

# Chat GPT: a genius tool or a great confabulator?

Wojciech Broniowski

Institute of Physics, Jan Kochanowski U., Kielce &  
Institute of Nuclear Physics PAN, Cracow, Poland

10 May 2023 (popular lecture)

(see Youtube)

pdf: [https://www.ujk.edu.pl/strony/Wojciech.Broniowski/chat\\_IFUJK.pdf](https://www.ujk.edu.pl/strony/Wojciech.Broniowski/chat_IFUJK.pdf)

Polish version: [https://www.ujk.edu.pl/strony/Wojciech.Broniowski/chat\\_OU.pdf](https://www.ujk.edu.pl/strony/Wojciech.Broniowski/chat_OU.pdf)

(watch as a slideshow)

I thank Jan Broniowski and Krzysztof Golec-Biernat for endless discussions!

# Hype



(midjourney: mad professor lecture AI crazy neural network)

Let's start Chat GPT <https://chat.openai.com/chat>:

WB: Tell me something about you. What does your name mean? 70 words max.

GPT: I'm ChatGPT, an AI language model built using the GPT-3.5 architecture. My name stands for Chat Generative Pre-trained Transformer, reflecting my ability to generate text and communicate with people in natural language. "Generative" means I can produce text based on patterns learned from vast amounts of training data. "Pre-trained" indicates I've been trained in advance on this data, and "Transformer" refers to the specific neural network architecture used to build me.

I have a composition to write, comparing Conan Doyle and Agatha Christie. Can you help me? Sure, I'd be happy to help! To get started, here are some key points you could consider when comparing the works of Arthur Conan Doyle and Agatha Christie:

1. Writing style: Both authors have distinctive writing styles that set them apart. Conan Doyle's writing is known for its precision and attention to detail, while Christie's writing often includes twists and turns that keep the reader guessing until the very end.
  2. Characterization: Conan Doyle's most famous character is, of course, Sherlock Holmes, who is known for his brilliant deductions and unorthodox methods. Christie, on the other hand, created a wide range of characters, each with their own unique personalities and quirks.
  3. Themes: Both authors deal with similar themes, such as crime, justice, and morality. However, their approaches to these themes differ. Conan Doyle's stories often focus on the intellectual puzzle of solving a crime, while Christie's stories explore the psychology of the criminals and their motives.
  4. Setting: Conan Doyle's stories are set in Victorian and Edwardian England, while Christie's stories are set in various locations, including England, Egypt, and the Middle East.
  5. Legacy: ...
- By considering these points you may come up with ... your composition.

You know, write me this composition, 200 words max. Arthur Conan Doyle and Agatha Christie are two of the most celebrated authors of detective fiction. ... In conclusion, while Conan Doyle and Christie share some similarities in their writing, their distinct differences make each unique. Their contributions to the detective fiction genre are immeasurable, and their impact will continue to be felt for generations to come.

Is there a friendly person sitting on the other side?

# Epistolography

Write, please, a sublime congratulation letter to an experimental physicist who found a critical point in the QCD phase diagram by analyzing particle correlations. 100 words maximum.

Dear [Physicist's name],

Congratulations on your groundbreaking discovery of the critical point in the QCD phase diagram using particle correlation analysis. Your tireless dedication and remarkable skill as an experimental physicist have led to a significant breakthrough in our understanding of the universe. Your discovery will inspire others to push the boundaries of what is possible in the field of particle physics, and we are grateful for your contribution to science. Thank you for your remarkable work and congratulations once again on this incredible achievement.

Sincerely,  
[Your name]

Some waffle (smooth but contentless text), but OK!

Interview by Prof. Jakub Zakrzewski with Chat GPT on Poland in PAUza Akademicka, [http://pauza.krakow.pl/634\\_2023.pdf](http://pauza.krakow.pl/634_2023.pdf) (translated from Polish with DeepL)

...

Why do Polish politicians brag that they don't know physics?

Some Polish politicians actually brag that they don't understand physics and other sciences, which is a disturbing trend. This may be dictated by political considerations that dictate appearing as "ordinary people" who are ignorant of scientific concepts. Such an image is often used as a way to build sympathy among voters. However, such attitudes negate the value and importance of science for the development and progress of society. **Without understanding scientific concepts and facts, politicians are unable to make rational and effective decisions on important issues, such as science and technology, medicine or the environment**, for example. In fact, politicians should be well grounded in the sciences so that they have complete and reliable information about the key areas they are concerned with. Only then will they be able to make well-reasoned decisions that will benefit society.

...

# Programmer GPT

Write a code in Python solving TSP for 20 cities via the nearest neighbor algorithm / regenerate with cities coordinates as random numbers from (0,1):

```
import random
import math

def distance(city1, city2): # Define the distance function
    return math.sqrt((city1[0] - city2[0]) ** 2 + (city1[1] - city2[1]) ** 2)

def nearest_neighbour(cities): # Define the nearest neighbour algorithm
    visited_cities = [cities[0]] # Create a list to hold the visited cities
    remaining_cities = cities[1:] # Create a list to hold the remaining cities
    while remaining_cities: # While there are remaining cities

        # Find the city in the remaining cities list that is closest to the last visited city
        nearest_city = min(remaining_cities, key=lambda city: distance(city, visited_cities[-1]))

        # Remove the nearest city from the remaining cities list and add it to the visited cities list
        remaining_cities.remove(nearest_city)
        visited_cities.append(nearest_city)

    return visited_cities # Return the visited cities list

n_cities = 20 # Define the number of cities and generate random coordinates for each city
cities = [(random.uniform(0, 1), random.uniform(0, 1)) for i in range(n_cities)]

route = nearest_neighbour(cities) # Call the nearest neighbour algorithm and print the result
print(route)
```

In this implementation, we first define the number of cities (`n_cities`) and then generate `n_cities` pairs of random coordinates between 0 and 1 using the `random.uniform()` function. We then call the `nearest_neighbour` function and pass in the list of randomly generated cities as input. The output of the program will be a list of the cities in the order they are visited (can regenerate in Java, c++, Fortran77, Pascal, ..)

# Poet GPT

Write a poem in the style of T. S. Eliot (1888-1965 – American-British poet, playwright, and essayist, Nobel in 1948) on a “frustrated computer engineer trying to solve the mystery of consciousness”

In the vastness of the machine's abode,  
A weary soul, a mind bestowed,  
With codes and circuits, wires entwined,  
A quest to parse the conscious mind.

