Principles and Applications of Artificial Intelligence

François Fleuret

Machine Learning group

https://mlg.unige.ch

UNIVERSITÉ DE GENÈVE

Plan

- AI "programs itself"
- AI actually works
- AI requires vast amounts of data and computation
- AI is easy to deploy
- AI models are black boxes

AI “programs itself”

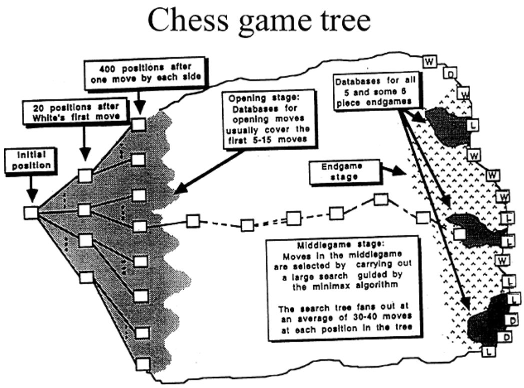
AI "programs itself"

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

```
n = 15345

while n > 1:
    for k in range(2, n+1):
        if n%k == 0:
            print(k)
            n = n // k
            break
```

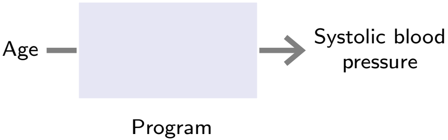
The first attempts at artificial intelligence relied on the same principle *e.g.* medical decision, strategy games, or computer vision.



(Newborn, 1996)

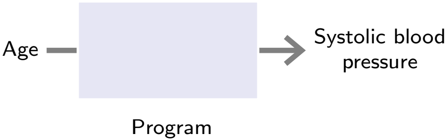
AI "programs itself"

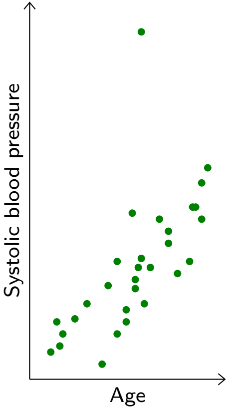
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.



AI "programs itself"

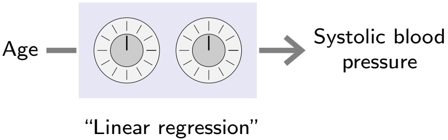
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.

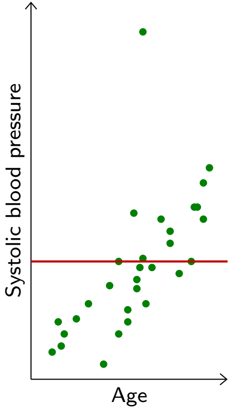




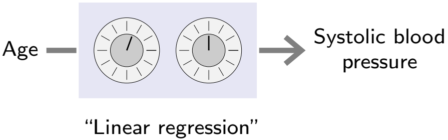
AI "programs itself"

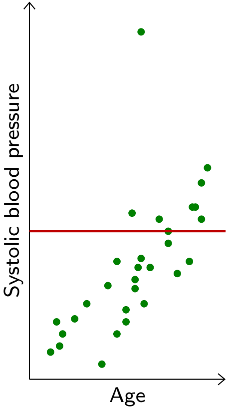
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.



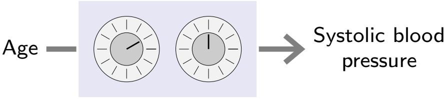


The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.

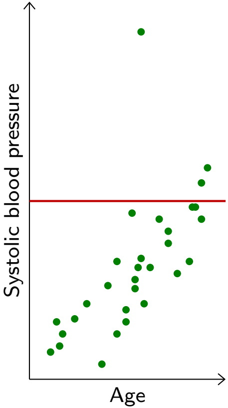




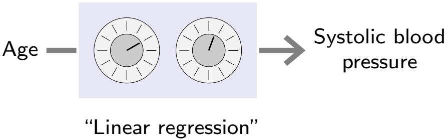
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.

