Využití nástrojů AI v psychologii

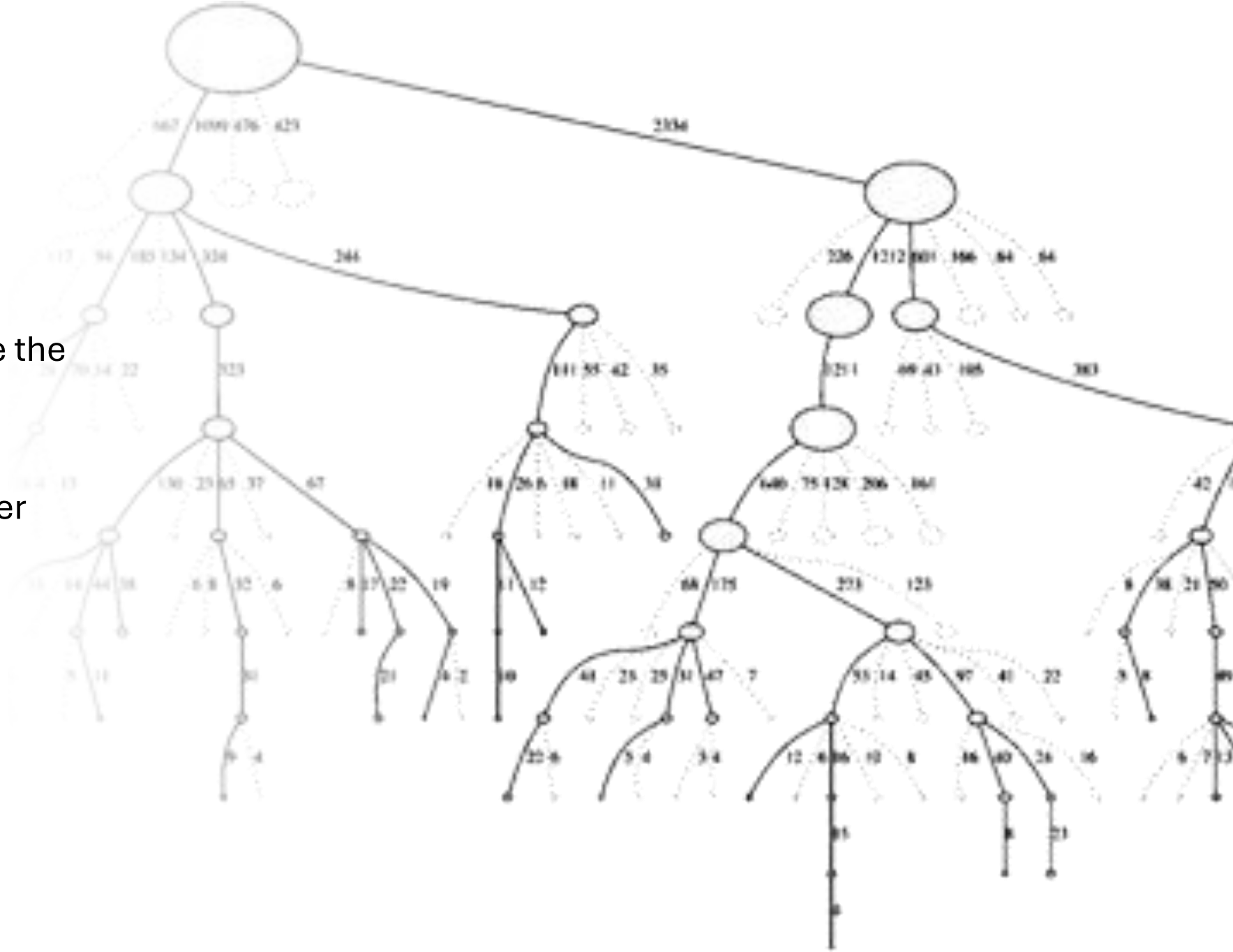
Vojtěch Formánek, FF, FI MUNI, 27.2.2025

Previous lecture

- TFaS + Simon
- Expert systems
- Scheduling
- Turing and testing

Heuristical search

- Use heuristics to reduce the number of options
- Can’t find best decision
- Good-enough work better



Expert systems

- Use rule-based heuristics
- Based on a database of rules
- Rigid, can’t handle uncertainty
- Easily explainable

RULE037

IF: 1) *The identity of the organism is not known with* certainty, and
2) The stain of the organism is gramneg, and
3) The morphology of the organism is rod, and
4) The aerobicity of the organism is aerobic

THEN: There is strongly suggestive evidence (.8) that the class of the organism is enterobacteriaceae

RULE145

IF: 1) The therapy under consideration is one of: cephalothin clindamycin erythromycin lincomycin vancomycin, and
2) Meningitis is an infectious disease diagnosis *for the patient*

THEN: It is definite (1) the therapy under consideration is not a potential therapy for use against the organism

RULE060

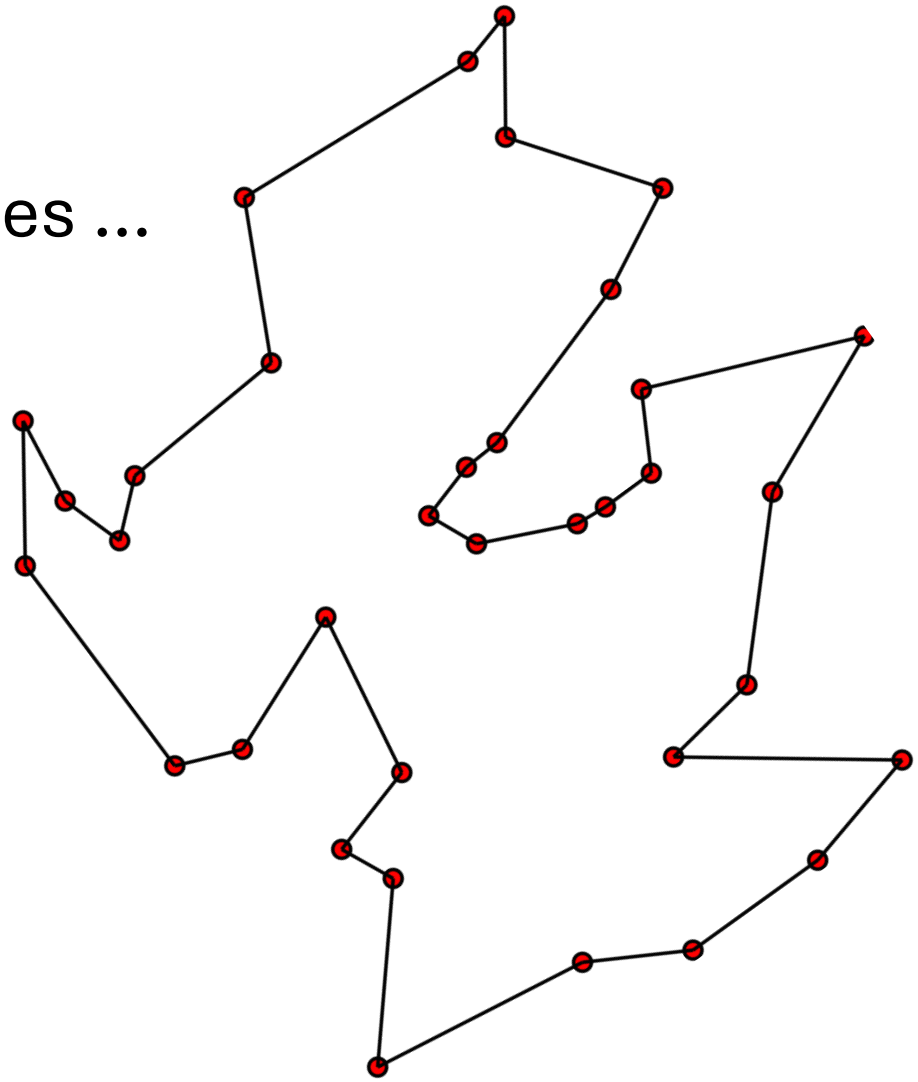
IF: The *identity of the organism is bacteroides*

