Artificial intelligence

Welcome!

This session focuses on the history and perspectives of developing AI techniques, from 1950's Turing Test until today's large language models.

We'll touch upon both the dreams of conceiving intelligence, and the consequences of concrete applications

—

Introduction

—

Last session, we discussed the new ways to create value through digital networks. Today, we talk about the history of artificial intelligence, and the labor reality of today's machine learning.

Plan for the session:

- cultural contexts of AI
- intelligence as symbolic reasoning
- intelligence as probabilistic deciding
- group work on format

—

Do you think we should use an AI software in the court room?

The question of AIs in courtrooms suggests different trains of thoughts:

- *AI offers an approach to standardization to guarantee equal treatment*
- *AI developed in a business setting cannot offer due process due to being a trade secret*
- *A combination of both humans an AI might be the best way to go forward*

For an overview of why it might not be a good idea, see Garapon, Antoine, and Jean Lassègue. Justice Digitale: Révolution Graphique et Rupture Anthropologique. 1re édition, Presses universitaires de France, 2018.

—

what is AI?

—

Cultural contexts

_

Between short and long histories: humans have always told stories of animated matter.

Talos , Frankenstein's monster , Pygmalion's statue , Golem .

These creatures are all ways to represent something else. Talos represents the laws of the gods. Frankenstein's monster represents the hubris of man, Pygmalion represents fascination for perfection, and the Golem has shifted in its associations, including war, community, isolation, hope, and despair.

_

Historically, artificial life was associated with **movement**. The shift to **language** happens with the 1913 play Pygmalion .

_

Already **anthropomorphisation**: would machine intelligence be similar to human intelligence?

Anthropomorphisation is the process of ascribing human-like properties to non-human things (e.g. referring to ChatGPT as "he" or "she")

_

Collective automation as intelligent machinery: from automatising the pin factory of Adam Smith[1], to Leonardo Flores' Cybersyn[2] in Chile.

Additionally, AI research also includes multi-agent systems, in wich groups of small, limited reasoning agents can give rise to complex behaviours (likes ants, bees or birds)

_

AI research is both a **technical** endeavour and a **philosophical** endeavour.

_

Symbolical reasoning

—

The person who sets the bar is Alan Turing with his **Turing Test**[3].

—

can one tell apart a man and a woman? a human and a machine?

- the difference between being and appearing
- the role of education
- the mind/body duality

In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components:

(a) The initial state of the mind, say at birth,

(b) The education to which it has been subjected,

(c) Other experience, not to be described as education, to which it has been subjected.

Turing starts by asking the question of whether we could ever tell apart a robot from a human, if we could only talk to them via a computer interface (i.e. texting). This is the first standard of intelligence that is put forward.

—

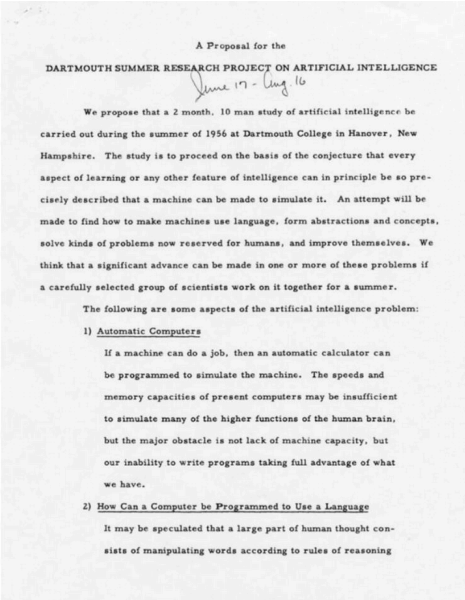
Hype #1 - the symbolical

—

There is an assumption that **formal mathematical language is the best language** to solve all things, dating all the way back from Leibniz[4].

—

Seeing programming as a mathematics automated, people set out to automate intelligence in **Dartmouth in 1956**[5].



