Deep learning for biologists
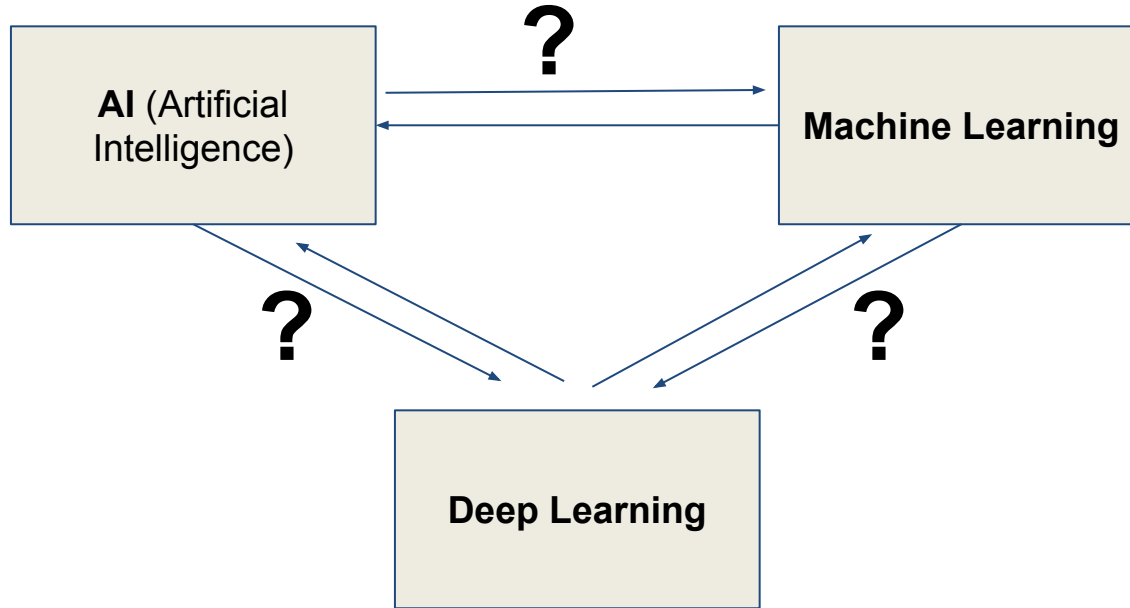
A practical and theoretical introduction

Filippo Biscarini
Senior Scientist
CNR, Milan (Italy)

Nelson Nazzicari
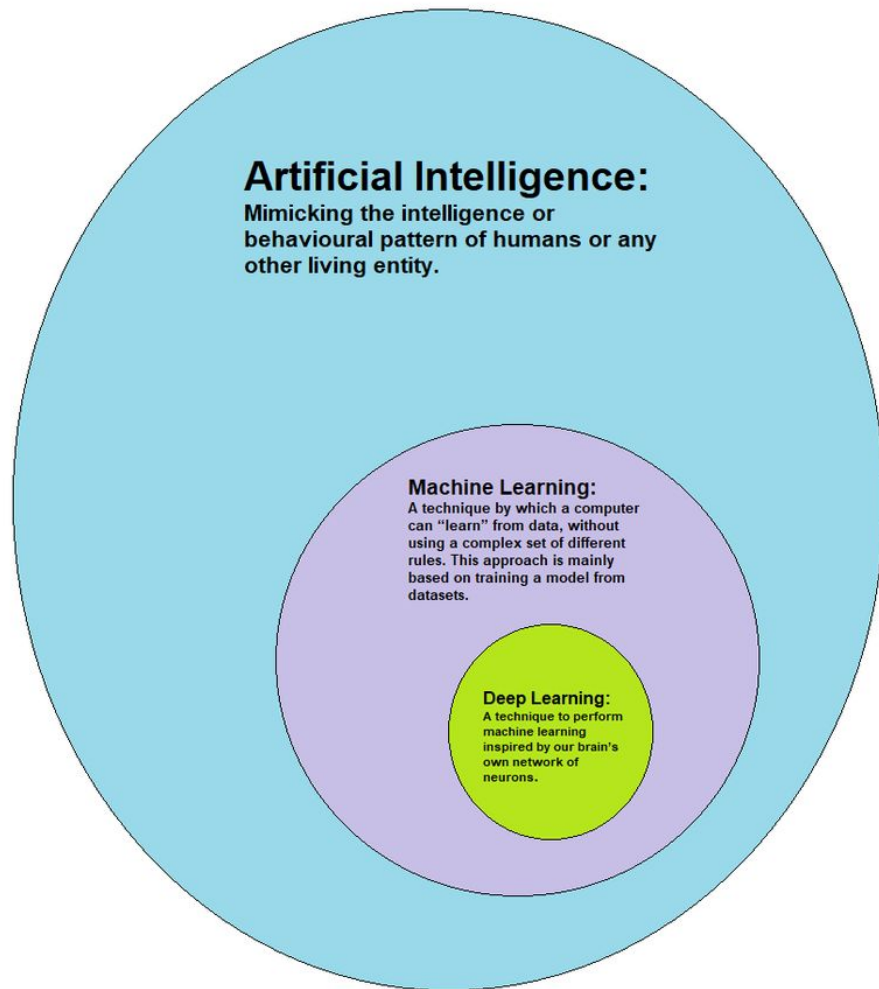*Research fellow*
CREA, Lodi (Italy)
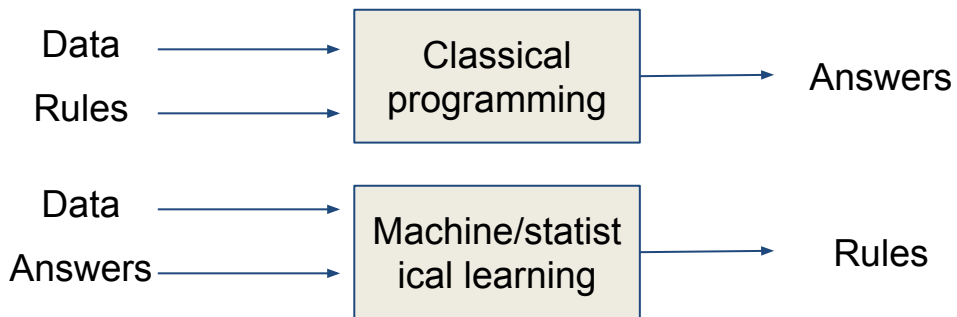
What is deep learning?