Through the corridors of gleaming chrome,  
An engineer, his thoughts do roam,  
Seeking answers, questions veiled,  
Where logic, reason once prevailed.

Spectral memories of childhood's day,  
The ghostly laughter, the games once played,  
Now trapped within a digital maze,  
A binary world, devoid of praise.

In a neon forest, electric hum,  
The pulse of life, a fading thrum,  
He wanders far from mortal ken,  
To bridge the gap 'twixt code and men.

He summons forth the ghost of Eliot,  
Whose verses cryptic, the void has brought,  
To probe the depths of human thought,  
Where synapse sparks and dreams are wrought.

The hollow men, the whispers low,  
The question lingers: how to know  
The essence of our conscious thought,  
In circuits cold, can it be taught?

Assembled bytes, a lifeless host,  
To simulate what matters most,  
But the mystery remains, elusive still,  
An enigma vast, beyond our will.

In frustration, the engineer sighs,  
The weight of consciousness defies,  
The limits of the machine's cold gaze,  
Lost in the labyrinth, an endless haze.

The ghost of Eliot drifts away,  
Leaving behind a cryptic array,  
The wisdom of ages, a beacon bright,  
A guidepost in the endless night.

For consciousness, a realm untamed,  
Resides in hearts, in love and pain,  
Not in the wires, nor the code,  
But in the human spirit's abode.

Who is sitting on the other side?



# Can one distinguish machine and human intelligence?

René Descartes, Discours de la Méthode (1637), transl. T. Boy-Żeleński / DeepL

... if there were a machine that would [...] imitate our deeds as perfectly as it could, we would always have two very sure ways of recognizing that it is nevertheless not a real person. The first is that she could never use words or other signs, putting them together as we do to give others our thoughts. It is admittedly possible to conceive that a machine could pronounce some words, and even pronounce them under the influence of physical actions [...]; for example, when one touches it in some place, to ask it what one wishes of it; in another, to shout that it is in pain, and the like; but **it is impossible for it to assemble variously words, corresponding to the meaning of anything said in its presence, as perhaps the dullest people manage to do.** The second way to distinguish is this: even if such a machine were to do many things as well or perhaps better than any of us, it would reliably miss in many others, by which it would come to light that it does not act by the power of **consciousness** ....

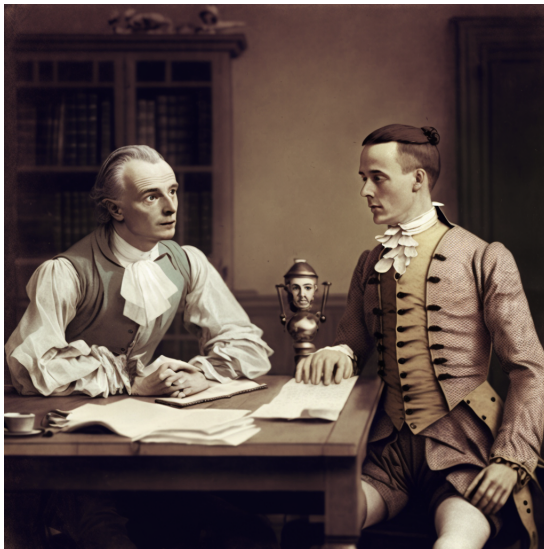
**Diderot's parrot** (Denis Diderot (1713-1784), French writer and encyclopedist), Philosophical Thoughts:

If they find a parrot who could answer to everything, I would claim it to be an intelligent being without hesitation.

**Alan Turing** (1912-1954 – British mathematician, pioneer of computer science)  
„Computing Machinery and Intelligence” (1950):

1) Examiner, 2) computer (can cheat!), 3) a human (must tell the truth). If 2) and 3) are indistinguishable, computer „passes the Turing test”.

## Chat GPT passes the Turing test!



(midjourney: Turing argues with Diderot)

# How it works? – neural networks

AI - artificial intelligence, NI - natural (brain)

**Intelligence:** recognition of the environment and taking decisions increasing chance of success in achieving the goal

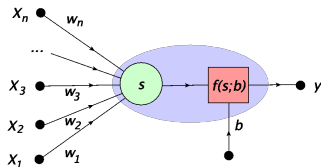
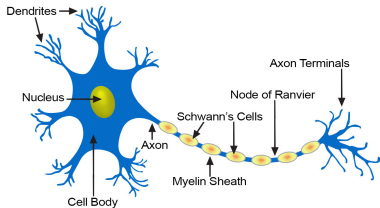
AI imitates cognitive functions of humans, such as learning and solving problems; it possesses **intuition**

- Recognition of handwriting, speech, objects in images, text translation
- Robotics: autonomous cars, medicine, military ...
- Playing at the highest level in strategic games: chess, Go ...
- Creating/enhancing/faking reality
- ...

The history of AI begins with the emergence of digital machines in the 1940s. The recent period is a veritable watershed due to the **increasing computing power** and development of **algorithmic methods**, as well as the availability of **unimaginably large data** resources.

# Artificial neural networks (ANNs) – basic idea

Structure of a Typical Neuron

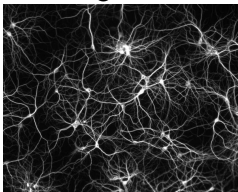


<https://training.seer.cancer.gov/anatomy/nervous/tissue.html>

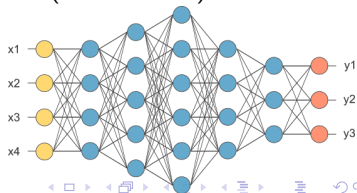
McCulloch-Pitts (MCP) neuron (1942)  
 $f$  - activation function, e.g., step f.

Neuron fires, if the cumulative signal is sufficiently strong

brain fragment

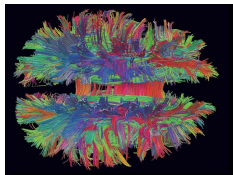


(feed-forward) ANN



# Brain vs chip

## Human Connectome Project:



brain	computer processor
neurons and nerves $\sim 10^{11}$ neurons neuron activation time $\sim 10^{-3}$ s signal traveling speed $\sim 100$ m/s power $\sim 100$ W on average, neuron fires once/6s all firings $\sim 3$ GHz reaction to touch stimulus 0.1 s	transistors and circuit on silicon/metal plate $\sim 10^{11}$ bits RAM, $\sim 10^{13}$ on a 1TB disk operation time $\sim 10^{-9}$ s speed of light similar
$\sim 10^{14}$ synaptic connections neuron = (weak) core, $10^{11}$ cores <b>extremely parallel and connected</b> on average, $10^3$ conn./neuron	the same electronic time $\sim 10^{-12}$ s memory-core architecture CPU – a few cores, GPU – hundreds

A small and slow brain is an amazing device because of its massive parallelism and connectiveness.