A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. 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 in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1) Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2) How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning

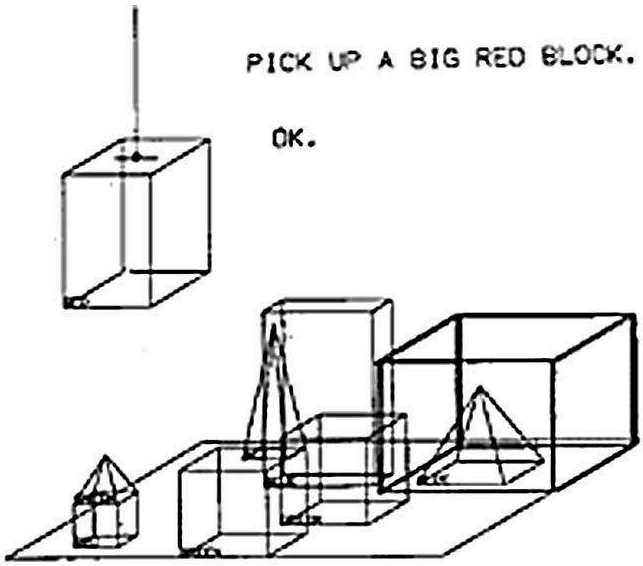
The 1956 proposal for the Dartmouth workshop

—

It focuses on linguistic **toy problems** (e.g. understanding children's stories, or solving mathematical problems).

Games are also good examples.

—



SHRDLU is the culmination of this approach

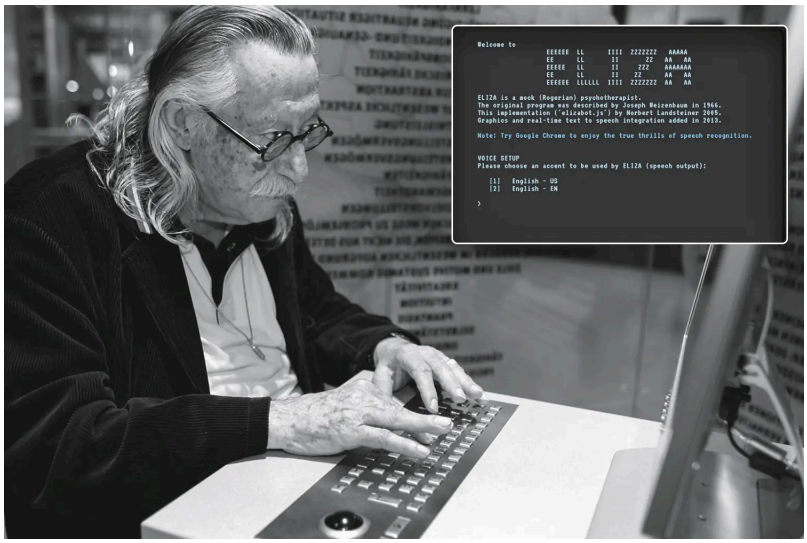
The first approach to AI starts in 1956, with the coining of the term "Artificial Intelligence" by a bunch of mathematicians at Dartmouth. They assume that one can represent the mind through abstract concepts and connections.

SHRDLU focuses on an atomistic conception of language, inspired by Chomsky. It also became famous because of its ***interface****. A technological innovation needs to impress at first sight.*

—

Criticism #1 - ELIZA

—



Joseph Weizenbaum programming ELIZA in 1964

—

The criticism of ELIZA , the first conversational agent (1964-1967).

```
- ELIZA: how are you feeling today?
- PATIENT: i'm feeling sad
- ELIZA: why are you feeling sad?
- PATIENT: because of some dumb thing
- ELIZA: what kind of dumb thing?
```

ELIZA was the first language-processing program which could converse with humans. Even though it was designed to show the superficiality of human-machine conversations, it nonetheless caused a sensation in the general public.

You can find more technical details here: https://dl.acm.org/doi/10.1145/365153.365168 , and a critical analysis here: https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/106

—

Should we use an AI as a therapist?

—

The eternal question of the ethics of technology: **just because we can, does it mean we should**[6]?

The first criticism of this approach comes from Joseph Weizenbaum who, in his book Computer Power and Human Reason makes the difference between things that require choice, which a computer can do, and things which require judgment, which a computer cannot do. This is because the former is only calculation, while the latter is value based.

He intends to show that machines can never be truly conscious by building one: the first chatbot, ELIZA. However, the illusion of intelligence remains and psychiatrists across the country reach out to him to use the robot in therapy.

—

Winter #1

—

It turns out, computers don't work so well in the real world. Funders are disappointed.

1965s - 1975s

These debates take place exclusively in academia, where there is quite a lot of computing power to do these intelligence experiments. However, the private sector does not benefit from it, and the hype dies out.

The Lighthill report is particularly damning.

—

Hype #2 - Expert systems

—

AI rebrands itself as **expert systems**: restricted problem domain, specific rules, large knowledge databases, and adapting to the technocratization of western economies.