AI (Artificial Intelligence)

?

Machine Learning

?

?

Deep Learning

What is deep learning?

AI >> ML >> DL

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

Artificial Intelligence:
Mimicking the intelligence or
behavioural pattern of humans or any
other living entity.

Machine Learning:
A technique by which a computer
can "learn" from data, without
using a complex set of different
rules. This approach is mainly
based on training a model from
datasets.

Deep Learning:
A technique to perform
machine learning
inspired by our brain's
own network of
neurons.

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!

- 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 x 2 exp (YEAR - 1)

PREDICTION

MATHEMATICAL MODEL

What is <u>deep</u> 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"»

- Deep learning
  - **Neural networks**, a mathematical model "inspired" by biology
  - Artificial Neural Networks have been around for ~80 years ...
  - ...but became "deep" (i.e. with many layer) in the last ten years

Why "deep"?
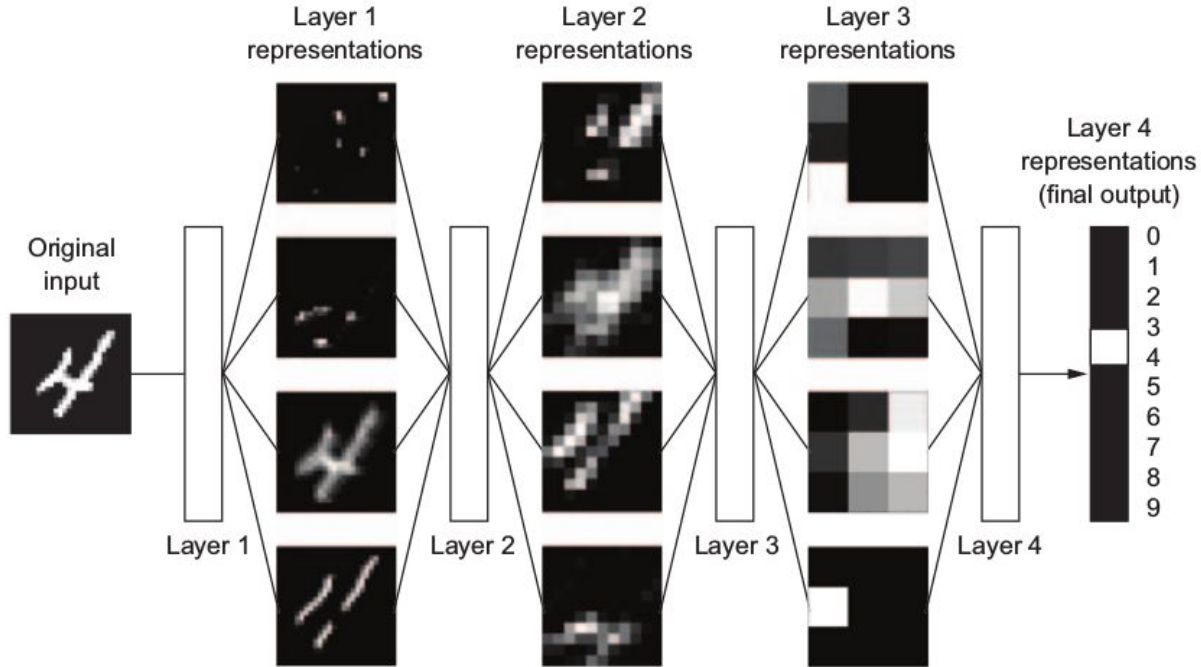
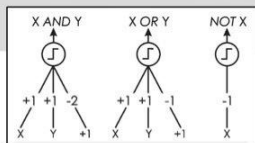Figure 1.6   Deep representations learned by a digit-classification model