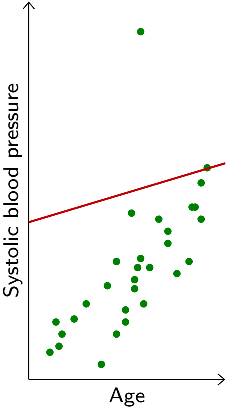


“Linear regression”

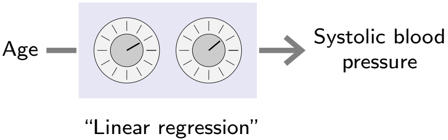


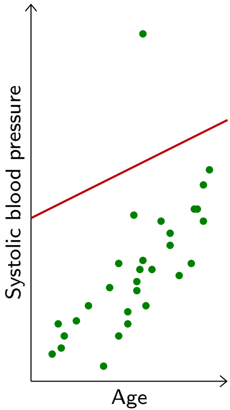
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.





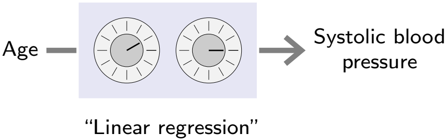
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.

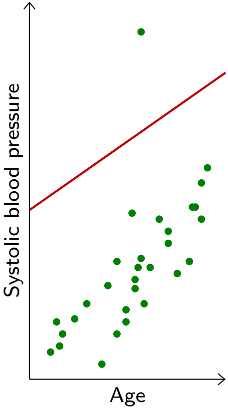




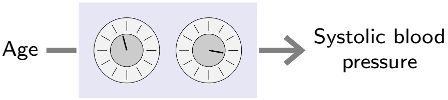
AI “programs itself”

The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.

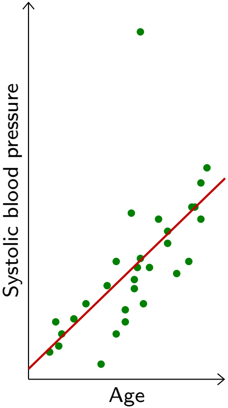




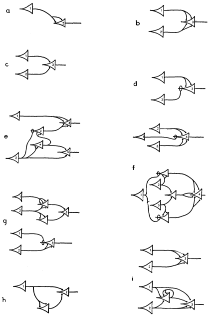
The fundamental idea of machine learning is to automatically tune a program to make it work well on known examples.



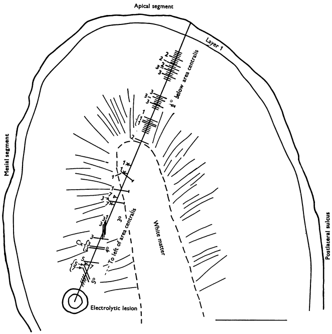
"Linear regression"



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

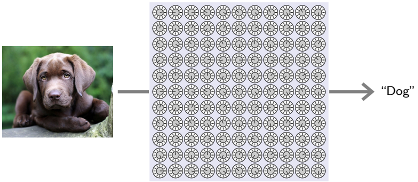


(McCulloch and Pitts, 1943)

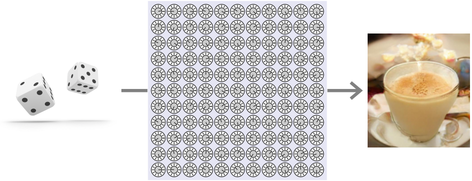


(Hubel and Wiesel, 1962)

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

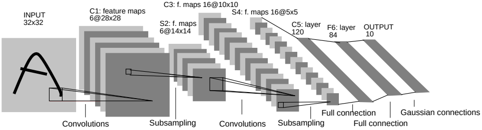


or to synthesize a complex signal



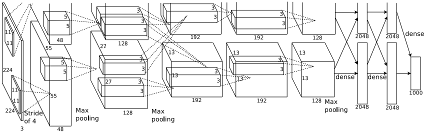
AI “programs itself”

