Demystifying AI

“Learn how to use neural networks with your own data”

A quest for information-driven applied science

HOGESCHOOL ROTTERDAM

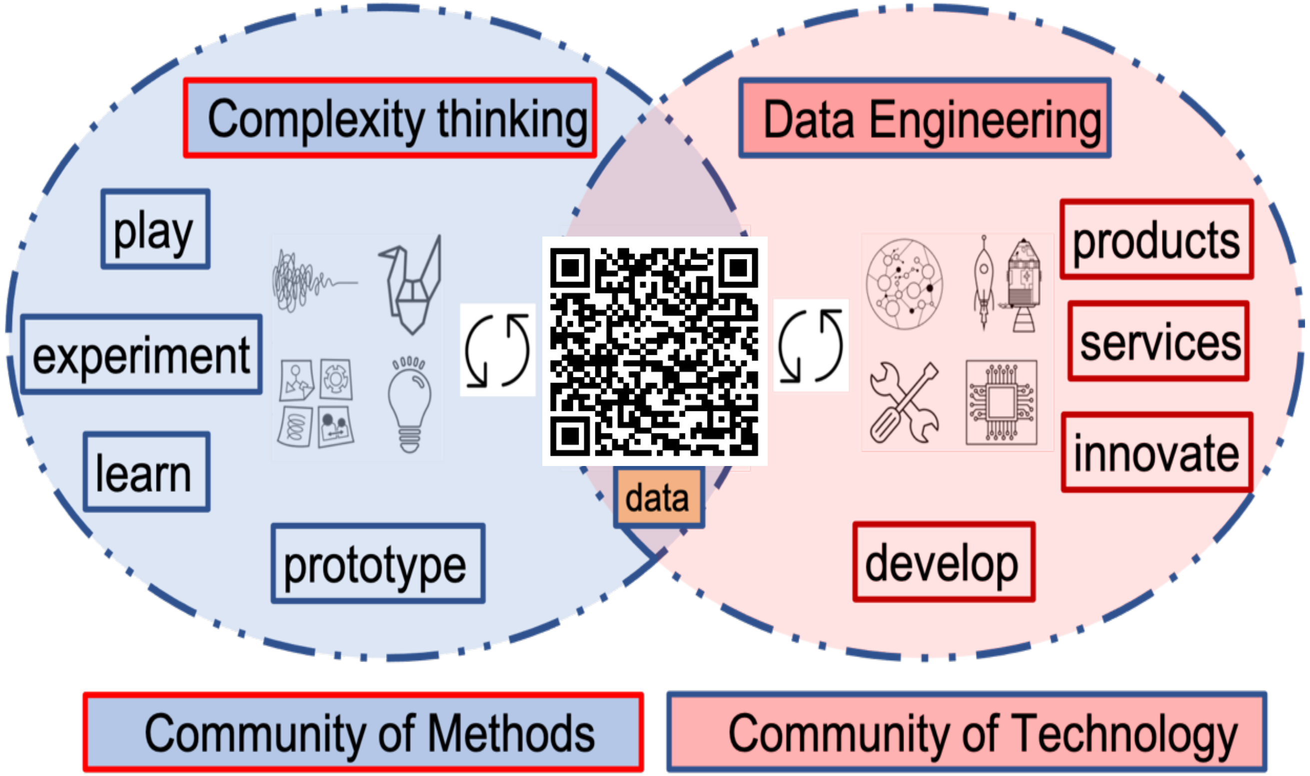
Lunch-Lezing CMD MEI 2022

Living-Lab: *AiRA,*
Hub voor AI: Responsible & Applied
Rob van der Willigen

{Envisioning a hands-on Data Lab}

PoP-UP Data Lab: PROMETHEUS

Opbouwen van kennis & expertise
door *hands-on* seminatie van datatechnologie



https://robfvdw.medium.com/a-generic-approach-to-data-driven-activities-d85ad558b5fa

This talk is about providing a
Biology-Inspired Data Science {DS}
framework of how to

(i) demystify &
(ii) make {AI} tangible

within a higher educational setting

https://robfvdw.medium.com/the-world-wide-web-ai-safari-b2e4f7f90647

{A working hypothesis}

HOGESCHOOL ROTTERDAM

Barn owls: why give a hoot?

by Kara Rosania

SCIENTIFIC NAME
Tyto alba

TAXONOMY
PHYLUM: Chordata
CLASS: Aves
ORDER: Strigiformes
FAMILY: Tytonidae

npg

Physical description

The barn owl is known for its striking appearance: two almond-shaped black eyes with distinctive brown streaks in the inner corners are surrounded by a snowy white, heart-shaped face. The white feathers of the face and undersides of the body and wings contrast sharply with the orangey-brown color of the head, back, upper wings and tail. Their long, rounded wings and short tails give them a uniquely buoyant, loping flight style. This flight is noiseless, aided by the soft plumage that helps to muffle the sound of their feathers when flying, ensuring them a silent approach when honing in on their prey[1]. Measuring 29–44 cm in length and weighing 250–480 g, barn owls are short-lived, with an average life expectancy of 1–2 years in the wild. Females are often heavier than males and are somewhat showier, with a more reddish color and a more heavily spotted chest.

Research résumé

Strictly nocturnal, barn owls hunt by flying slowly over open fields at night or dusk. They specialize in hunting small ground mammals, and the vast majority of their food consists of small rodents such as voles, pocket gophers, shrews, mice and rats. The barn owl has excellent low-light vision and can easily find prey at night by sight, but these birds search for prey primarily using sound localization. The barn owl's ability to locate prey by sound alone is the best of any animal that has ever been tested and rivals that of humans[2]. Barn owls lack ear tufts, and their ears are asymmetrically positioned on their head, allowing them to use the interaural time difference, or the difference in the arrival time of the sound at each ear, to determine the position of a sound source in space[3]. Additionally, the satellite-dish shape of a barn owl's face helps it to gather and amplify sounds from its surroundings.

Neuroethologist Masazaku Konishi was the first to speculate that the owl brain might contain a map of auditory space, constructed on the basis of information about the arrival times, intensity and frequency of sounds perceived by each ear to construct a map of the location of sound sources. This would allow the owls to direct their strikes accurately towards prey making noise but hidden from sight[4]. In the late 1970s, Knudsen and Konishi localized this auditory space map to a region of the owl's brain called the inferior colliculus[5]. Further studies of the owl's inferior colliculus have shown that the development of the auditory space map is highly dependent on early auditory experiences[6] and that it shifts adaptively as the locations of sound sources move through space[7].

Kim Caesar/Nature Publishing Group

The barn owl also provides an exceptional model to study stereovision, or the ability to see in three dimensions, because it displays one of the highest degrees of binocular specialization. Like humans, owls have two frontally placed eyes. As a result, owls can simultaneously compare images in the left and right eye in order to discriminate between objects and background[8]. Interestingly, the owl's visual experience also plays a crucial role in the formation and maintenance of the auditory space map by increasing auditory responses to sounds that are accompanied by visual information[9].

1. Jaworski, J. & Peake, N. Vortex noise reductions from a flexible fiber model of owl down. 66th Annual Meeting of the American Physical Society's (APS) Division of Fluid Dynamics. 24–26 November 2013, Pittsburgh, PA.
2. Knudsen, E.I. Instructed learning in the auditory localization pathway of the barn owl. *Nature* **417**, 322–328 (2002).
3. Fischer, B.J., Christianson, G.B. & Peña, J.L. Cross-correlation in the auditory coincidence detectors of owls. *J. Neurosci.* **28**, 8107–8115 (2008).
4. Auditory processing, plasticity, and learning in the barn owl. *ILAR J.* **51**, 338–352 (2010).
5. Knudsen, E.I. & Konishi, M. A neural map of auditory space in the owl. *Science* **200**, 795–797 (1978).
6. Efrati, A. & Gutfreund, Y. Early life exposure to noise alters the representation of auditory localization cues in the auditory space map of the barn owl. *J. Neurophysiol.* **105**, 2522–2535 (2011).
7. Witten, I.B., Bergan, J.F. & Knudsen, E.I. Dynamic shifts in the owl's auditory space map predict moving sound location. *Nat. Neurosci.* **9**, 1439–1445 (2006).
8. van der Willigen, R.F. Owls see in stereo much like humans do. *J. Vis.* **11**, 10 (2011).
9. Bergan, J.F. & Knudsen, E.I. Visual modulation of auditory responses in the owl inferior colliculus. *J. Neurophysiol.* **101**, 2924–2933 (2009).

Sensory Specialist like Owls, Monkeys and Humans can teach data scientists:

"How to make the application of deep neural networks more intuitive *{high Affordance}*, *less* labor-intensive & costly"!

[Barn owls: why give a hoot? | Lab Animal (nature.com)]

WHY DESMISTFYING AI?

To make sense of the world

“Sense-making is the way that humans choose between multiple possible explanations of sensory input.”

– Dave Snowden

http://kwork.org/Stars/Snowden/snowden3.html#Simplicity

"The inescapable resurgence of {AI}
on the world wide web {WWW}
— along with the arrival of Internet-of-Things {IoT} —
has expanded the scope of the
digital world into the realm of cybernetics"

Cybernetics studies communication & control of information in living beings + the machines built by humans

======> Feedback & Reinforcement <======

The cybernetic foundation of {AI} explains its insatiable hunger for Big data, with the promise to solve societal challenges ranging from:

Health - Climate Change - Safety
up to Cyber Physical Systems {CPSs}:
Robotics & Driverless Cars