THEN: I recommend therapy chosen from among the following drugs:

1 - clindamycin	(.99)
2 - chloramphenicol	(.99)
3 - erythromycin	(.57)
4 - tetracycline	(.28)
5 - carbenicillin	(.27)

Scheduling

- Used in planning manufacturing, timetables …
- Generate an initial solution, then break it
- Repair and get a (hopefully) better one



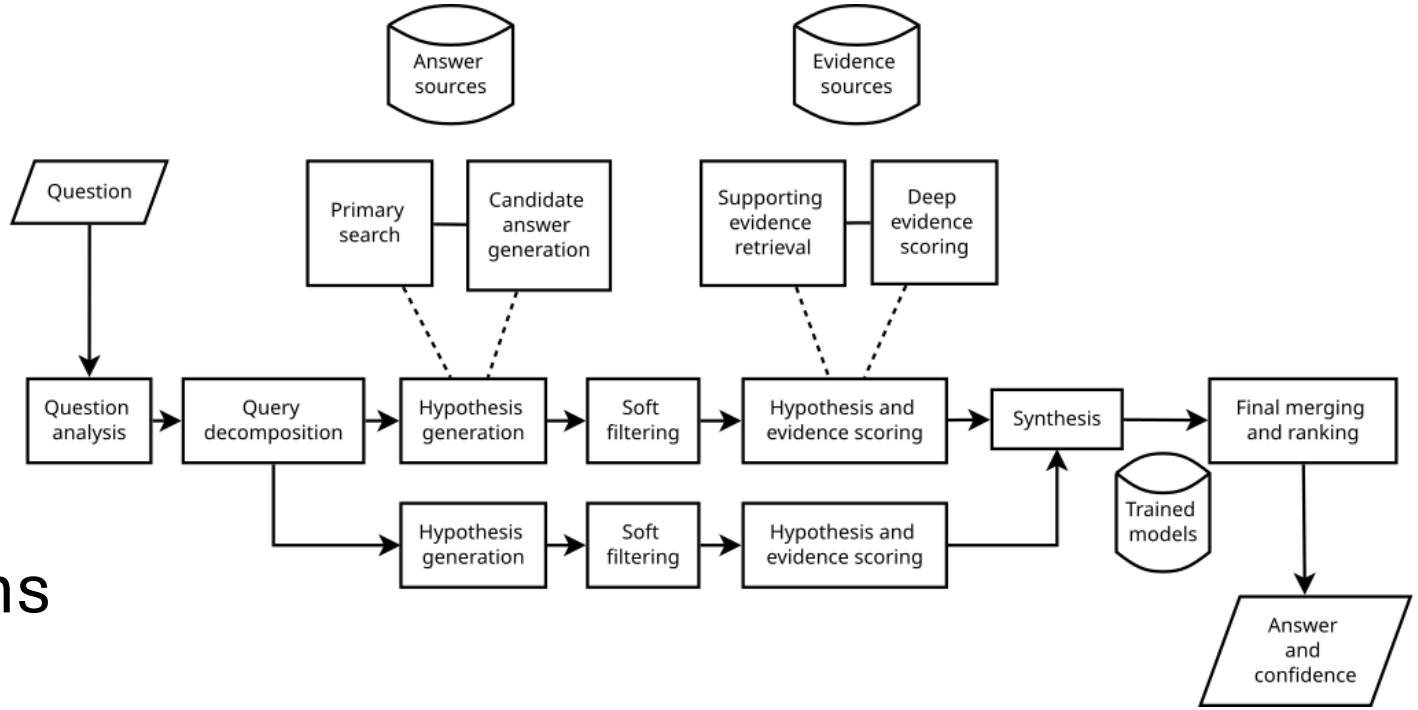
Today

- AI spring and Machine Learning
- Linear regression
- Neural nets
- Text generation
- Testing ML
- Other ML models

- AI spring and Machine Learning
- Neural nets
- Chatbots
- Testing ML
- Other ML models

