

Deep learning for biologists

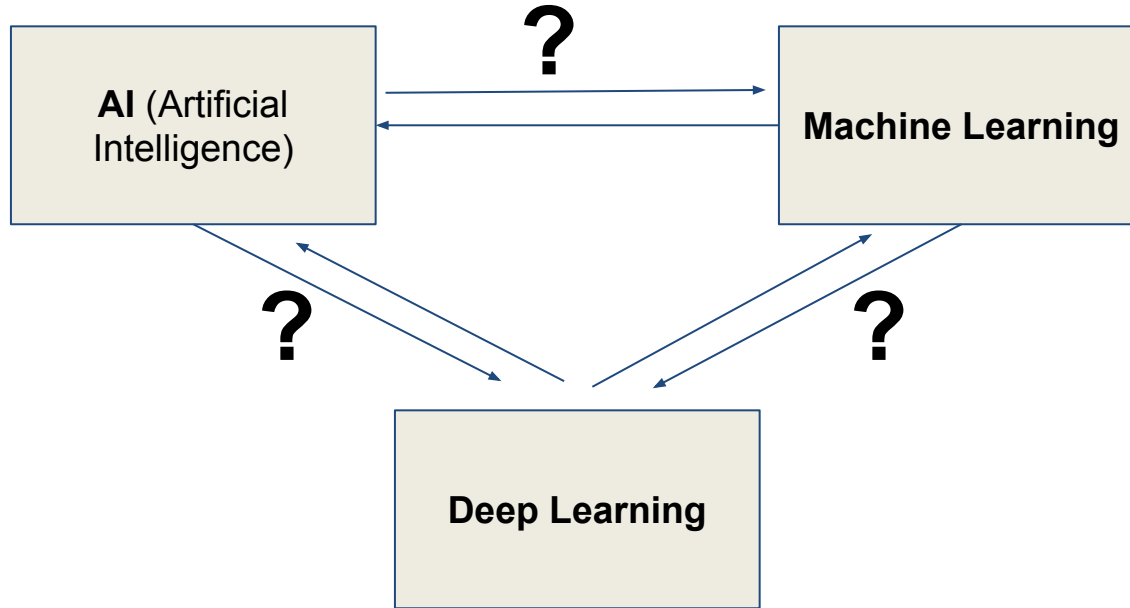
A practical and theoretical introduction

Filippo Biscarini
Senior Scientist
CNR, Milan (Italy)

Nelson Nazzicari
Senior Scientist
CREA, Lodi (Italy)



What is deep learning?

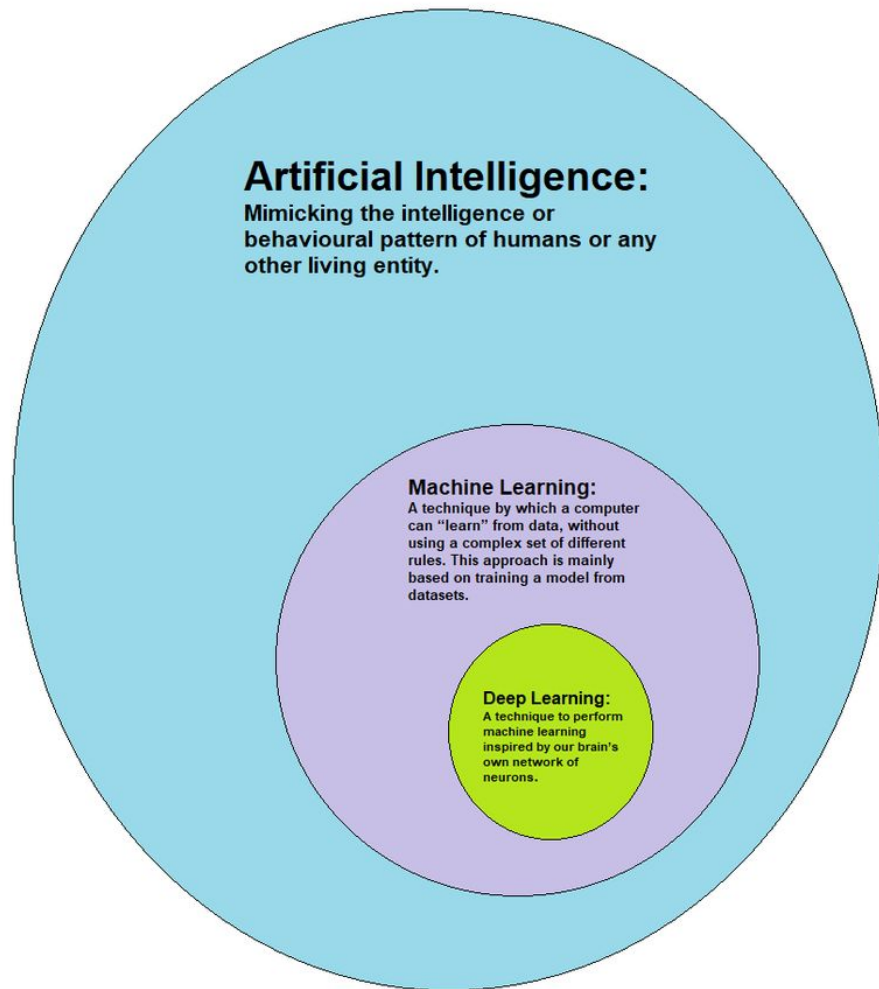


What is deep learning?

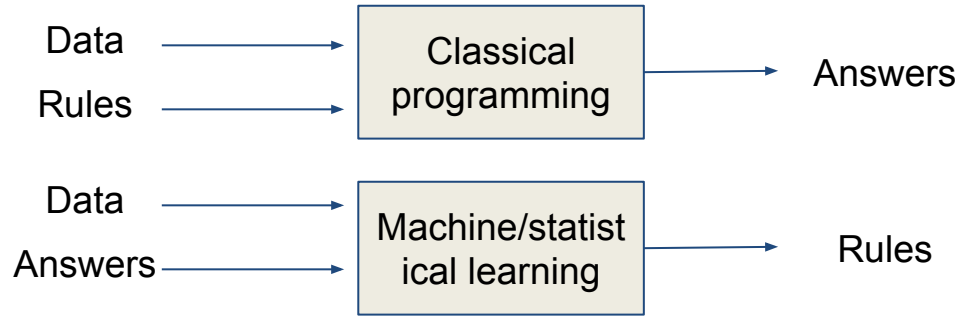
AI >> ML >> DL

Source:

https://en.wikipedia.org/wiki/Deep_learning#/media/File:AI-ML-DL.png



What is (deep) learning?



- (machine; statistical) learning
 - tune a **mathematical model** using some **training data** to **make predictions** on unknown, new data
 - a machine/statistical learning model is **trained** rather than explicitly programmed



What is (deep) learning?

(machine; statistical) learning

1. Input data (e.g. sound recordings, images)
2. Output examples (e.g. sound transcripts, image-tags)
3. Performance measure: how well is the algorithm working →
adjustment steps → **learning**



You can do (statistical) learning in your head!

- The first edition of this course gets 10 students
- The second edition gets 20 students
- The third edition gets 40 students
- The fourth edition gets 80 students
- How many students in the sixth edition?



You can do (statistical) learning in your head!

TRAINING DATA

- The first edition of this course gets 10 students
- The second edition gets 20 students
- The third edition gets 40 students
- The fourth edition gets 80 students
- How many students in the sixth edition?

NEW, UNKNOWN DATA

STUDENTS IN SIXTH EDITION = 320

$STUD = 10 \times 2^{\text{YEAR} - 1}$

PREDICTION

MATHEMATICAL
MODEL



What is deep learning?