Today it seems we only receive ambiguous promises and paradoxical stories

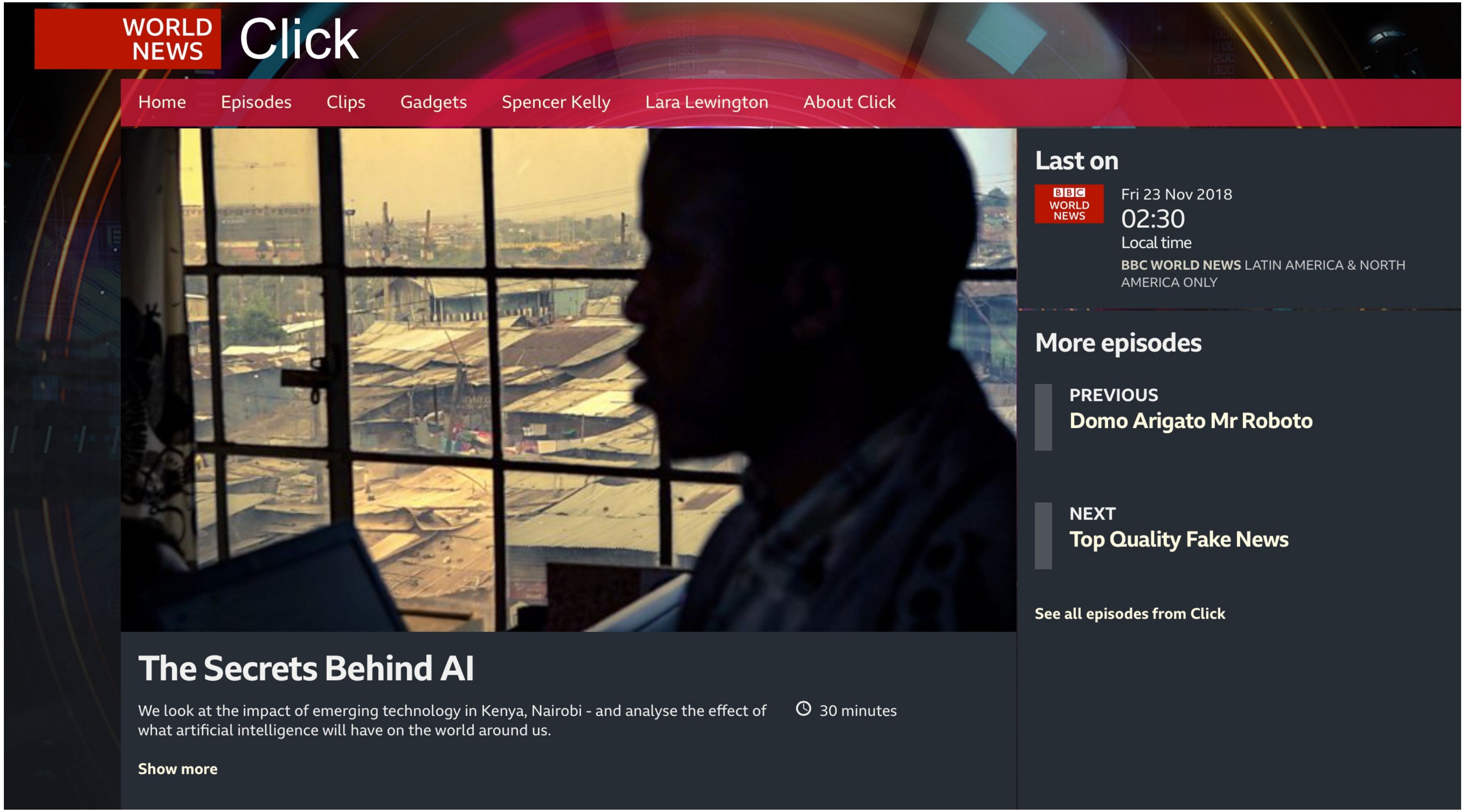
{AI} has revealed itself to us as a double-edged sward:

Dangerous *yet* Supportive
All-Consuming *yet* Liberating

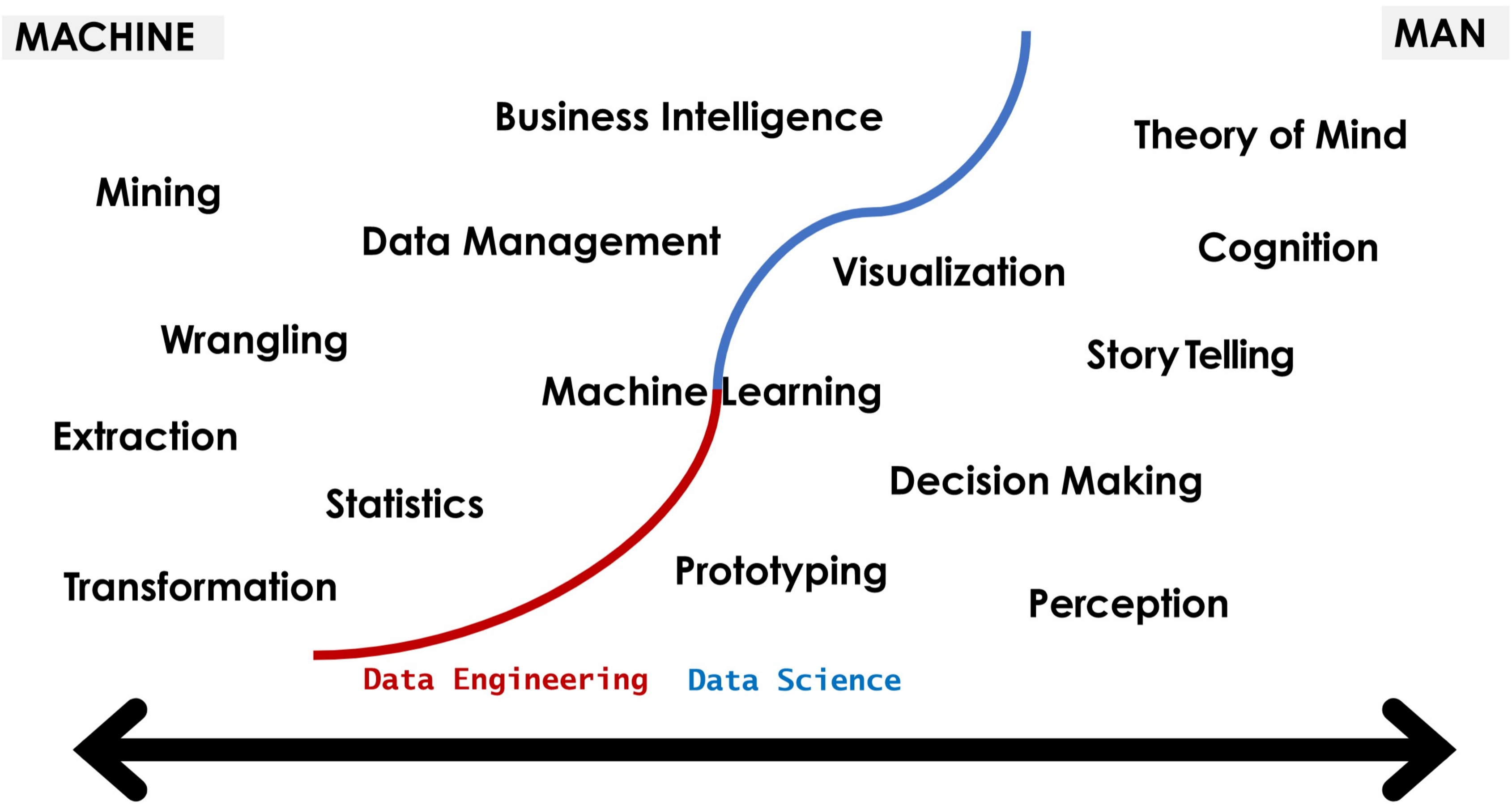
{AI} is dominated by labour intensive, wasteful & costly Brute-Force practices using neural networks

{The Secrets Behind AI}

HOGESCHOOL ROTTERDAM



{AI integrates two Scientific Disciplines}

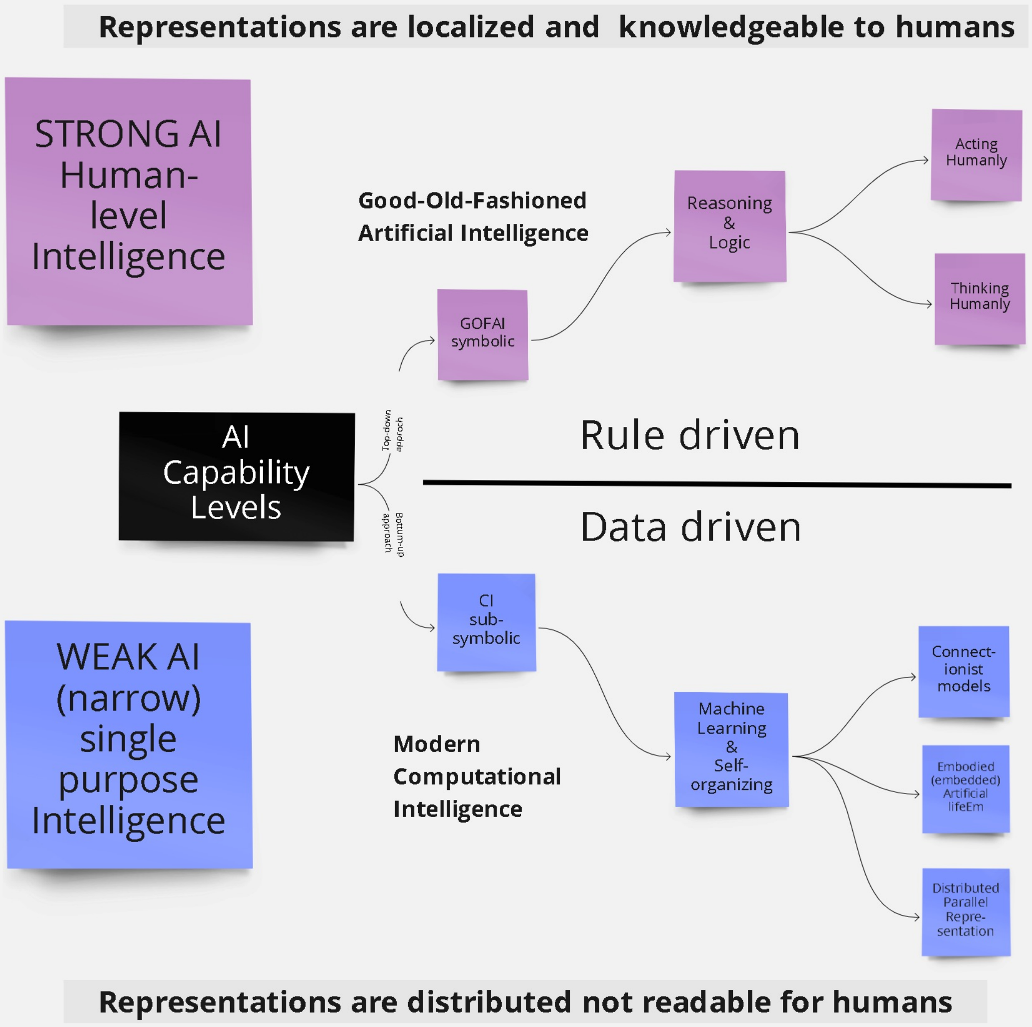


Inspired by Daniel Keim, "Visual Analytics: Definition, Process, and Challenges"