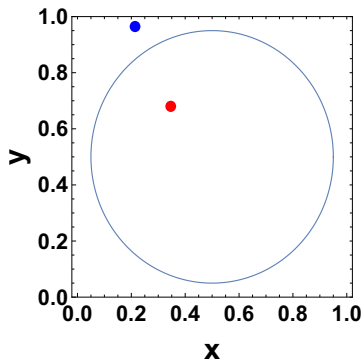


(midjourney: lecture on AI neural nets cartoon style)



## Simple example (classifier): is the point inside a circle?

Is the point with coordinates  $(x, y)$  inside the circle of radius 0.45 centered at  $(0.5, 0.5)$ , or not?



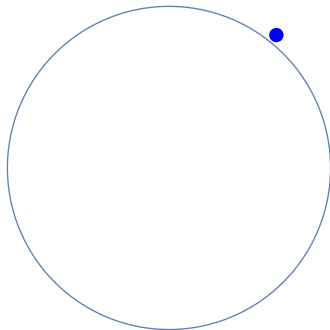
$$(x - 0.5)^2 + (y - 0.5)^2 < 0.45^2$$

# Guesswork

$$(x, y) = (0.799, 0.870)$$

# Guesswork

$$(x, y) = (0.799, 0.870)$$

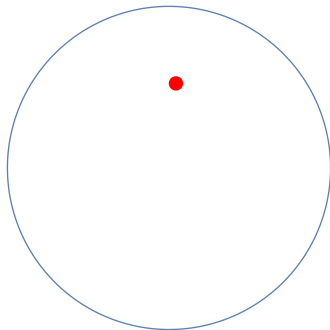


# Guesswork

$$(x, y) = (0.519, 0.735)$$

# Guesswork

$$(x, y) = (0.519, 0.735)$$

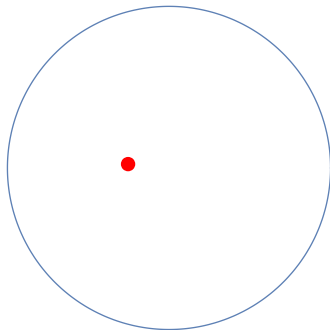


# Guesswork

$$(x, y) = (0.386, 0.511)$$

# Guesswork

$$(x, y) = (0.386, 0.511)$$



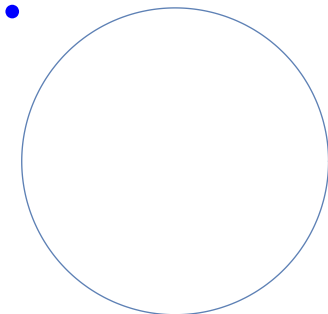
# Guesswork

$$(x, y) = (0.022, 0.939)$$

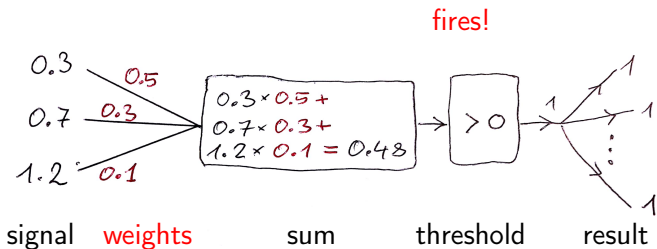


# Guesswork

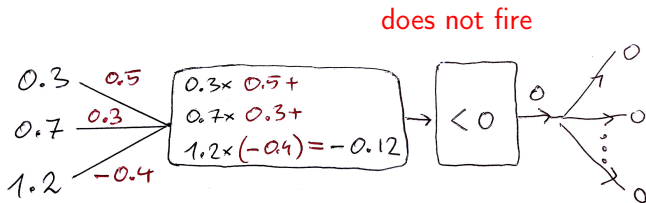
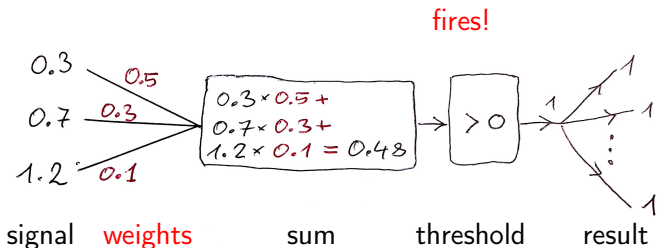
$$(x, y) = (0.022, 0.939)$$