- (machine; statistical) learning
 - tune a **mathematical model** using some **training data** to make **predictions** on unknown, new data
 - «If you make a bunch of random changes to your program until it sort-of works, that's "hacky" and "bad coding practice". But if you do it really fast, it's "machine learning"» [Steve Maine, Software Engineer]
- Deep learning
 - **Neural networks**, a mathematical model “inspired” by biology
 - Artificial Neural Networks have been around for ~80 years ...
 - ... but became “deep” in the last ten years

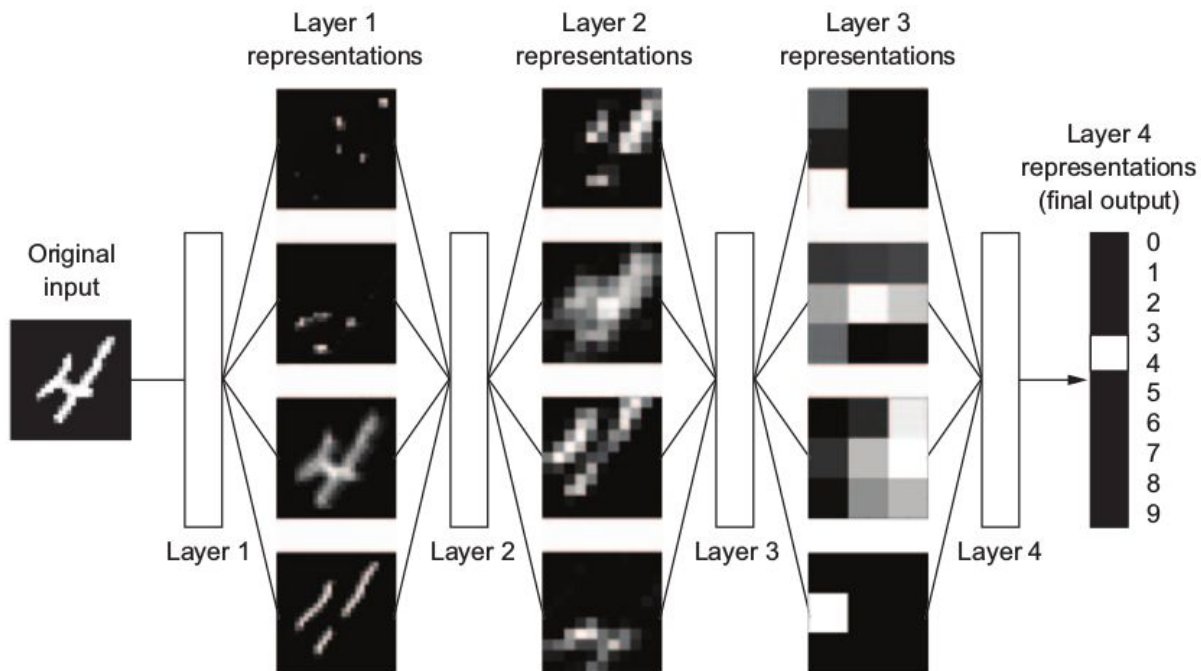


Why “deep”?

What do you think?



Why “deep”?



deep neural networks
(i.e. with many layers)

Figure 1.6 Deep representations learned by a digit-classification model

Source: François Chollet and J.J. Allaire “Deep learning with R” (2018)

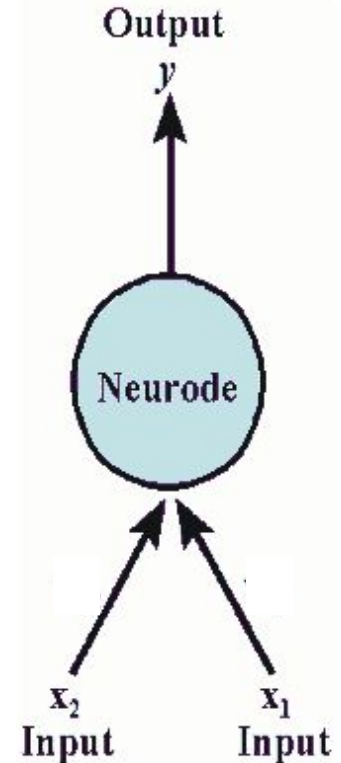


A brief history of DL



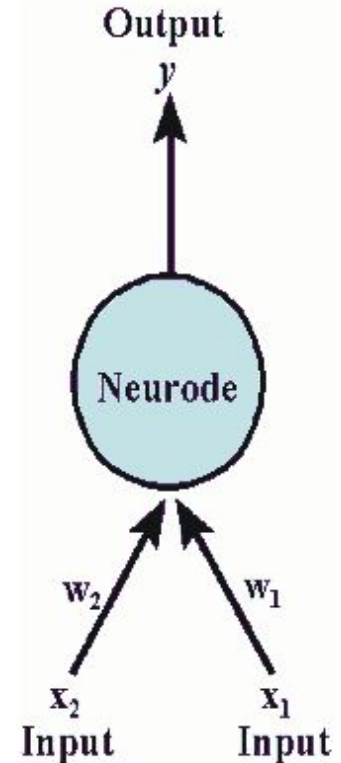
How it all began

- **1943, McCulloch & Pitts:**
 - first mathematical model of a neural network
 - only binary input and output (0/1),
 - only used the threshold step activation function
 - did not incorporate weighting the different inputs
- **1950s:** pioneers started asking whether computers could be made to “think” → (A.I.)



How it all began

- **1943, McCulloch & Pitts:**
 - first mathematical model of a neural network
 - only binary input and output (0/1),
 - only used the threshold step activation function
 - did not incorporate weighting the different inputs
- **1950s:** pioneers started asking whether computers could be made to “think” → (A.I.)
- **1957, Rosenblatt:** introducing the **perceptron**:
 - weighting inputs
 - additional activation functions (e.g. sigmoid)



From symbolic AI to ML

- this was all **no learning AI**
 - e.g. early chess programs based on hardcoded rules crafted by programmers
 - experts believed that human-level A.I. could be achieved by handcrafting a sufficiently large set of explicit rules → **symbolic AI**
- dominant paradigm from the 1950s to the late 1980s
- **expert systems boom** of the 1980s
- symbolic AI suitable to solve well-defined, logical problems (like chess); impossible to figure out explicit rules for solving more complex, fuzzy problems like image classification, speech recognition etc.
- **enter learning**: ML emerged to take the place of symbolic AI
 - the missing piece was an efficient way to train large neural networks
 - → enter the **backpropagation algorithm!**
 - **1989, Yann LeCun**: first application of NN to the recognition of handwritten digits

4



Beware of the hype!



This happened twice in the past:

1. **symbolic AI in the 1960s:**

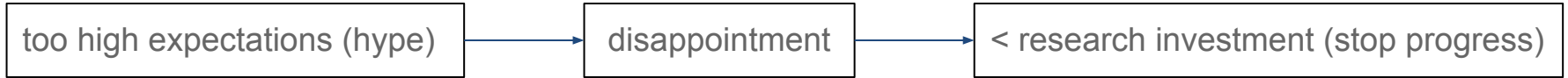
- a. expectations: 1967, Marvin **Minsky**: “the problem of creating ‘artificial intelligence’ will be solved soon”; Minsky, 1970: “in three to eight years we will have a machine with the general intelligence of an average human being”;
- b. failure: researchers and funds turned away from the field → **first AI winter**

2. **1980s, expert systems:**

- a. initial success stories attracted investment, companies developed expert systems
- b. expert systems were expensive to maintain, difficult to scale, limited in scope: interest declined → **second AI winter**



Beware of the hype!



This happened twice in the past:

1. **symbolic AI in the 1960s:**

- a. expectations: 1967, Marvin **Minsky**: “the problem of creating ‘artificial intelligence’ will be solved soon”; Minsky, 1970: “in three to eight years we will have a machine with the general intelligence of an average human being”;
- b. failure: researchers and funds turned away from the field → **first AI winter**

2. **1980s, expert systems:**

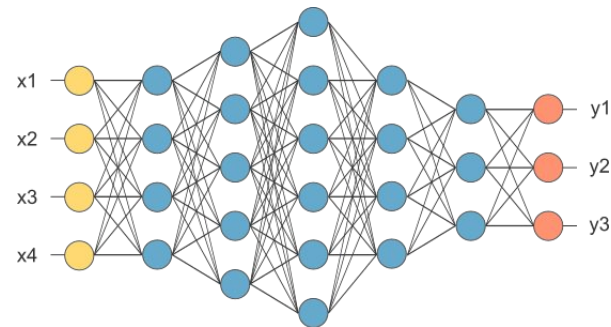
- a. initial success stories attracted investment, companies developed expert systems
- b. expert systems were expensive to maintain, difficult to scale, limited in scope: interest declined → **second AI winter**

What now? Third cycle of AI hype and disappointment? In spite of possibly unrealistic short-term expectations, the long-term picture is bright: **no short-term hype → long-term vision instead!**



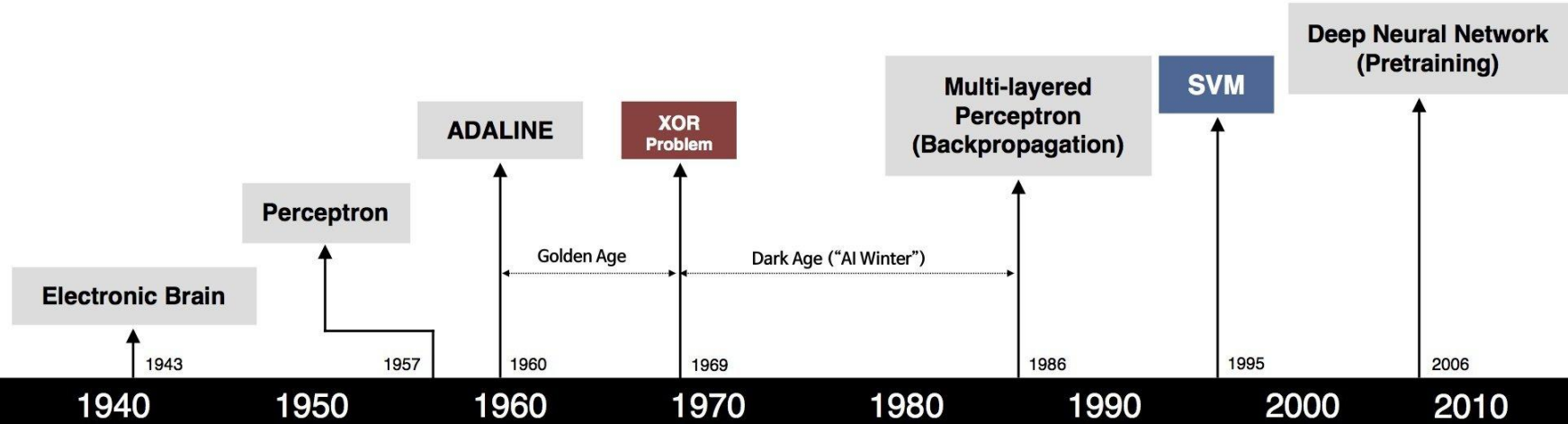
Rationing NN through winter, before summer!

