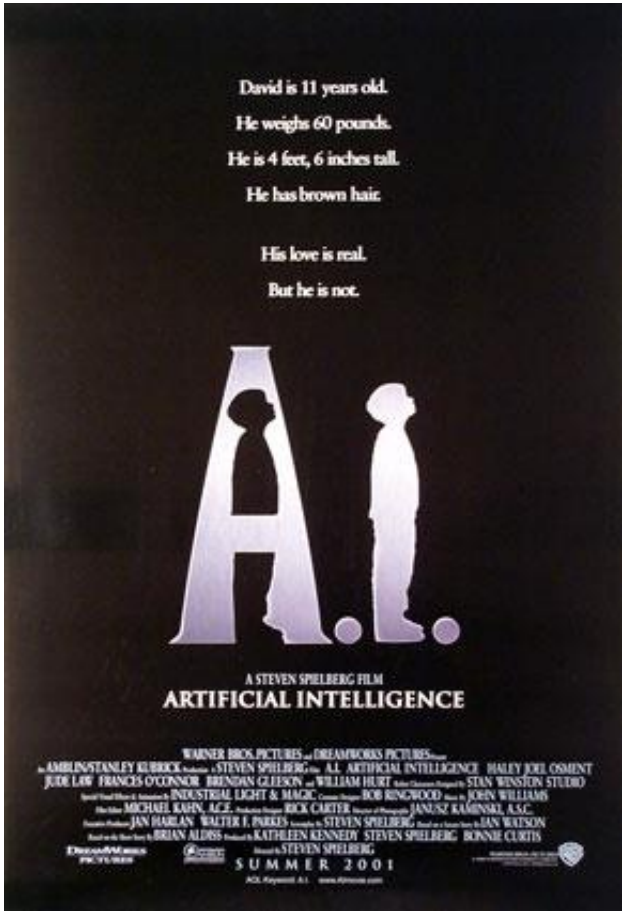
What is AI?



How are people different from computers?

Can computers think?

These questions have been thought about from the earliest days of computers

What is intelligence?

class answers:
- rationality
- self awareness
- learning from past experiences (introspection)
- abstract thinking

What is intelligence?

- capacities to remember, to reason, to plan, to solve problems, to think abstractly, to comprehend ideas, to use language, and to learn <== we tend to focus on this one
- cooperative problem solving?
- capacity to understand the intentions, motivations, fears and desires of other people? emotional intelligence
- capacity to understand oneself; to appreciate one's own intentions, motivations, fears and desires? introspection
- creativity? personality? character? morality? knowledge? wisdom? aesthetics?

A consensus definition

Intelligence is a general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on", "making sense" of things, or "figuring out" what to do.

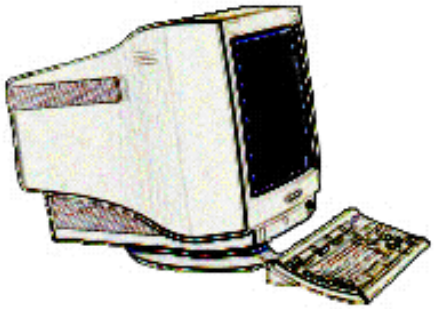
Gottfredson, L. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence* 24(1):13–23, 1997.

Analogy

drove AI for a long time, we've since shifted away from it

The mind as software:
manipulating symbols

like symbols in logic class



The brain as hardware

Thinking is symbolic reasoning

- Models of computation
  - Turing machine (Alan Turing): writes symbols on an infinitely long tape
  - λ-calculus (Alonzo Church): rewriting symbolic formulas
- Church-Turing thesis (conjecture):

  Any effectively computable (i.e., following well-defined operations) function can be carried out on a Turing machine (or other equivalent formalism)

all it can do is read and write from locations

anything modern day technology can do (in terms of things that can be computed), it can be done on a Turing machine

Thinking is symbolic reasoning

- Can humans compute functions that are not Turing computable?
- Physical symbol system:
  - Symbol: meaningful pattern that can be manipulated
  - Physical symbol: physical object that is part of the real world
- Newell-Simon hypothesis (conjecture):