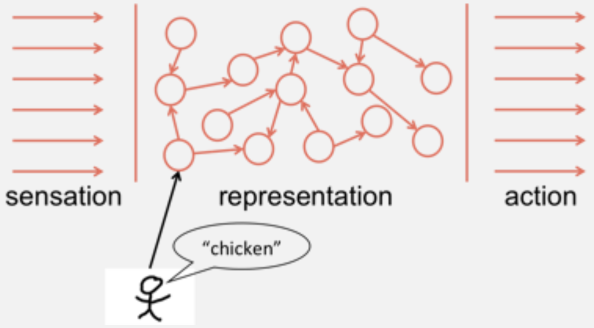
https://robfvdw.medium.com/a-generic-approach-to-data-driven-activities-d85ad558b5fa

{The taxonomy of AI is complex}

HOGESCHOOL ROTTERDAM

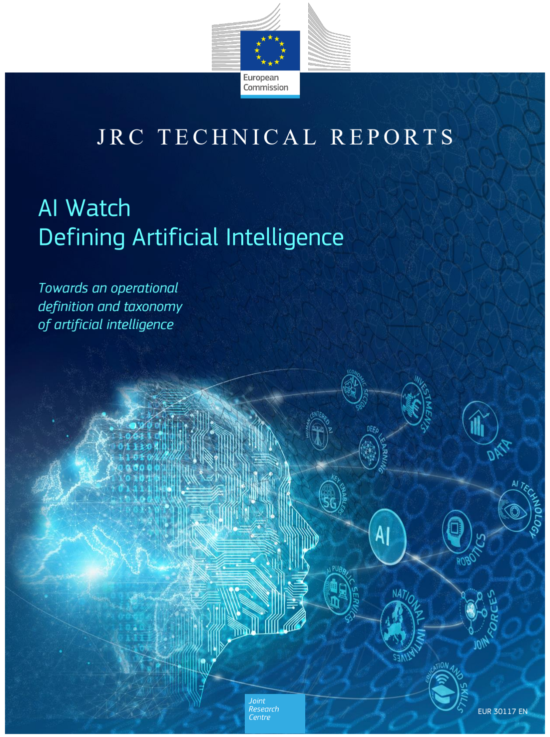


Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.



{The taxonomy of AI is complex}

HOGESCHOOL ROTTERDAM



https://publications.jrc.ec.europa.eu/repository/handle/JRC118163

	AI taxonomy	
	AI domain	**AI subdomain**
Core	**Reasoning**	Knowledge representation
		Automated reasoning
		Common sense reasoning
	Planning	Planning and Scheduling
		Searching
		Optimisation
	Learning	Machine learning
	Communication	Natural language processing
	Perception	Computer vision
		Audio processing
Transversal	**Integration and Interaction**	Multi-agent systems
		Robotics and Automation
		Connected and Automated vehicles
	Services	AI Services
	Ethics and Philosophy	AI Ethics
		Philosophy of AI

http://dx.doi.org/10.2760/382730%20

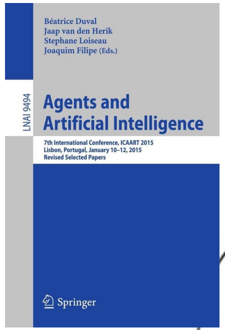
{AI favors Agent-Based Models}

HOGESCHOOL ROTTERDAM

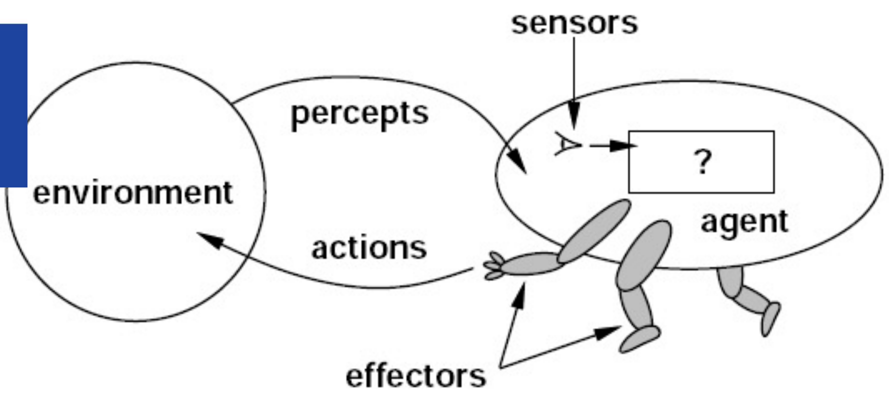
Agents

are abstractions of the real world {models} that can perceive their environment through sensors (input) and act upon that environment through effectors (output), combined with learning capabilities.

As a result, agent behaviour is desirable from an AI-viewpoint



Model



Real-world

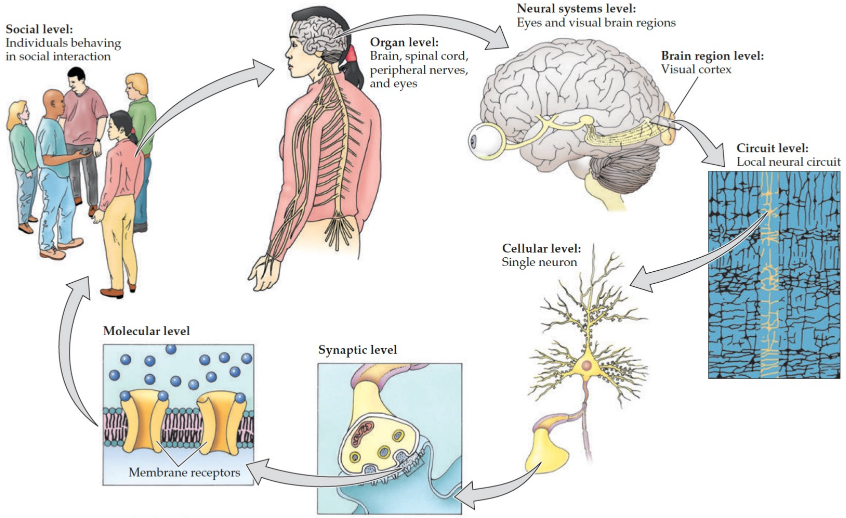
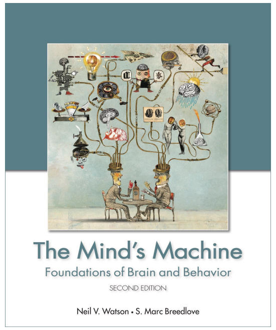


FIGURE 1.10 Levels of Analysis in Biological Psychology

Animal research is an essential part of life sciences research, including biological psychology



https://pdfs.semanticscholar.org/0629/51ac18d311efbfb4a019b1b5c513f1208de5.pdf

{Brute Force}

HOGESCHOOL ROTTERDAM

CREATING

State-of-the-Art {SOTA} {AI} relies
on a brute-force approach:

High-performance GPU or TPU computing power

Black-Box, poorly understood solutions

Training with massive data-sets

Deep Neural Network architecture

Billions of parameters

Lengthy Process & Huge budgets

to produce cognitive abilities
that are on par with Human-Level Performance.

{Chat-Bot Meena}

https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html
https://arxiv.org/pdf/2001.09977.pdf

HOGESCHOOL ROTTERDAM

CREATING

Google research 2020:

"Meena is a conversational agent capable of chatting convincingly about any topic that is meaningful to humans"

HPC: 2,048 TPU v3 cores, 16GB DDR5 memory

trained on 40 billion words (61B BPE tokens) [341 GB]

Evolved Transformer NAS

32 decoder layers, deep neural network

2.6 [Billion] parameters

30 days of training, 4[Million] tokens per second

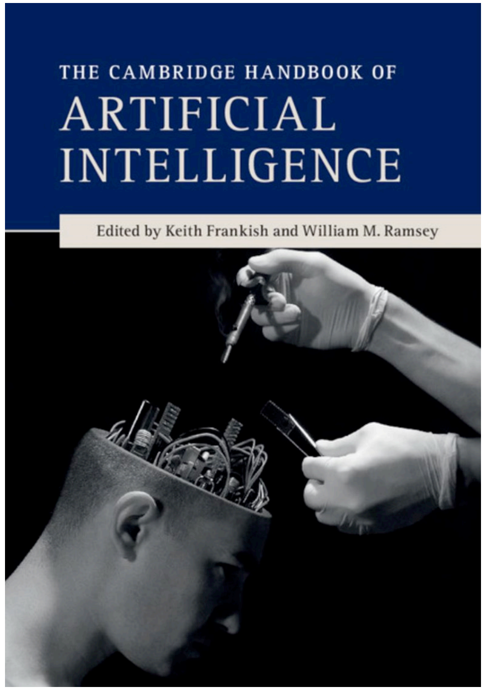
Training cost: 1,500,000$

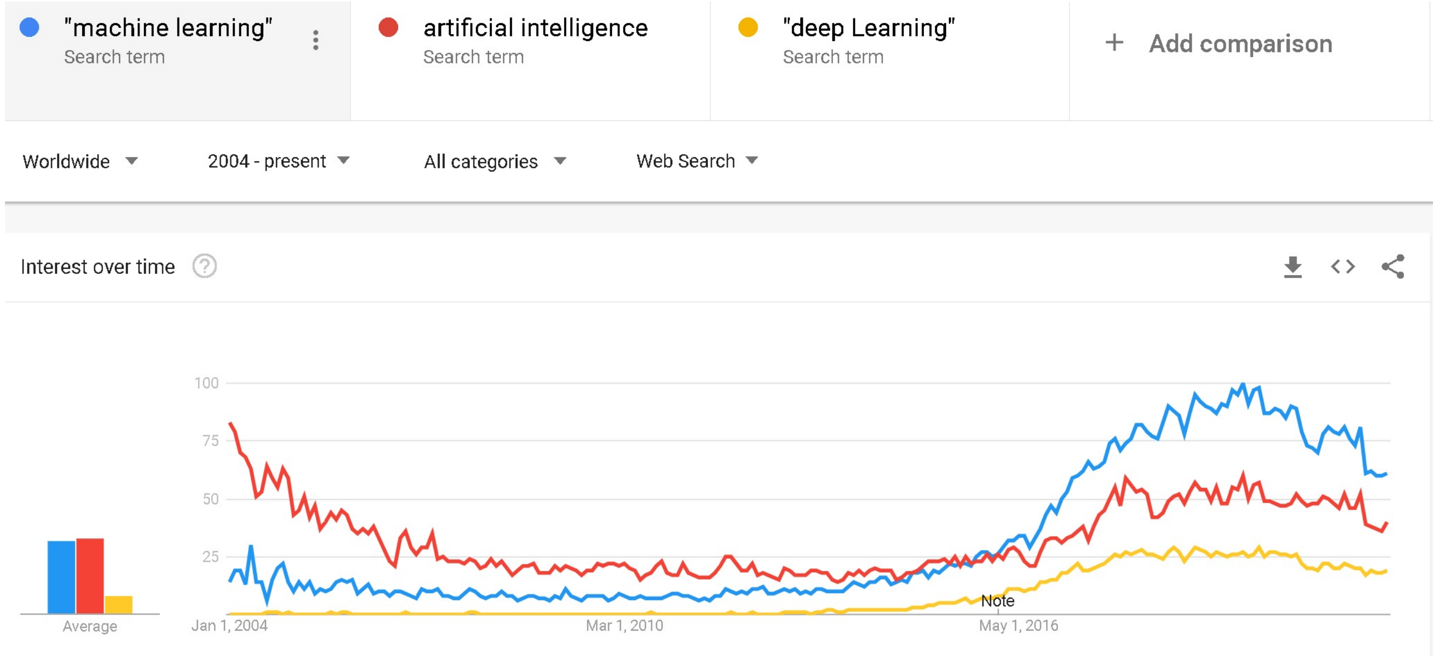
https://cloud.google.com/tpu/pricing

What constitutes a {Deep} Neural Net?