- 1990s: **SVM** (Vladimir Vapnik, 1995)
- 2000s: **Random Forest, Gradient Boosting** (tree-based methods)
- **deep NN** (DL at last!):
 - 2011, Dan Ciresan: **GPU-trained deep neural networks** for image classification
 - 2012, Alex Krizhevsky & Geoffrey Hinton: **deep convolutional neural network** → 83.6% accuracy at image classification
 - DL automates feature extraction and engineering

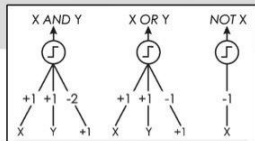


[From:
<https://github.com/ashishpate126/Andrew-NG-Notes/blob/master/andrewng-p-1-neural-network-deep-learning.md>]

A little timeline of Neural Networks



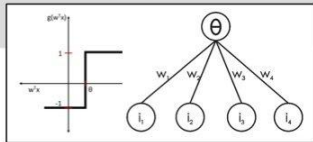
S. McCulloch – W. Pitts



- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



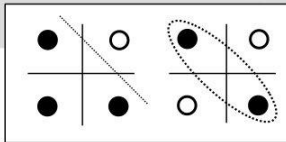
- Learnable Weights and Threshold



B. Widrow – M. Hoff



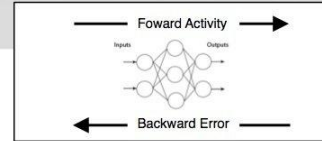
M. Minsky – S. Papert



- XOR Problem



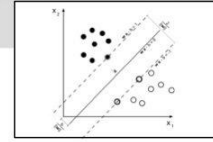
D. Rumelhart – G. Hinton – R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



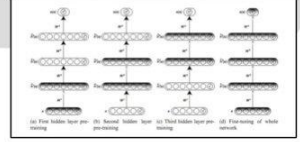
V. Vapnik – C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention



G. Hinton – S. Ruslan

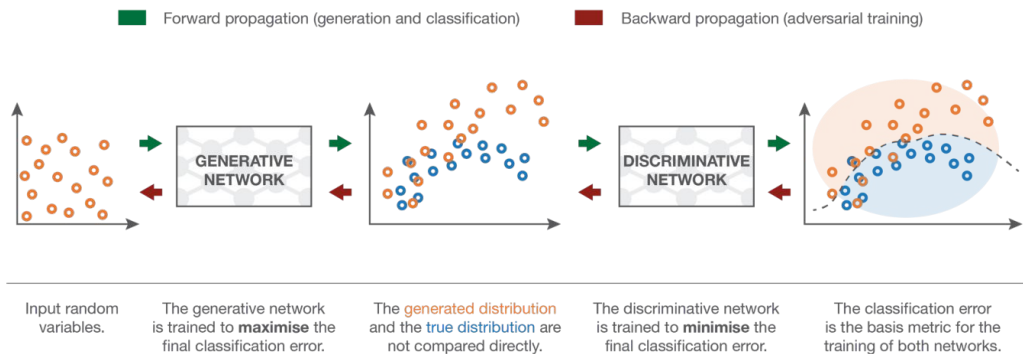


- Hierarchical feature Learning

Latest developments - what in the 2010/20s?



- DL and gradient boosting **dominate Kaggle competitions** since 2012
- **unsupervised / semi-supervised learning** → *Google Brain*, [the cat experiment](#) (2012, [Arxiv](#))
- **Generative Adversarial Networks** (GANs): [Ian Goodfellow et al.](#), 2014: generative models trained “indirectly” by trying to fool another network that is trained at the same time to distinguish “generated” data from “true” data
- **GRU** (gated-recurrent units): [Cho et al.](#), 2014: breakthrough for **RNN**
- **LLMs**: large language models (more on this later)
- Self-attention (“[Attention is all you need](#)”) → transformers
- GNN



[From: [towardsdatascience.com](#)]



[REF] History of deep learning

- McCulloch and Pitts, *A logical calculus of the ideas immanent in nervous activity*, 1943.
<https://link.springer.com/article/10.1007/BF02478259>
- Wang and Raj, *On the Origin of Deep Learning*, 2017 <https://arxiv.org/pdf/1702.07800.pdf>
- Andrew Beam, *Deep Learning 101 - Part 1: History and Background*, 2017
https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html
- Chung et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling → GRU layers <https://arxiv.org/abs/1412.3555>
- Hochreiter & Schmidhuber 1997. Long short-term memory → LSTM layers
<https://ieeexplore.ieee.org/abstract/document/6795963>



Why now?

Innovations in:

- Hardware
- Big Data
- Algorithms
- Infrastructure



Why now?

Innovations in:

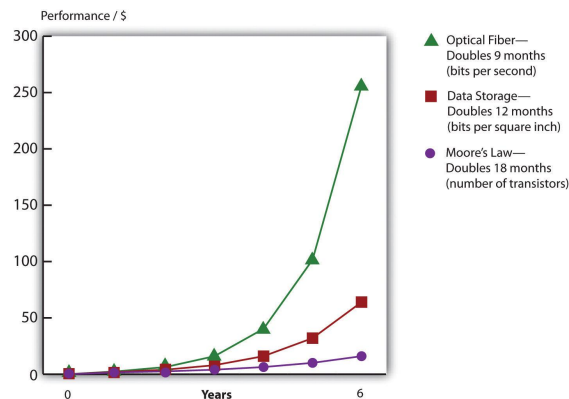
- Hardware:
 - CPUs now 5,000 times faster than 25 years ago
 - GPUs (thanks to videogames!)
 - TPUs (Tensor Processing Units: designed specifically for deep learning)
 - Active R&D on specialised A.I. processors/accelerators: e.g. [Google's TPU](#), [Intel's Loihi](#), [IBM's TrueNorth](#), [NVIDIA T4 GPU](#) ...



Why now?

Innovations in:

- Big Data:
 - Internet: search engines, social media, image datasets etc.
 - IoT: sensors, interconnected devices
 - Data storage (Moore's law)
 - GPT-3 was trained on 570 GB of text data (september 2021)
 - [Google Search index](#): 1e+8 GB



<https://2012books.lardbucket.org/books/getting-the-most-out-of-information-systems-v1.2/s09-moore-s-law-fast-cheap-computi.html>



Why now?

Innovations in:

- Algorithms:
 - backpropagation/gradient propagation (efficient ways to solve deep learning models)
 - better activation functions (e.g. ReLU)
 - better optimizers (e.g. RMSProp and ADAM)
 - specialized layers (e.g. GRU, LSTM)
 - advanced architectures (e.g. transformers, GAN, GNN etc.)



Why now?

Innovations in:

- Infrastructure:
 - scaling-up of computation frameworks (e.g. cloud computing)
 - distributed computing (and storage)
 - programming frameworks



Deep learning is a matter of

- Scale
 - Available hardware (GPU: thanks gamers)
 - Available big data (e.g. massive databases of labeled images)
 - Available infrastructure
 - Available pre-trained model (transfer learning)
- Theoretical breakthrough
 - ReLU activation functions
 - Back propagation
 - Gradient descent and other solvers/optimizers



State of the art: nerd stuff



Image generation

DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research

in the style of ukiyo-e as a one-line drawing in ancient Egypt



DALL·E 2



Zero-Shot Text-to-Image Generation

Aditya Ramesh¹ Mikhail Pavlov¹ Gabriel Goh¹ Scott Gray¹
Chelsea Voss¹ Alec Radford¹ Mark Chen¹ Ilya Sutskever¹

Abstract

Text-to-image generation has traditionally focused on finding better modeling assumptions for training on a fixed dataset. These assumptions might involve complex architectures, auxiliary losses, or side information such as object part labels or segmentation masks supplied during training. We describe a simple approach for this task based on a transformer that autoregressively mod-



Credits: Dall-E 2 <https://openai.com/dall-e-2/>

Play with the small version: Dall-E Mini (then train your own on Colab)

- <https://dallemini.com/>
- <https://www.crayon.com/>
- <https://github.com/borisdavyma/dalle-mini>
- https://colab.research.google.com/github/borisdavyma/dalle-mini/blob/main/tools/inference/inference_pipeline.ipynb

Natural Language Processing (NLP)

The hard thing with deep learning is

having the chance to make something happen.

to learn the system.

having an open mind.

