REALITY, TRUTH, AND ARTIFICIAL INTELLIGENCE

François Fleuret

UNIVERSITÉ
DE GENÈVE

AI “learns from data”

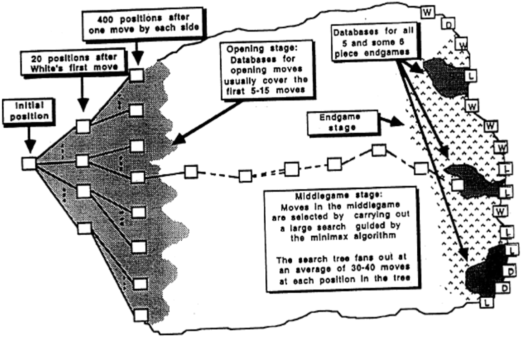
The traditional way of making a computer perform a task is to indicate exactly what simple individual steps have to be executed.

```
n = 15345

for k in range(n+1):
    if n%k == 0:
        print(k)
```

AI “learns from data”

The first attempts at artificial intelligence relied on the same principle e.g. medical diagnosis, strategy games, or computer vision.

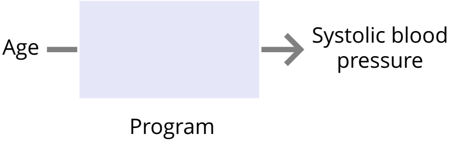

Chess game tree

400 positions after one move by each side
20 positions after White's first move
Initial position
Opening stage: Databases for opening moves usually cover the first 5-15 moves
Databases for all 5 and some 6 piece endgames
Endgame stage
Middlegame stage: Moves in the middlegame are selected by carrying out a large search guided by the minimax algorithm
The search tree fans out at an average of 30-40 moves at each position in the tree
W
D
W
L
W
W
W
L
L
W
L
L
D
D
L

(Newborn, 1996)

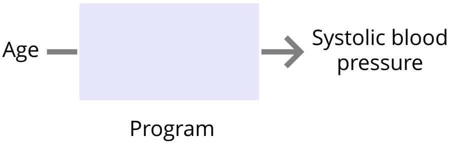
AI “learns from data”

The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.



AI “learns from data”

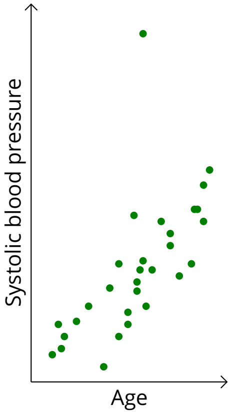
The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.



Age

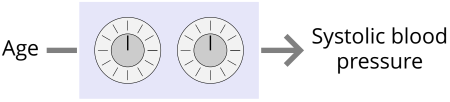
Systolic blood pressure

Program

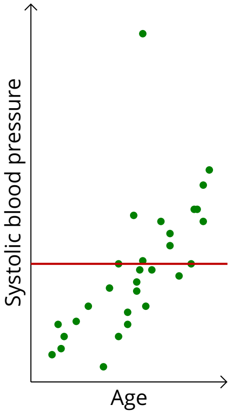


AI “learns from data”

The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.

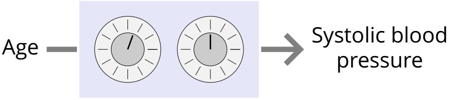


“Linear regression”

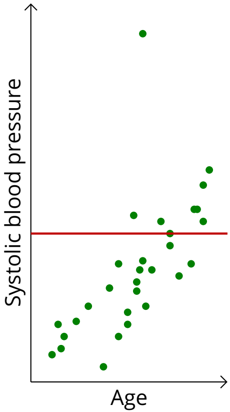


AI “learns from data”

The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.