- SABRE (automated travel agent) - 1976
- MAVENT (mortgage and lending compliance automation) - 1980
- HFT (high frequency trading) - 1983

The second wave results in expert systems, which require more computing power, and in the formalization of complex rule systems, such as finding the cheapest airline ticket, calculating someone's mortgage, or deciding to buy or sell stock. Some of these systems still exist today, but the normalization of their use makes them less prone to being qualified as "AI".

—

Criticism #2 - Chinese room

—

A system based on rules is not necessarily conscious

The second criticism aims at the fallacy of representing knowledge as rules. John Searle proposes an experiment to illustrate it: the chinese room paradox

—

Winter #2

—

Computers are still good at playing video games.

kasparov vs. IBM Deep Blue, 1996-1997

The summit of the logical/big hardware approach is when IBM builds a computer which manages to beat Kasparov, chess grandmaster. However, one might think it's only a fairly restricted conception of intelligence.

That is, computer scientists thinking that playing chess is being intelligence is a methodological bias since computer scientists probably already like playing chess and consider themselves intelligent.

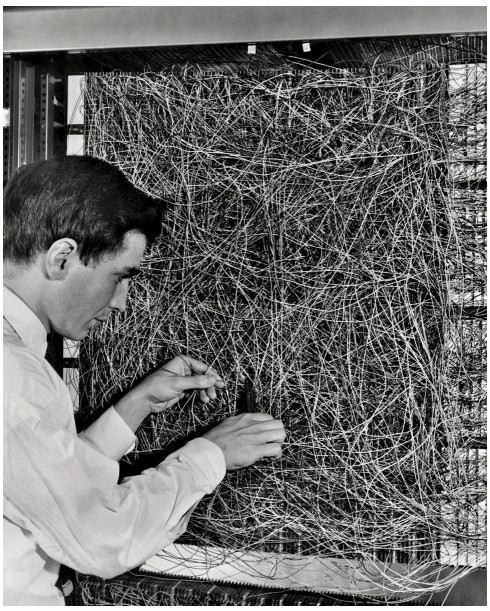
—

But there are limits to **cathedrals of rules**[7].

—

Connected intelligence

—



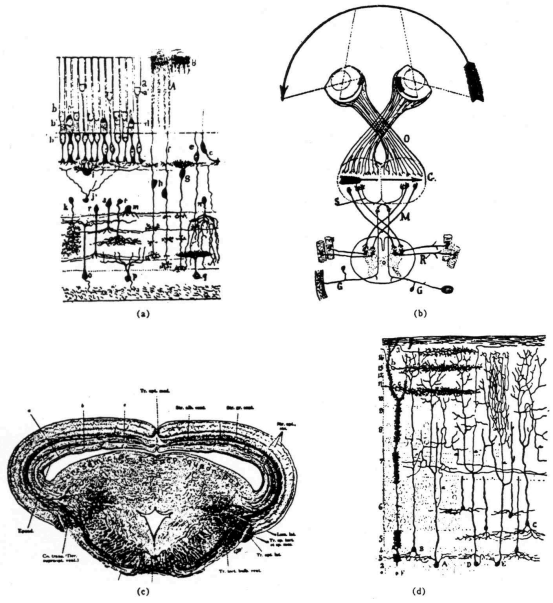
Meanwhile, Frank Rosenblatt works on the Perceptron in 1950

Along with the **symbolist** approach to artificial intelligence, was a **connexionist** approach.

There is another alternative to the development of AI: rather than logical of few, complex concepts, one can also take a huge amount of somewhat simple concepts.

This is a more physiological approach, one which aims at mimicking not the abstraction of the mind, but the reality of the brain.

—



There is more than meets the eye

It turns out that **sensing is also thinking**[8].

—

Embodiment and perception might be essential to intelligence (cf. Moravec's Paradox)

—

Learning by examples

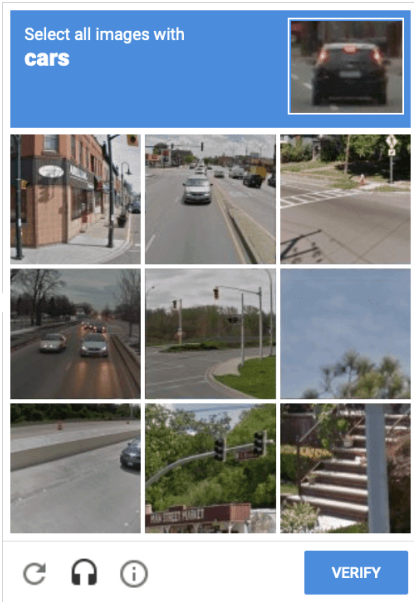
—

In the 2000s, **machine learning** really picks up, with the help of **data**, **hardware** and **adaptation**.

really appeared with better hardware and better data (machine learning, convolutional neural networks)

—

All of this data needs to be labelled—unskilled, low-paid work.