{DL}

(Pre-trained) Deep Learning {DL} Neural Networks {NNs} are the most advanced, successful & fastest growing Artificial Intelligence {AI} technology





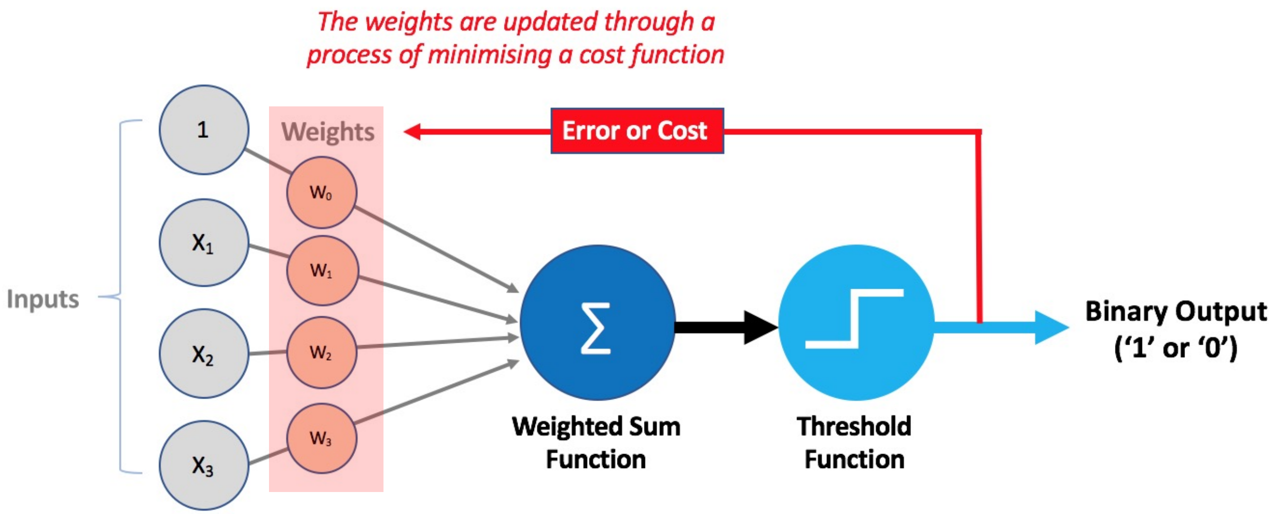
https://trends.google.com/trends/explore?date=all&q=%22machine%20learning%22,artificial%20intelligence,%22deep%20Learning%22

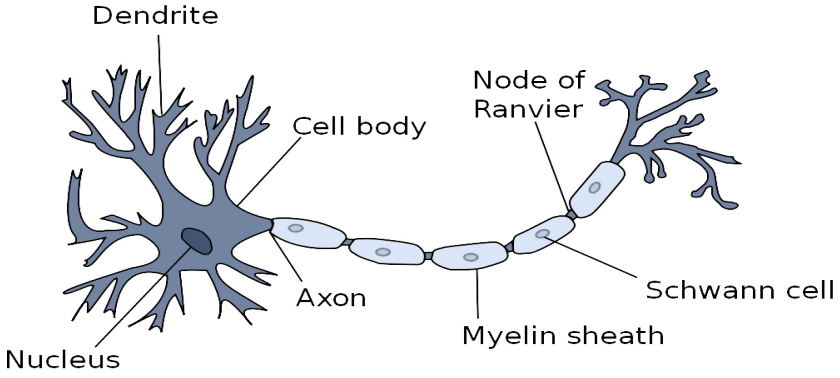
HOGESCHOOL ROTTERDAM
CREATING

{Artificial Neurons}

Deep Neural Nets {DNNs} harbor vast amounts of "artificial neurons" ➔smallest computational unit🡐

Names for
Artificial Neurons
{unit}
{cell}
{node}
{perceptron}



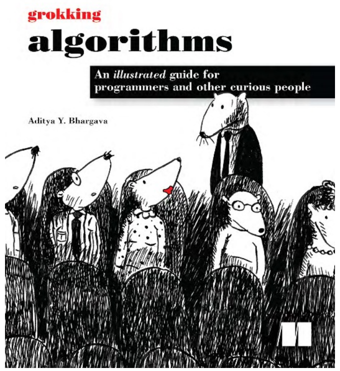


{Algorithm}

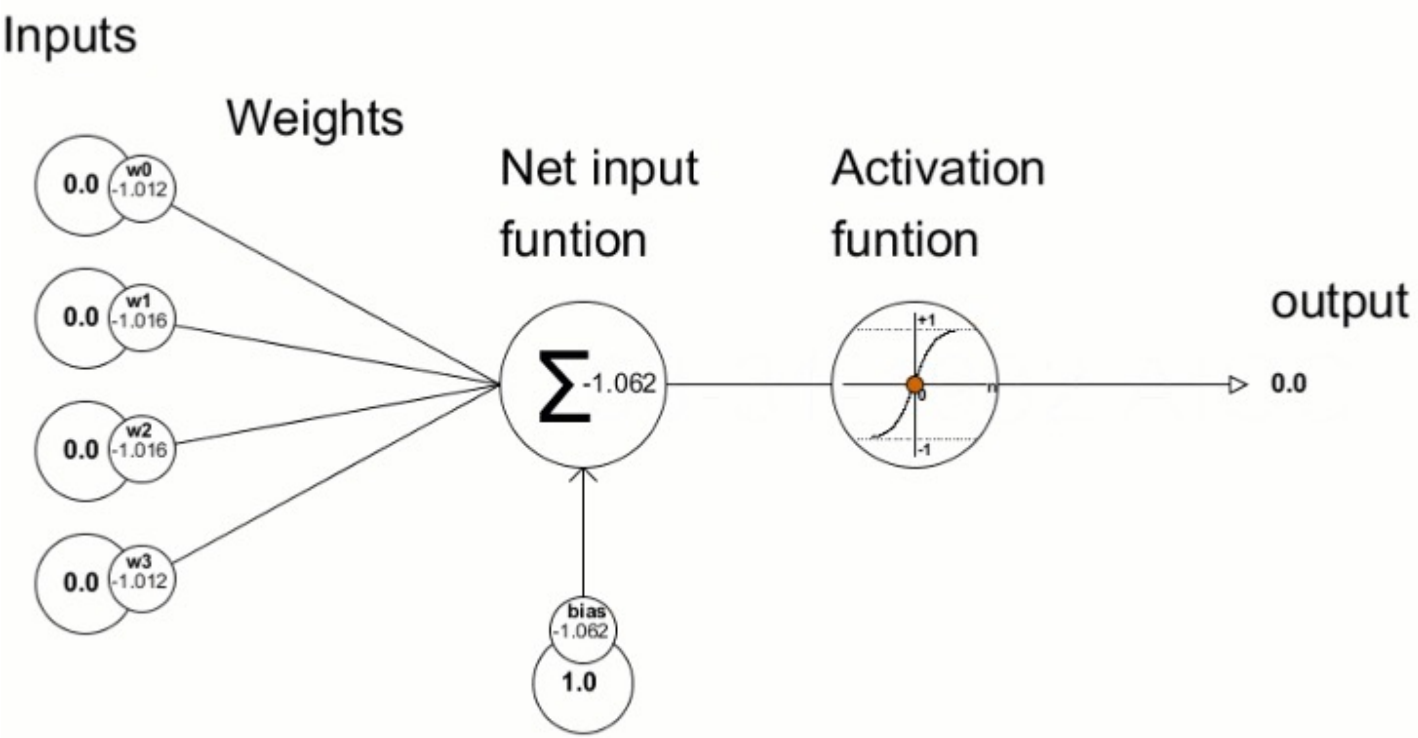
HOGESCHOOL ROTTERDAM

CREATING

Step by step **process or recipe** describing
how to **solve a problem** and/or **complete a task**,
which will always give
identical **end results**