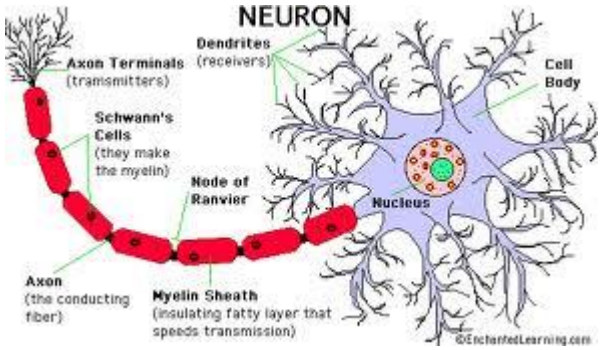
  A physical symbol system has the necessary and sufficient means for general intelligent action
  - An intelligent agent is necessarily a physical symbol system
  - A physical symbol system is all that is needed for intelligence

lots of people think that there are things humans can do that computers never will be able to, but some people claim we just haven't gotten there yet but will one day

Thinking is neurons firing

lower level

- Inspired by how brains and their neurons work



15 instructions per second, which is pretty slow, but there are lots and lots of stuff happening in parallel which makes up for it

- Neural networks
  - Turing complete: equivalent in power to λ-calculus and Turing machines

Analogy

Don’t have to mimic naturally occurring solutions

Can we directly engineer intelligence?

planes don’t fly like birds, but they do fly

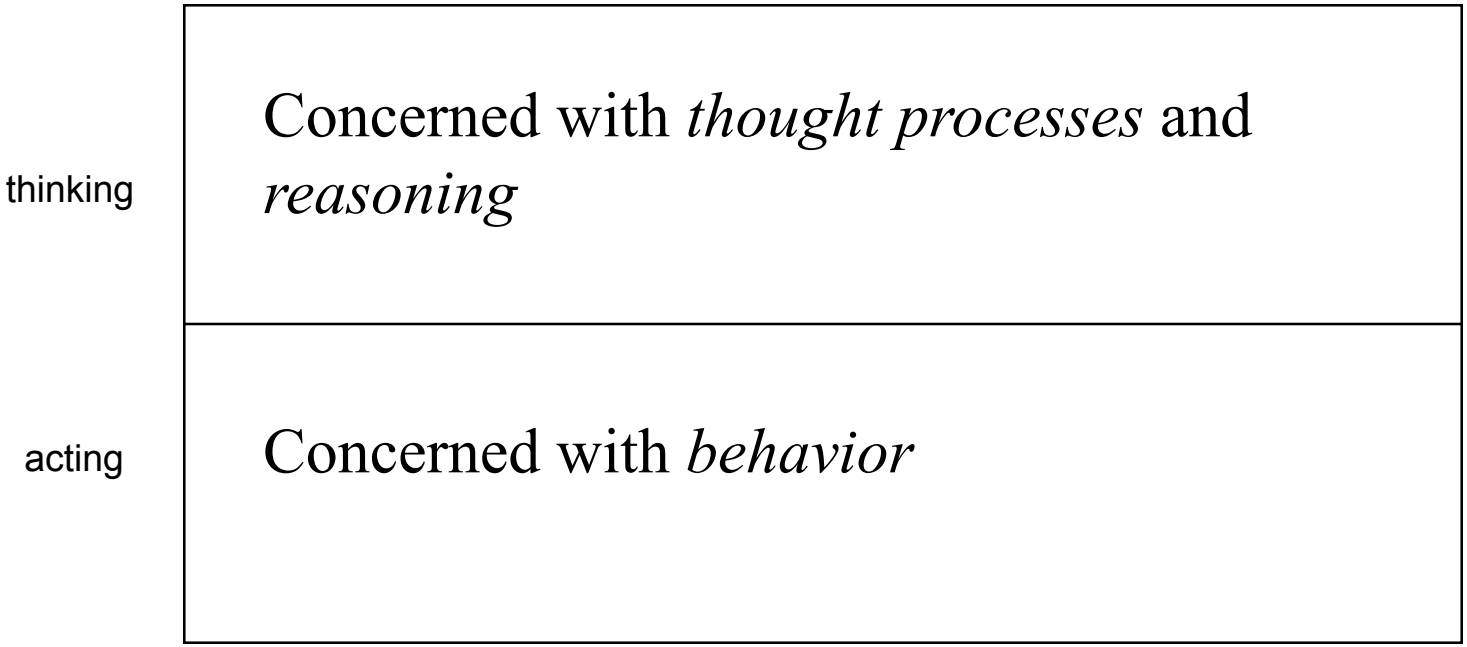
let’s study the principles of intelligence and see if we can build something that could do the same thing