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:



Dataset:



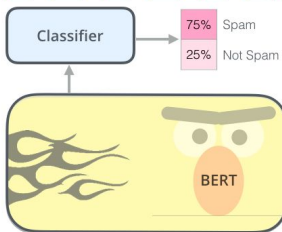
Objective:

Predict the masked word (language modeling)

2 - **Supervised** training on a specific task with a labeled dataset.

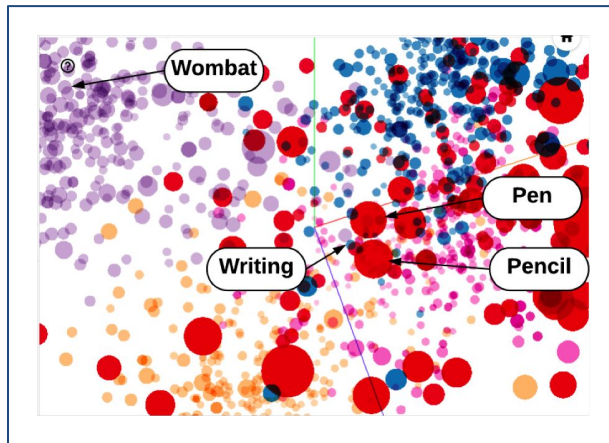
Supervised Learning Step

Model:
(pre-trained in step #1)



Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached...	Not Spam



Credits:

<https://mc.ai/whats-new-in-deep-learning-research-facebook-meta-embeddings-allow-nlp-models-to-choose-their/>

<http://jalammar.github.io/illustrated-bert/>

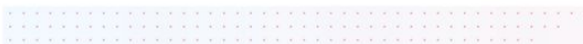
Text generation: GPT2, 3, 3.5, 4...



BY JESSE SUMRAK • 2020-11-24



What Is GPT-3: How It Works and Why You Should Care



arXiv > cs > arXiv:2005.14165

Search...
Help | Advanced Search

Computer Science > Computation and Language

[Submitted on 28 May 2020 (v1), last revised 22 Jul 2020 (this version, v4)]

Language Models are Few-Shot Learners

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions - something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

Comments: 40+32 pages

Subjects: Computation and Language (cs.CL)

Cite as: arXiv:2005.14165 [cs.CL]

(or arXiv:2005.14165v4 [cs.CL] for this version)

<https://doi.org/10.48550/arXiv.2005.14165>

Submission history

From: Tom B Brown [view email]

[v1] Thu, 28 May 2020 17:29:03 UTC (6,995 KB)

[v2] Mon, 1 Jun 2020 17:08:53 UTC (6,997 KB)

[v3] Fri, 5 Jun 2020 02:52:35 UTC (6,998 KB)

[v4] Wed, 22 Jul 2020 19:47:17 UTC (6,998 KB)

ENTERPRISE TECH

What Is GPT-3 And Why Is It Revolutionizing Artificial Intelligence?

Bernard Marr Contributor @

Oct 5, 2020, 12:21am EDT

Listen to article 9 minutes

There's been a great deal of hype and excitement in the artificial intelligence (AI) world around a newly developed technology known as GPT-3. Put simply, it's an AI that is better at creating content that

What is GPT-3 AI and How Can it Help



Adam Enfroy • Updated Aug 07, 2022

What is GPT-3 AI and How Can it Help Me?

In this article, I review GPT-3 AI and



GPT3/4 → ChatGPT

- Has anybody not heard of ChatGPT?
- Hype of the moment: here to stay?

→ <https://openai.com/blog/chatgpt/>
→ <https://chat.openai.com/auth/login>

RTutor

- <http://rtutor.ai>
- uses OpenAI's powerful text-davinci-003 language model to translate natural language into R code

Microsoft Bing

- <https://www.bing.com> → chat
- Microsoft has beaten the competitors and started a partnership with OpenAI



User help me write a short note to introduce myself to my neighbor

ChatGPT Hi there,

My name is [Your Name] and I just moved in next door. I wanted to reach out and introduce myself. I'm looking forward to getting to know you and the neighborhood. If you ever need anything or just want to chat, please don't hesitate to stop by or give me a call.

Best,
[Your Name]

User can you make it more formal?

ChatGPT Dear Neighbor,

I am writing to introduce myself. My name is [Your Name] and I have recently moved in next door. I wanted to take a moment to reach out and say hello. I am looking forward to getting to know you and the neighborhood.

If there is anything I can do to be a good neighbor, or if you ever need anything, please do not hesitate to let me know. I am happy to help in any way that I can.

Sincerely,
[Your Name]

User can you send it for me?

ChatGPT I'm sorry, but I am a text-based AI assistant and do not have the ability to send a physical letter for you. I can only provide suggestions and advice on what to write. It is up to you to take that advice and send the letter yourself.



OpenAI models are directly available (if you pay)



```
response = openai.Completion.create(  
    model="text-davinci-003",  
    prompt="Tell me a joke about AI",  
    temperature=0.6  
)
```

- Models: <https://platform.openai.com/docs/models/overview>
- Cost is token-dependent, see <https://platform.openai.com/tokenizer>
- For actual application you'll need fine tuning:
<https://platform.openai.com/docs/guides/fine-tuning>
- [QUESTION] A course for a deeper dive?



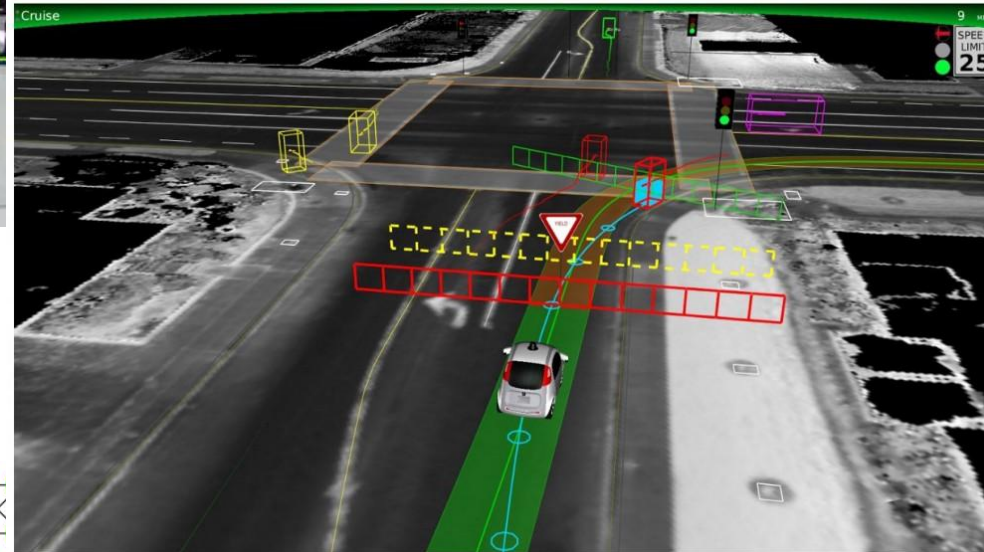
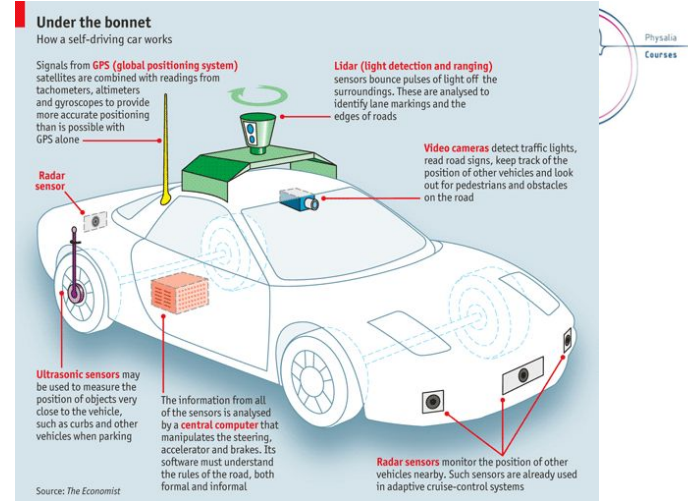
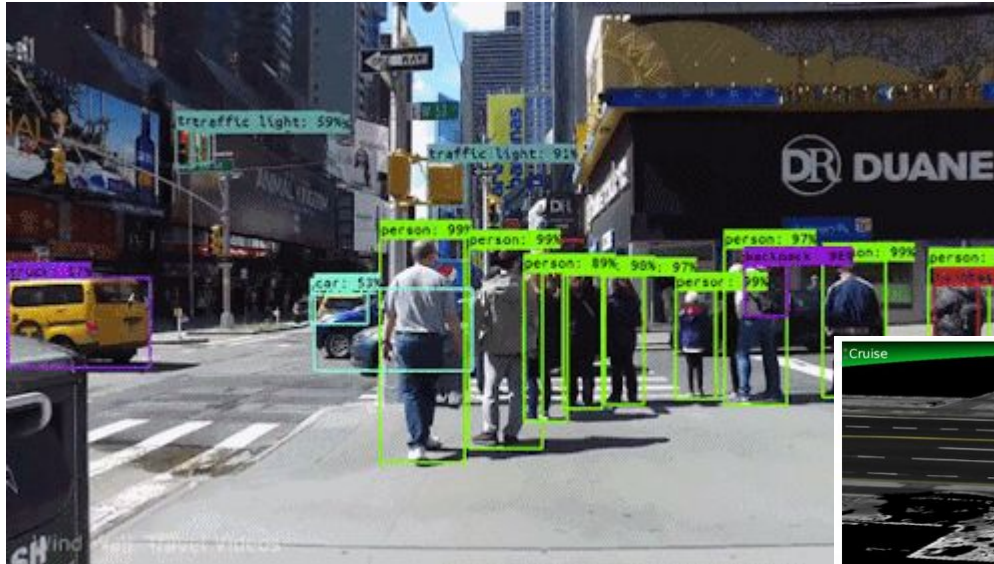
Alternatives to ChatGPT