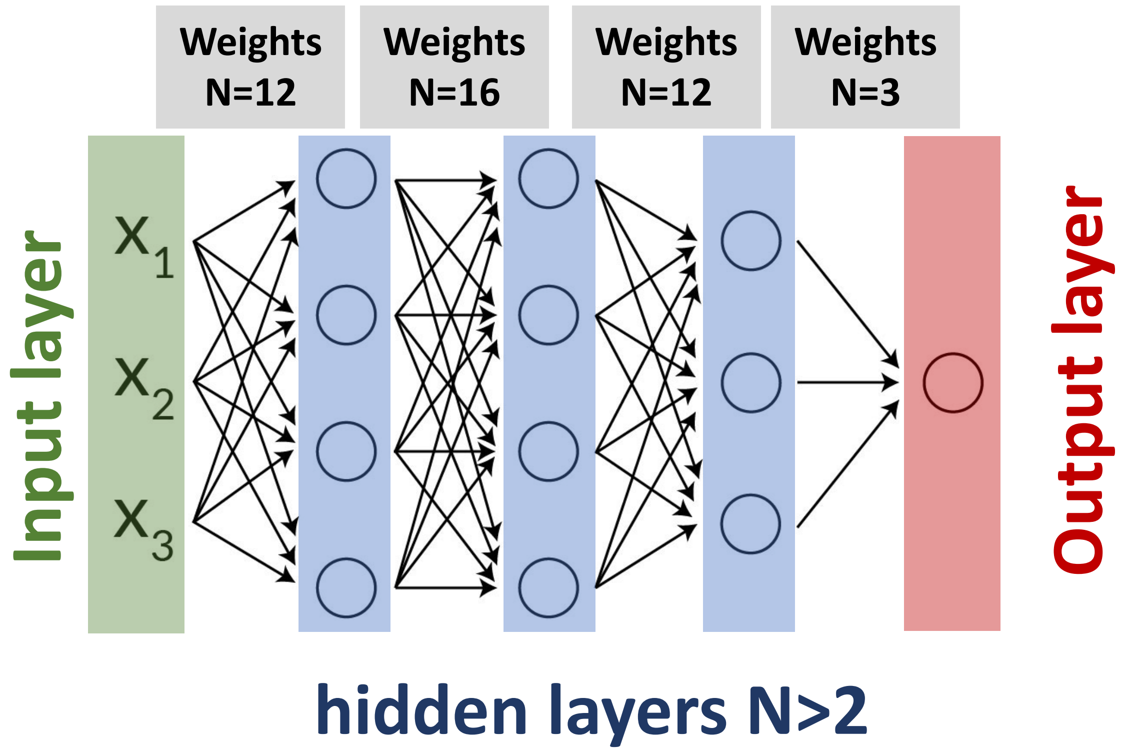


{Artificial Neurons}

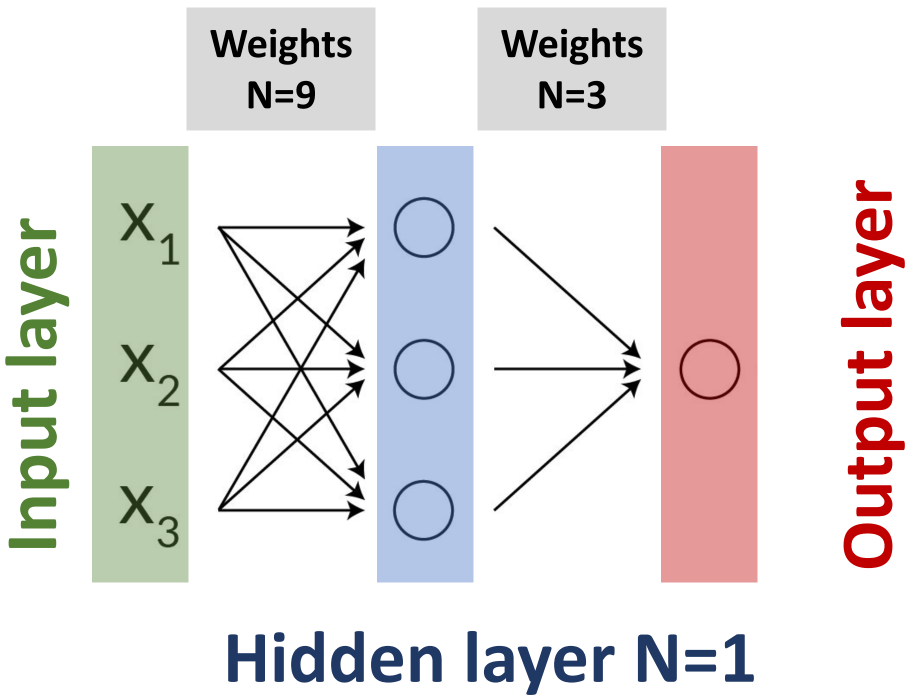


{NN Layers}

Neural Network {NN} Layer Architecture



Deep network

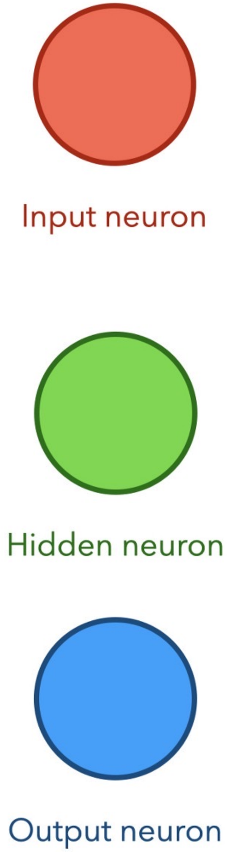


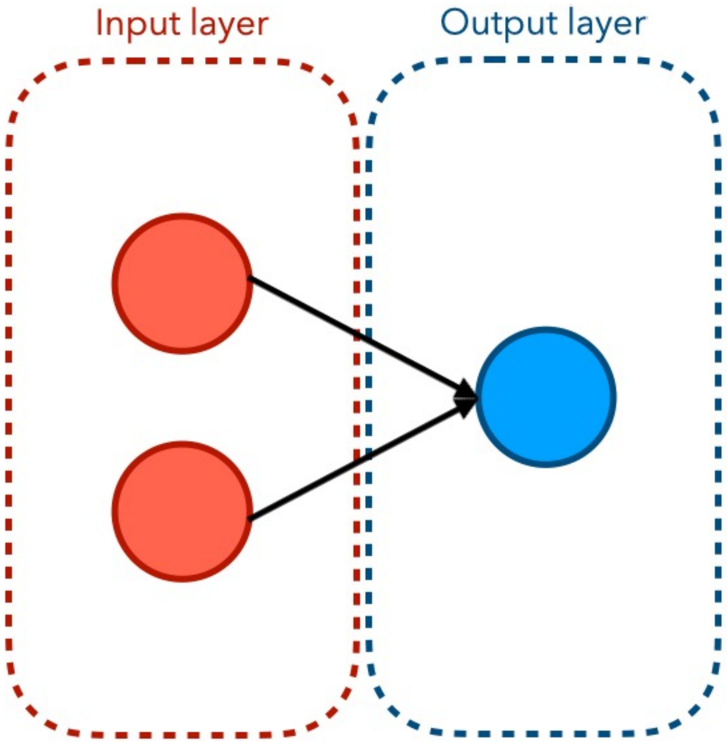
Shallow network

{NN Layers}

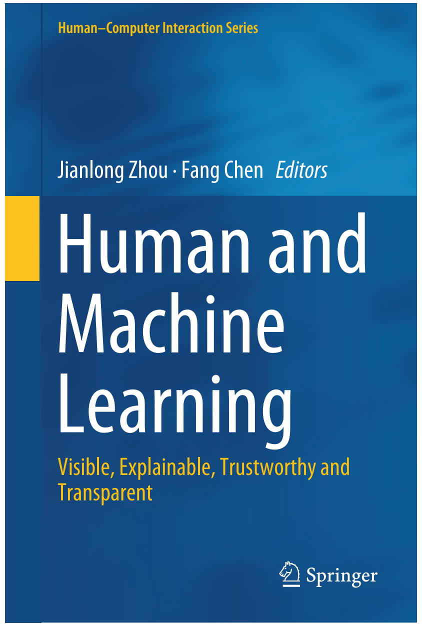
HOGESCHOOL
ROTTERDAM

Neural Network {NN} Layer Architecture





{Human-in-the-Loop}



$$AI = ML + TD + HITL$$

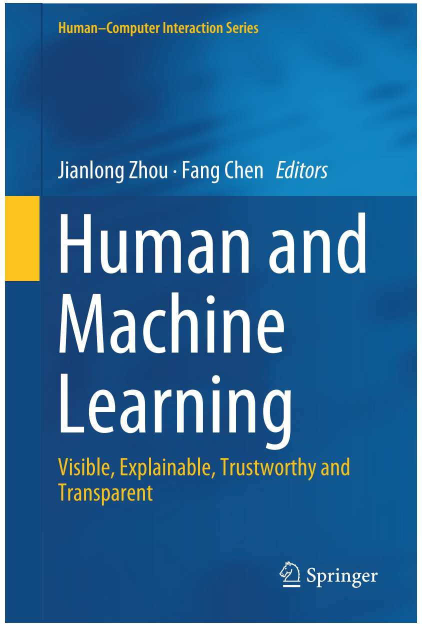
Artificial Intelligence: in contrast to natural intelligence, it is *the ability of computer systems to perform tasks or actions that would normally require a human*

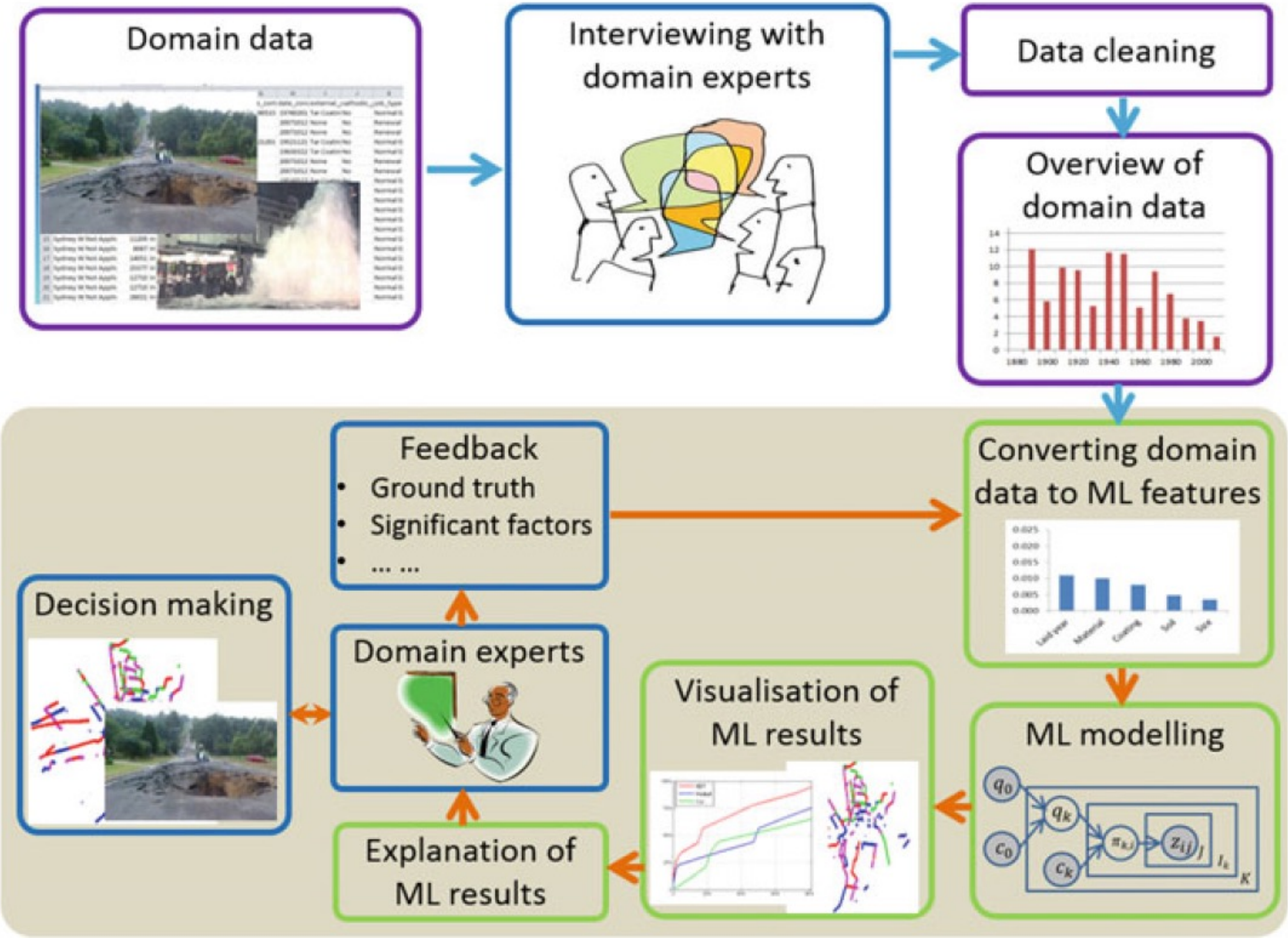
Machine Learning: *the ability of computer systems to use algorithms and statistical models to perform tasks without explicit instruction, through patterns and inferences*