# Recall the MCP neuron

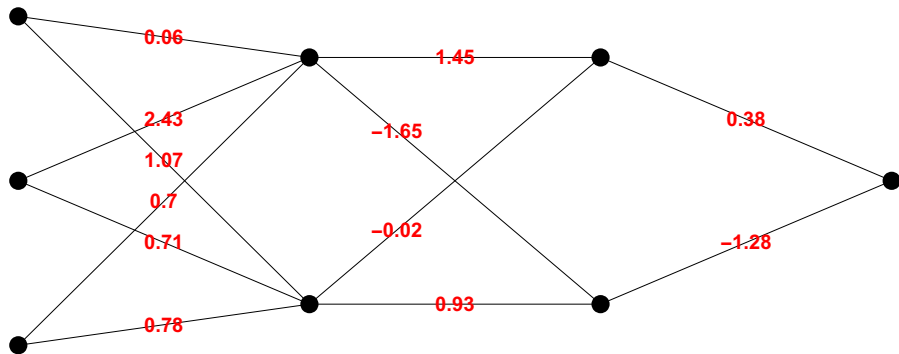


# Recall the MCP neuron

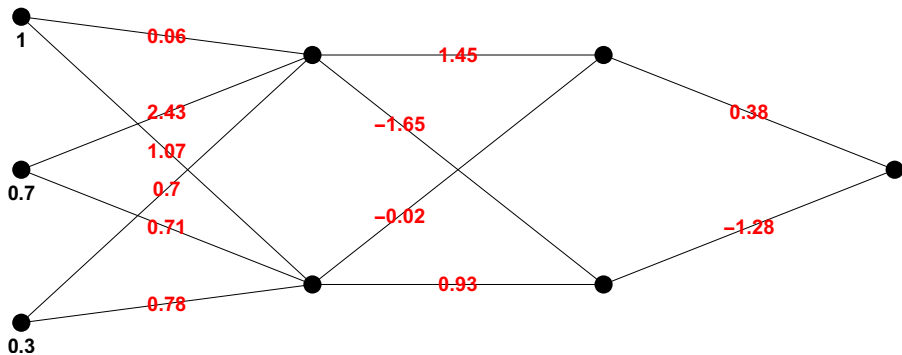


The **weights** define the functionality of the neuron

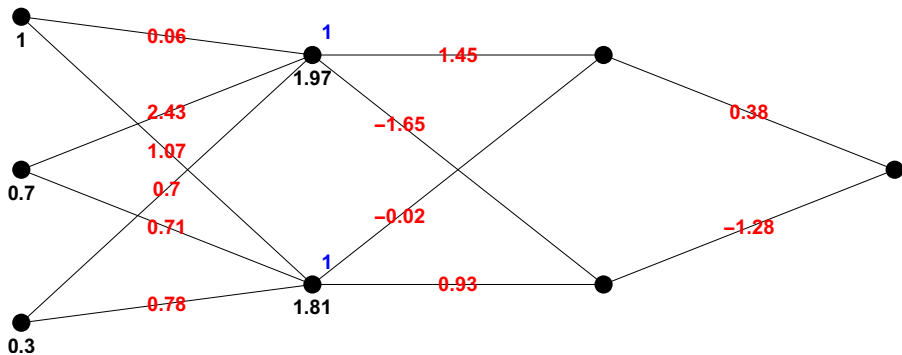
# Action of a sample ANN



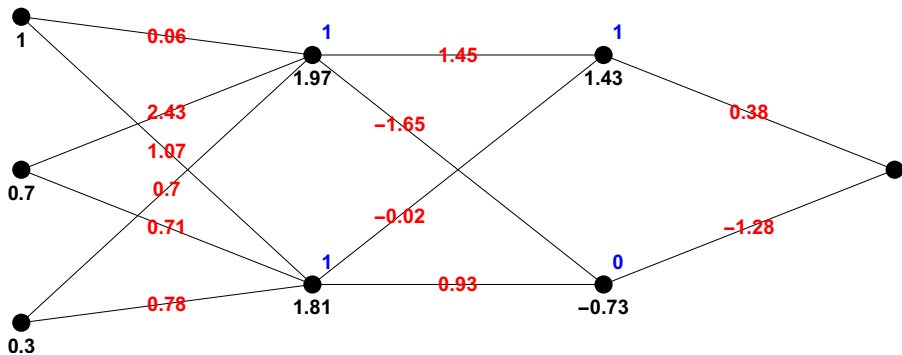
# Action of a sample ANN



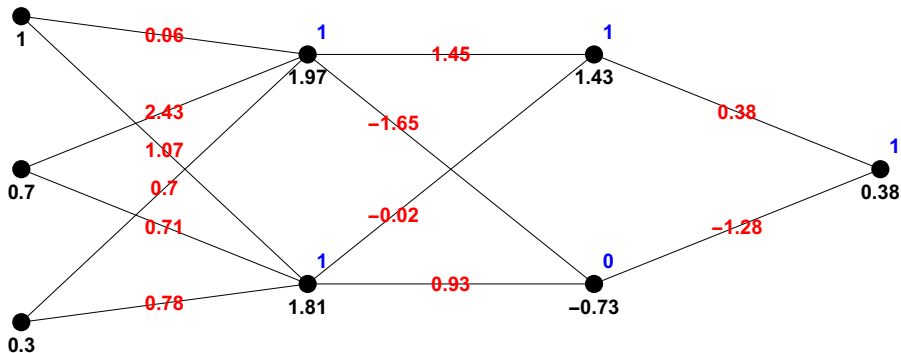
# Action of a sample ANN



# Action of a sample ANN

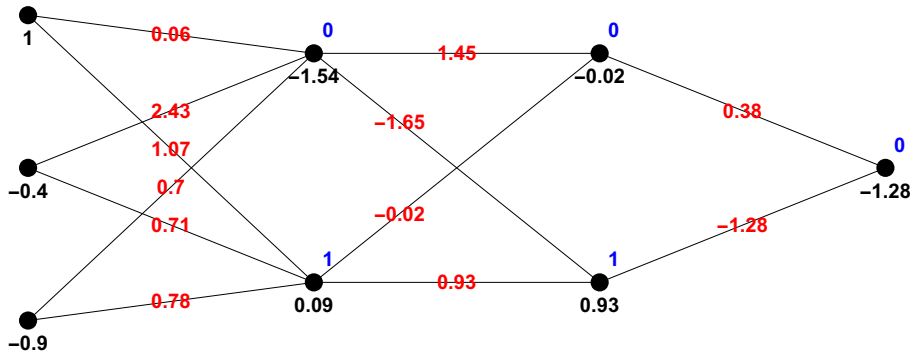


# Action of a sample ANN

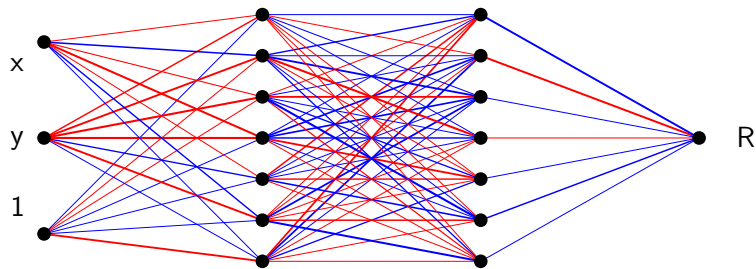




# Action of a sample ANN



# ANN for our “point in the circle” classifier



Result  $R$  depends, for given  $(x, y)$ , on the weights

We want  $R = 1$  for a point inside, and  $R = 0$  outside the circle

**ANN's answer can be correct or not!**

# Training (supervised)

[see my executable book for details:

<https://bronwojtek.github.io/neuralnets-in-raw-python/docs/index.html>]

ANN learns by mistakes!