- Google's Gemini:
 - <https://gemini.google.com>
 - More updated to recent world event
 - Less capable of complex tasks
- Microsoft's Copilot
 - <https://copilot.microsoft.com>
- Phind
 - <https://www.phind.com/>
 - Supports you for writing code
 - Not a chatbot

These all appear to be lagging a bit behind ChatGPT ...



Self driving cars

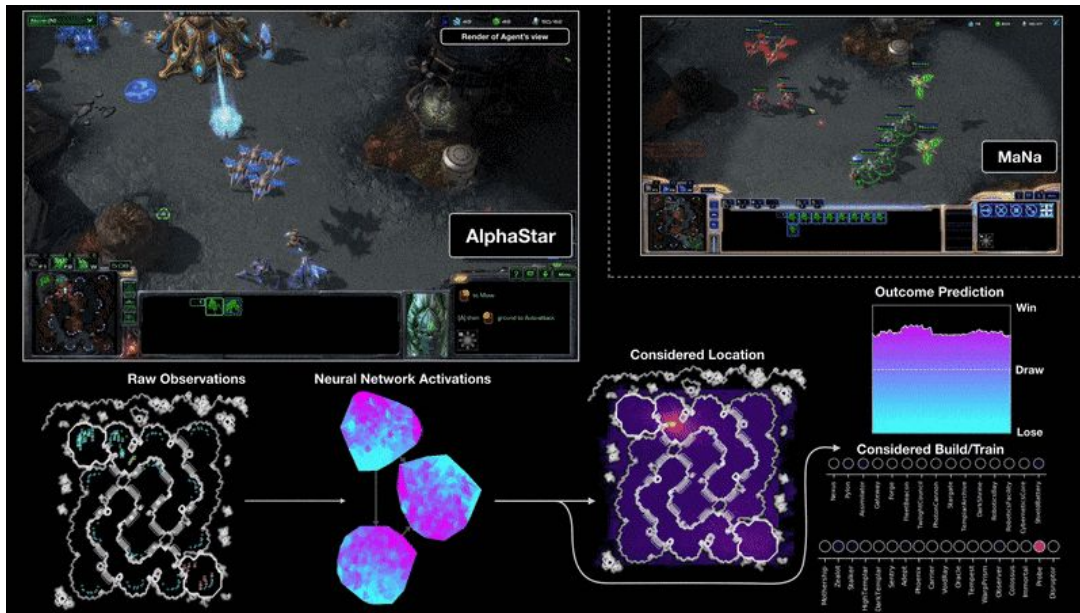


Credits:

<https://medium.com/@feiqi9047/the-data-science-behind-self-driving-cars-eb7d0579c80b>



Games & Videogames



Generating believable videos (deepfake)



Source Sequence



Our Reenactment
(Full Head)



Averbuch-Elor et al. 2017



Living portraits



Credits:

<https://www.gizmodo.co.uk/2018/06/deepfake-videos-a-re-getting-impossibly-good/>

<https://www.sciencealert.com/samsung-s-ai-can-now-generate-talking-heads-from-a-single-image>

YouTube @ birbfakes

And many, many more...

- [News Aggregation and Fraud News Detection](#)
- [Virtual Assistants](#)
- [Entertainment](#)
- [Visual Recognition](#)
- [Fraud Detection](#)
- [Healthcare](#)
- [Personalisations](#)
- [Detecting Developmental Delay in Children](#)
- [Colourisation of Black and White images](#)
- [Adding sounds to silent movies](#)
- [Automatic Machine Translation](#)
- [Automatic Handwriting Generation](#)
- [Language Translations](#)
- [Pixel Restoration](#)
- [Photo Descriptions](#)
- [Demographic and Election Predictions](#)
- [Deep Dreaming](#)



[REF] Deep learning applications

- State of the art for self driving cars: <https://neurohive.io/en/state-of-the-art/self-driving-cars/> and <https://www.bloomberg.com/features/2020-self-driving-car-race/>
- Waymo blog on their autonomous vehicles: <https://blog.waymo.com/>
- Updated repository of NLP state of the art <https://github.com/sebastianruder/NLP-progress>
- NLP transformers: <https://github.com/huggingface/transformers>
- NLP BERT
<https://medium.com/analytics-vidhya/text-classification-with-bert-using-transformers-for-long-text-inputs-f54833994df>
- Deepmind Agent on mastering Atari Games
<https://deepmind.com/blog/article/Agent57-Outperforming-the-human-Atari-benchmark>
- Deepmind AlphaGO on mastering the game of game of GO
<https://deepmind.com/research/case-studies/alphago-the-story-so-far>
- Deepming Alphastar on mastering Real Time Strategy videogame Starcraft II
<https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii>
- Generating living portraits from few shots <https://arxiv.org/abs/1905.08233>
- Deep video portraits original paper - ACM TOG 2018 conference
<https://dl.acm.org/doi/abs/10.1145/3197517.3201283>

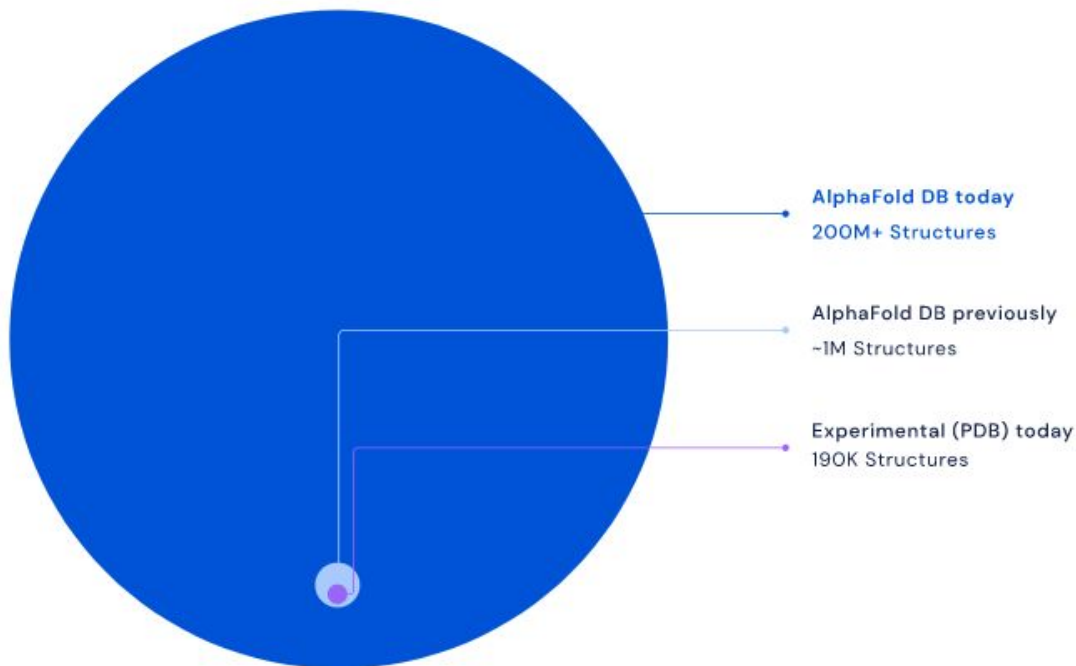


State of the art/2: wet stuff



Protein folding + public DB

Number of Protein Structures



Images credit: AlphaFold reveals the structure of the protein universe
<https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe>



nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 15 July 2021](#)

Highly accurate protein structure prediction with AlphaFold

[John Jumper](#) , [Richard Evans](#), ... [Demis Hassabis](#)  [+ Show authors](#)

Nature **596**, 583–589 (2021) | [Cite this article](#)

745k Accesses | **2740** Citations | **3148** Altmetric | [Metrics](#)

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort^{1,2,3,4}, the structures of around 100 000 unique proteins have been determined⁵, but this represents a

THE JOURNAL OF
PHYSICAL
CHEMISTRY
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB



[Review Article](#)

AlphaFold, Artificial Intelligence (AI), and Allostery

Published as part of *The Journal of Physical Chemistry* virtual special issue "Protein Folding and Dynamics: An Overview on the Occasion of Harold Scheraga's 100th Birthday".