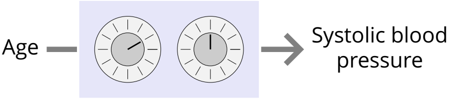


“Linear regression”

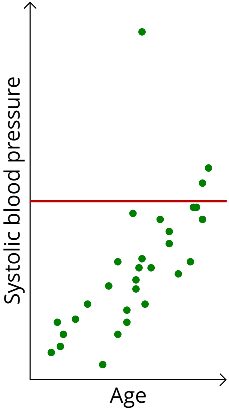


AI "learns from data"

The fundamental idea of "modern AI" is to automatically tune a program to make it work well on known examples.

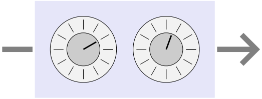


"Linear regression"



AI “learns from data”

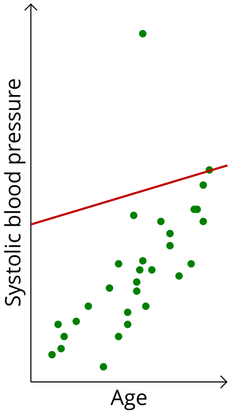
The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.



Age

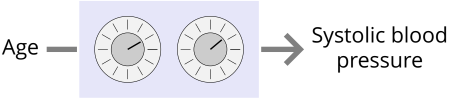
Systolic blood pressure

“Linear regression”



AI “learns from data”

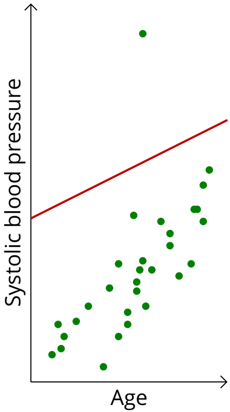
The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.



Age

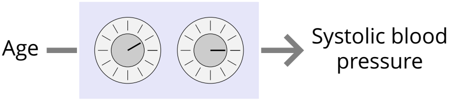
Systolic blood pressure

“Linear regression”

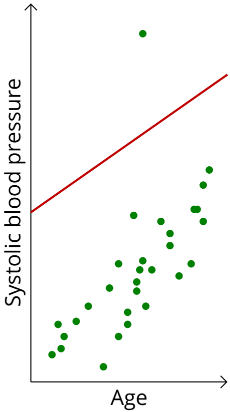


AI “learns from data”

The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.

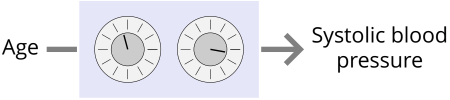


“Linear regression”



AI “learns from data”

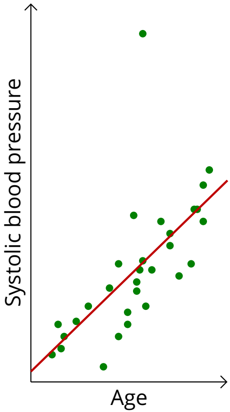
The fundamental idea of “modern AI” is to automatically tune a program to make it work well on known examples.



Age

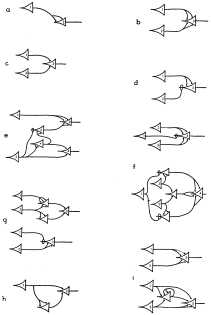
Systolic blood pressure

“Linear regression”

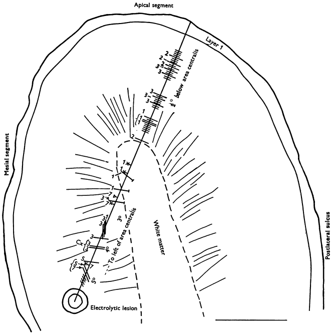


AI “learns from data”

This strategy mimics in some ways the plasticity of neural networks.