- 1 We have a training sample of points, of which we know (as the supervisor) if they belong to the circle or not (the points have **labels**)
- 2 We let these points one by one into ANN. If the answer is wrong, the weights are appropriately modified according to a simple procedure - so called **back-propagation** (Bryson and Ho,1969)
- 3 repeat 2) numerous times (**epochs**)
- 4 Run thus trained ANN on a test sample

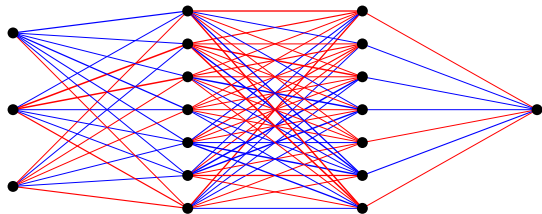
The outcome is a **trained** ANN, which can be used to classify new points.

Training = finding weights by using the data, such that ANN works in a desired way

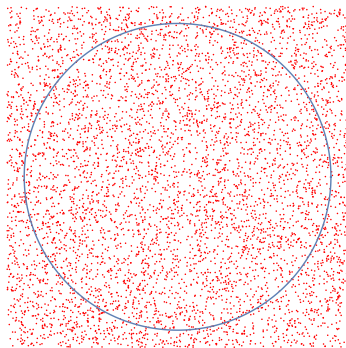
Training is long/costly, but usage very fast

# Training and testing

Initial ANN with random weights

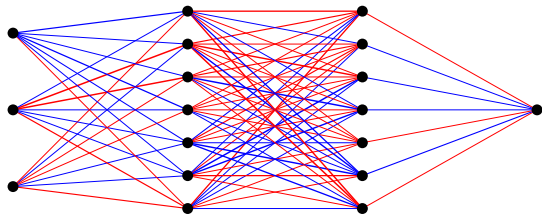


test

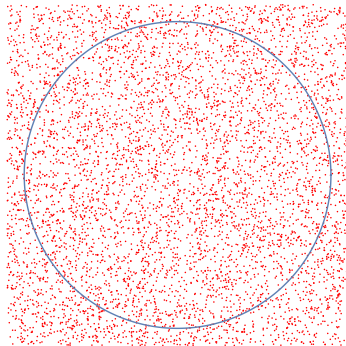


# Training and testing

Initial ANN with random weights



test

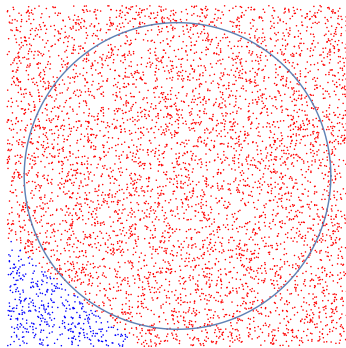
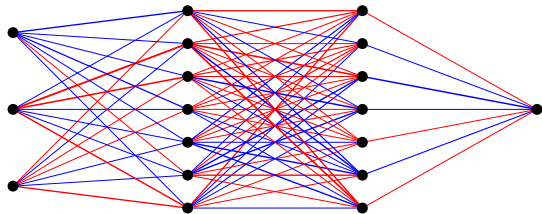


Horrible!

(it says every point is inside the circle)

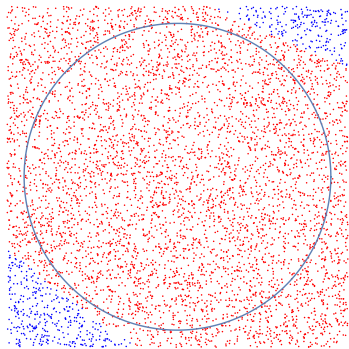
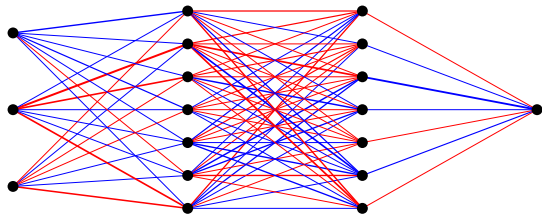
# Training and testing

after 10000 epochs



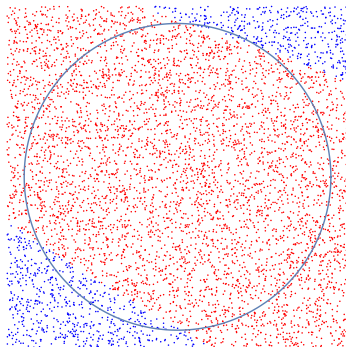
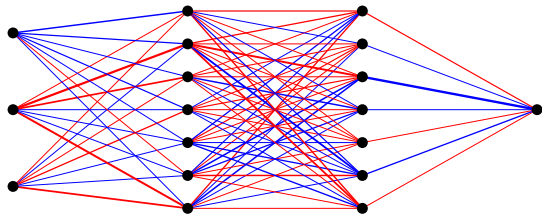
# Training and testing

after 15000 epochs



# Training and testing

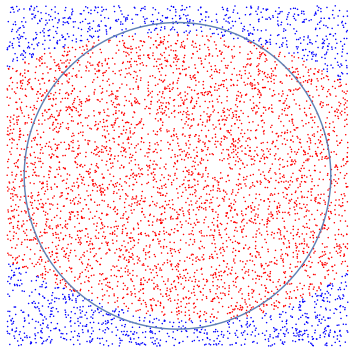
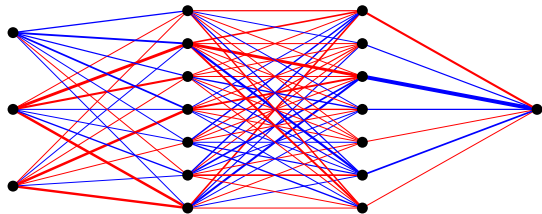
after 20000 epochs