Ruth Nussinov,* Mingzhen Zhang, Yonglan Liu, and Hyunbum Jang

 **Cite This:** *J. Phys. Chem. B* 2022, 126, 6372–6383

 [Read Online](#)

[ACCESS |](#)

 [Metrics & More](#)

 [Article Recommendations](#)

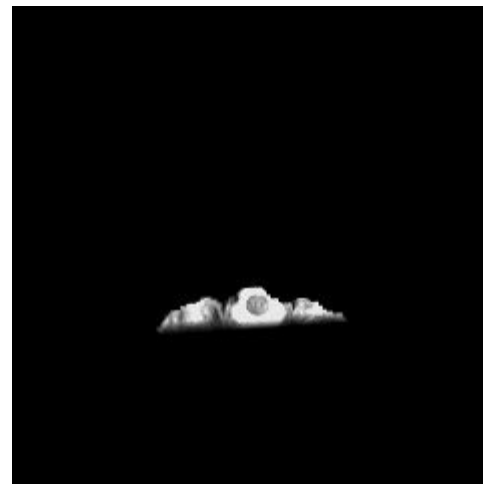
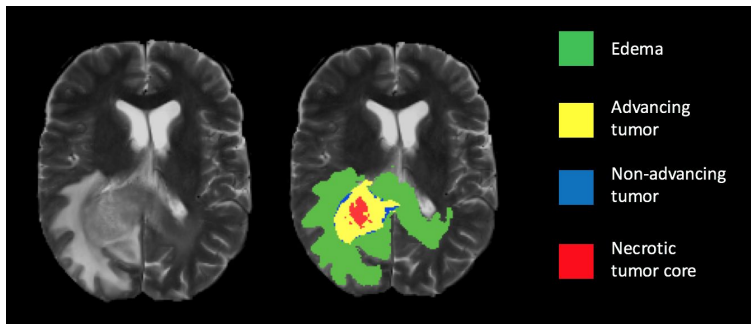
ABSTRACT: AlphaFold has burst into our lives. A powerful algorithm that underscores the strength of biological sequence data and artificial intelligence (AI). AlphaFold has appended projects and research directions.






AI in Structural Biology

Prediction

Brain Tumor Segmentation



RANK	METHOD	DICE SCORE	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	OM-Net + CGAp	87%	×	One-pass Multi-task Networks with Cross-task Guided Attention for Brain Tumor Segmentation			2019
2	CNN + 3D filters	85%	✓	CNN-based Segmentation of Medical Imaging Data			2017

Images credit: Brain Tumor Segmentation with Deep Neural Networks https://github.com/naldeborgh7575/brain_segmentation

Classification of medical images

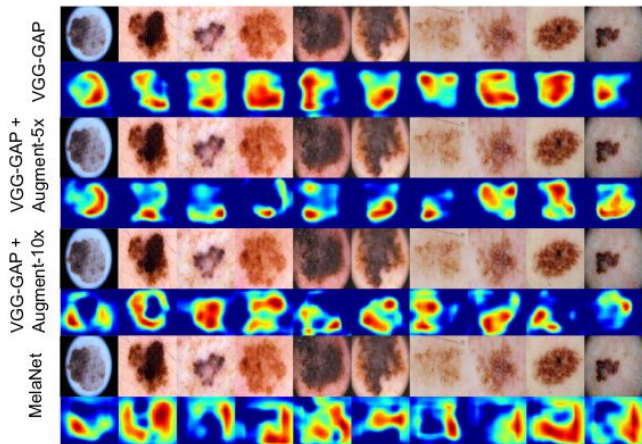


Figure 8: Grad-CAM heat maps for the correctly classified malignant cases by MelaNet and baseline methods.

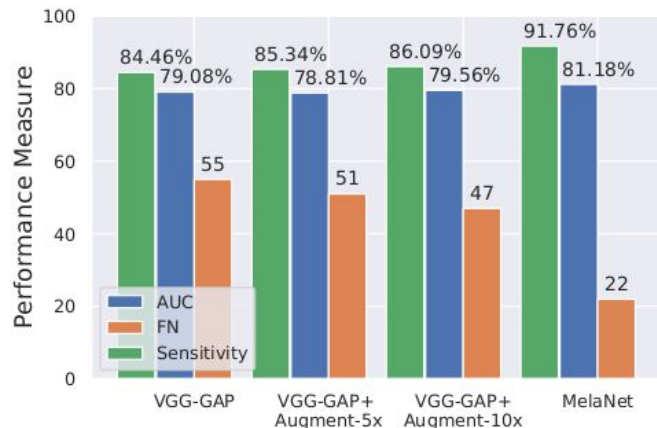


Figure 5: Classification performance of MelaNet and the baseline methods using AUC, FN and Sensitivity as evaluation metrics on the ISIC-2016 test set.

Source: Zunair and Hamza, 2020. Melanoma Detection using Adversarial Training and Deep Transfer Learning.



Drug discovery/optimization

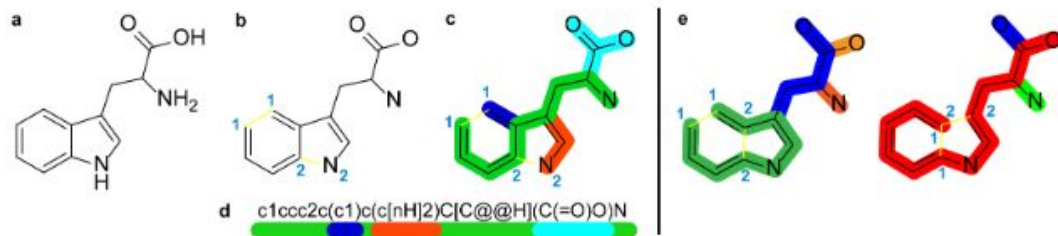
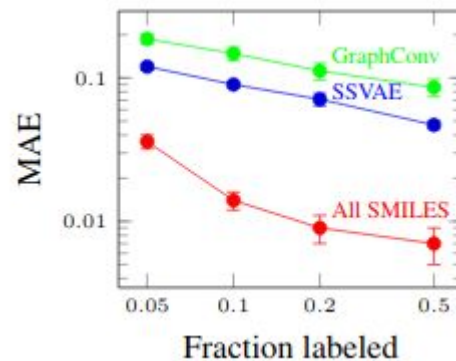


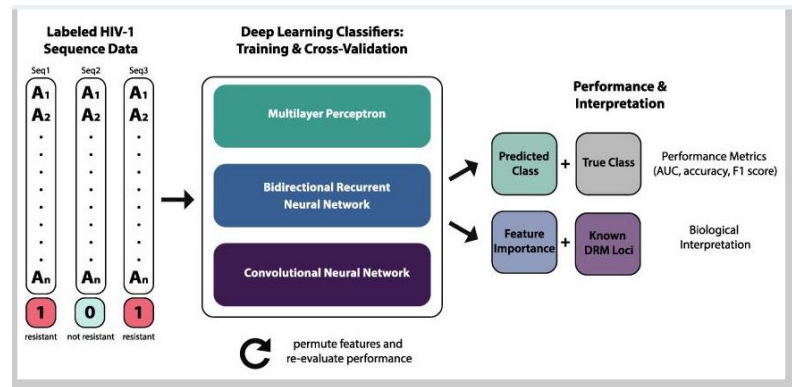
Figure 1: The molecular graph of the amino acid Tryptophan (a). To construct a SMILES string, all cycles are broken, forming a spanning tree (b); a depth-first traversal is selected (c); and this traversal is flattened (d). The beginning and end of intermediate branches in the traversal are denoted by (and) respective. The ends of broken cycles are indicated with matching digits. The full grammar is listed in Appendix D. A small set of SMILES strings can cover all paths through a molecule (e).



Source: Alperstein et al, 2019. All SMILES Variational Autoencoder



Drug resistance prediction



[Viruses](#). 2020 May; 12(5): 560.