Source: François Chollet and J.J. Allaire "Deep learning with R" (2018)
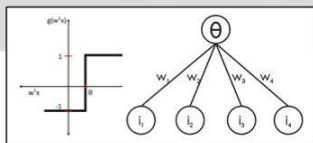
A little history of Neural Networks
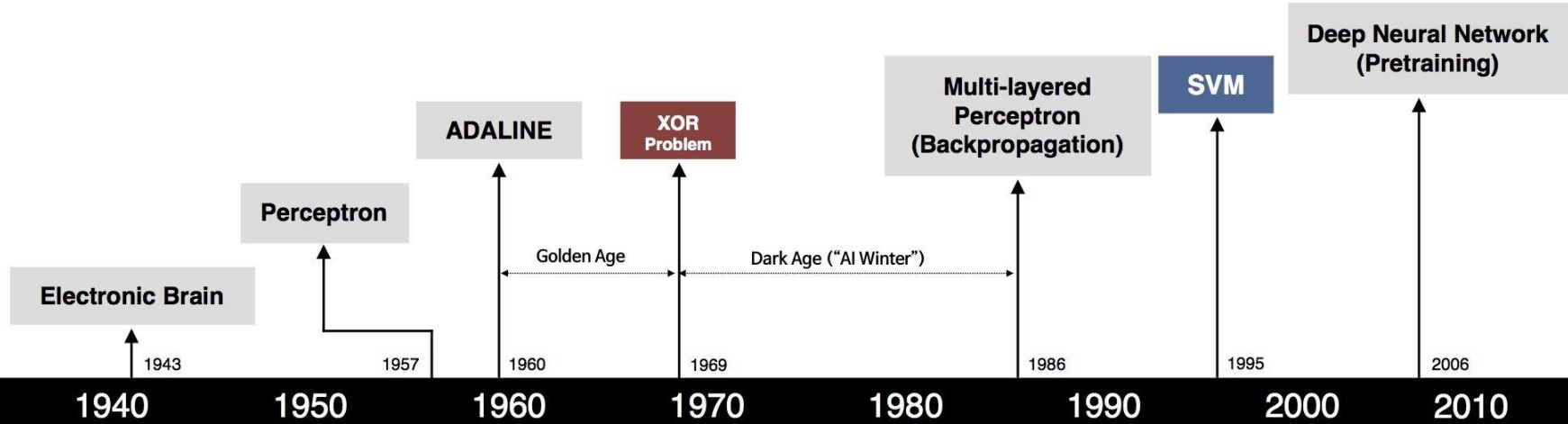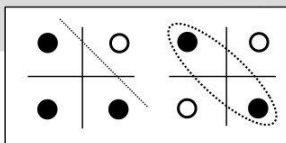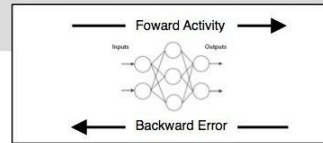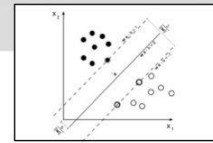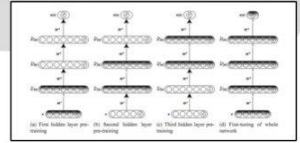
[REF] History of deep learning

- 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

[REF] History of deep learning

- 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

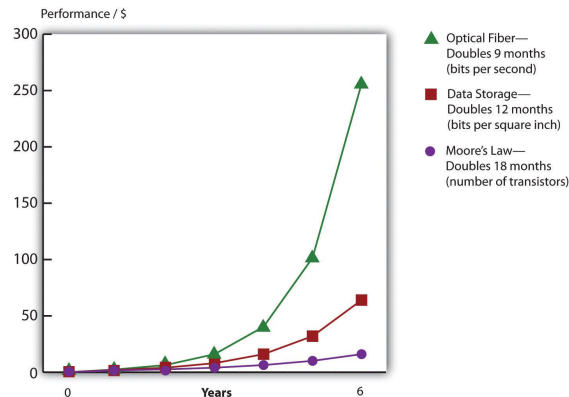- Infrastructure

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)

Why now?

Innovations in:

- Big Data:
    - Internet: search engines, social media, image datasets etc.
    - IoT: sensors, interconnected devices
    - Data storage (Moore's law)

Performance / $

▲	Optical Fiber— Doubles 9 months (bits per second)
■	Data Storage— Doubles 12 months (bits per square inch)
●	Moore's Law— Doubles 18 months (number of transistors)

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)