Training Data: *the data used to train a machine learning algorithm to perform a task in supervised machine learning*

Human in the Loop: *the involvement of a human in training a machine learning algorithm*

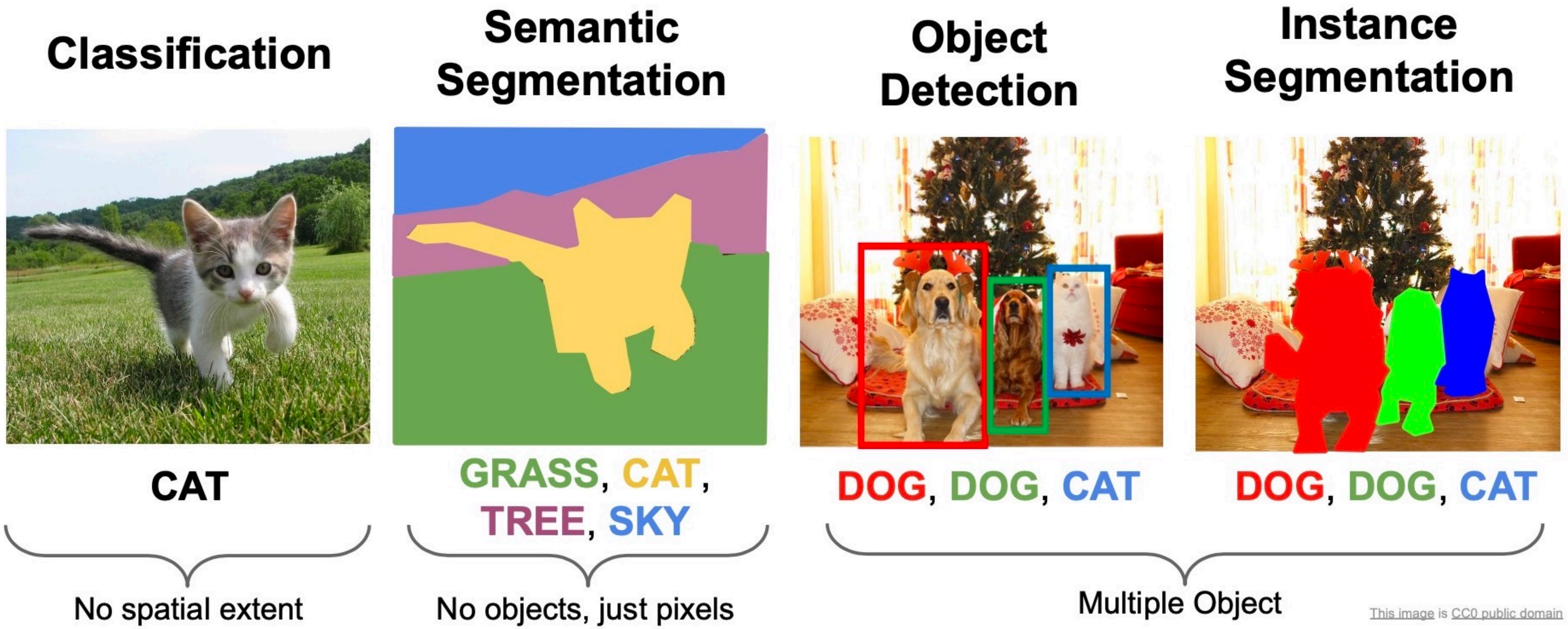
{Modeling}


Human–Computer Interaction Series
Jianlong Zhou · Fang Chen Editors
Human and Machine Learning
Visible, Explainable, Trustworthy and Transparent
Springer


Domain data
Interviewing with domain experts
Data cleaning
Overview of domain data
Converting domain data to ML features
ML modelling
Visualisation of ML results
Explanation of ML results
Domain experts
Feedback
• Ground truth
• Significant factors
• … …
Decision making

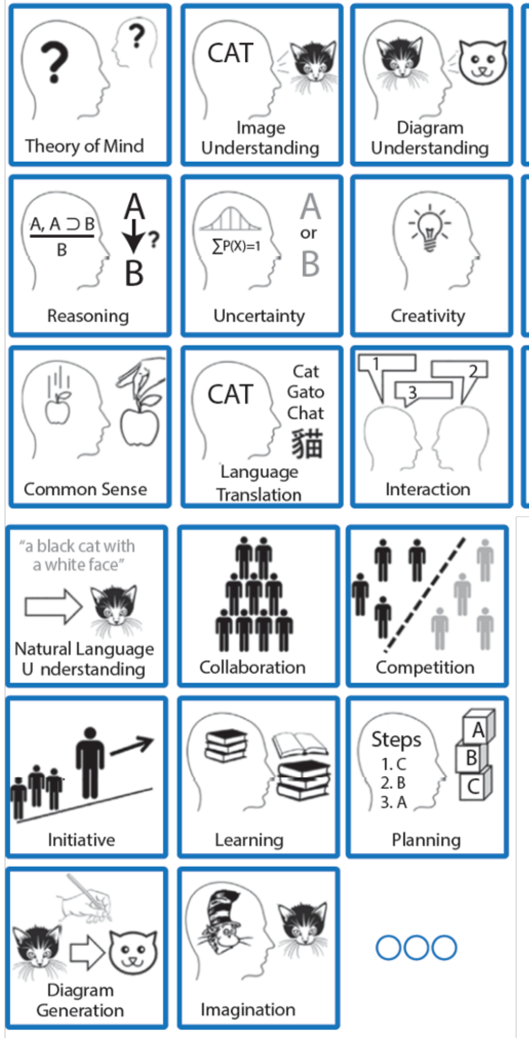
{Modeling}

HOGESCHOOL
ROTTERDAM

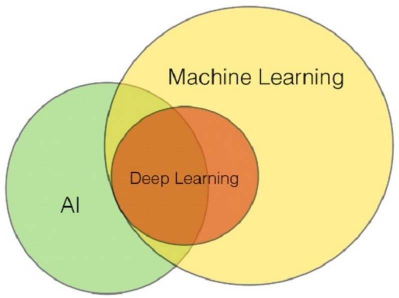


{AI=ML=DL}

https://doi.org/10.1609/aimag.v37i1.2643



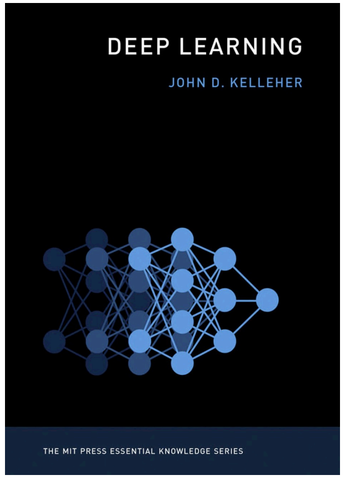
AI enabled through {DL} must be understood as any form of Machine Learning {ML} technology mimicking & automatizing tasks which otherwise require