… but planes are much bigger, faster, stronger than birds…

Four categories of AI definitions

Systems that think like humans	Systems that think rationally (as we know, humans are not fully rational)
Systems that act like humans	Systems that act rationally

One distinction



Another distinction

Measure success in terms of *human* performance does this AI compare well to a human, capture what a human does?	Measure success in terms of an *ideal* concept of intelligence ultimate rational agent

Thinking humanly: The cognitive modeling approach

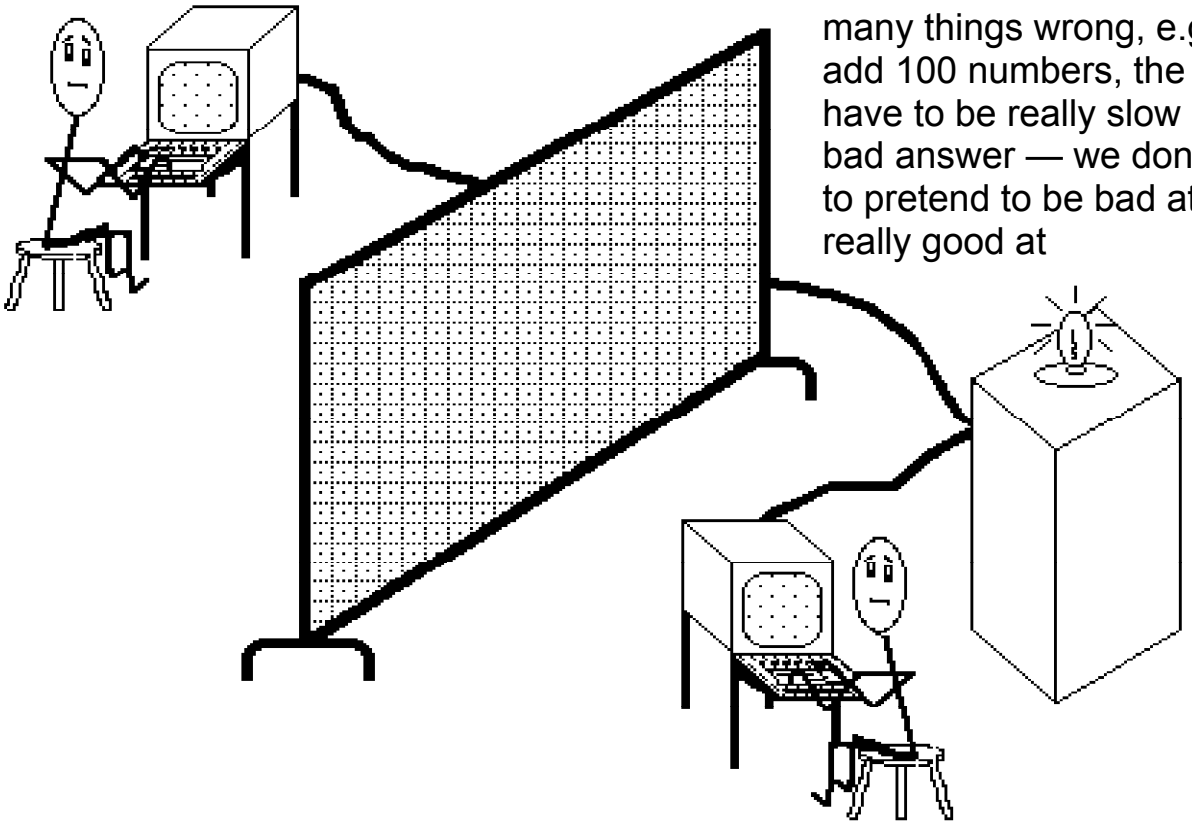
- Determine how humans think
  - introspection?
  - psychological experimentation
- Computational theories of the mind
  - symbolic: beliefs, goals, reasoning steps
  - non-symbolic
- Goal: model that exhibits all and only the behavior observed in humans or animals

we don’t do this in CS — check out cog psych courses

Acting humanly: The Turing Test approach

not used much anymore

many things wrong, e.g. if they ask to add 100 numbers, the computer would have to be really slow and then give a bad answer — we don't want computers to pretend to be bad at things they're really good at



can the judge tell the difference between a person and an AI machine?