Published online 2020 May 19. doi: [10.3390/v12050560](https://doi.org/10.3390/v12050560)

Drug Resistance Prediction Using Deep Learning Techniques on HIV-1 Sequence Data

Margaret C. Steiner,^{1,*} Keylie M. Gibson,¹ and Keith A. Crandall^{1,2}

► Author information ► Article notes ► Copyright and License information [Disclaimer](#)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7290575/>



And the slack channel (thanks Pleuni!)

Breast Cancer detection

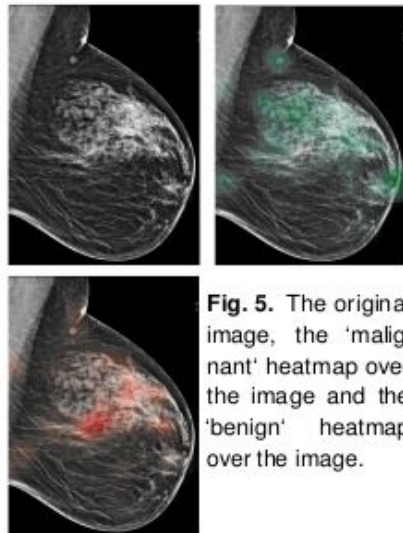
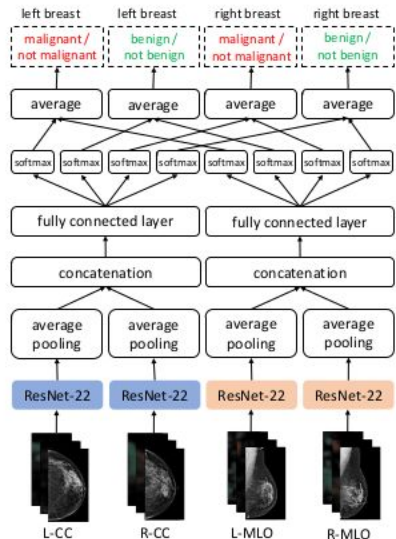
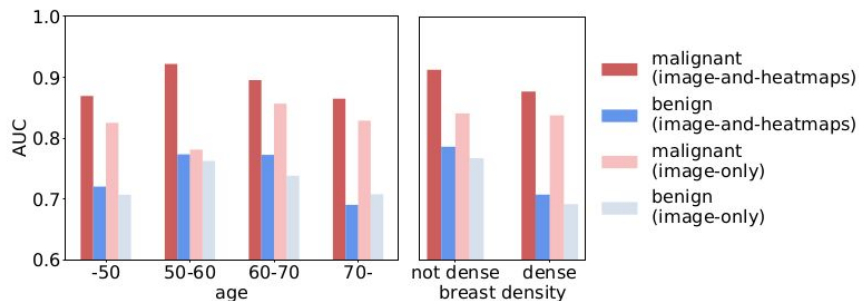


Fig. 5. The original image, the 'malignant' heatmap over the image and the 'benign' heatmap over the image.



Source: Wu et al, 2019. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening



EEG interpretation/analysis

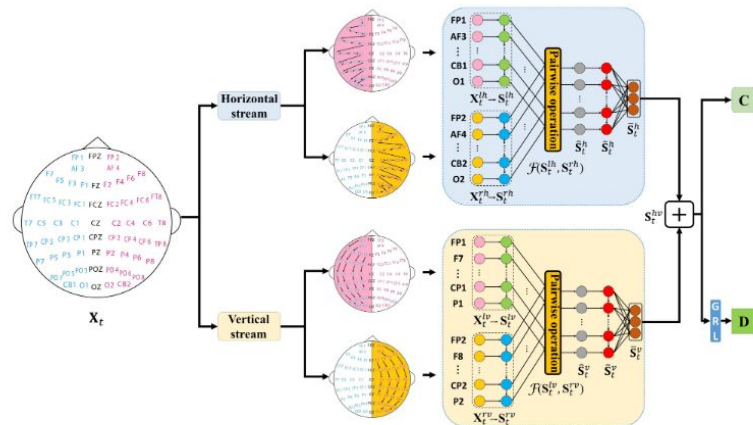
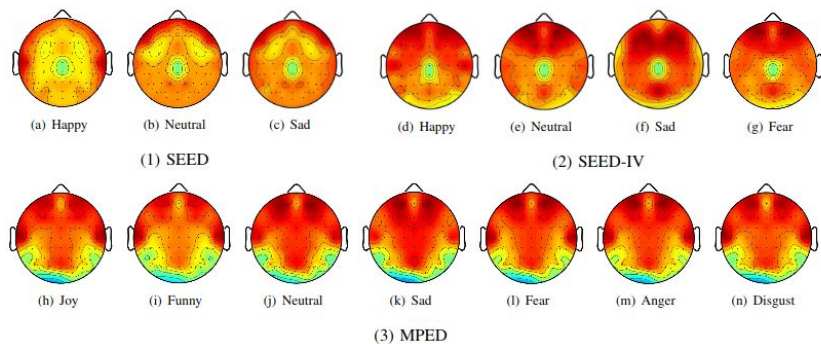


Fig. 1: The framework of BiHDM. BiHDM consists of four RNN modules to capture each hemispheric EEG electrodes' information from horizontal and vertical streams. Then all the electrodes' data representations interact and construct the final vector for the classifier and discriminator.

Source: Li et al., 2019. A Novel Bi-hemispheric Discrepancy Model for EEG Emotion Recognition



[REF] Deep learning state of the art

- Continuously updated applications of DL, divided by topic (Computer Vision, NPL, medical...)
<https://paperswithcode.com/sota>
- MIT Deep learning state of the art 2020 seminar and course <https://deeplearning.mit.edu/>
- Deep Learning Papers Reading Roadmap
<https://github.com/floodsung/Deep-Learning-Papers-Reading-Roadmap>



[REF] Conferences

- Deep Learning World is the premier conference covering the commercial deployment of deep learning
<https://www.deeplearningworld.com/>
- Machine Learning Week Europe Berlin - October 5-6, 2022 <https://machinelearningweek.eu/>



Keywords

- Classification
- Regression
- Data representation
- Model/Method/Algorithm/Software
- Data regularization
- Accuracy, error, correlation
- Overfitting
- Training set, test set



Can deep learning predict it all?



Difficult or easy?

- “Prediction is very difficult, especially about the future”
(*Niels Bohr*)

Or not? (with deep learning)

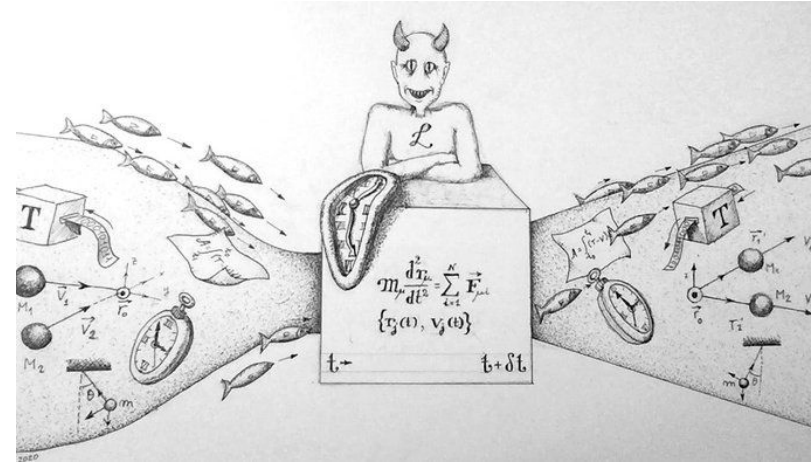
- spam filters accuracy: 99.9% (1 out of 1000 emails escapes) [[here](#)]
- OCR (optical character recognition) accuracy: 99% [[here](#)]
- retinal disease diagnosis accuracy: 94.5% [[here](#)]
- LLMs, Chat-GPT etc.



Laplace's demon

It has the ability to know the exact position and state of motion of **every particle in the universe**, and it fully understands the laws of physics that describe how they interact with each other.

In a **deterministic universe**, this all-knowing demon could work out how the universe will evolve over time and predict its state in the future.

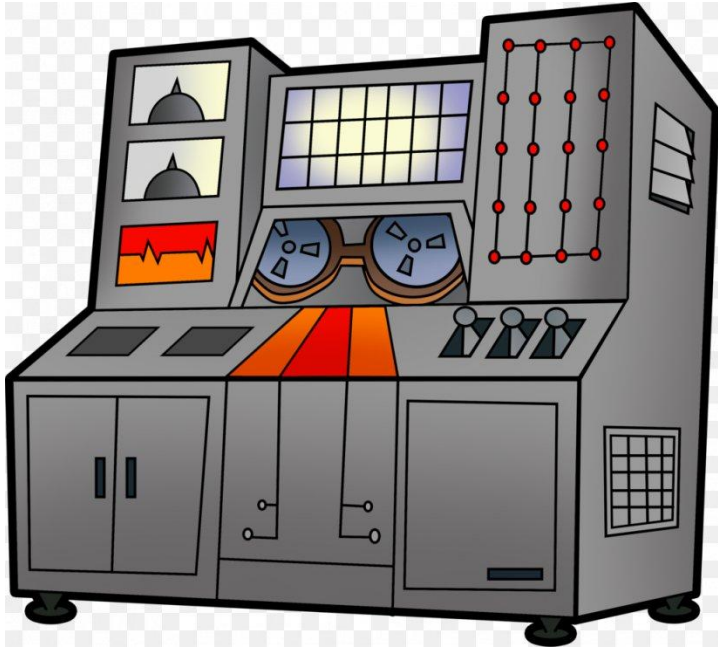


[source]

« Une **intelligence** qui, à un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était suffisamment vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome ; rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux. »



Laplace's demon



[[source](#)]

Laplace's really believed in the deterministic universe and that his demon could predict everything.