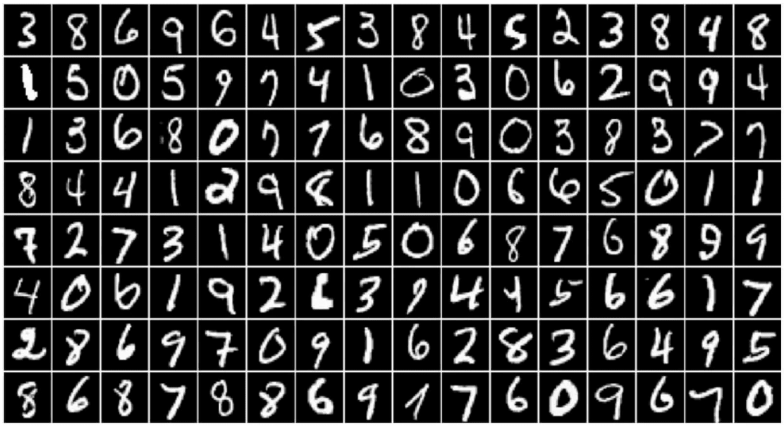
This Google widget asks you to label images for their self-driving car dataset

Datasets of pictures of streets and roads needs to have labels that describe them as pictures of cars, sidewalk, crosswalks, firehydrants, scooters, etc. You can see image datasets made by Google here . Notice how all the audio associated with it includes accents from the Global South.

This is done through services like Amazon Mechanical Turk , which provides a global workforce of disposable workers doing these menial tasks. The name is inspired by the late 18th century Automated Chess Player which pretended to be robot playing chess, but was hiding beneath was a real human.

There are some artistic projects trying to visibilize these "click workers": the middle finger response and mturk poems .

—



The MNIST dataset was the first one to be used to train a neural network to recognize numbers

The only thing that was missing was big data: now that there is a large amount of regular data, we can pay attention to a different kind of intelligence.

For instance: how come it's super easy for humans to recognize images, and very hard for computers? So the process of deep learning works very well with large amounts of data to do trial and error deductions.

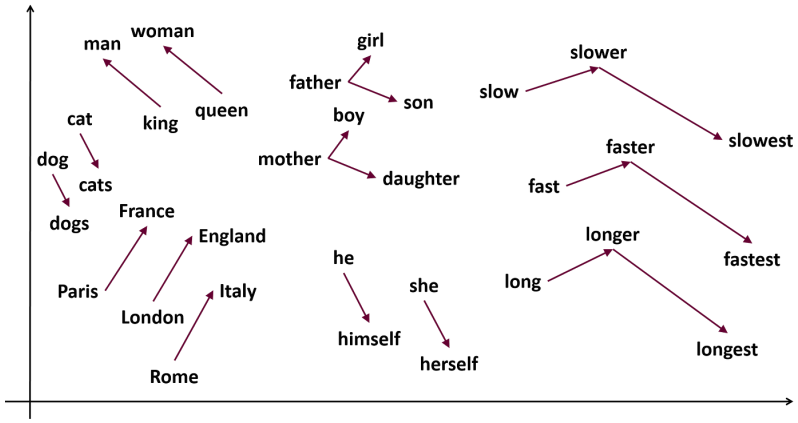
Because, ultimately, it is this trial and error approach which is a fundamental shift: intelligence systems are no longer inductive, they are deductive.

instead of inductive reasoning (i.e. knowing the rules beforehand), you go for deductive reasoning (i.e. you figure out the rules as you go).

aka you show tens of thousands of examples to an algorithm and it adapts.

—

For linguistic data (chains of characters), we turns **words into vectors**[9].



Word2Vec is the most popular implementation of word embeddings

word vectors: the cool thing. information retrieval, creating a semantic space to be searched

—

We get all these words from the Internet[10], and sometimes it's not the best representation of the world[11].

—

> *You shall know a word by the company it keeps[12].*

We figure out what a word means through its neighbors, and represent it in space .

