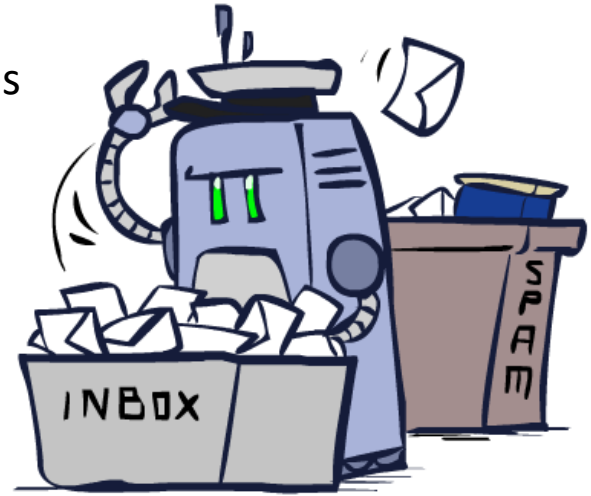


Images from UC Berkeley, Boston Dynamics, RoboCup, Google

AI everywhere…

- Search engines
- Route planning, e.g. maps, traffic
- Logistics, e.g. packages, inventory, airlines
- Medical diagnosis, machine diagnosis
- Automated help desks
- Spam / fraud detection
- Smarter devices, e.g. cameras
- Product recommendations
- Assistants, smart homes

- … Lots more!



Future

- We are doing AI…
  - To create intelligent systems
    - The more intelligent, the better
  - To gain a better understanding of human intelligence
  - To magnify those benefits that flow from it
    - E.g., net present value of human-level AI ≥ $13,500T
    - Might help us avoid war and ecological catastrophes, achieve immortality and expand throughout the universe
- What if we succeed?

BALTIMORE
Post-Examiner

Artificial Intelligence could spell the end of the human race

BY PAUL CROKE · JUNE 9, 2015 · NO COMMENTS

What's bad about better AI?

- AI that is incredibly good at achieving something other than what we really want
- AI, economics, statistics, operations research, control theory all assume utility to be ***exogenously specified***

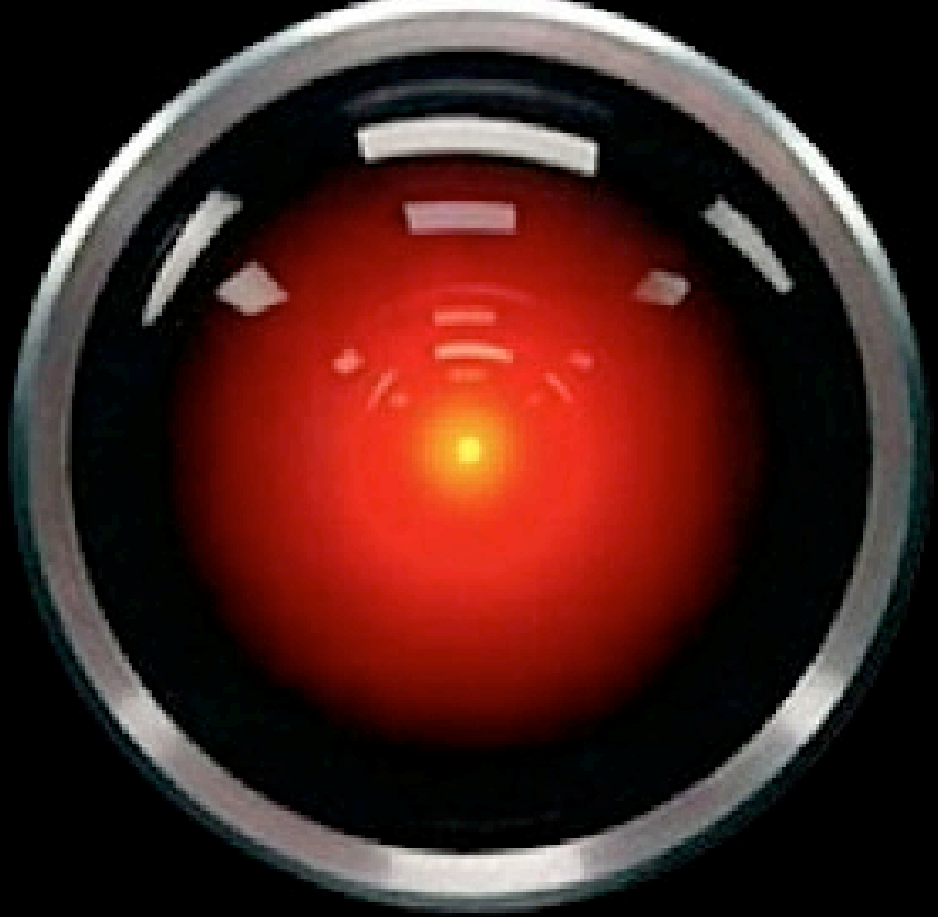
Value misalignment

- E.g., “Calculate pi”, “Make paper clips”, “Cure cancer”
- Cf. Sorcerer’s Apprentice, King Midas, genie’s three wishes

> ***We had better be quite sure that the purpose put into the machine is the purpose which we really desire***
>
> Norbert Wiener, 1960

Instrumental goals

- For ***any primary goal***, the odds of success are improved by
  1) Maintaining one's own existence
  2) Acquiring more resources
- With value misalignment, these lead to obvious problems for humanity

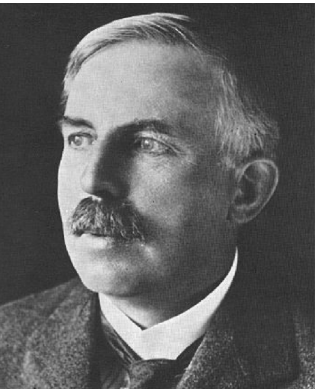


I’m sorry, Dave, I’m afraid I can’t do that

Towards human-level AI

- Still missing:
  - Real understanding of language
  - Integration of learning with knowledge
  - Long-range thinking at multiple levels of abstraction
  - Cumulative discovery of concepts and theories
- **Date unpredictable**

Unpredictability



Sept 11, 1933: Lord Rutherford addressed BAAS:
"Anyone who looks for a source of power in the transformation of the atoms is talking moonshine."

Sept 12, 1933: Leo Szilard invented neutron-induced nuclear chain reaction

"We switched everything off and went home. That night, there was very little doubt in my mind that the world was headed for grief."

AI as computational rationality

- Humans are intelligent to the extent that our actions can be expected to achieve our objectives
- Machines are intelligent to the extent that their actions can be expected to achieve their objectives
  - Control theory: minimize cost function
  - Economics: maximize expected utility
  - Operations research: maximize sum of rewards
  - Statistics: minimize loss function
  - AI: all of the above, plus logically defined goals
- **We don't want machines that are intelligent in this sense**
- Machines are ***beneficial*** to the extent that ***their*** actions can be expected to achieve ***our*** objectives
- We need machines to be ***provably beneficial***

Provably beneficial AI

1. The machine's only objective is to maximize the realization of human preferences
2. The robot is initially uncertain about what those preferences are
3. Human behavior provides evidence about human preferences

The standard view of AI is a special case, where the human can exactly and correctly program the objective into the machine

So, if all this matters…..

- Can we affect the future of AI?
  - Can we reap the benefits of superintelligent machines and avoid the risks?
- ***"The essential task of our age."***

  Nick Bostrom, Professor of Philosophy, Oxford University.