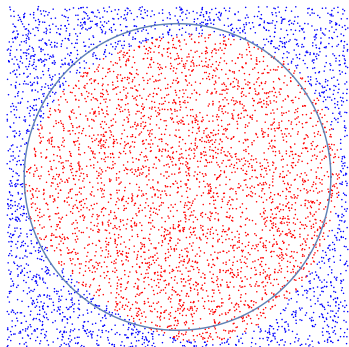
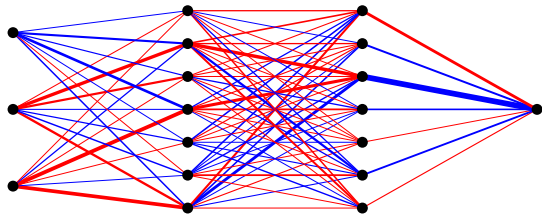
# Training and testing

after 30000 epochs



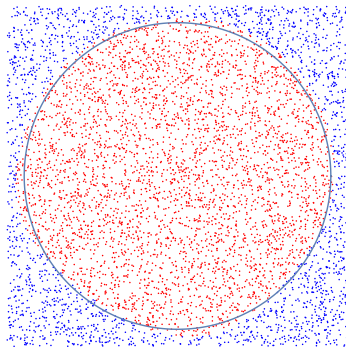
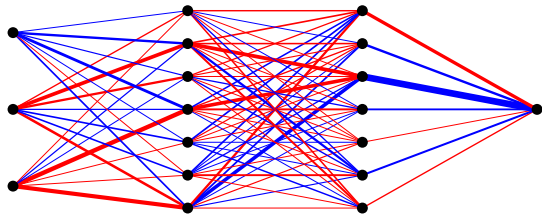
# Training and testing

after 40000 epochs



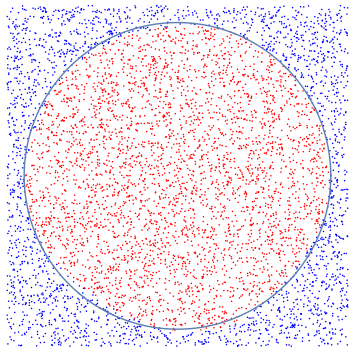
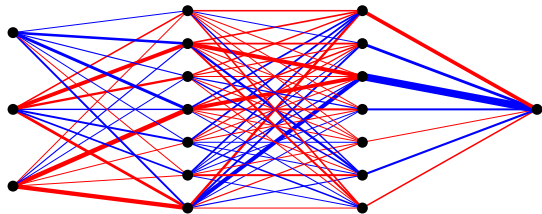
# Training and testing

after 50000 epochs



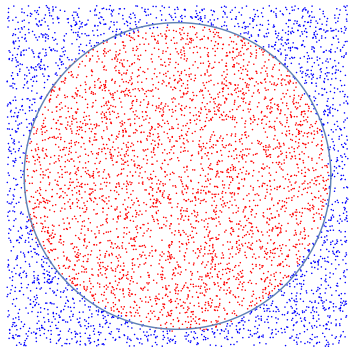
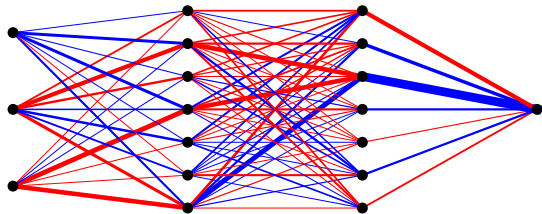
# Training and testing

after 100000 epochs



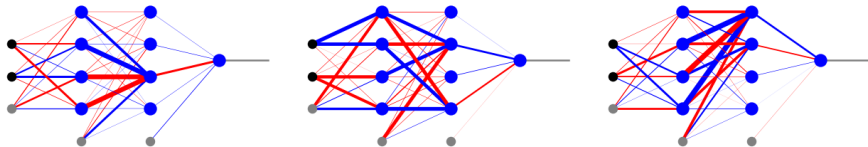
# Training and testing

after 500000 epochs



# Local minima

Each time you train, you get a different but equally good set of weights!



# Comments

- The trained network recognizes if a point is in a circle, without having a clue about geometry! Pure **intuition**, based on prior experience. Looking at the ANN's weights, you can't figure out what the network is doing!
- Learning laborious, but use very fast.
- A single neuron is very simple, but thanks to their large number and numerous connections, we experience **emergence** of a highly complex behavior.
- No coding in a "functional" sense, only adjusting the weights.

---

[our classifier example](#) → advanced ANNs

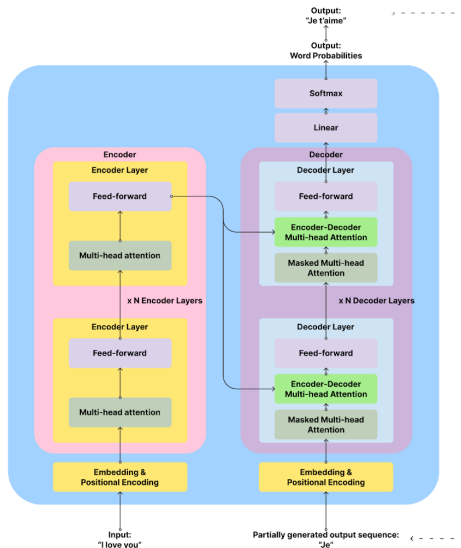
Many neurons and layers, dedicated architecture, training on huge data bases, different training strategies, advanced algorithms

A breakthrough in recent years through the use of ANNs with multiple layers (**deep learning**), trainable with current computing power and new algorithms. The basic idea is, however, similar our example. The architecture of advanced networks is assembled from simpler pieces, just like Lego bricks.

# How does GPT look like?



# Architecture of GPT



(source 2 slides down)

- ver. 3.5: 175 B ( $B = 10^9$ ) parameters (weights), a few hundred layers
- data 500 B words (mostly Common Crawl, book corpora, Wikipedia: 3 B)
- max. query length: about 50 pages of text
- training:  $3 \times 10^{23}$  FLOP operations (GPU farms)
- training cost: \$50 M
- trained model about 10 times too big to run on a single GPU (on a laptop)

With a typical asset of a large lattice QCD collaboration of 15 TeraFLOPs, to compute  $3 \times 10^{23}$  FLOP operations would take over 600 years!

# How GPT works?

<https://medium.com/design-bootcamp/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9>

**Task:** given a string of words (query), generate the most appropriate next word in the answer, then the next, and so on.

- The network learns (sets its weights) "on its own" (in the so-called "unsupervised" way) on a huge corpus of texts/websites.
- Then it is trained in a **supervised** way on a sample of data, where queries have good answers (labels) chosen by people.
- In order to further "tailor" the network for a given functionality, remove inappropriate content, etc., another human-based retraining takes place: assigning a reward/punishment (known as **reinforcement learning**).
- The **transformer** has **attention**, i.e., knows which words are more important and how they correlate with each other. This is technically very important/groundbreaking.
- The trained network predicts the next word based on the previous words of the conversation and its parameters (weights).
- Randomness of choice - the answer is different each time you run the same query.
- **The GPT is not connected to the Internet.** The knowledge gained through learning (up to the cut-off date) is encoded in the "rigid" parameters (weights) of the network. Updating the data requires retraining everything from scratch!

blond

blond girl

# Correlations

blond girl /woman

# Correlations

blond girl /woman /man

# Correlations

blond girl /woman /man entered /smiled / ...