Thinking rationally: The laws of thought approach

- Emphasis on *thinking* the right thing
  - performing correct inferences
- Thinking right means following the laws of thought
- Logic and probabilities as normative theories

if you believe a -> b and you believe a, if you're logical you'd believe b

Acting rationally: The rational agent approach

- Emphasis on *doing* the right thing
  - taking the best possible action
- An agent is just something that perceives and acts
- Acting rationally means acting so as to achieve one's goals, given one's beliefs
- Decision theory

this slide was skipped

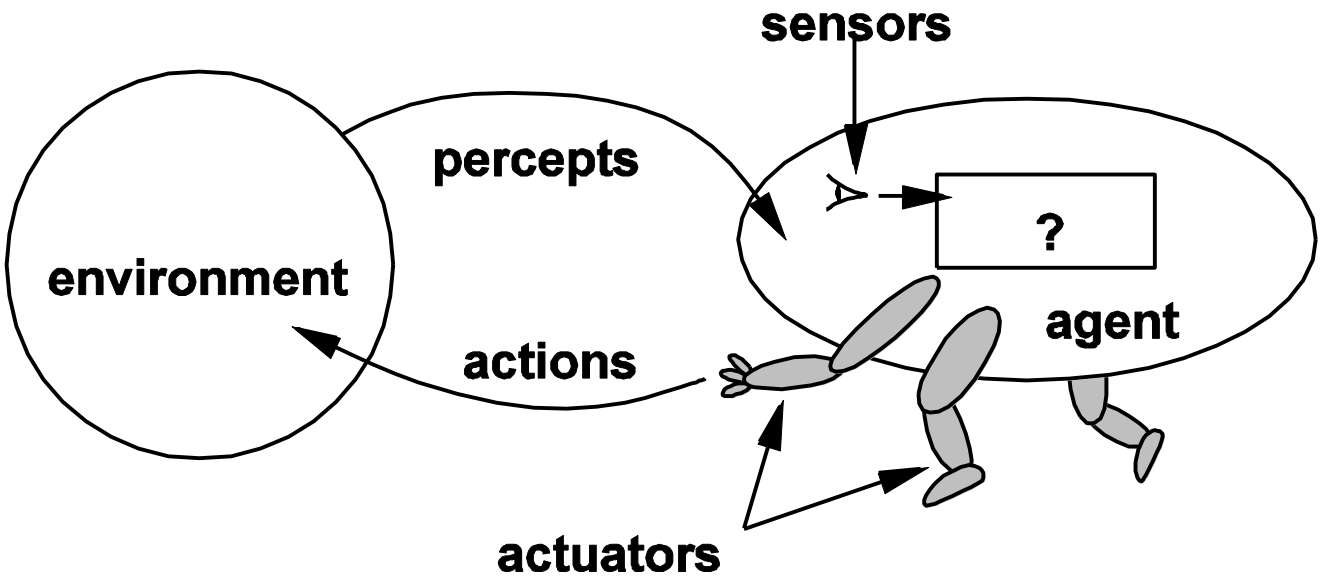
Unifying theme: Intelligent agents

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

Concept of agent is a tool for designing and analyzing systems

this slide was skipped

Agents and environments



Agents include humans, robots, softbots, thermostat, …

this slide was skipped

Design space of AI agents

- Modularity or decomposability
  - Flat, modular, hierarchical
- Representation scheme
  - States, features, relations
- Planning horizon
  - Non-planning, finite stage, infinite stage

this slide was skipped

Design space of AI agents (con't)

- Uncertainty
  - Sensing uncertainty: fully observable, partially observable
  - Effect uncertainty: deterministic, stochastic
- Preferences
  - Goals, complex preferences
- Number of agents
  - Single agent, multiple agents

this slide was skipped

Design space of AI agents (con't)

- Learning
  - Knowledge is given, knowledge is learned
- Computational limits
  - Perfect rationality, bounded rationality