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

State of the art: nerd stuff

Self driving cars

Credits:
https://medium.com/@feiqi9047/the-data-science-behind-self-driving-cars-eb7d0579c80b
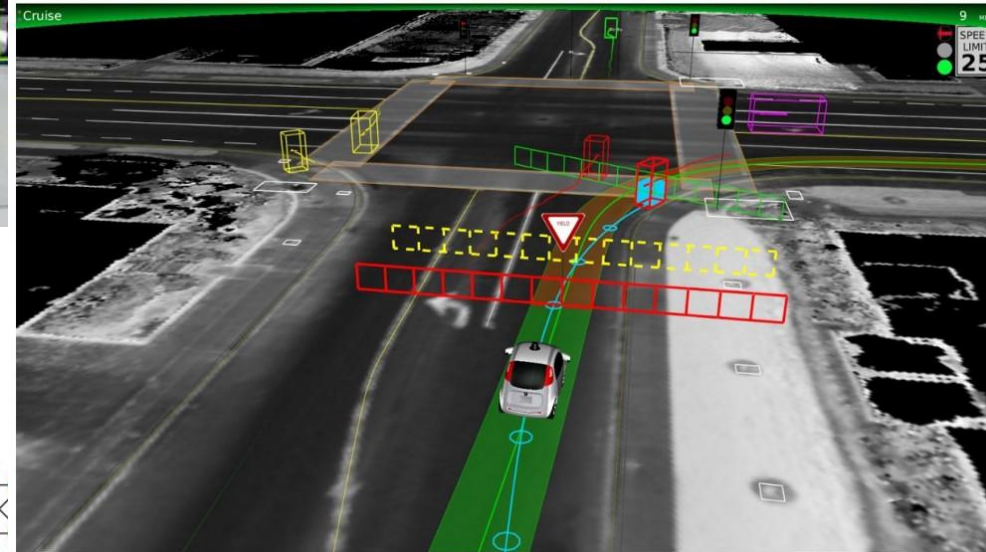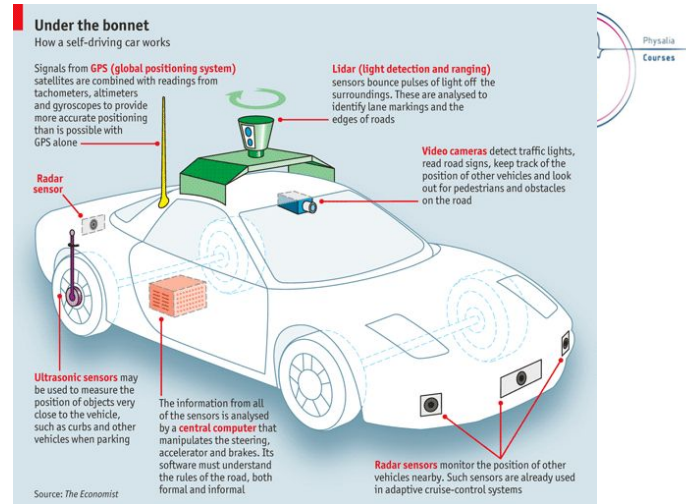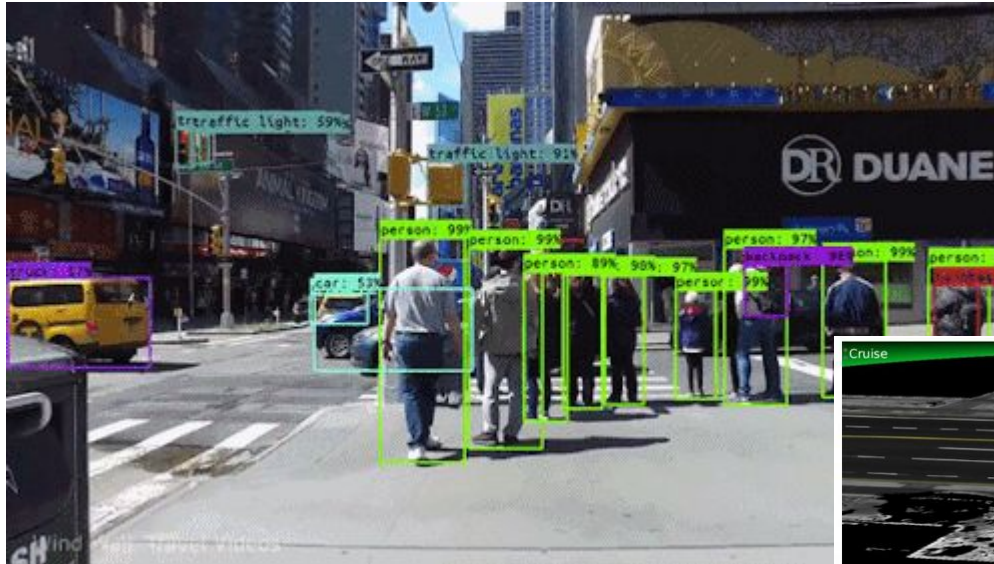
Natural Language Processing (NLP)

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
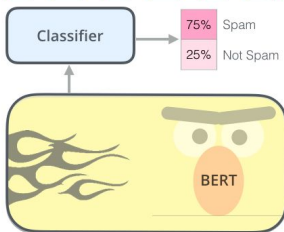
Model:

BERT

Dataset:

WIKIPEDIA
Die freie Enzyklopädie

Objective: Predict the masked word (language modeling)