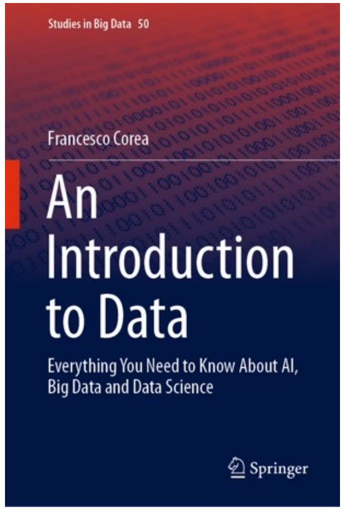


human perception, cognition and/or motor skills

https://robfvdw.medium.com/the-world-wide-web-ai-safari-b2e4f7f90647

{DL + DNNs}

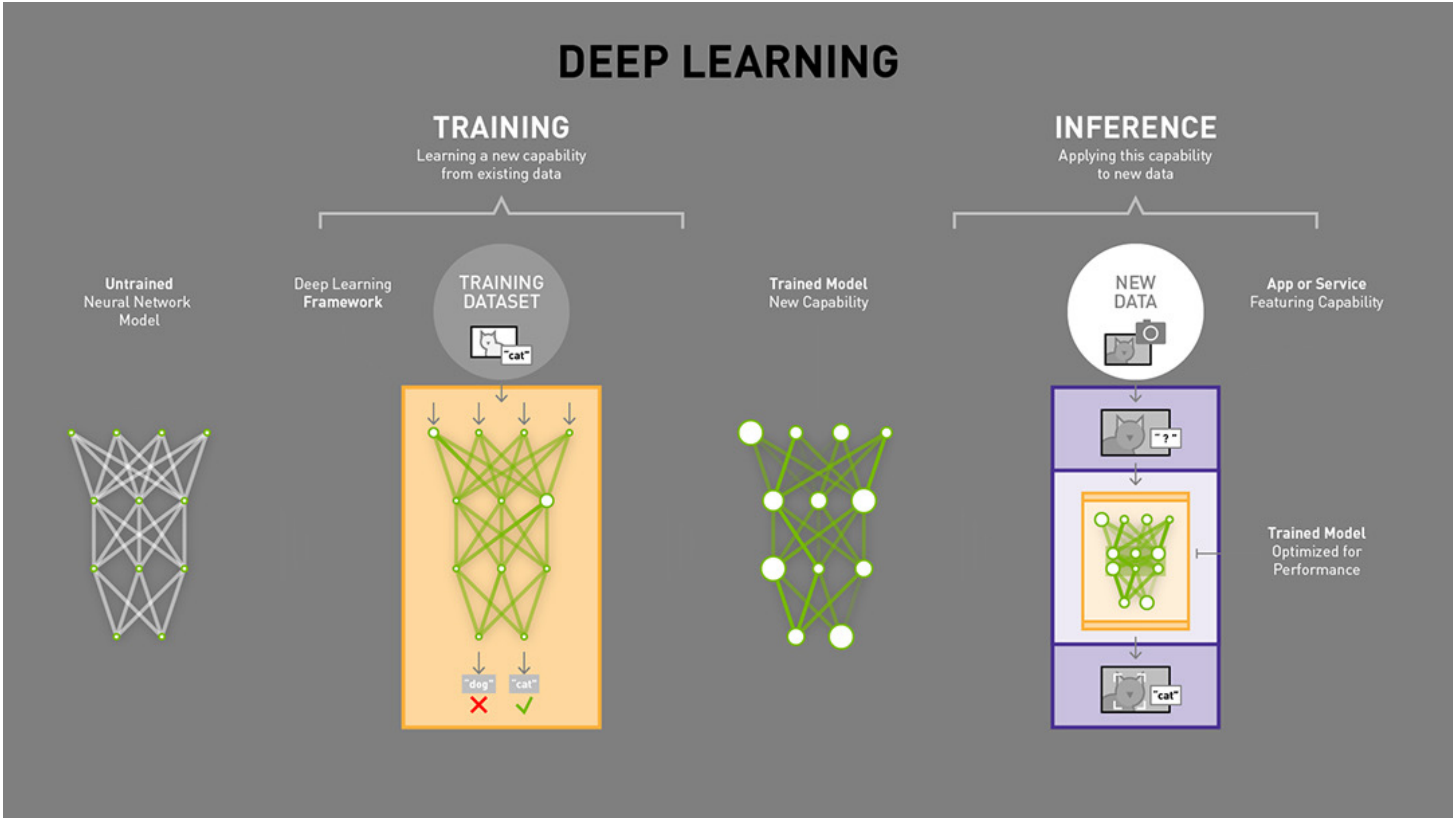




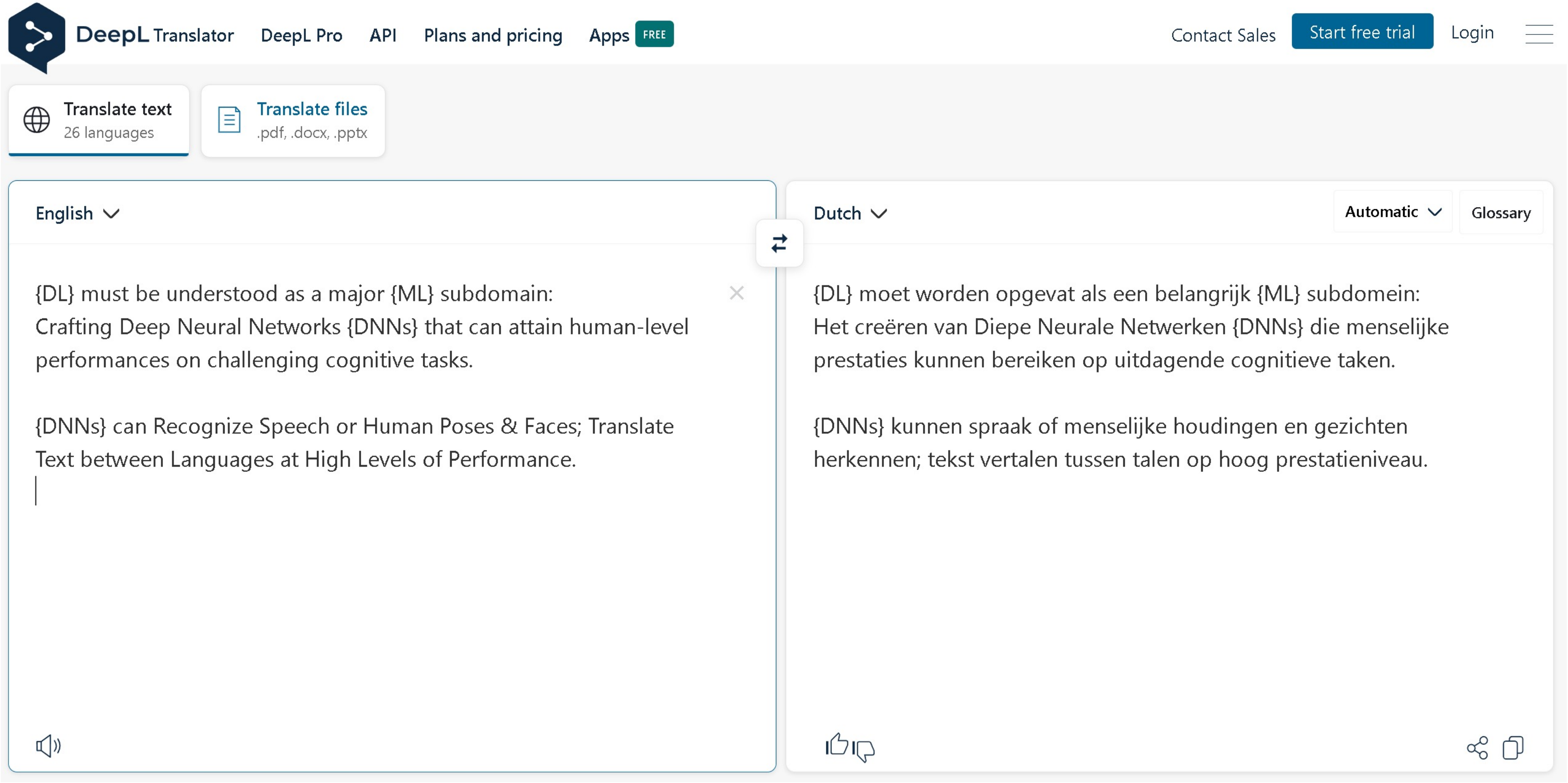
Deep learning {DL} must be understood as a major Machine Learning {ML} subdomain:

Crafting Deep Neural Networks {DNNs} that can attain human-level performances on challenging cognitive tasks.

{DNNs} can Recognize Speech or Human Poses & Faces; Translate Text in real time at High Levels of Performance.


DEEP LEARNING
TRAINING
Learning a new capability from existing data
INFERENCE
Applying this capability to new data
Untrained Neural Network Model
Deep Learning Framework
TRAINING DATASET
"cat"
"dog"
"cat"
Trained Model New Capability
NEW DATA
App or Service Featuring Capability
"?"
Trained Model Optimized for Performance
"cat"

{Human-level performance}



https://www.deepl.com/translator

{Human-level performance}

Probeer Speech to Text uit met deze demo-app, ontwikkeld op basis van onze JavaScript-SDK

Taal

Dutch (Netherlands)

Automatische interpunctie

Spreken

Bestand uploaden

Uw spraakgegevens worden niet opgeslagen

Ontdek hoe u Speech to Text in uw apps en producten gebruikt >

Verken meer aspecten van uw Speech to Text-uitvoer met het programma zonder code in Speech Studio >

Kies de knop Spreken aan de linkerkant en begin met spreken. De spraakservice retourneert herkenningsresultaten terwijl u spreekt. Als u verschillende talen spreekt, kunt u elke taal uitproberen die door de spraakservice wordt ondersteund. U kunt ook bestanden uploaden om de spraakservice voor uw specifieke gebruiksscenario's te testen. Raadpleeg onze documentatie en ontdek hoe u de spraak-naar-tekstfunctie in uw oplossingen inbouwt.

https://azure.microsoft.com/nl-nl/services/cognitive-services/speech-to-text/#features

{DL} represents an
{AI} breakthrough

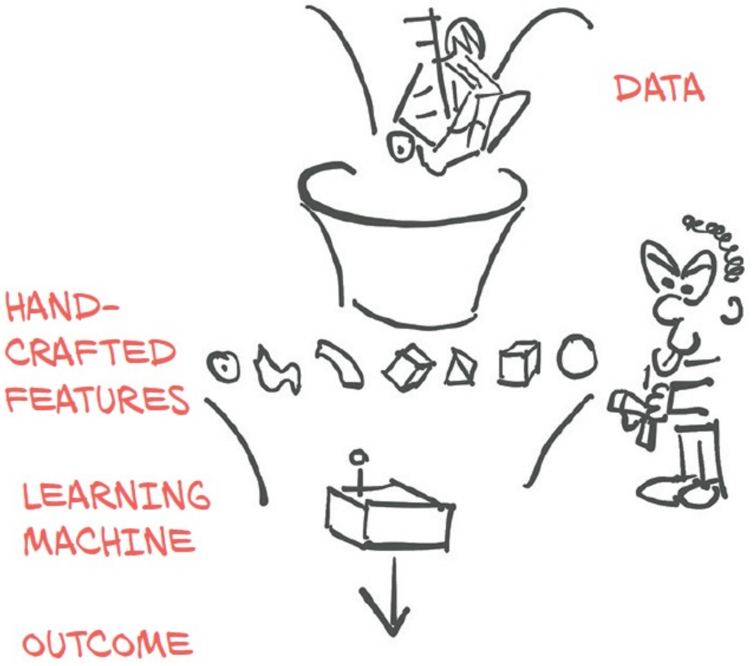
Paradigm-Shift

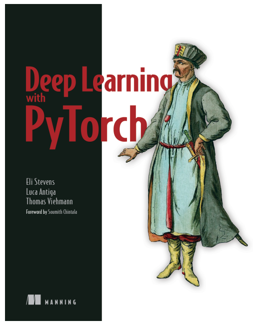
{AI Paradigm-shift}

HOGESCHOOL ROTTERDAM

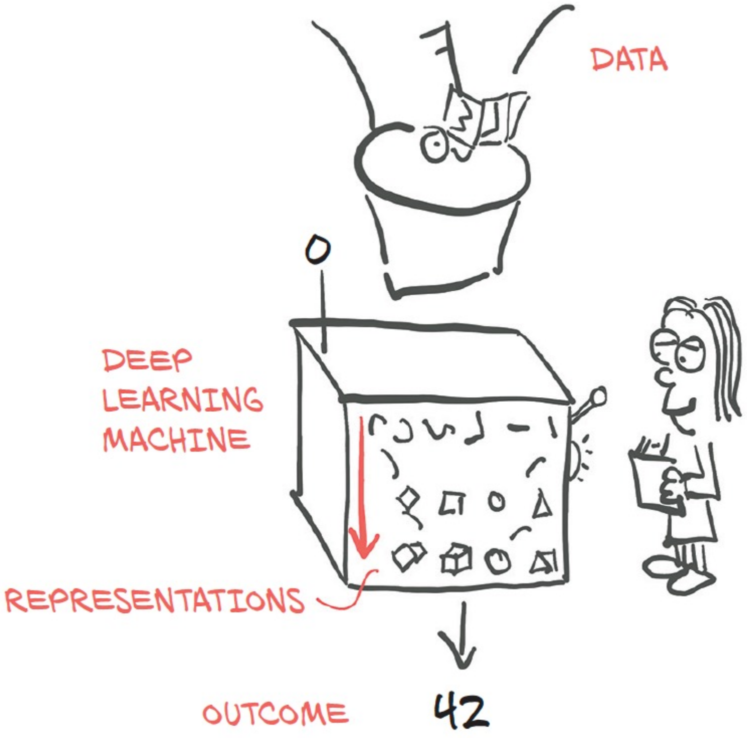
More data, parameters & computing power | Less human-in-the-loop