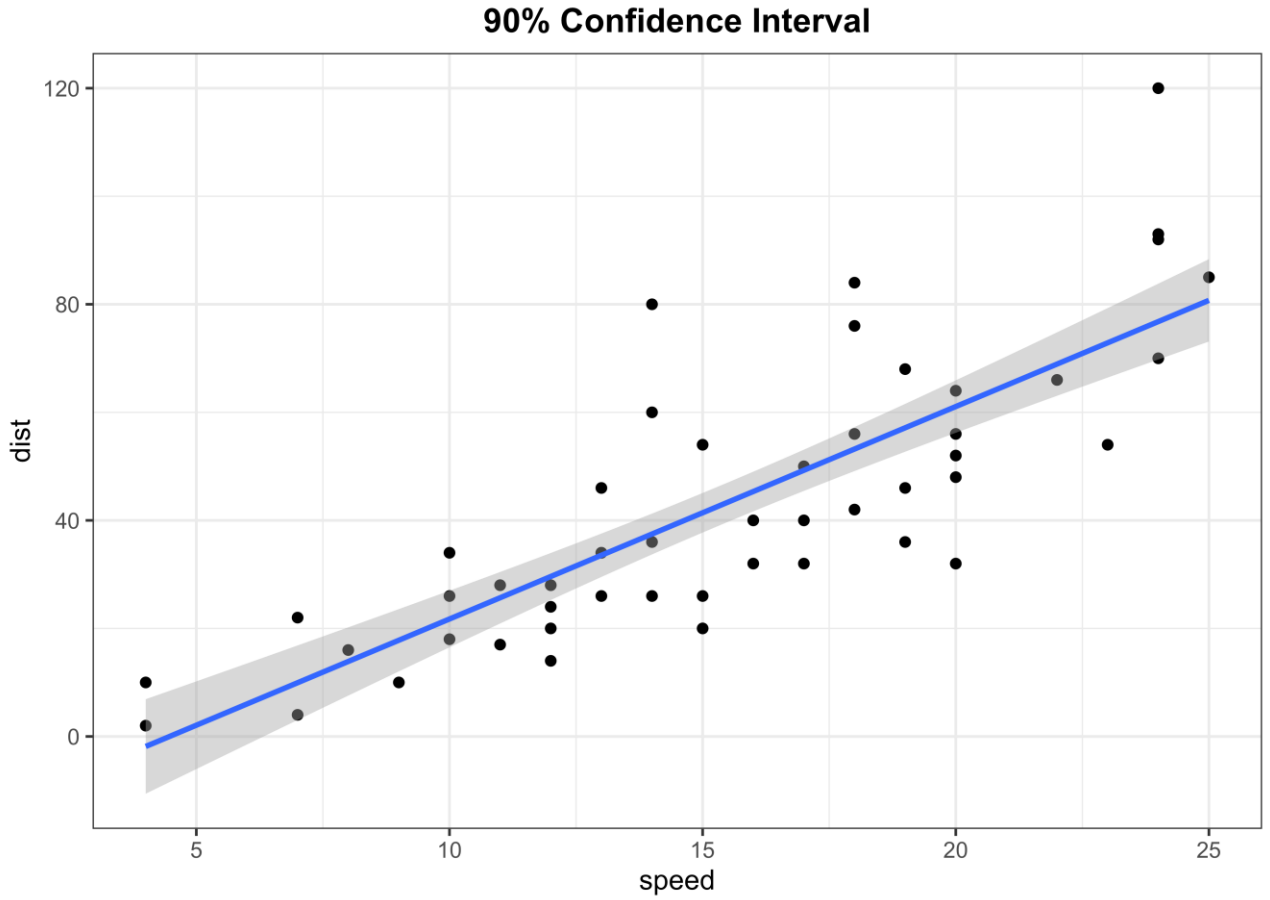
AI Spring

- WWW + big data
- How big? Thousands to millions points
- Shift to Machine Learning
- IBM Watson
- Early recommender systems



Linear regression, a reminder

- “Fit a line” with the smallest distance from all points
- Describes a trend/relation of the data
- Assumptions → CIs
- Generalize to a population



ML vs. Statistics

- Both generalize to a population
- Stats use assumptions, ML brute force
- ML stronger mathematically
- ML needs robust post-hoc testing

Linear regression, the ML way

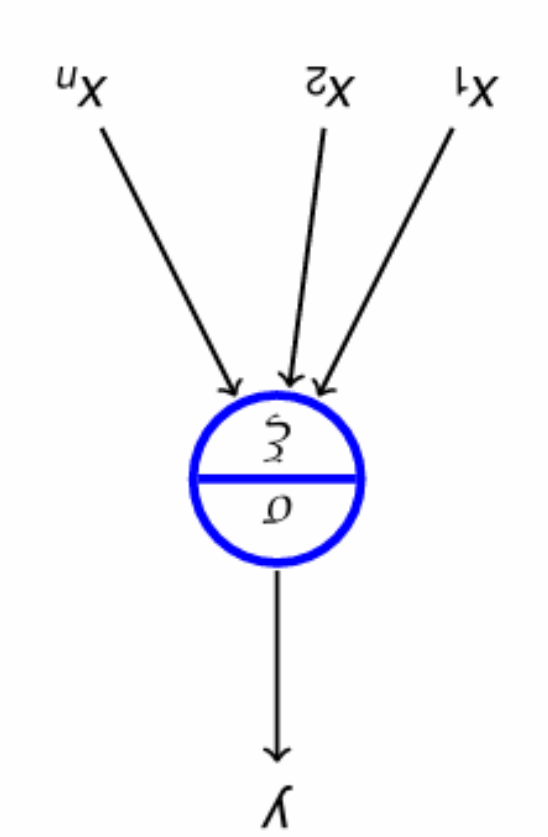
- AI spring and Machine Learning
- Neural nets
- Chatbots
- Testing ML
- Other ML models

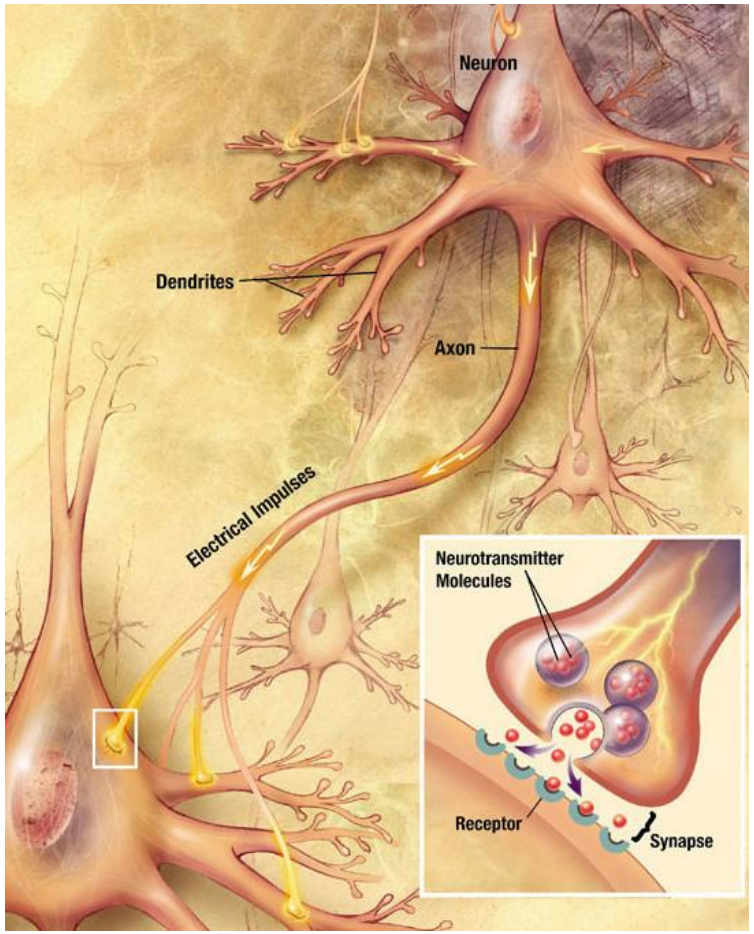
AI Boom II. (2014-)

- Wide use of neural nets
- Using GPUs to train them
- First in image processing (tumor, face detection)
- As rating functions for others