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

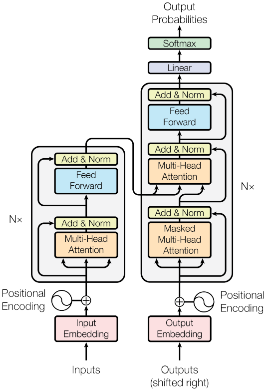


LeNet (1989)

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

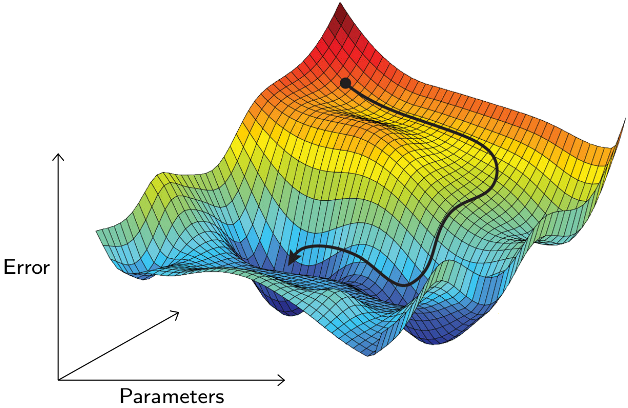


AlexNet (2012)



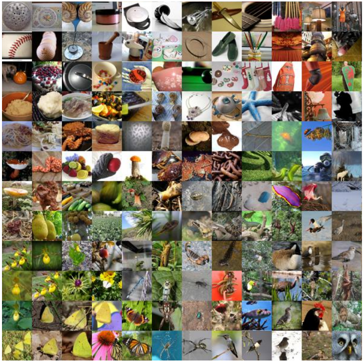
Transformer (2018)

Training an AI model consists of very progressively modifying its parameters to reduce its error on the training examples, so that performance on unseen examples will follow.

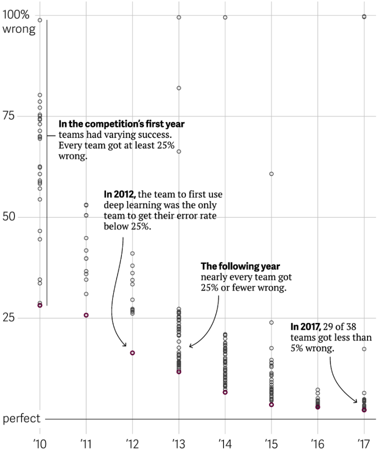


AI actually works

AI actually works

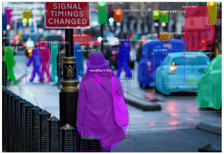


ImageNet

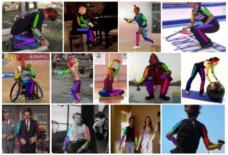


(Gershgorn, 2017)

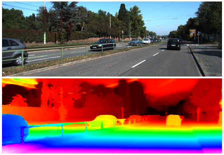
AI actually works



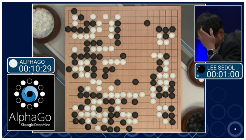
Scene understanding



Pose estimation



Geometry estimation



Goal planing

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

AI actually works

Human-level performance :

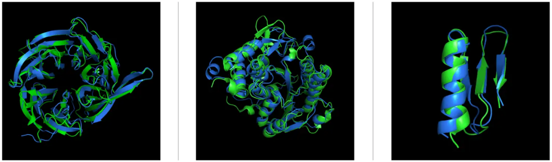
- Skin cancer detection.
- Speech processing.

Super-human performance :

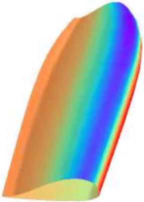
- Image recognition.
- Road sign detection.
- Reconnaissance de visages.
- Go and chess ("from first principles"), poker.
- Video games from the 80s.