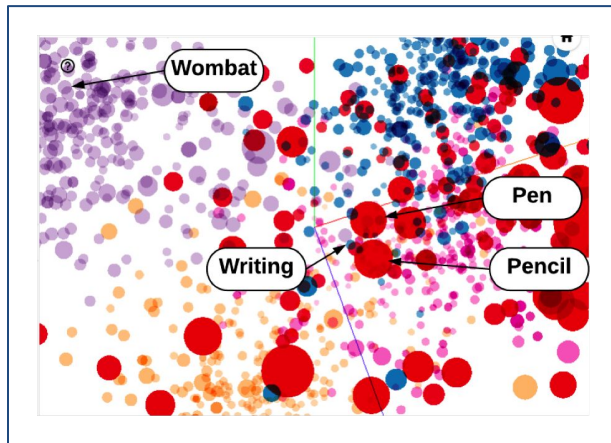2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step

Classifier

| 75% | Spam |
| 25% | Not Spam |

Model:
(pre-trained in step #1)

BERT

Dataset:

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

Wombat

Pen

Writing

Pencil

Games & Videogames
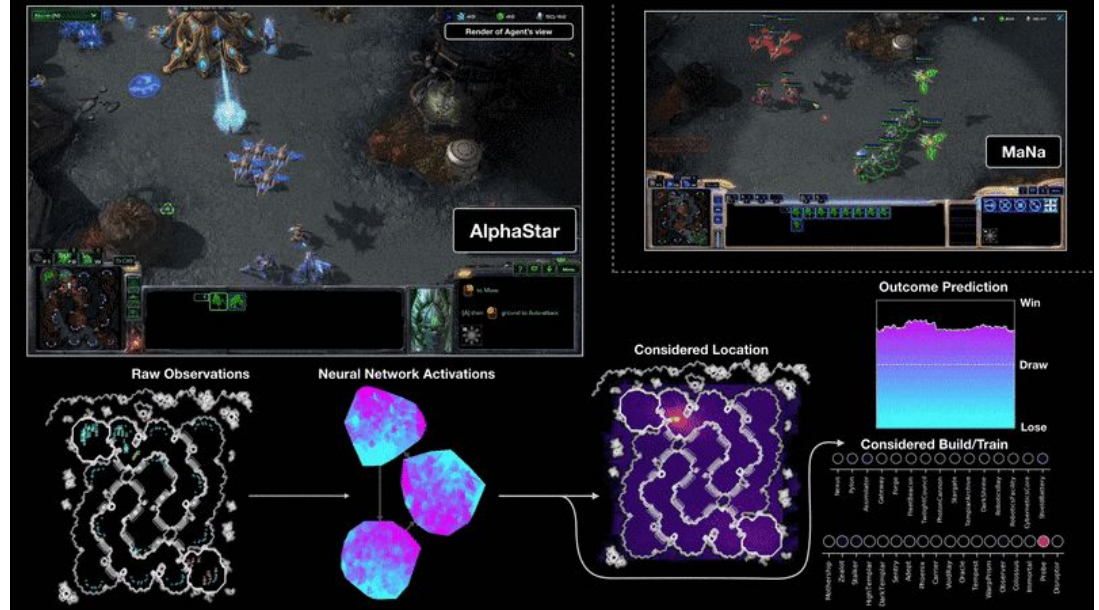
Generating believable videos (deepfake)

Source Sequence → Our Reenactment (Full Head) → Averbuch-Elor et al. 2017

Living portraits

Credits:
https://www.gizmodo.co.uk/2018/06/deepfake-videos-are-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**
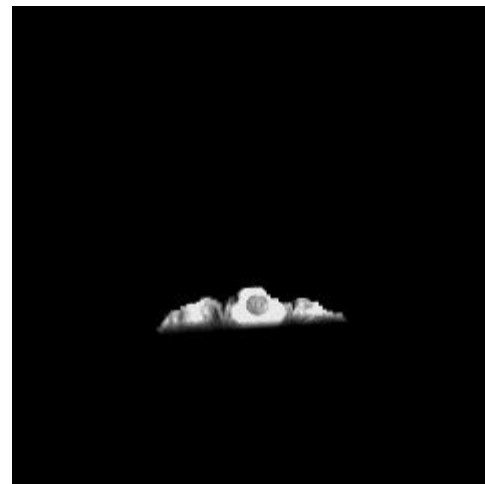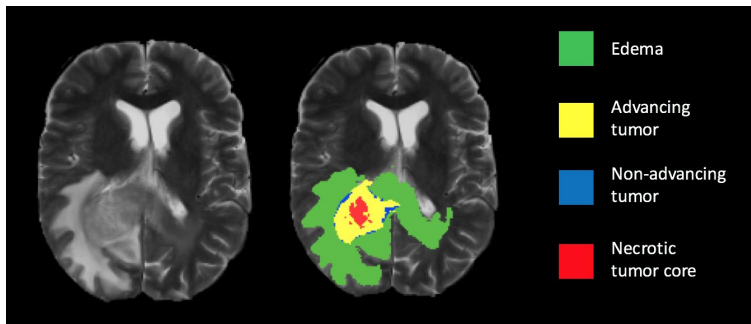- **Deep Dreaming**