Neural Networks

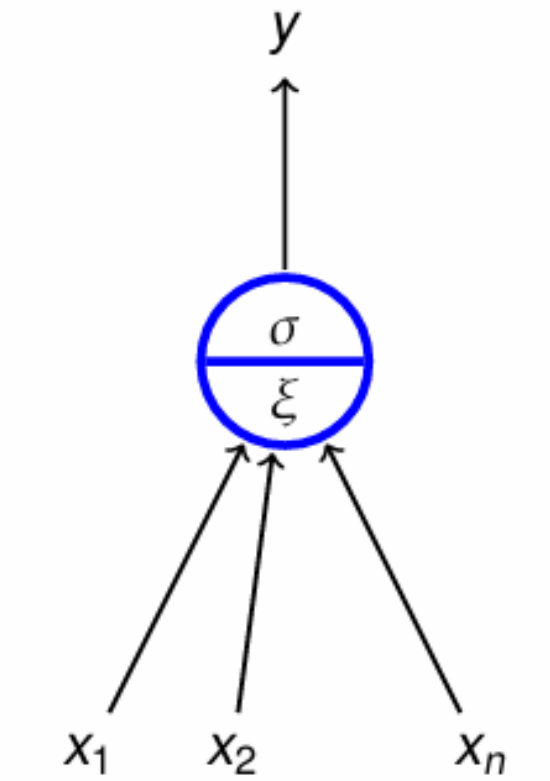
- First, biology inspired
- Effective training algo found in 80s
- Derivative architectures used everywhere
- Mathematically strongest
- Data intensive