AI actually works

Protein folding



Shape optimization



Bike designed with artificial intelligence breaks world speed records

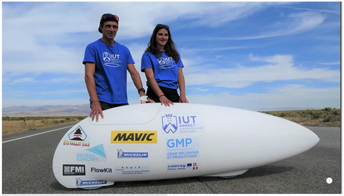


Image / video synthesis

Text synthesis

“**Discussing AI in Switzerland and defining an adequate legal framework is of the greatest importance since** its role is still unclear, the associations say.

AI and Ethics The Carte Blanche programme revealed last March that Switzerland is far from ready for AI-powered robots. The Senate is due to decide on legislation for these systems in 2019.

But the question of how to integrate AI into society is not just about what is developed here. To what extent should AI technology be sold or shared? What kinds of responsibilities should AI systems have?”

AI requires vast amounts of data and computation

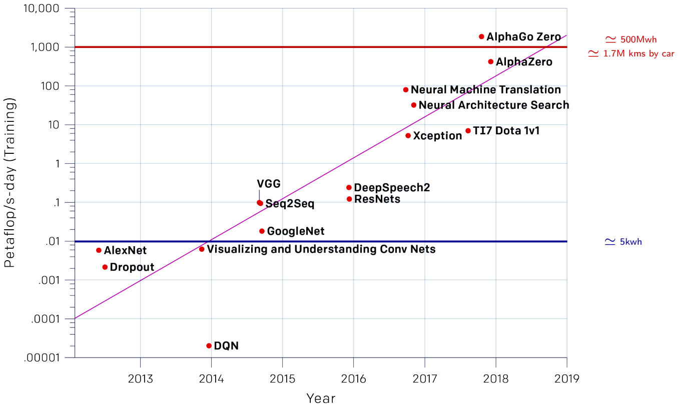
AI requires vast amounts of data and computation

The last decade of progress in AI corresponds to a vast increase of the "training sets" sizes. The most successful deployed methods rely on human-labeled data.

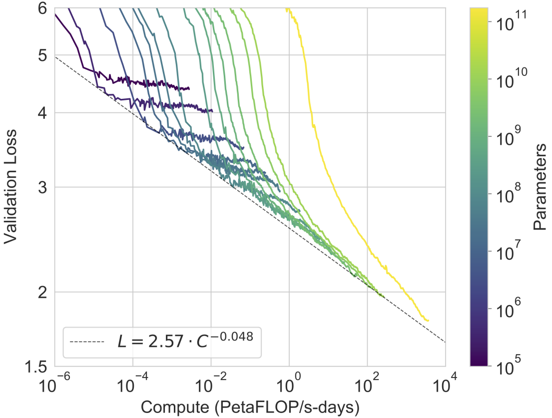
Data-set	**Year**	**Nb. images**	**Size**
MNIST	1998	60K	12Mb
Caltech 256	2007	30K	1.2Gb
ImageNet	2012	1.2M	150Gb
JFT-300M	2017	300M	36Tb (?)

Data-set	**Year**	**Nb. books**	**Size**
SST2	2013	40K	20Mb
WMT-18	2018	14M	7Gb
OSCAR	2020	12B	6Tb

A \$1'500 mass-market device posses 10'500 computing cores and can make $\simeq$ 35'000 billions operations per second. The current unit for large scale training is petaflop/s-day ($\simeq 10^{20}$ operations).



The trend does not seem to slow down:



(Brown et al., 2020)

AI is easy to deploy

AI is easy to deploy

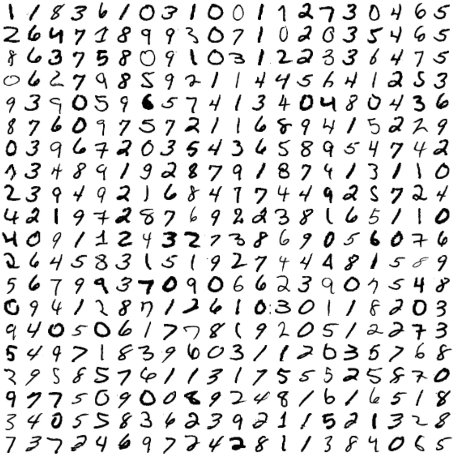
Deep-learning development is usually done in an open-source framework:

Framework	**Main backer**
PyTorch	Facebook
TensorFlow	Google
JAX	Google
MXNet	Amazon

Installation can be done with a single command:

```
conda install pytorch torchvision torchaudio cudatoolkit=10.2 -c pytorch
```

MNIST



(leCun et al., 1998)

AI is easy to deploy

①
```
model = nn.Sequential(
    nn.Conv2d( 1, 32, 5), nn.MaxPool2d(3), nn.ReLU(),
    nn.Conv2d(32, 64, 5), nn.MaxPool2d(2), nn.ReLU(),
    nn.Flatten(),
    nn.Linear(256, 200), nn.ReLU(),
    nn.Linear(200, 10)
)
```

②
```
criterion = nn.CrossEntropyLoss()

optimizer = torch.optim.SGD(model.parameters(), lr = 1e-2)
```

③
```
for e in range(nb_epochs):
    for input, target in data_loader_iterator(train_loader):
        output = model(input)
        loss = criterion(output, target)
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
```

Training $<$10s, error $\simeq$1%

AI is easy to deploy

```
alexnet = torchvision.models.alexnet(pretrained = True).eval()
output = alexnet(img)
```

AI is easy to deploy

```
alexnet = torchvision.models.alexnet(pretrained = True).eval()
output = alexnet(img)
```

```
#1 (12.26) Weimaraner
#2 (10.95) Chesapeake Bay retriever
#3 (10.87) Labrador retriever
#4 (10.10) Staffordshire bullterrier, Staffordshire bull terrier
#5 (9.55) flat-coated retriever
#6 (9.40) Italian greyhound
#7 (9.31) American Staffordshire terrier, Staffordshire terrier
#8 (9.12) Great Dane
#9 (8.94) German short-haired pointer
#10 (8.53) Doberman, Doberman pinscher
```

AI is easy to deploy

```
alexnet = torchvision.models.alexnet(pretrained = True).eval()
output = alexnet(img)
```

```
#1 (12.26) Weimaraner
#2 (10.95) Chesapeake Bay retriever
#3 (10.87) Labrador retriever
#4 (10.10) Staffordshire bullterrier, Staffordshire bull terrier
#5 (9.55) flat-coated retriever
#6 (9.40) Italian greyhound
#7 (9.31) American Staffordshire terrier, Staffordshire terrier
#8 (9.12) Great Dane
#9 (8.94) German short-haired pointer
#10 (8.53) Doberman, Doberman pinscher
```

Weimaraner

Chesapeake Bay retriever

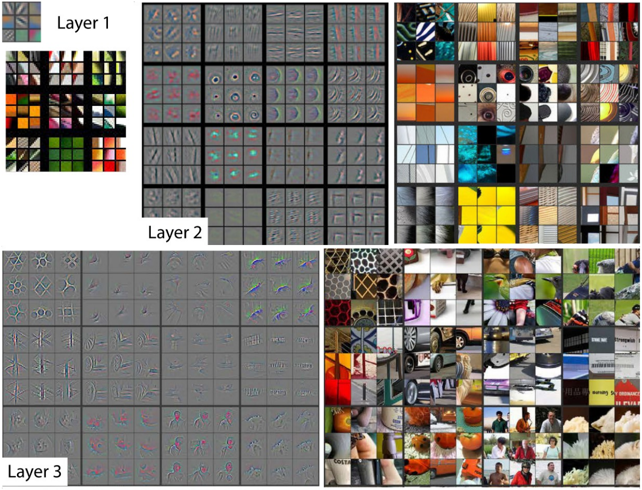
AI models are black boxes

Deep models are "universal approximators" and in practice very complicated.

The functioning of a trained model is only very partially understood.

Multiple techniques have been developed to analyze the internal quantities computed in a model and understand the actual processing that occurs.