A brief history of AI

- 1950's – 1969: Enthusiasm and expectations
  - Samuel's checkers program, neural networks, machine translation, theorem proving, planning, …
- 1966 – 1973: A dose of reality, AI winter
  - need for domain knowledge, high computational complexity, computational limits of perceptrons
- 1969 – 1979: Knowledge-based systems AI got worse, less learning
- 1980 – 1988: Expert system industry booms getting better again
- 1988 – 1993: A dose of reality, another AI winter
  - over-sold, did not meet expectations
- 1986 – present: Return of connectionism
- 1988 – present: AI becomes a science
  - improved methodology and theoretical frameworks

it's like a roller coaster - things get better, we get really high expectations, we don't meet them, it becomes less hyped — we're in a hype phase right now

Progress

- *Little* success on the grand goal
- Lots of success in restricted domains

Example: Checkers

Chinook is the World Man-Machine Checkers Champion. It has defeated the best players in the world

if it's playing its best, the best you can do is tie the machine (this is proven true)

Example : Chess

Kasparov
(Courtesy IBM/Equitable Building)

Kasparov concedes defeat
(Courtesy IBM/Equitable Building)

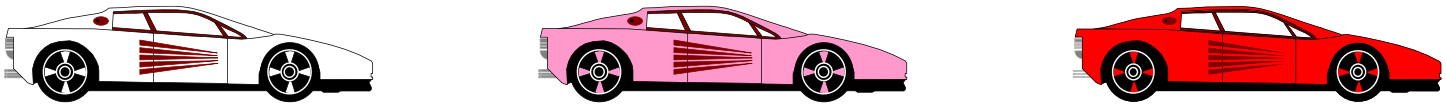
NEW YORK (CNN, 1997) -- He had never lost a chess match. But that all changed after 19 moves Sunday against the Deep Blue IBM computer...

Example : Go

DeepMind, a London-based artificial intelligence (AI) company bought by Google for $400m in 2014, challenged Lee Sedol to a five-game Go match, and won the best-of-five series with four wins and one loss.

Example: assembly line sequencing

In what order should the cars be manufactured?



if you're building for north america, you don't build on Monday (Monday cars are exported) — the quality isn't as good because people are more sleepy

red cars shouldn't be made at the same time as white - the red will mix in with white to make pink cars

Example: credit card authorization

Handle majority of transactions autonomously.

Pass exceptional cases (e.g., unusual travel patterns) to a human along with recommendation and reasons.

Example: mortgage authorization

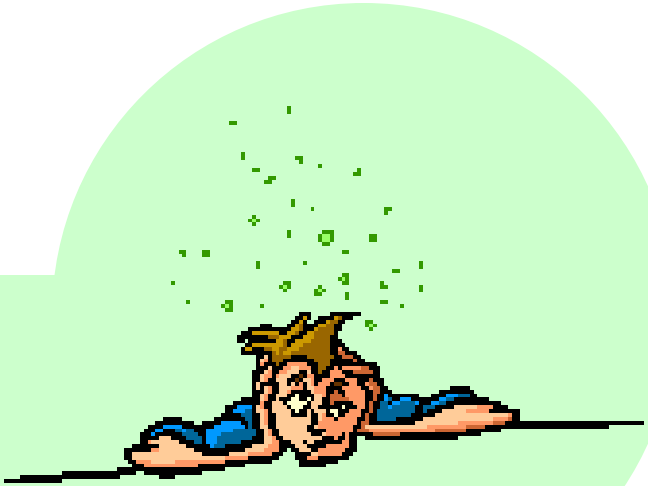
One of the largest mortgage companies in U.S.

Refer hard cases to human underwriter

this helps reduce bias in decisions to lend money :D

Example: medical diagnostic systems

Interactive medical decision support software for consumers and health care providers.



E.g., Alcoholism

Do you ever drink in the morning?

Yes / No

Have you been drunk or experienced a hangover in the past month?

Once / Twice or more / No

...