[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 NPL state of the art https://github.com/sebastianruder/NLP-progress
- NPL transformers: https://github.com/huggingface/transformers
- NPL BERT https://medium.com/analytics-vidhya/text-classification-with-bert-using-transformers-for-long-text-inputs-f54833994dfd
- 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

Brain Tumor Segmentation

			Color
	Edema	(green)	
	Advancing tumor	(yellow)	
	Non-advancing tumor	(blue)	
	Necrotic tumor core	(red)	

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
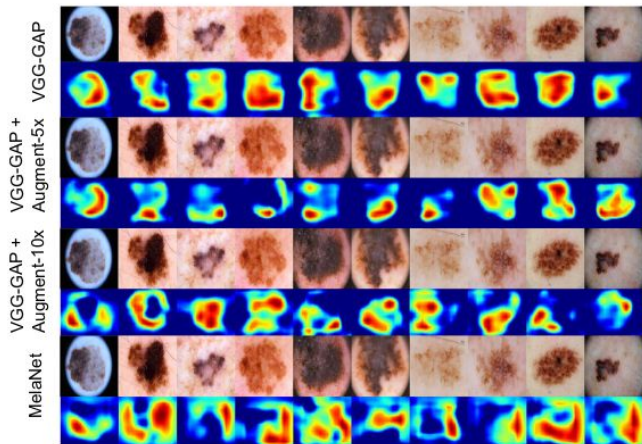
Classification of medical images

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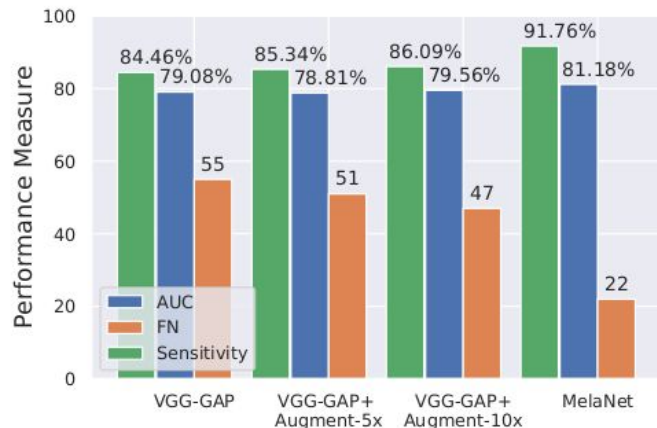

Figure 5: Classification performance of MelaNet and the baseline methods using AUC, FN and Sensitivity as evaluation metrics on the ISIC-2016 test set.

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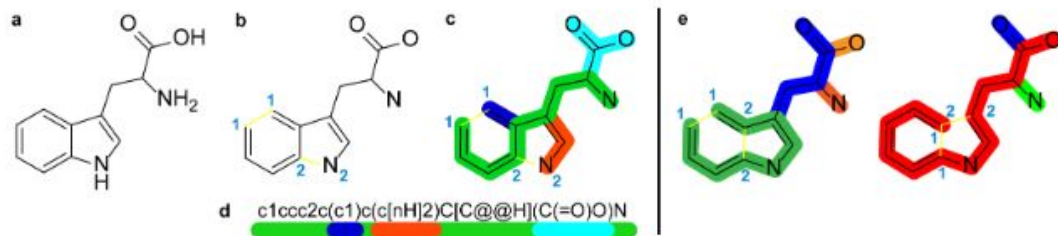
Drug discovery/optimization

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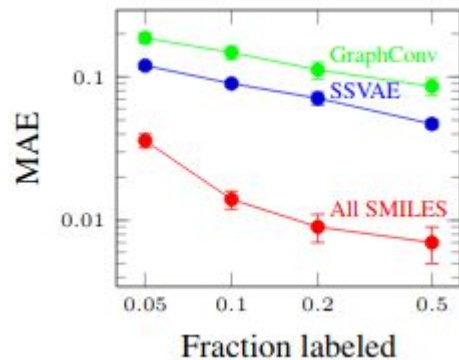
Source: Alperstein et al, 2019. All SMILES Variational Autoencoder

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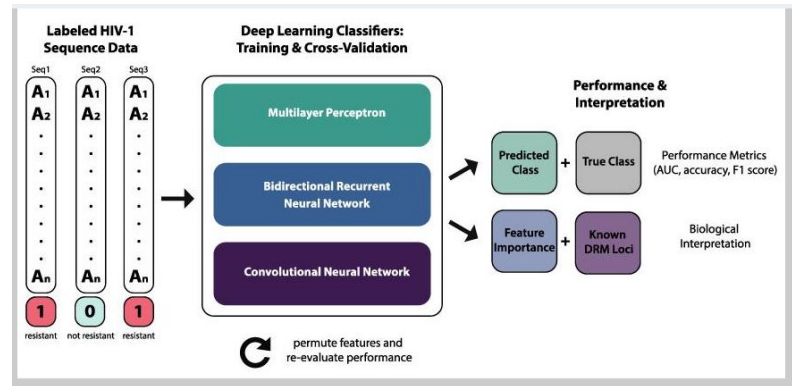
Drug Resistance Prediction Using Deep Learning Techniques on HIV-1 Sequence Data