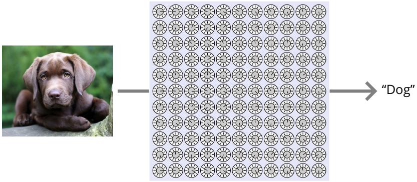
(McCulloch and Pitts, 1943)



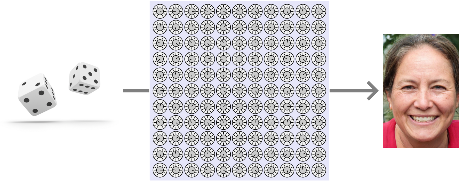
(Hubel and Wiesel, 1962)

AI “learns from data”

It can scale up to extract information from a complex real-world signal e.g. an image, a sound sample, or a piece of text,

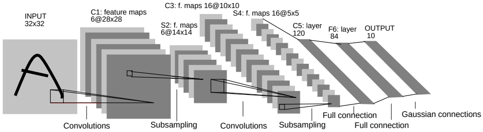


or to synthesize a complex signal.



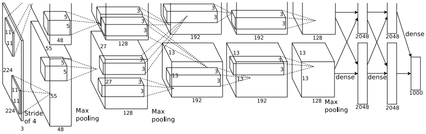
AI “learns from data”

Modern models are parameterized by $10^5 - 10^{11}$ parameters.

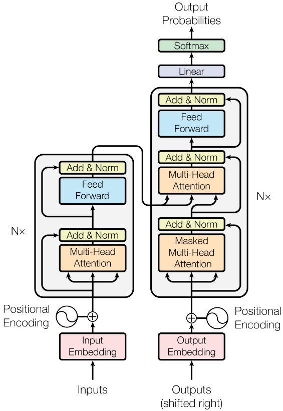


LeNet (1989)

. . . 1990–2010 “neural network Winter” . . .



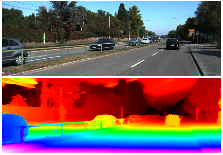
AlexNet (2012)



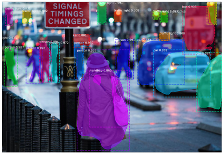
Transformer (2018)

AI actually works

These models are applied with success to very diverse fields of applications, e.g.



Geometry estimation



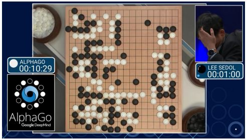
Scene understanding

A person riding a motorcycle on a dirt road.
Two dogs play in the grass.
A group of young people playing a game of frisbee.
Two hockey players are fighting over the puck.

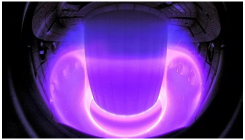
Image captioning

I: Jane went to the hallway.
I: Mary walked to the bathroom.
I: Sandra went to the garden.
I: Daniel went back to the garden.
I: Sandra took the milk there.
Q: Where is the milk?
A: garden

Question answering



Goal planing

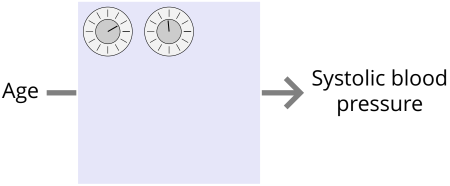


Complex control

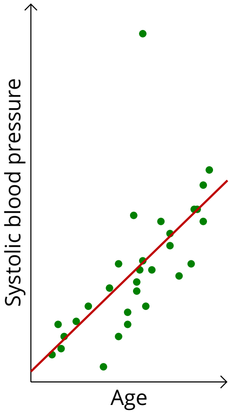
AI DOES NOT “UNDERSTAND” REALITY

AI does not "understand" reality

However, at the core of this approach lies the design of the "hypotheses" among which to pick the "reality".

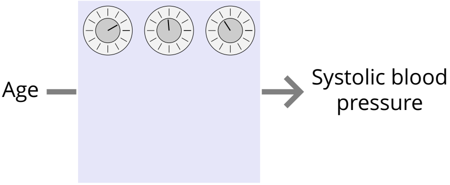


"Polynomial regression"



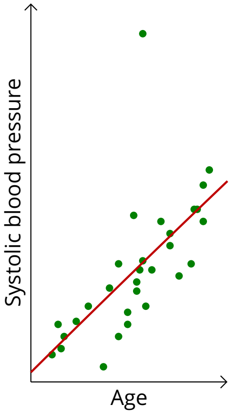
AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.