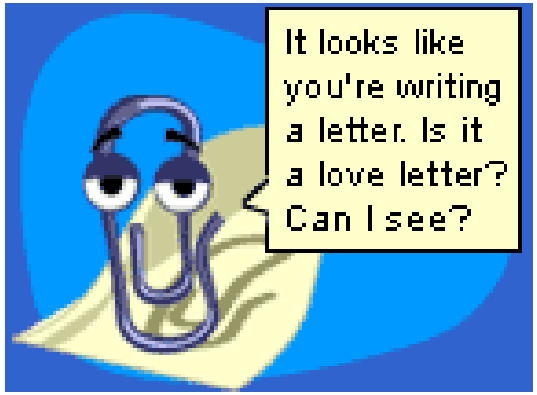
doesn't replace the doctor, but gives the doctor extra information/second opinion

Example: customized help

Microsoft's Clippy

Context-specific help with using software

Tailored to individual user



"Clippy used AI techniques, but don't hold that against AI"

Example : interactive entertainment

Agent Modeling: modeling and reasoning about your opponent's goals, plans, knowledge, capabilities, and emotions or ...

allowing your creature to make its own decisions and inferences is important for designing realistic games

Example: automating chores

Robomower

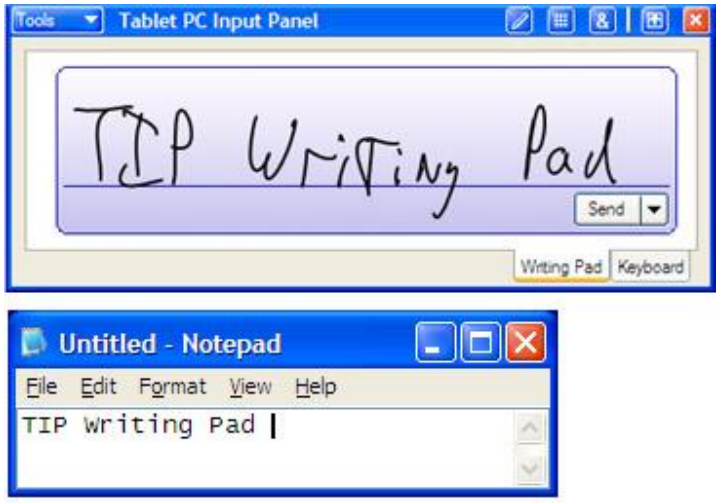
Roomba

Example: recommender agents

Amazon.com

Recommendations for purchases tailored to an individual user

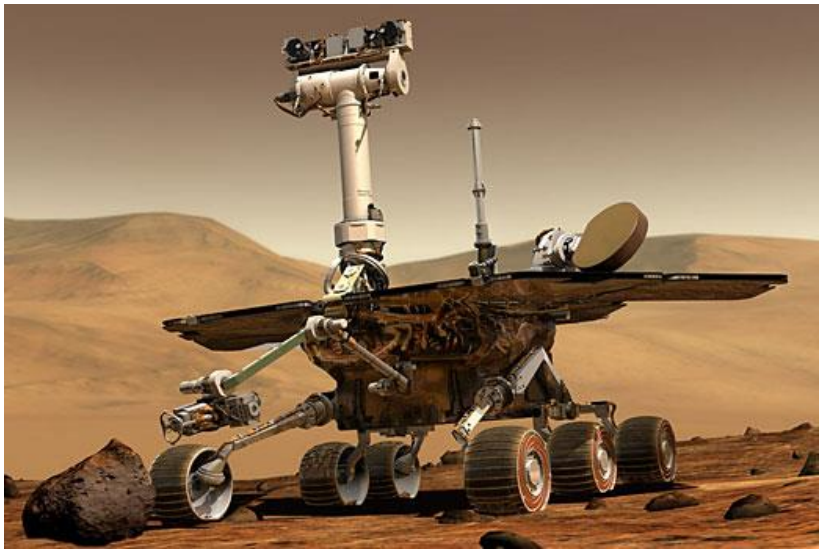
Example: handwriting recognition



Example: interplanetary exploration

NASA Mars Rover

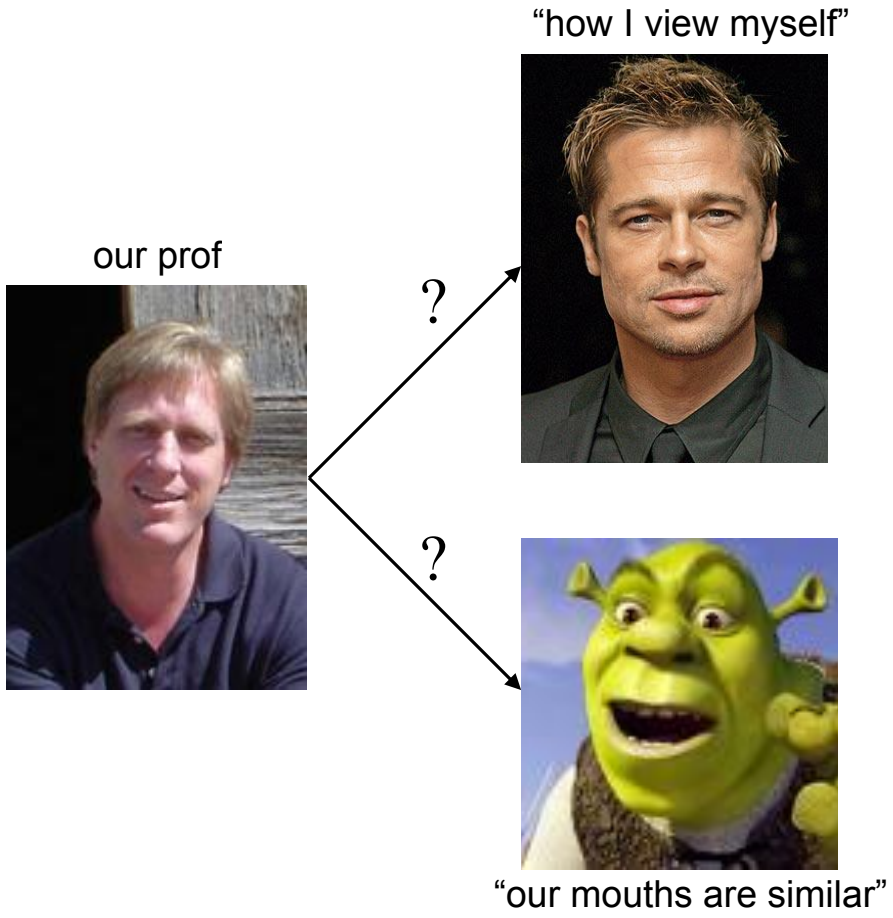
Autonomous decisions in response to the environment