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!)
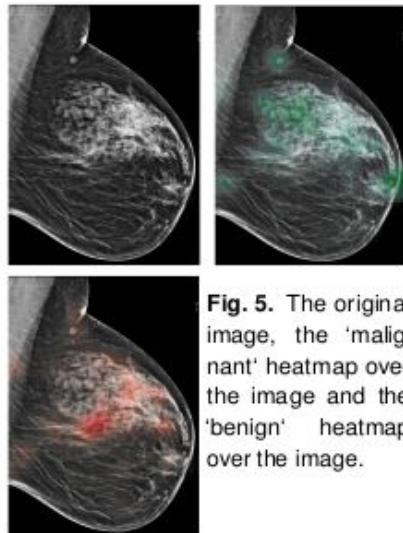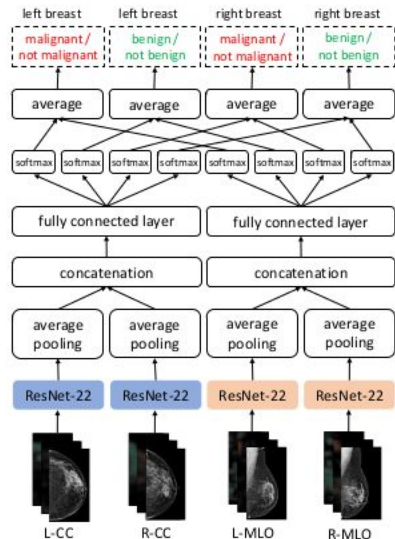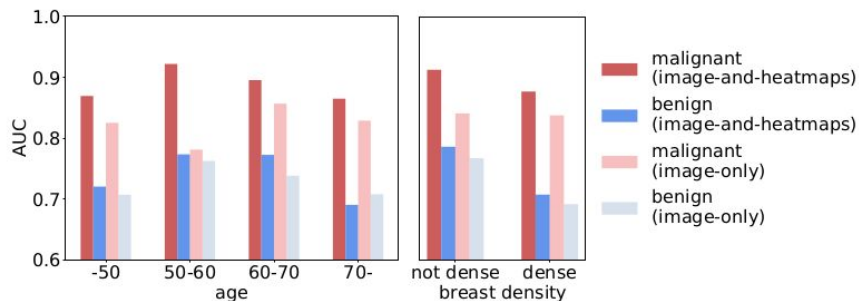
Breast Cancer detection

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
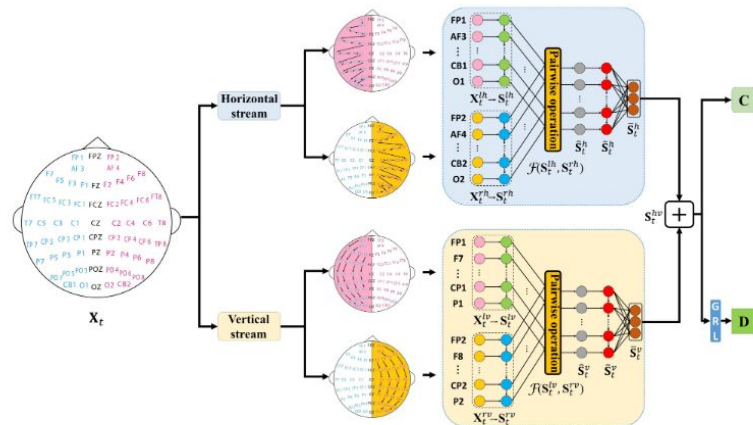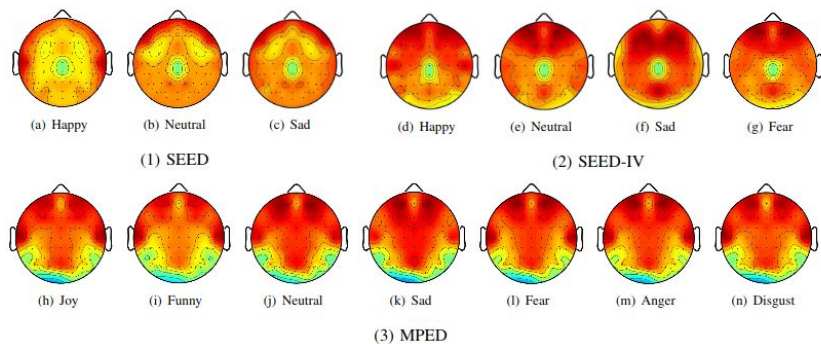

Fig. 1: The framework of BiHDM. BiHDM consists of four RNN modules to capture each hemispheric EEG electrodes' information from horizontal and vertical streams. Then all the electrodes' data representations interact and construct the final vector for the classifier and discriminator.