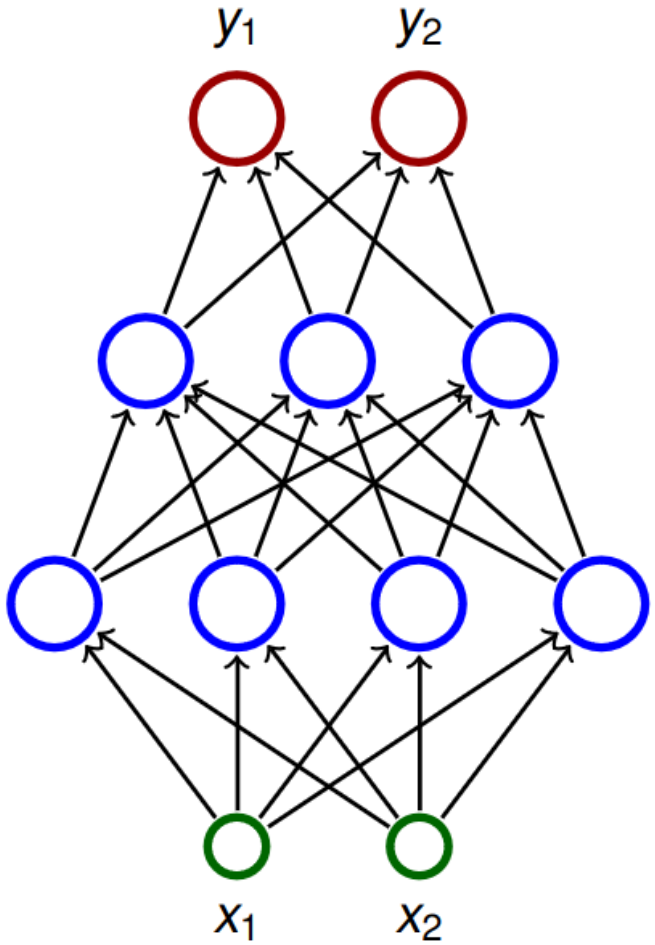


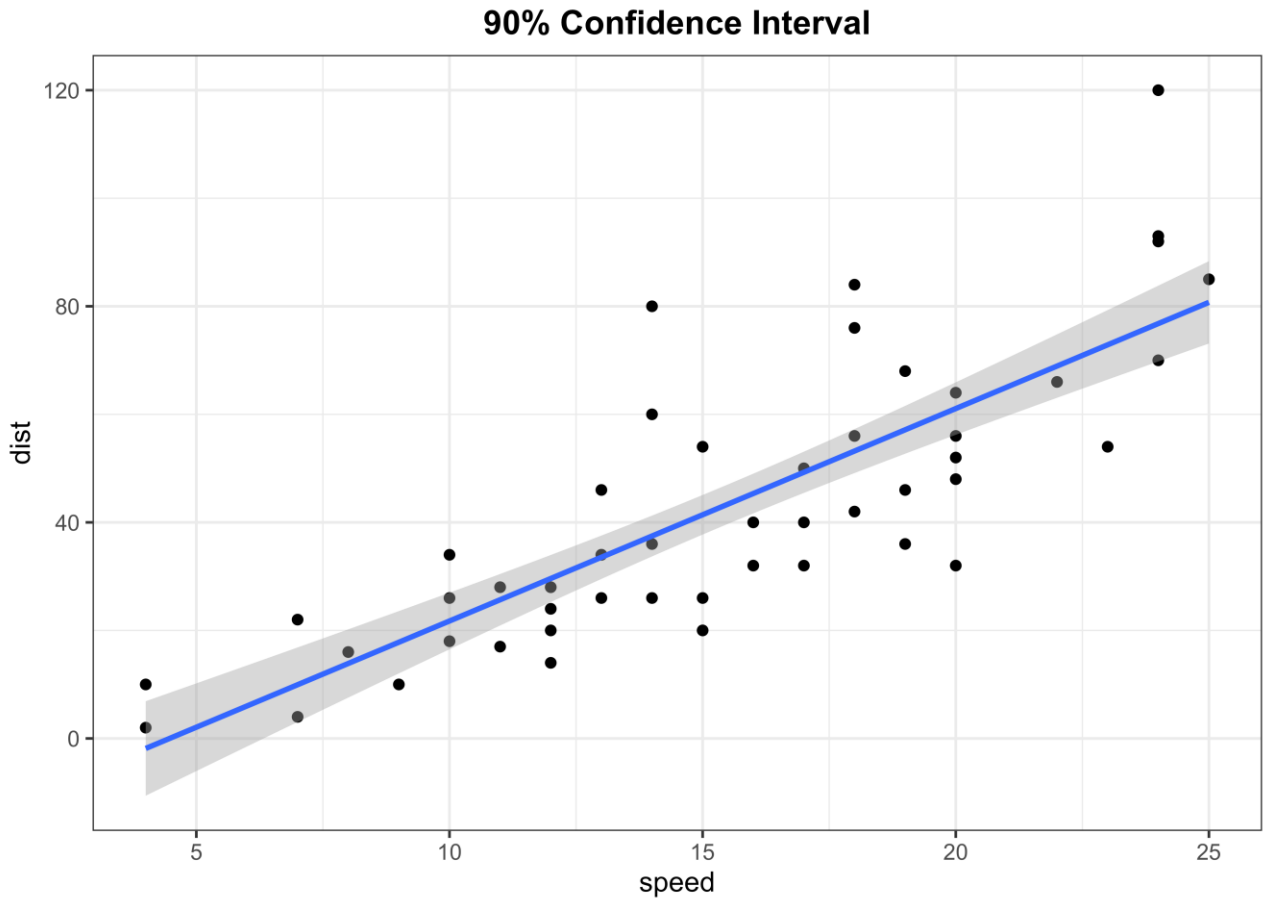


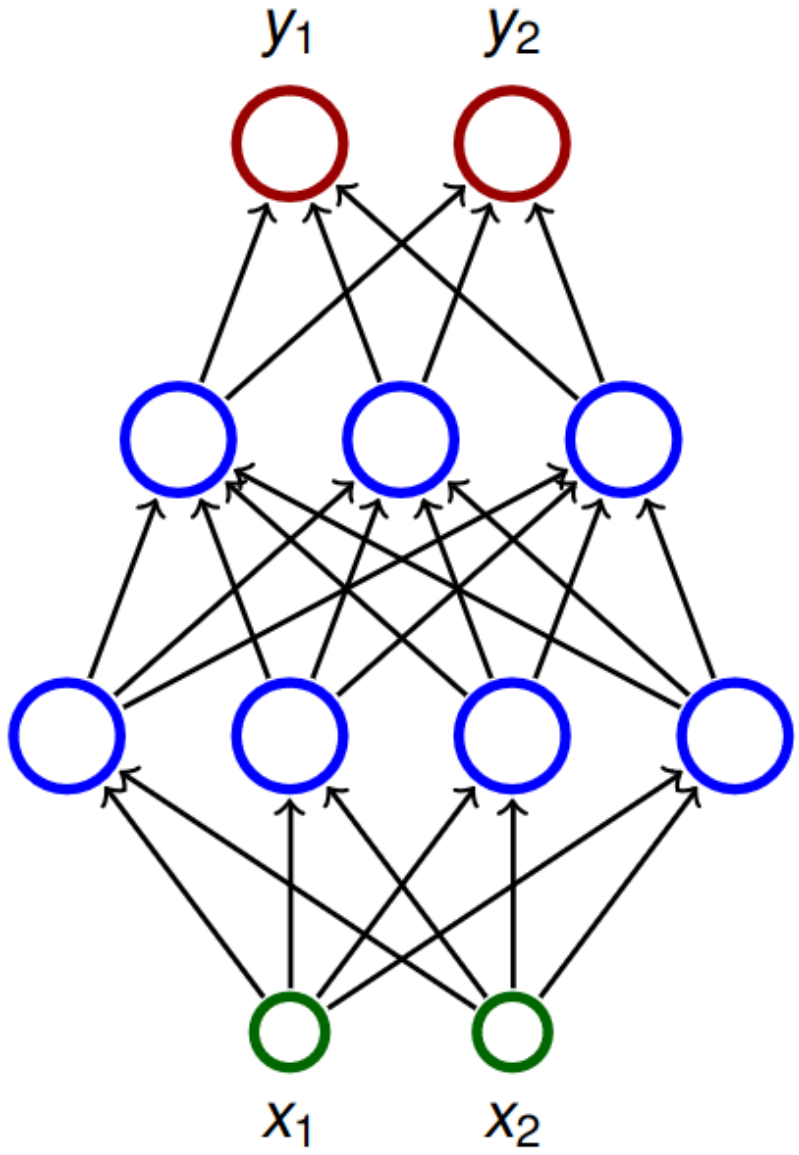
NNs and LR

- Like regression
- … on the results of regression
- Over and over
- A Neural Network Playground