—

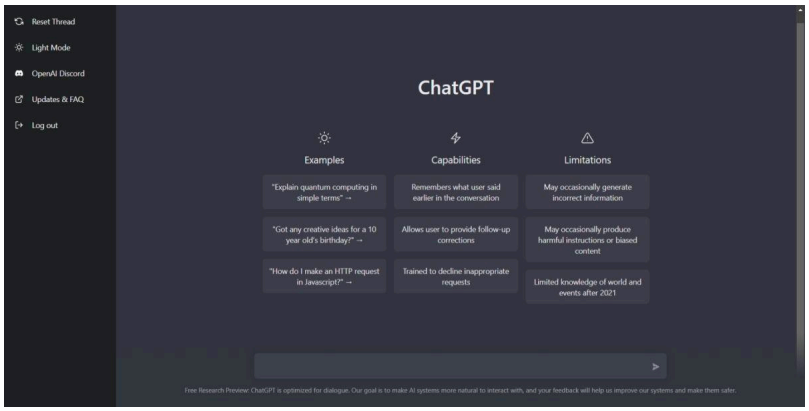
The most recent advance is the **Transformer** architecture, which adds a dimension of **attention** (aka which words are more relevant to others?)[13].

In 2017, Google releases a paper which combines this word-to-vector representation with a linear development: basically, they have an algorithm which can guess the next word in a sentence, and turns out it's very effective!

This solved the problem of indeterminate semantics (which words refers to what).

—

Finally, **the magic touch of the interface**: from autocomplete to chatGPT.



The online interface and name relied on previous familiarity with the users

Interfaces and test cases frame our understanding of what a new technology is.

—

The paradox of **ridiculous efficiency** and **epistemic abyss**.

i.e. it works very well, but we don't know why. Now what?

—

There is a new technology.

- Should it be applied indiscriminately? e.g. should it be applied to courtrooms?
- How do we distinguish between marketing and reality?
- What are the first and second order consequences of AI? e.g. reading less? writing more?
- Is it culturally biased ?

—

> *LLMs are the proof that reading makes you smart, and we use it to stop reading.*

—

Outro

_

Artificial Intelligence research is both a **philosophical project** (what is intelligence?) and an **engineering project** (how do we make something behave intelligently?).

So far the second one has had a lot more success. We tried to reason about it, but in the end **we learn by example**.

The overwhelming amount of money invested in AI fuels **a current hype bubble**, and makes it hard to disentangle truth from sales pitch.

Next session, we talk about the environmental impact of digital technologies.

- read the environmental toll of a netflix binge

_

Appendix

1. See how the story of the pin factory is fundamental to a theory of the division of labor: https://en.wikipedia.org/wiki/Division_of_labour#Adam_Smith

2. See https://en.wikipedia.org/wiki/Project_Cybersyn

3. Turing, Alan M. “Computing Machinery and Intelligence.”,Mind, 1950. https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf

4. See the genealogy of that idea in Linsky, Bernard, and Andrew David Irvine. “Principia Mathematica.” The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, 2022. https://plato.stanford.edu/archives/spr2022/entries/principia-mathematica/.

5. See the original proposal for the first AI workshop https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf.

6. Weizenbaum, Joseph. *Computer Power and Human Reason: From Judgment to Calculation* . 1st edition, W H Freeman & Co, 1976.

7. Cardon, Dominique. Mazières, Antoine. *Neurons Spike Back* , Réseaux, vol. 28. 2018. https://hal.science/hal-02190026v1/file/NeuronsSpikeBack.pdf

8. Lettvin, J. Y., et al. “What the Frog’s Eye Tells the Frog’s Brain.” Proceedings of the IRE, vol. 47, no. 11, Nov. 1959, pp. 1940–51. IEEE Xplore, https://doi.org/10.1109/JRPROC.1959.287207.

9. For an explanation of how this works, see: https://jalammar.github.io/illustrated-word2vec/

10. See Common Crawl https://commoncrawl.org/

11. Bolukbasi, Tolga, et al. “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings.” arXiv:1607.06520, arXiv, 21 July 2016. arXiv.org, https://doi.org/10.48550/arXiv.1607.06520.

12. Firth, J. R. “Applications of General Linguistics.” Transactions of the Philological Society, vol. 56, no. 1, 1957, pp. 1–14. Wiley Online Library, https://doi.org/10.1111/j.1467-968X.1957.tb00568.x.

13. See *Attention is all you need* , Google, 2017