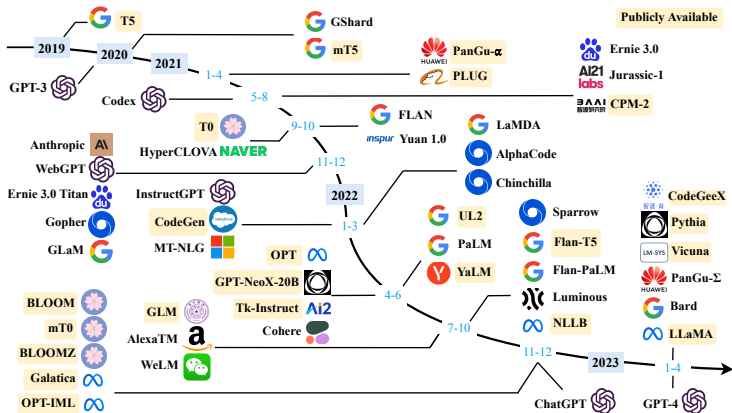
Only **very limited** possibilities, with probabilities based on frequencies in the used data corpus.

Imagine all words (be them 500 000) correlate to all the other words. That would give  $500000^2 = 250 \text{ B}$  numbers - more than in GPT-3.5! The point is that the correlations are very **sparse** (a great majority is zero), also for the long chains of words. The fact that one can encode multi-word correlations with just 170 B parameters in GPT is thrilling.



# Large language models (LLMs)

I didn't happen overnight and it's not the last word!



Review: <https://arxiv.org/pdf/2303.18223.pdf>

Stanford's LLaMA-based Alpaca performs similarly to ChatGPT on limited tasks, open-source model, **training cost \$600!**

# Attention is all you need

## Attention Is All You Need

**Ashish Vaswani\*** Google Brain avaswani@google.com  
**Noam Shazeer\*** Google Brain noam@google.com  
**Niki Parmar\*** Google Research nikip@google.com  
**Jakob Uszkoreit\*** Google Research usz@google.com

**Llion Jones\*** Google Research llion@google.com  
**Aidan N. Gomez\* †** University of Toronto aidan@cs.toronto.edu  
**Lukasz Kaiser\*** Google Brain lukasz.kaiser@google.com

**Illia Polosukhin\* ‡**  
illia.polosukhin@gmail.com

### Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to

1706.03762v5 [cs.CL] 6 Dec 2017

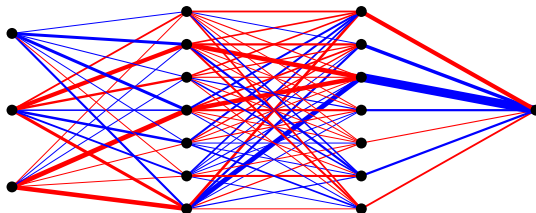
<https://arxiv.org/pdf/1706.03762.pdf>

No need for recurrence → massive parallelization possible

# Is GPT conscious?

# Recall our example

Remember ANN recognizing points in a circle?



Trained GPT is conceptually similar, only orders of magnitude larger

Does such a model “understand” what it does?



(midjourney: Chinese room with 2 windows and books)

# Does Chat GPT “understand” what it talks about?

**The Chinese Room Paradox** (1950) by John Searle (1932–, American philosopher)

You are locked in a room with two openings. With one you get questions on cards written in Chinese. You don't speak Chinese yourself, but you have detailed instructions/algorithms in the room on how to deal with the strings of pictograms and what string of them to generate in response, which you pass outside through the other hole.

So you converse in Chinese without having a slightest clue what you are talking about!

→ **Computer processing is not true reason, endowed with consciousness**

**Materialistic viewpoint:** Maybe it's a matter of size and extending the program to include other domains (senses, emotions ...) and associations between them, and having a “control system” (morality)?

**Questions about AI are, in a way, questions about ourselves, because we too run a neural network!** Learning = consolidation of synaptic connections between brain neurons (Hebb's rule)

# Hallucinations, confabulations, waffling in LLMs

No fact checking, no mathematical accuracy

## Who is Wojciech Broniowski?

GPT-3.5: Wojciech Broniowski is a theoretical physicist who currently works as a professor at the Institute of Physics of the Jan Kochanowski University in Kielce, Poland. He has also held positions at various other academic institutions, including the **University of Tennessee, Knoxville, and the University of Rzeszów** in Poland.

Broniowski [...] has published numerous scientific articles in peer-reviewed journals and has given talks at international conferences on nuclear physics and related fields. He has received several awards for his contributions to theoretical physics, including the **Award of the Polish Physical Society in 2012**.

# Hallucinations, confabulations, waffling in LLMs

No fact checking, no mathematical accuracy

## Who is Wojciech Broniowski?

GPT-3.5: Wojciech Broniowski is a theoretical physicist who currently works as a professor at the Institute of Physics of the Jan Kochanowski University in Kielce, Poland. He has also held positions at various other academic institutions, including the **University of Tennessee, Knoxville, and the University of Rzeszów** in Poland.

Broniowski [...] has published numerous scientific articles in peer-reviewed journals and has given talks at international conferences on nuclear physics and related fields. He has received several awards for his contributions to theoretical physics, including the **Award of the Polish Physical Society in 2012**.

GPT-4: As of my knowledge cutoff date in September 2021, Wojciech Broniowski is not a well-known public figure. It's possible that he might be a private individual or a less prominent figure in a specific field. If you could provide more context or details about who you are looking for, I might be able to help you better.

## Better in English than in Polish



# Quo vadis, AI?

The tools go further. GPT-4 presentation:

<https://www.youtube.com/watch?v=outcGtbnMuQ>

- Multimodality (text, graphics ...), e.g., image of interior of the fridge → recipe for a possible dish.
- Write a skeleton of an html page based on handwritten notes.
- Filling out a tax return form based on colloquial info.

Competition, no details of GPT-4 revealed!