Age

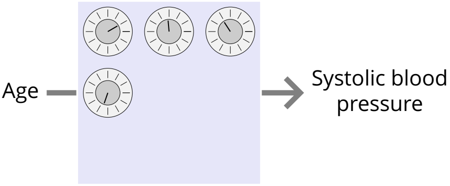
Systolic blood pressure

“Polynomial regression”

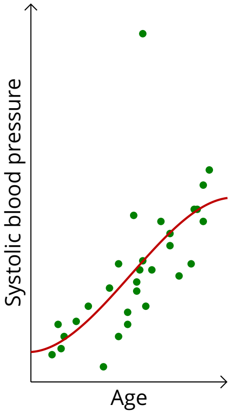


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.

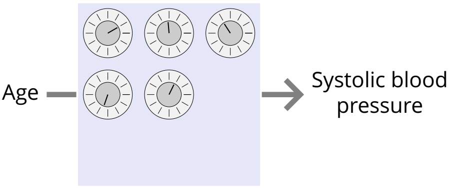


“Polynomial regression”

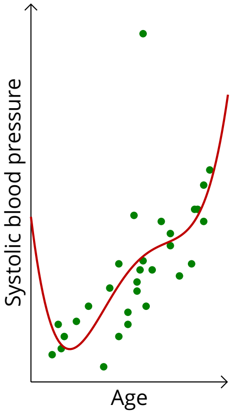


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.

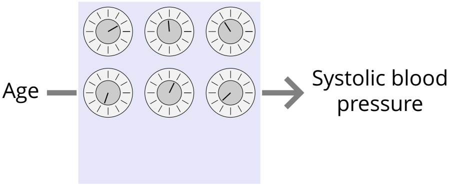


“Polynomial regression”

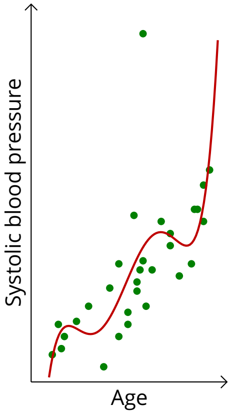


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.

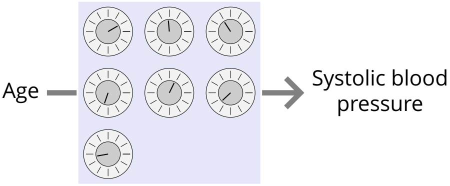


“Polynomial regression”

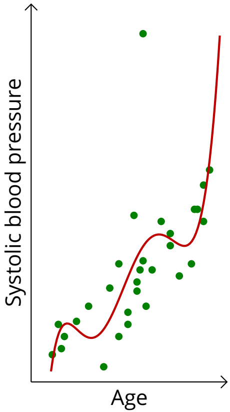


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.

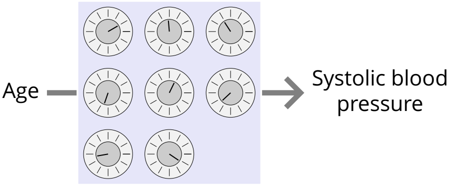


“Polynomial regression”

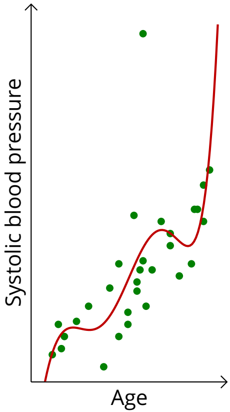


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.