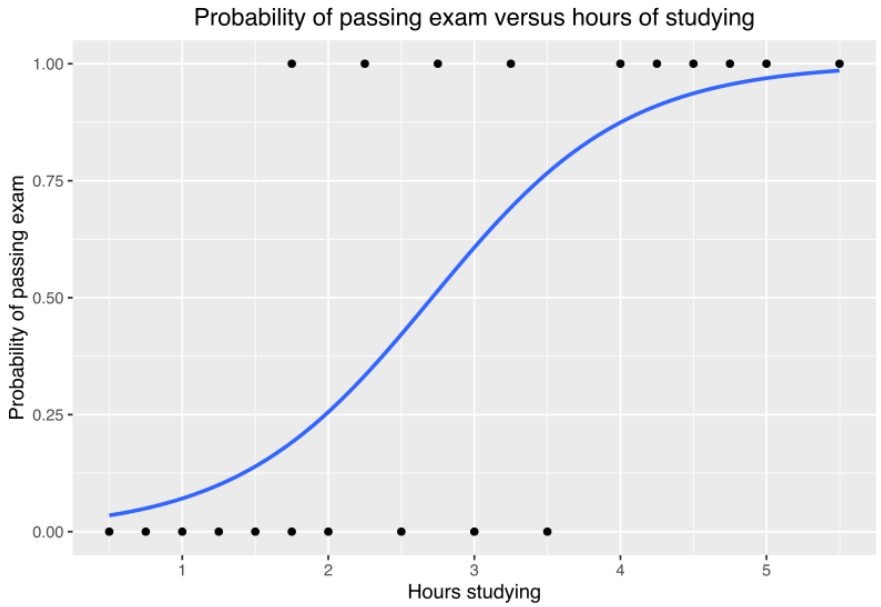


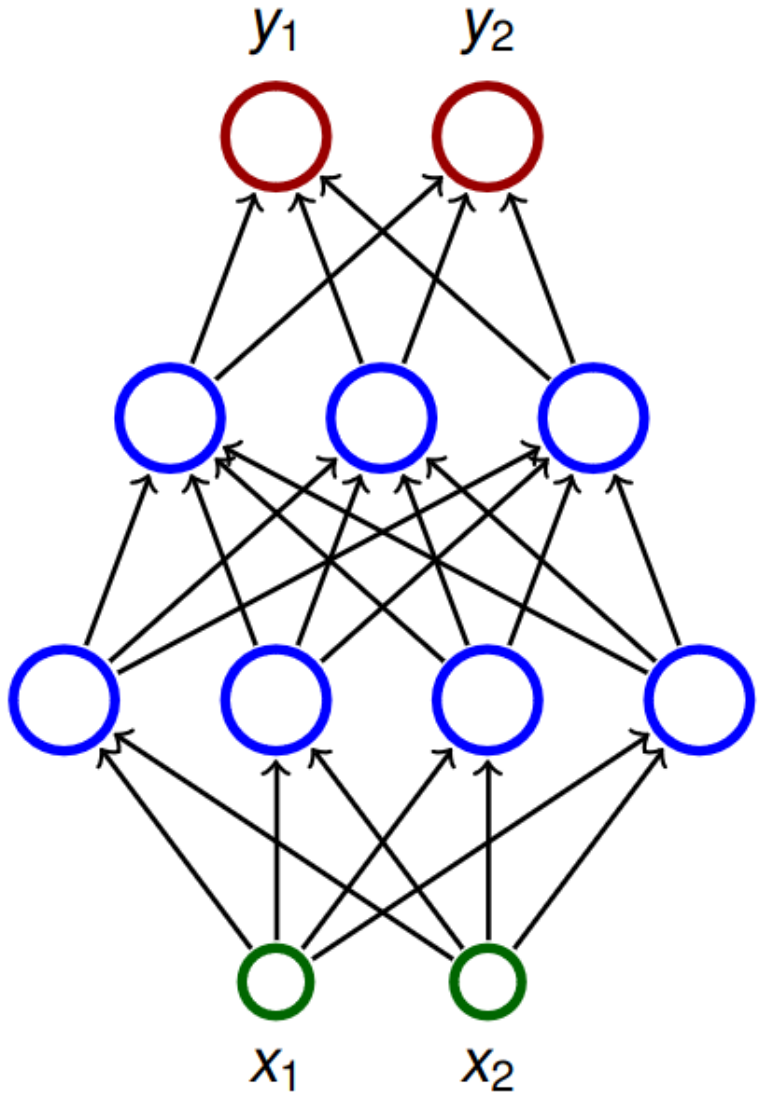

90% Confidence Interval
120
80
40
0
dist
5
10
15
20
25
speed


y_1
y_2
x_1
x_2

*about Activation Functions

- Individual layers differ by activation functions
- Which changes the curvature of the line
- Original NNs with sigmoid imitating biology
- Logistic > Linear





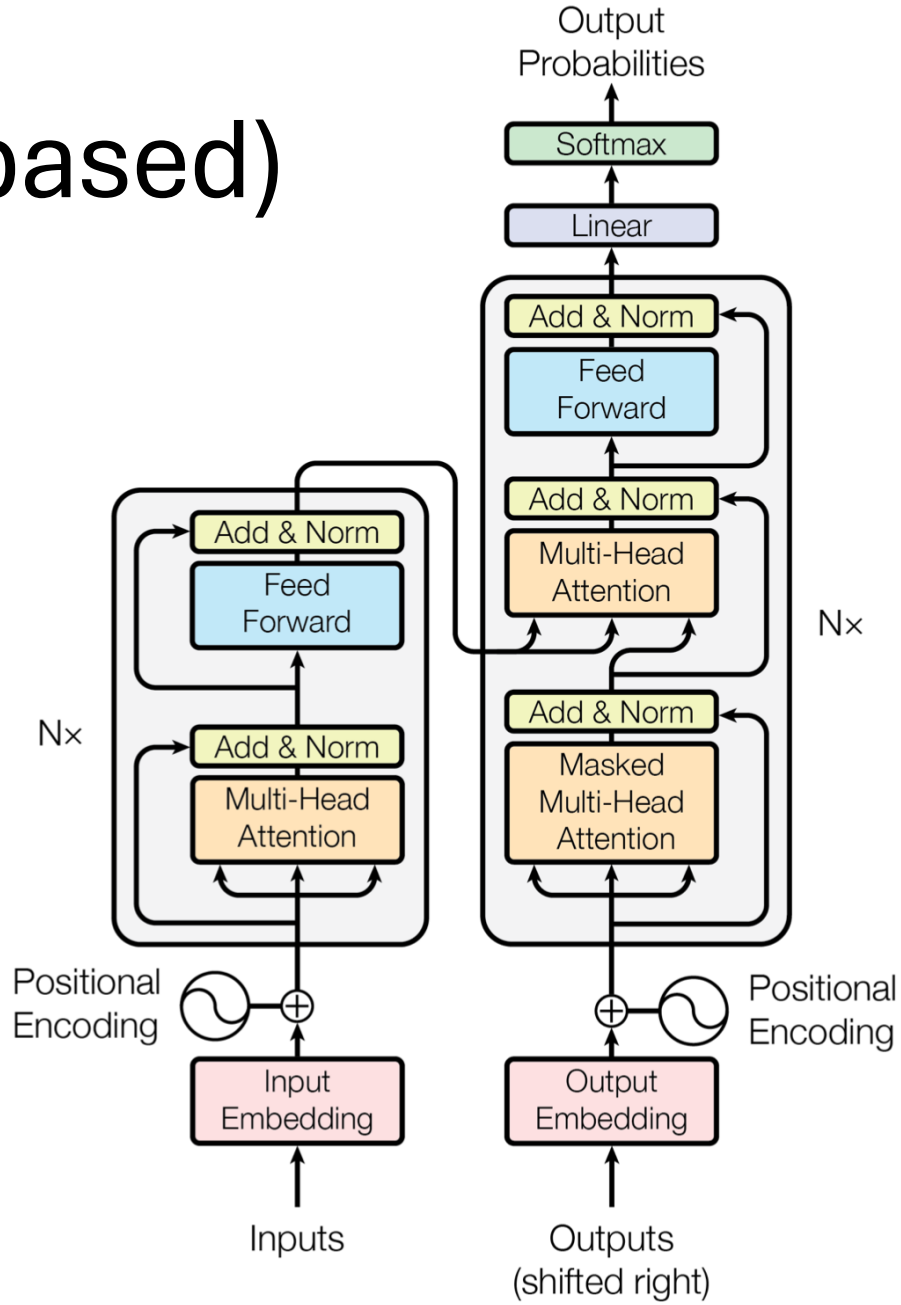
- AI spring and Machine Learning
- Neural nets
- Chatbots
- Testing ML
- Other ML models

AI Currently

- Recommender systems – Netflix, Youtube, TikTok …
- Image processing – Tumor segmentation, face recognition …
- Language models – ChatGPT, translation…
- Multimodal models
- Autonomous agents

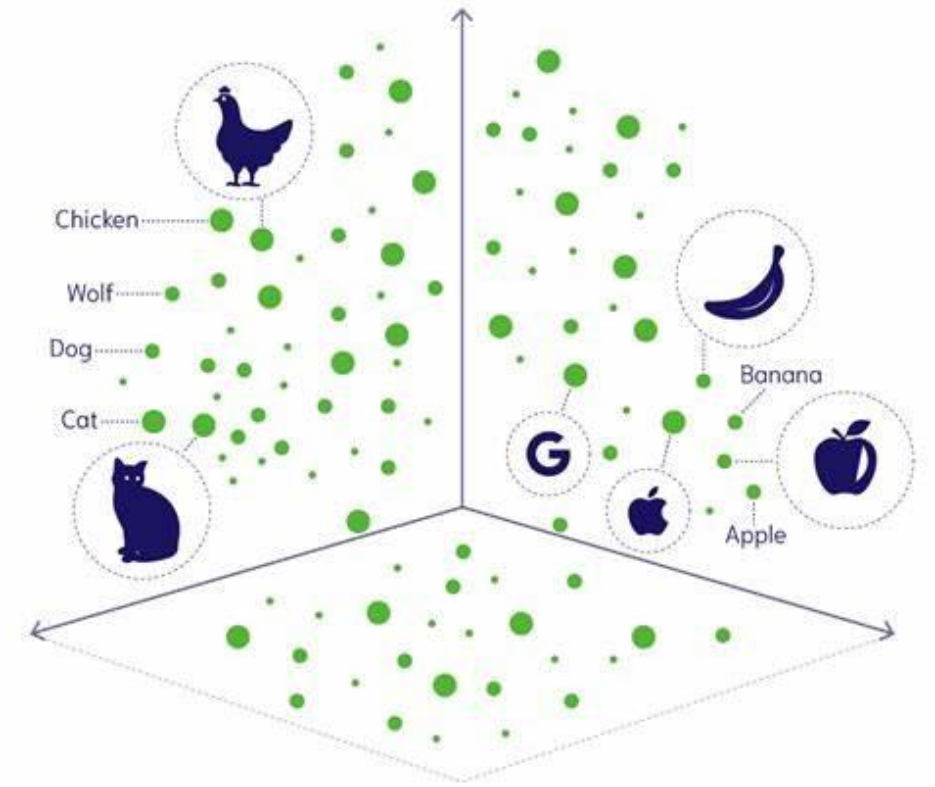
But What is a Chatbot (LLM-based)