But is this true?

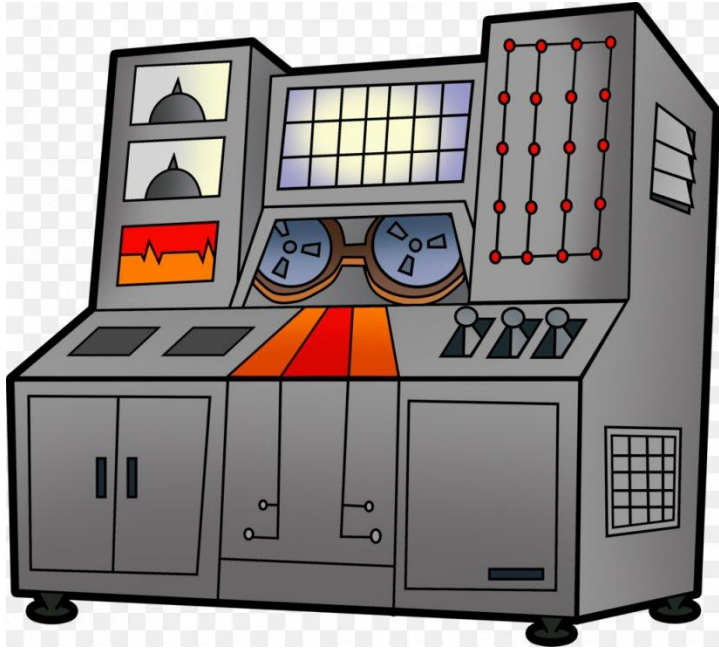
We can think of Laplace's demon as a **supercomputer** that contains all the information of the universe, and is able to compute precisely how the future will unfold

But then it could decide to use this information to change the course of events compared to its earlier prediction, thus making the prediction wrong (or a short-circuit may alter the course of events)

So can it or can it not predict everything?



Laplace's demon



[[source](#)]

Let's examine what it means for the supercomputer to have knowledge about its own decisions (or short-circuits)

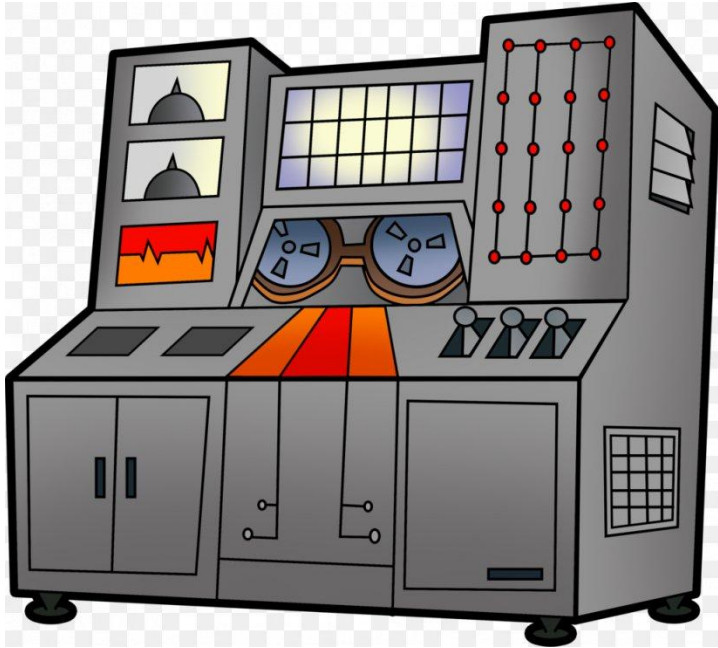
To predict its own actions (voluntary or not) the supercomputer would need to **include in the calculations also every details of its inner structure and functioning** (the state and motion of every atom and every electron it is made of)

But **this information would need to be stored in memory disks** which are themselves **made of atoms and particles arranged in a special way**, which would be **other pieces of information** to be stored, and so on in an **infinite recursion**

The demon can't therefore include itself in its calculations, hence **its knowledge of the universe is incomplete** and **its predictions imperfect**



Laplace's demon



We conclude that this omniscient supercomputer can predict everything about the universe outside it, but not what happens inside it

Max theoretical accuracy is not 1, but maybe 0.99999

[[source](#)]



But is the universe deterministic?

- Isaac Newton and the **clockwork universe** → determinism
- The probabilistic nature of the **quantum world**:
 - The behaviour of single particles is unpredictable (e.g. radioactive decay)
 - The average behaviour of many particles is predictable



But is the universe deterministic?

- Isaac Newton and the **clockwork universe** → determinism
- The probabilistic nature of the **quantum world**:
 - The behaviour of single particles is unpredictable (e.g. radioactive decay)
 - The average behaviour of many particles is predictable

Max theoretical accuracy further down, from 0.99999 to say 0.99988



And is the universe predictable?

- Konrad Lorenz and the butterfly effect: tiny changes in the **initial conditions** lead to huge differences in the end results:
 - e.g. weather, or tossing a coin in the exact same manner
 - to make 100% accurate predictions we would need to know the initial conditions to infinite accuracy (which is not possible)
- Chaos theory and nonlinear dynamics:
 - simple cause → complex effect (non linear) (e.g. turbulent fluids)
 - determinism can generate randomness: a complex system obeys the laws of physics, but can become disordered and **unpredictable**
 - but also: from a chaotic system order and structure may emerge (e.g. A.I.)



And is the universe predictable?

- Konrad Lorenz and the butterfly effect: tiny changes in the **initial conditions** lead to huge differences in the end results:
 - e.g. weather, or tossing a coin in the exact same manner
 - to make 100% accurate predictions we would need to know the initial conditions to infinite accuracy (which is not possible)
- Chaos theory and nonlinear dynamics:
 - simple cause → complex effect (non linear) (e.g. turbulent fluids)
 - determinism can generate randomness: a complex system obeys the laws of physics, but can become disordered and **unpredictable**
 - but also: from a chaotic system order and structure may emerge (e.g. A.I.)

More reasons to reduce the max theoretical accuracy from 0.99999 to say 0.99988



To sum up

- part of the universe may be **inherently random / probabilistic**
- even if our universe is deterministic:
 - Laplace's demon **can never know it all**
 - sensitivity to the initial conditions and chaos theory lead to **unpredictability**
- “Prediction is **still** very difficult, **not only** about the future”
(adapted from: *Niels Bohr*)



Additional pills: the limits of learning

Adversarial Policies Beat Superhuman Go AIs

Tony T. Wang, Adam Gleave, Tom Tseng, Kellin Pelrine, Nora Belrose, Joseph Miller, Michael D. Dennis, Yawen Duan, Viktor Pogrebniaik, Sergey Levine, Stuart Russell

[View PDF](#)

We attack the state-of-the-art Go-playing AI system KataGo by training adversarial policies against it, achieving a >97% win rate against KataGo running at superhuman settings. **Our adversaries do not win by playing Go well.** Instead, they trick KataGo into making serious blunders. Our attack transfers zero-shot to other superhuman Go-playing AIs, and is

<https://arxiv.org/abs/2211.00241>



Additional pills: the limits of learning

Adversarial Policies Beat Superhuman Go AIs

Tony T. Wang, Adam Gleave, Tom Tseng, Kellin Pelrine, Nora Belrose, Joseph Miller, Michael D. Dennis, Yawen Duan, Viktor Pogrebniaik, Sergey Levine, Stuart Russell

[View PDF](#)

We attack the state-of-the-art Go-playing AI system KataGo by training adversarial policies against it, achieving a >97% win rate against KataGo running at superhuman settings. **Our adversaries do not win by playing Go well.** Instead, they trick KataGo into making serious blunders. Our attack transfers zero-shot to other superhuman Go-playing AIs, and is

<https://arxiv.org/abs/2211.00241>



Lee Saedol
was avenged



Additional pills: the limits of learning

- Feed-forward DNN's execution time is:
 - linear with the depth (double the layers, double the execution time)
 - (sub) linear with the width (depending on the hardware)
- Some problems don't admit solutions in linear time (sorting, network analysis, chess/go/...)

ERGO:

Some problems are not solvable* by a feed-forward DNN

*in general terms, for all the cases