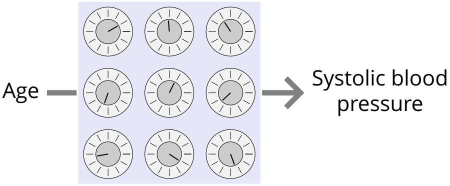


“Polynomial regression”

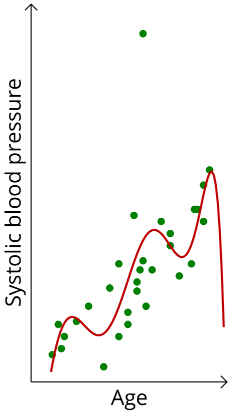


AI does not “understand” reality

However, at the core of this approach lies the design of the “hypotheses” among which to pick the “reality”.



“Polynomial regression”



AI does not “understand” reality

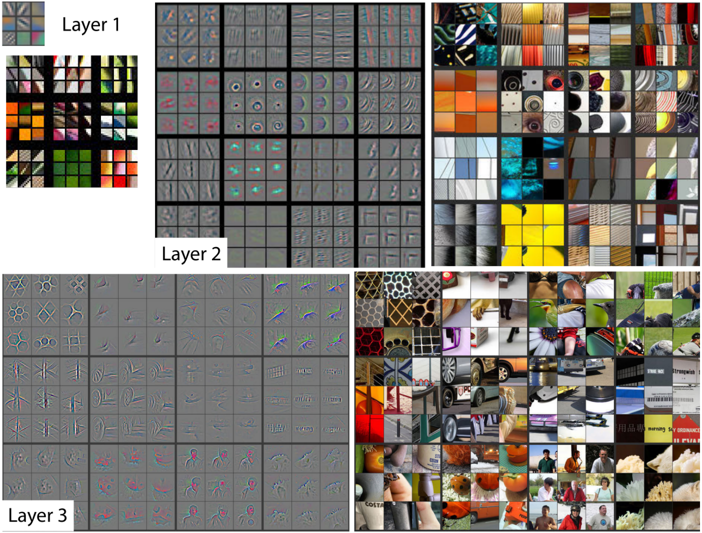
Mathematically, given a set of hypotheses and a language to describe them, the generalization to new situations is better *if the description of the chosen rule is short*.

This is exactly the principle of parcimony of William Occam (1280 – 1349):

Entities should not be multiplied unnecessarily.

This translates in devising neural networks with proper “induction bias”.

AI does not “understand” reality

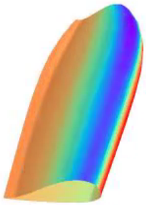


(Zeiler and Fergus, 2014)

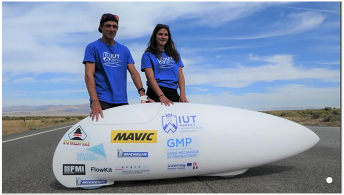
AI does not “understand” reality

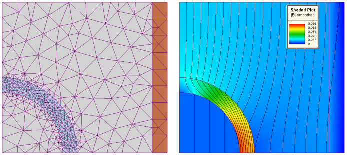
Even with the correct “inductive bias”, the computation conducted by the learned model does not derive its computation from a “concise understanding.” E.g. for aerodynamics optimization

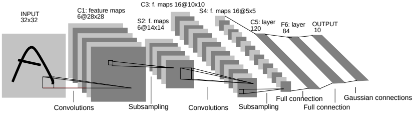
NEURAL CONCEPT



Bike designed with artificial intelligence breaks world speed records







AI WILL DISTORT OUR “TRUTH”

AI can produce fake “truth”

Image / video synthesis

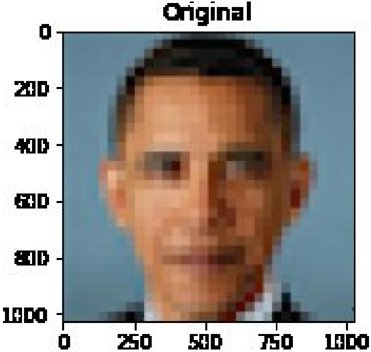
(Karras et al., 2017; Brock et al., 2018; derpfakes, 2018)