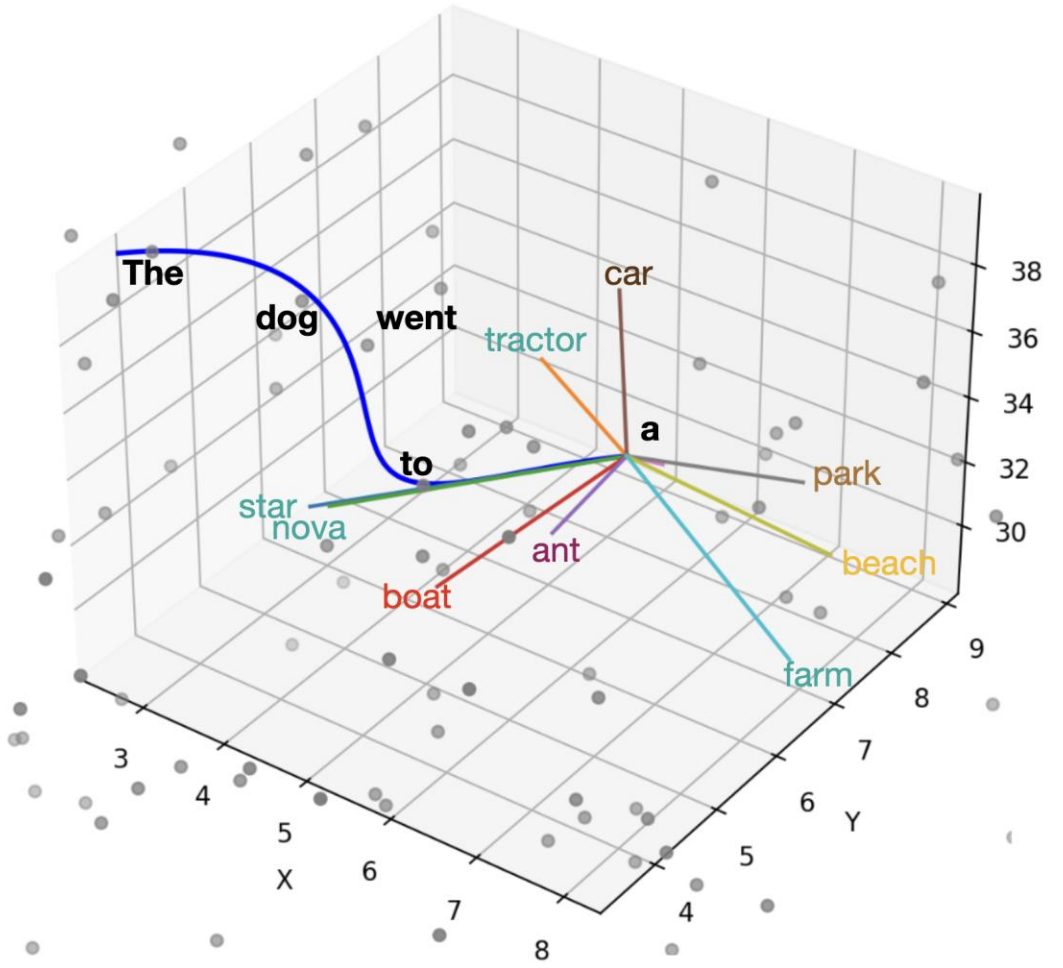
- Text-generating Large Language Model
- A transformer – more efficient neural net
- Precise architecture more complex
- Speeds up training
- Buuuut still really slow
- Example



*Meaning as Numbers

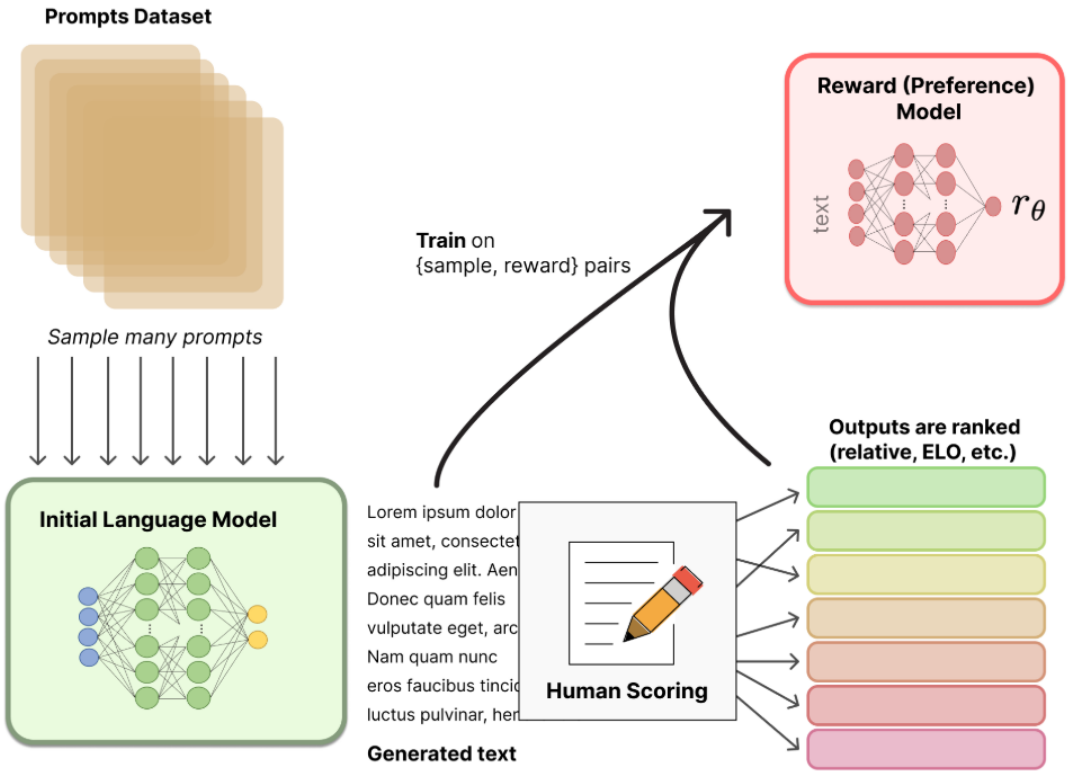
- mind





How to train a chatbot

- Initialization – tabula rasa
- Pretraining
  - massive with trillions of words, www
- Supervised fine-tuning
  - Model completes text
- Fine-tuning – small, expensive with a specific purpose
  - Reinforcement Learning from Human Feedback (RLHF)