Example: face recognition

Which celebrity do you most closely match?

Which parent do you look the most like?



Example: spam filtering

Mozilla.org

Mozilla Thunderbird email client filters spam

Example: autonomous driving

Automatic road following
(driverless driving)

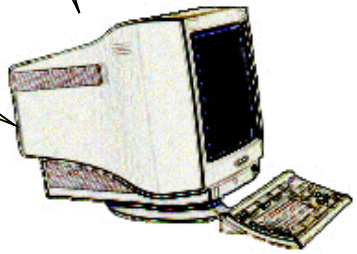
scary to think about military applications
this would also cost millions of people jobs

Example: question answering

"AI for $100, Alex."

"This popular TV quiz show is the latest challenge for IBM."

"What is Jeopardy?"

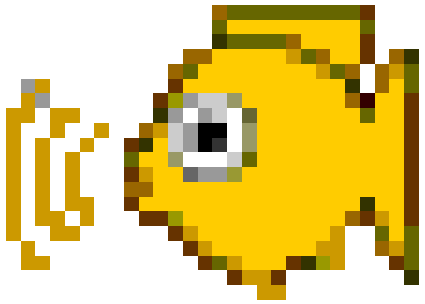


Example: language translation

Yahoos's Babel Fish,

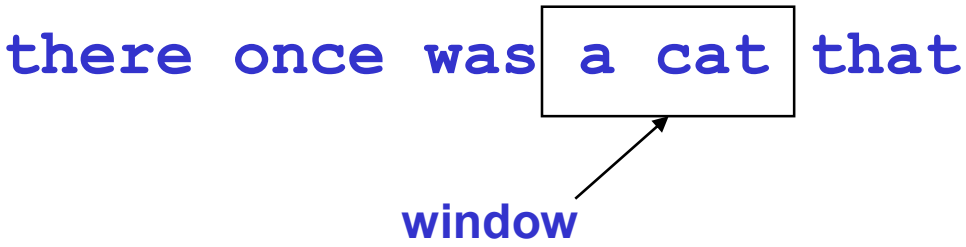
Google translate:

Translate text into many other languages



Example: speech generation

A window 7 characters wide is moved over the text. The network learns to pronounce the middle letter.



The sound of a letter depends on its context. For example, the letter “a” is pronounced differently in “mean”, “lamb”, and “class”.