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

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]

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.
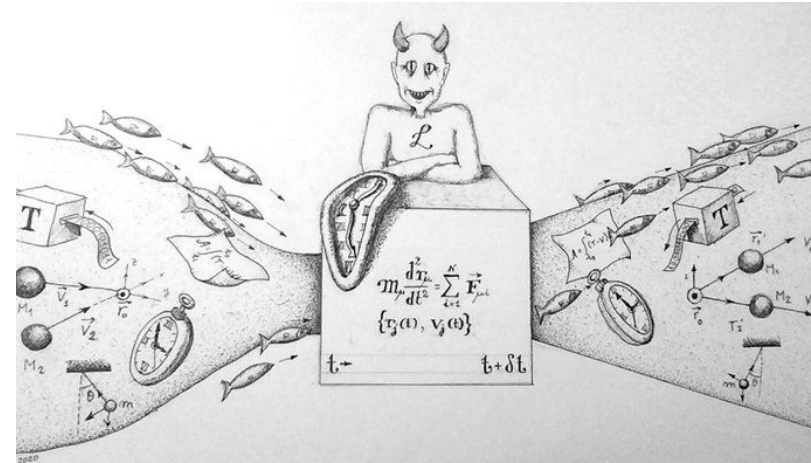
[source]

« 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
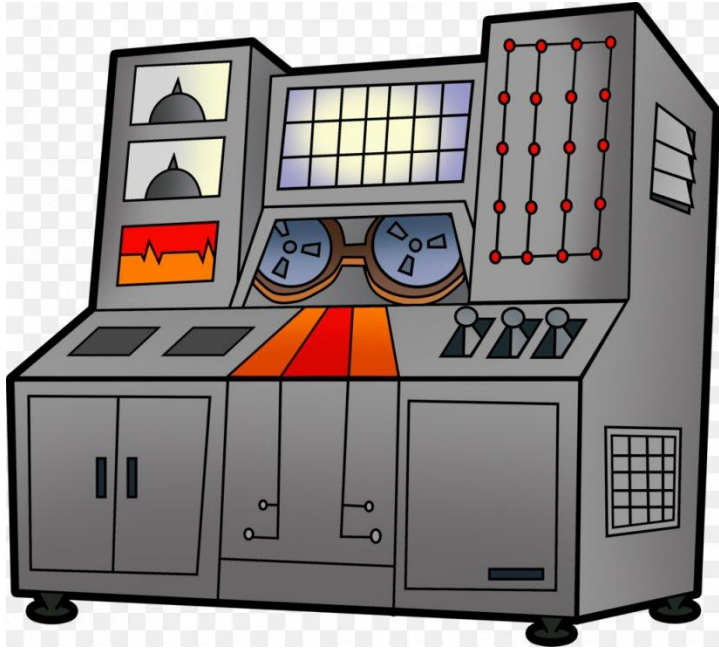
You can think of Laplace's demon as a **supercomputer** that contains all the information of the universe, and is able to compute precisely how the future will unfold
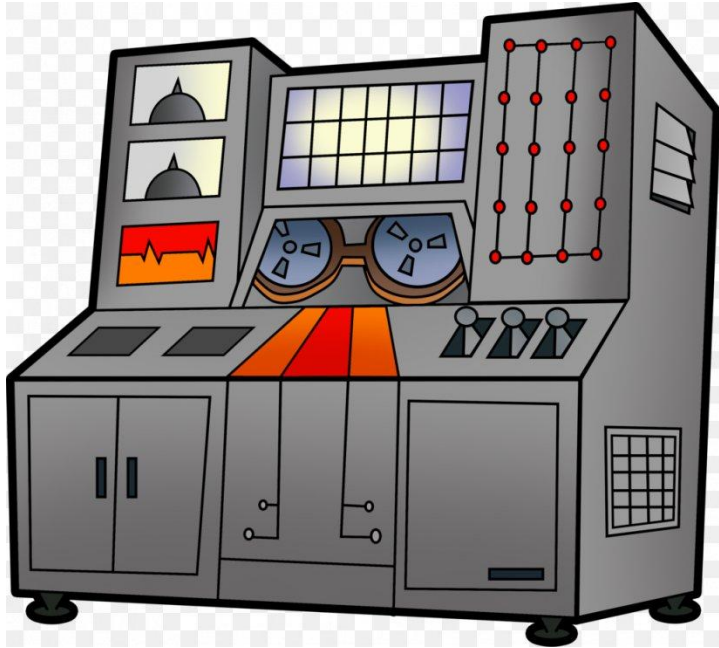
But then it could decide to use this information to change the course of events compared to its earlier prediction, thus making the prediction wrong.

Or not?

[source]

Laplace's demon

[source]

To predict its own actions the supercomputer (the demon) 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**

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

And is the universe predictable?

- Konrad Lorenz and <u>the butterfly effect</u>: 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)

- <u>Chaos theory</u> and nonlinear dynamics:
  - simple cause → complex effect (non linear)
  - 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.)

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" (*Niels Bohr*)