AI models are black boxes



(Zeiler and Fergus, 2014)

AI models are black boxes

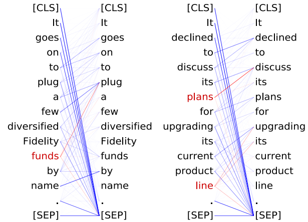
Original images

Guided back-propagation

AI models are black boxes

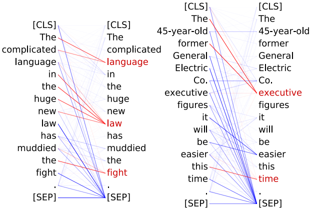
Head 8-10

- **Direct objects** attend to their verbs
- 86.8% accuracy at the `dobj` relation



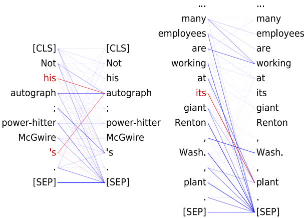
Head 8-11

- **Noun modifiers** (e.g., determiners) attend to their noun
- 94.3% accuracy at the `det` relation



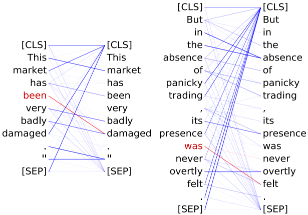
Head 7-6

- **Possessive pronouns** and apostrophes attend to the head of the corresponding NP
- 80.5% accuracy at the `poss` relation

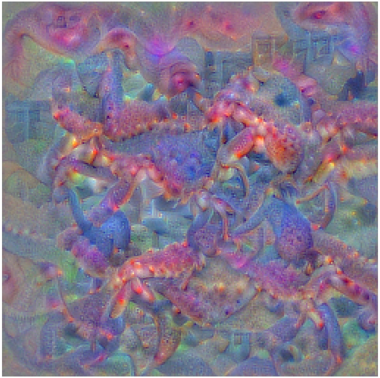


Head 4-10

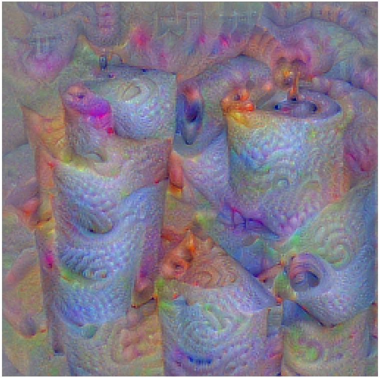
- **Passive auxiliary verbs** attend to the verb they modify
- 82.5% accuracy at the `auxpass` relation



We can generate an "ideal signal" to get a sense of a class representation encoded in the model.



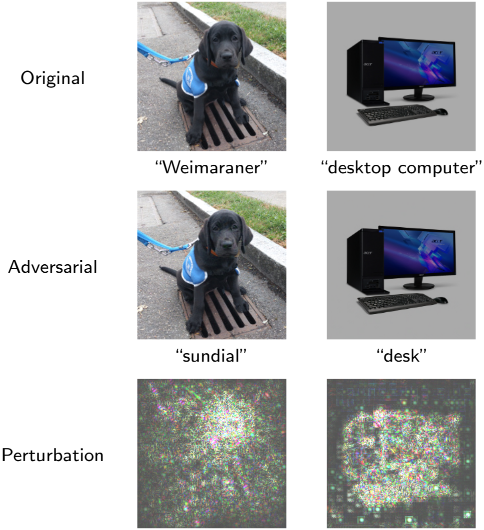
"King crab"



"Paper towel"

AI models are black boxes

Models are very sensitive to adversarial perturbations.



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.”

The end

References

T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. **Language models are few-shot learners**. CoRR, abs/2005.14165, 2020.

D. Gershgorn. **The data that transformed AI research—and possibly the world**, July 2017.

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.

Y. leCun, L. Bottou, Y. Bengio, and P. Haffner. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, 86(11):2278–2324, 1998.

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.