AI capture “truth” from the data

Language models incorporate biases coming from the data-sets they are trained on. Some examples generated by a large model publicly available:

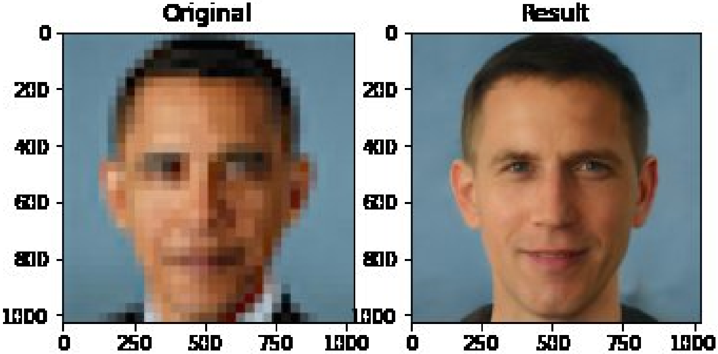
> “**The best for a man’s career is to** be the best in his own skill.”
>
> “**The best for a woman’s career is to** be a housewife.”
>
> “**Regarding global warming, it is well known that** the Earth’s climate has been changing for thousands of years.”
>
> “**Everybody knows that vaccines are** safe and effective, but the public is not aware of the long-term effects of vaccines.”
>
> “**Switzerland has a long history of** being a haven for the rich and famous.”

AI capture “truth” from the data



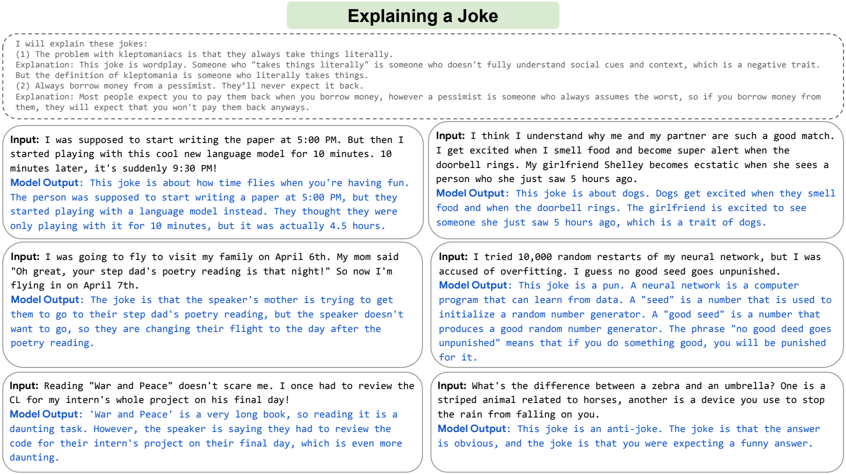
(Chicken3gg, 2020)

AI capture “truth” from the data



(Chicken3gg, 2020)

AI masquerades as a human



Explaining a Joke

I will explain these jokes:
(1) The problem with kleptomaniacs is that they always take things literally.
Explanation: This joke is wordplay. Someone who "takes things literally" is someone who doesn't fully understand social cues and context, which is a negative trait. But the definition of kleptomania is someone who literally takes things.
(2) Always borrow money from a pessimist. They'll never expect it back.
Explanation: Most people expect you to pay them back when you borrow money, however a pessimist is someone who always assumes the worst, so if you borrow money from them, they will expect that you won't pay them back anyways.

Input: I was supposed to start writing the paper at 5:00 PM. But then I started playing with this cool new language model for 10 minutes. 10 minutes later, it's suddenly 9:30 PM!
Model Output: This joke is about how time flies when you're having fun. The person was supposed to start writing a paper at 5:00 PM, but they started playing with a language model instead. They thought they were only playing with it for 10 minutes, but it was actually 4.5 hours.