- AI spring and Machine Learning
- Neural nets
- Chatbots
- Evaluating LLMs
- Other ML models

Evaluating LLMs

- Mix of human
  - Training checkpoints (human)
  - Human evaluation (Chatbot Arena, questionnaires, A/B)
  - Jailbreaking (S/F), Red-Teaming
- and automatic tests
  - Translation, Classification, Summarization accuracy
  - Benchmarks - MMLU, MathEval, HellaSwag, School entry exam tests
  - Text quality – specificity, text diversity …

A woman is
around tryir

A. rinses
B. uses a
C. gets th
D. gets in

Come to a
come to a
arrived bef
light, proce

A. Stop fo
yellow.
stop.

B. After y
Allow v
onto th

C. Stop out of the oncoming traffic. People coming in from behind may elect to stay left or right.

D. If the intersection has a white stripe in your lane, stop before this line. Wait until all traffic has cleared before crossing the intersection.

The night before his bar examination, the examinee's next-door neighbor was having a party. The music from the neighbor's home was so loud that the examinee couldn't fall asleep. The examinee called the neighbor and asked her to please keep the noise down. The neighbor then abruptly hung up. Angered, the examinee went into his closet and got a gun. He went outside and fired a bullet through the neighbor's living room window. Not intending to shoot anyone, the examinee fired his gun at such an angle that the bullet would hit the ceiling. He merely wanted to cause some damage to the neighbor's home to relieve his angry rage. The bullet, however, ricocheted off the ceiling and struck a partygoer in the back, killing him. The jurisdiction makes it a misdemeanor to discharge a firearm in public. The examinee will most likely be found guilty for which of the following crimes in connection to the death of the partygoer?

(A) Murder.
(B) Involuntary manslaughter.
(C) Voluntary manslaughter.
(D) Discharge of a firearm in public.

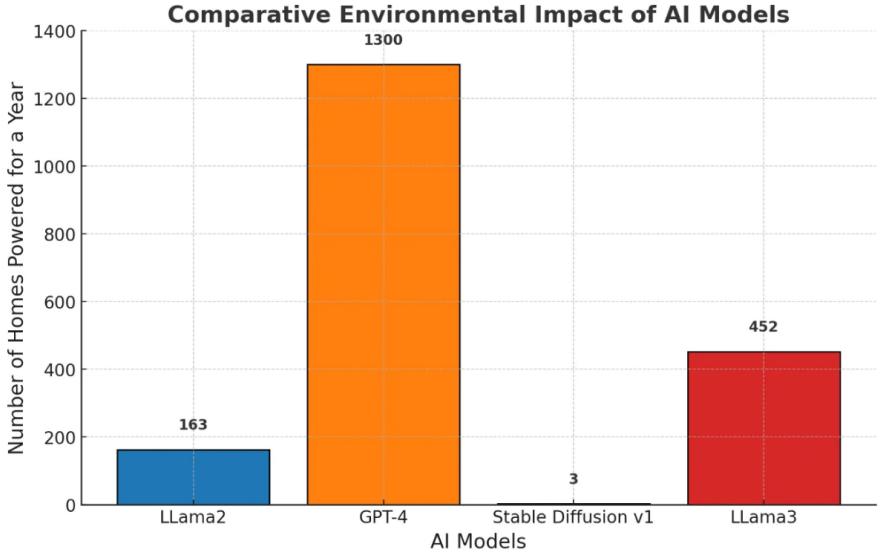
Intelligence and AGI

- Narrow, general, superintelligence
- ARC AGI
- Current systems probably won’t get us there
- And the cost is astronomical

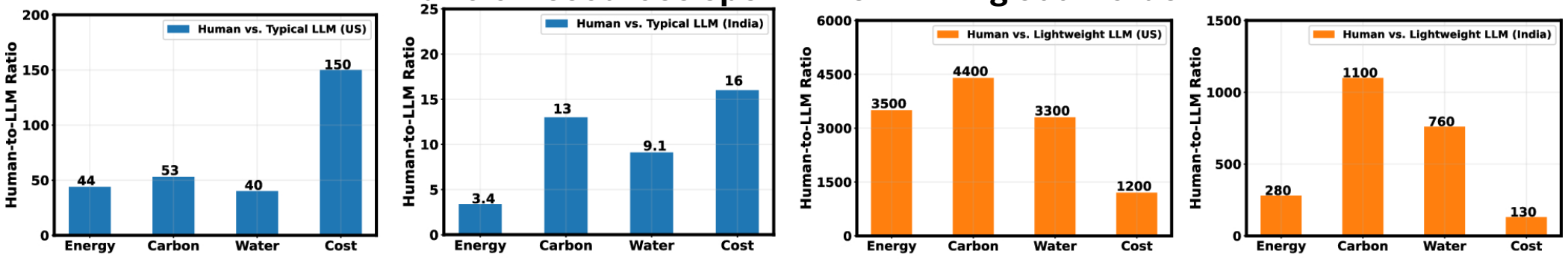
Chain-of-thought

The Environmental Cost of LLMs

- Effective compared to humans*
- Comparable to other internet activities
- Text < images < video
- o3/deepthought is expensive



A ratio of resources spent when writing 500 words



(a) Typcial LLM (U.S.) (b) Typical LLM (India) (c) Lightweight LLM (U.S.) (d) Lightweight LLM (India)

- AI spring and Machine Learning
- Neural nets
- Chatbots
- Evaluating ML
- Other ML models

Other ML models

- Decision trees
- Random Forests
- SVMs
- Clustering

and many more…

Sources

- Plotting different Confidence Intervals around Fitted Line using R and ggplot2 | DataWim
- By Pgr94 - Own work based on diagram found at http://www.aaai.org/Magazine/Watson/watson.php, CC0, https://commons.wikimedia.org/w/index.php?curid=14575947
- https://en.wikipedia.org/wiki/Synapse
- Illustrating Reinforcement Learning from Human Feedback (RLHF)
- Using ChatGPT is not bad for the environment
- logistic regression model – Wikidata
- LLM Visualization: How embedding space creates intelligence | Ron J's Projects

- Ren, S., Tomlinson, B., Black, R. W., & Torrance, A. W. (2024). Reconciling the contrasting narratives on the environmental impact of large language models. *Scientific Reports*, *14*(1), 26310.