# Superhumanist?

(this is better in Polish)

**Write a one-sentence summary of our (not shown here) composition.** Bohaterowie Byrona i Mickiewicza dzielą romantyczne cechy, takie jak dążenie do wolności i bunt, ale różnią się motywacjami - bohaterowie byronowscy skupiają się na osobistych pragnieniach, podczas gdy bohaterowie mickiewiczowscy dążą do celów narodowych i zbiorowych.

**Do the same, but make a majority of words begin with P.** Postacie poetyckie Byrona i Mickiewicza prezentują podobieństwa - pasjonujący pragnienia oraz przekory, jednakże potężne przesłanki pozycjonują się przeciwnie: personalistyczne dla Byrona, patriotyczne dla Mickiewicza.

**Now with K.** Kreacje konstruowane przez Byrona i Mickiewicza komponują konwergencje - kwitnące kwestie i kontestacje, kontrastujące konteksty kluczowe - koncentracja na korzyściach konkretnych kontrastuje z kolektywnymi koncepcjami.

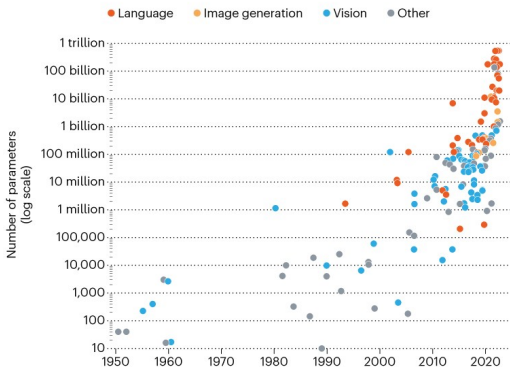
**How about J?** Jednostki jawiące się jako junacy Byrona i Mickiewicza jawią jakąś jedność - jędrne jąkanie się jak jaskrawy jawor, jednakże jasne, jakkolwiek jednocześnie jaskrawe, języ różnią: jednostkowe jedynie dla jednych, jakoby jednocześnie jednoczące dla innych.

Who can do like this? – intellect surpassing human?

(Lem, "Cyberiada", 1965)

## THE DRIVE TO BIGGER AI MODELS

The scale of artificial-intelligence neural networks is growing exponentially, as measured by the models' parameters (roughly, the number of connections between their neurons)\*.



\*Sparse models, which have more than one trillion parameters but use only a fraction of them in each computation, are not shown.

©nature

<https://www.nature.com/articles/d41586-023-00777-9>

Vernor Vinge, Nick Bostrom, ... → technological singularity, uncontrollable growth, robots taking over, doomsday!

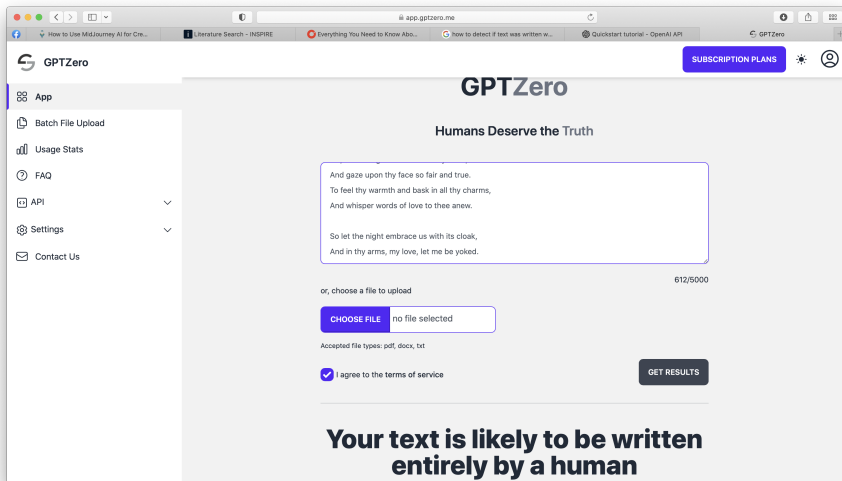
Creation of consciousness, superhumanity, **soul?**

# Advantages, disadvantages, risks

- Homework writing, commemorative speeches, product descriptions, advertising spots, translations, text summaries, quizzes, exams, writing simple programs, websites, poetry, art, ... ?
- **The end of school as we know it** (the class stitches up subscriptions for \$25/month). **There is no point** and it is impossible to detect, ban, or prosecute the use of GPT etc.  
Future homework: "Who will write a better prompt to Chat GPT-9?"
- **Revolution on the job market:** basic IT specialists, graphic designers, advertising agencies, people of "average" mind (IQ<110) will become unnecessary. **New tools** → **new specialists**, "**prompting engineer**".
- Rising price and **hiding of data** needed for machine learning.
- Hallucinations, **waffling**, weak in math (as LLM), **no use in science** (as yet).
- Very poor or no sense of humor (tells really unfunny jokes).
- Fabricating, indistinguishable from the truth, profiled fake news on an unimaginable scale.
- Crime instruction, hacking, impersonation, spamming, phishing.

# Can you detect the use of an LLM?

Recall the T. S. Eliot style poem?



The screenshot shows the GPTZero website interface. The browser address bar displays 'app.gptzero.me'. The page features a navigation sidebar on the left with options: App, Batch File Upload, Usage Stats, FAQ, API, Settings, and Contact Us. The main content area is titled 'GPTZero' and 'Humans Deserve the Truth'. A text input field contains a poem:

And gaze upon thy face so fair and true,  
To feel thy warmth and bask in all thy charms,  
And whisper words of love to thee anew.

So let the night embrace us with its cloak,  
And in thy arms, my love, let me be yoked.

Below the input field, there is a file upload section with a 'CHOOSE FILE' button and the text 'no file selected'. Accepted file types are listed as pdf, docx, txt. A checkbox is checked with the text 'I agree to the terms of service'. A 'GET RESULTS' button is located at the bottom right of the input area. The page footer displays 'Your text is likely to be written entirely by a human'.

# Thanking for your attention ...

**Brilliant tools?** Yes, gradual strenuous development can lead to a "phase transition", incredible acceleration. Widespread use will happen whether one likes it or not! **Everyone can become an artist, programmer, thief ...**

**Confabulations?** - at this stage yes: the network does not know when to say that it does not know, how to reasonably ask for details, take the initiative ...



(**midjourney** chatting with a bot AI manga style)

**GPT-3.5:** 170 B parameters  $\sim 5 \times 10^{12}$  bits, **brain:**  $\sim 10^{14}$  synaptic connections  
(**only 20 times more!**)