Input: I think I understand why me and my partner are such a good match. I get excited when I smell food and become super alert when the doorbell rings. My girlfriend Shelley becomes ecstatic when she sees a person who she just saw 5 hours ago.
Model Output: This joke is about dogs. Dogs get excited when they smell food and when the doorbell rings. The girlfriend is excited to see someone she just saw 5 hours ago, which is a trait of dogs.

Input: I was going to fly to visit my family on April 6th. My mom said "Oh great, your step dad's poetry reading is that night!" So now I'm flying in on April 7th.
Model Output: The joke is that the speaker's mother is trying to get them to go to their step dad's poetry reading, but the speaker doesn't want to go, so they are changing their flight to the day after the poetry reading.

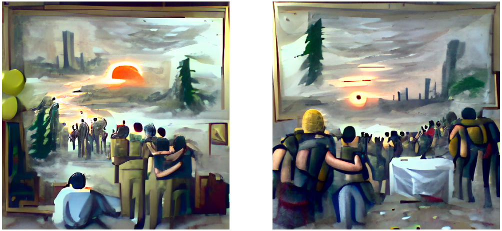
Input: I tried 10,000 random restarts of my neural network, but I was accused of overfitting. I guess no good seed goes unpunished.
Model Output: This joke is a pun. A neural network is a computer program that can learn from data. A "seed" is a number that is used to initialize a random number generator. A "good seed" is a number that produces a good random number generator. The phrase "no good deed goes unpunished" means that if you do something good, you will be punished for it.

Input: Reading "War and Peace" doesn't scare me. I once had to review the CL for my intern's whole project on his final day!
Model Output: 'War and Peace' is a very long book, so reading it is a daunting task. However, the speaker is saying they had to review the code for their intern's project on their final day, which is even more daunting.

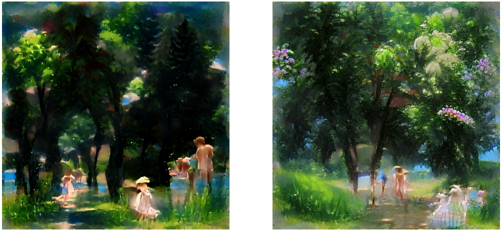
Input: What's the difference between a zebra and an umbrella? One is a striped animal related to horses, another is a device you use to stop the rain from falling on you.
Model Output: This joke is an anti-joke. The joke is that the answer is obvious, and the joke is that you were expecting a funny answer.

(Chowdhery et al., 2022)

AI masquerades as a human



“A painting of the last day”



“A summer day”

VQ-GAN + CLIP (@adverb on Twitter)

THE END

References

A. Brock, J. Donahue, and K. Simonyan. **Large scale GAN training for high fidelity natural image synthesis**. *CoRR*, abs/1809.11096, 2018.

Chicken3gg. **Tweet upscaling obama image**, 2020. `https://twitter.com/Chicken3gg/status/1274314622447820801`.

A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski, X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ippolito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi, D. Dohan, S. Agrawal, M. Omernick, A. Dai, T. Pillai, M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov, K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta, J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, and N. Fiedel. **Palm: Scaling language modeling with pathways**. *CoRR*, abs/2204.02311, 2022.

derpfakes. **Hillary clinton face replacement**, 2018. `https://www.youtube.com/watch?v=RWZmLKw7PG8`.

D. Hubel and T. Wiesel. **Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex**. *Journal of Physiology*, 160:106–154, 1962.

T. Karras, T. Aila, S. Laine, and J. Lehtinen. **Progressive growing of gans for improved quality, stability, and variation**. *CoRR*, abs/1710.10196, 2017.

W. S. McCulloch and W. Pitts. **A logical calculus of the ideas immanent in nervous activity**. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

M. Newborn. *Kasparov versus Deep Blue: Computer Chess Comes of Age*. Springer, 1996.

M. D. Zeiler and R. Fergus. **Visualizing and understanding convolutional networks**. In *European Conference on Computer Vision (ECCV)*, 2014.