Machine Learning Paradigm {ML}





Deep Learning Paradigm {DL}



{Big-data}

Big-data is needed to avoid hand-crafted feature extraction

HOGESCHOOL ROTTERDAM

CREATING

A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg
Paul G. Allen School of Computer Science
University of Washington
Seattle, WA 98105
slund1@cs.washington.edu

Su-In Lee
Paul G. Allen School of Computer Science
Department of Genome Sciences
University of Washington
Seattle, WA 98105
suinlee@cs.washington.edu

Abstract

Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy in many applications. However, the highest accuracy for large modern datasets is often achieved by complex models that even experts struggle to interpret, such as ensemble or deep learning models, creating a tension between *accuracy* and *interpretability*. In response, various methods have recently been proposed to help users interpret the predictions of complex models, but it is often unclear how these methods are related and when one method is preferable over another. To address this problem, we present a unified framework for interpreting predictions, SHAP (SHapley Additive exPlanations). SHAP assigns each feature an importance value for a particular prediction. Its novel components include: (1) the identification of a new class of additive feature importance measures, and (2) theoretical results showing there is a unique solution in this class with a set of desirable properties. The new class unifies six existing methods, notable because several recent methods in the class lack the proposed desirable properties. Based on insights from this unification, we present new methods that show improved computational performance and/or better consistency with human intuition than previous approaches.

1 Introduction

The ability to correctly interpret a prediction model's output is extremely important. It engenders appropriate user trust, provides insight into how a model may be improved, and supports understanding of the process being modeled. In some applications, simple models (e.g., linear models) are often preferred for their ease of interpretation, even if they may be less accurate than complex ones. However, the growing availability of big data has increased the benefits of using complex models, so bringing to the forefront the trade-off between accuracy and interpretability of a model's output. A wide variety of different methods have been recently proposed to address this issue [5, 8, 9, 3, 4, 1]. But an understanding of how these methods relate and when one method is preferable to another is still lacking.

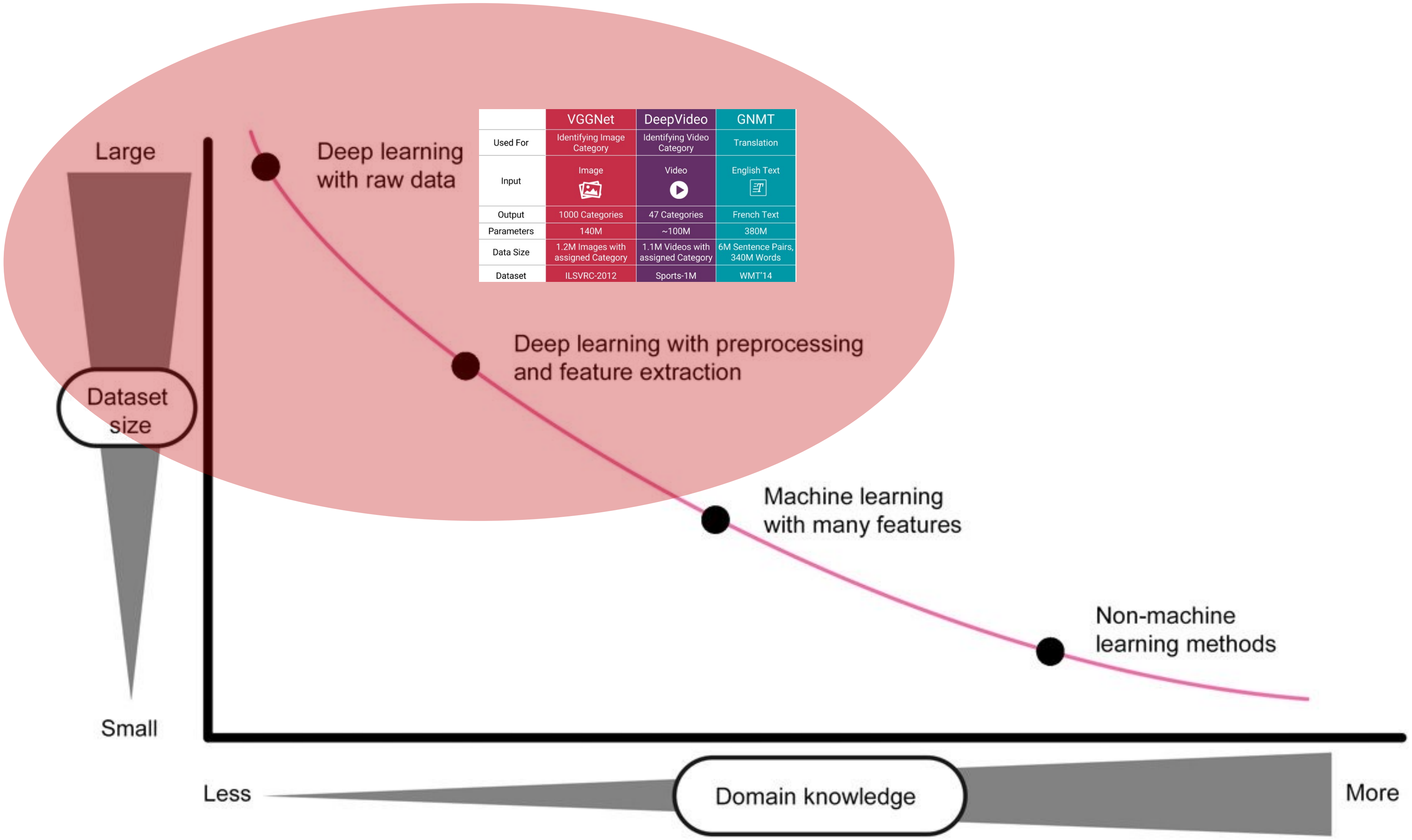
Here, we present a novel unified approach to interpreting model predictions.[1] Our approach leads to three potentially surprising results that bring clarity to the growing space of methods:

1. We introduce the perspective of viewing *any* explanation of a model's prediction as a model itself, which we term the *explanation model*. This lets us define the class of *additive feature attribution methods* (Section 2), which unifies six current methods.

[1] https://github.com/slundberg/shap

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

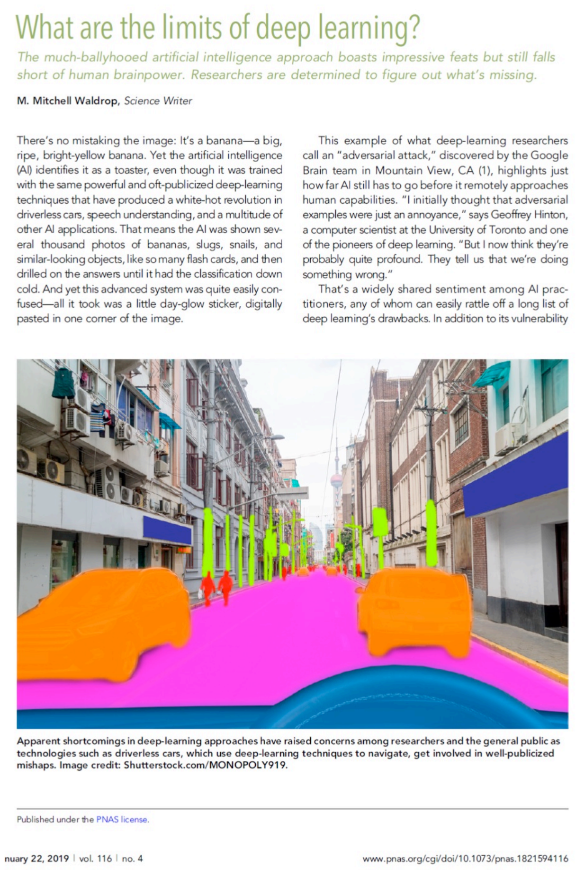
https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf



	VGGNet	DeepVideo	GNMT
Used For	Identifying Image Category	Identifying Video Category	Translation
Input	Image	Video	English Text
Output	1000 Categories	47 Categories	French Text
Parameters	140M	~100M	380M
Data Size	1.2M Images with assigned Category	1.1M Videos with assigned Category	6M Sentence Pairs, 340M Words
Dataset	ILSVRC-2012	Sports-1M	WMT'14

{Top-down} Top-down Encoding Capacity increases by adding hidden layers

HOGESCHOOL ROTTERDAM



What are the limits of deep learning?

The much-ballyhooed artificial intelligence approach boasts impressive feats but still falls short of human brainpower. Researchers are determined to figure out what's missing.

M. Mitchell Waldrop, *Science Writer*

There's no mistaking the image: It's a banana—a big, ripe, bright-yellow banana. Yet the artificial intelligence (AI) identifies it as a toaster, even though it was trained with the same powerful and oft-publicized deep-learning techniques that have produced a white-hot revolution in driverless cars, speech understanding, and a multitude of other AI applications. That means the AI was shown several thousand photos of bananas, slugs, snails, and similar-looking objects, like so many flash cards, and then drilled on the answers until it had the classification down cold. And yet this advanced system was quite easily confused—all it took was a little day-glow sticker, digitally pasted in one corner of the image.

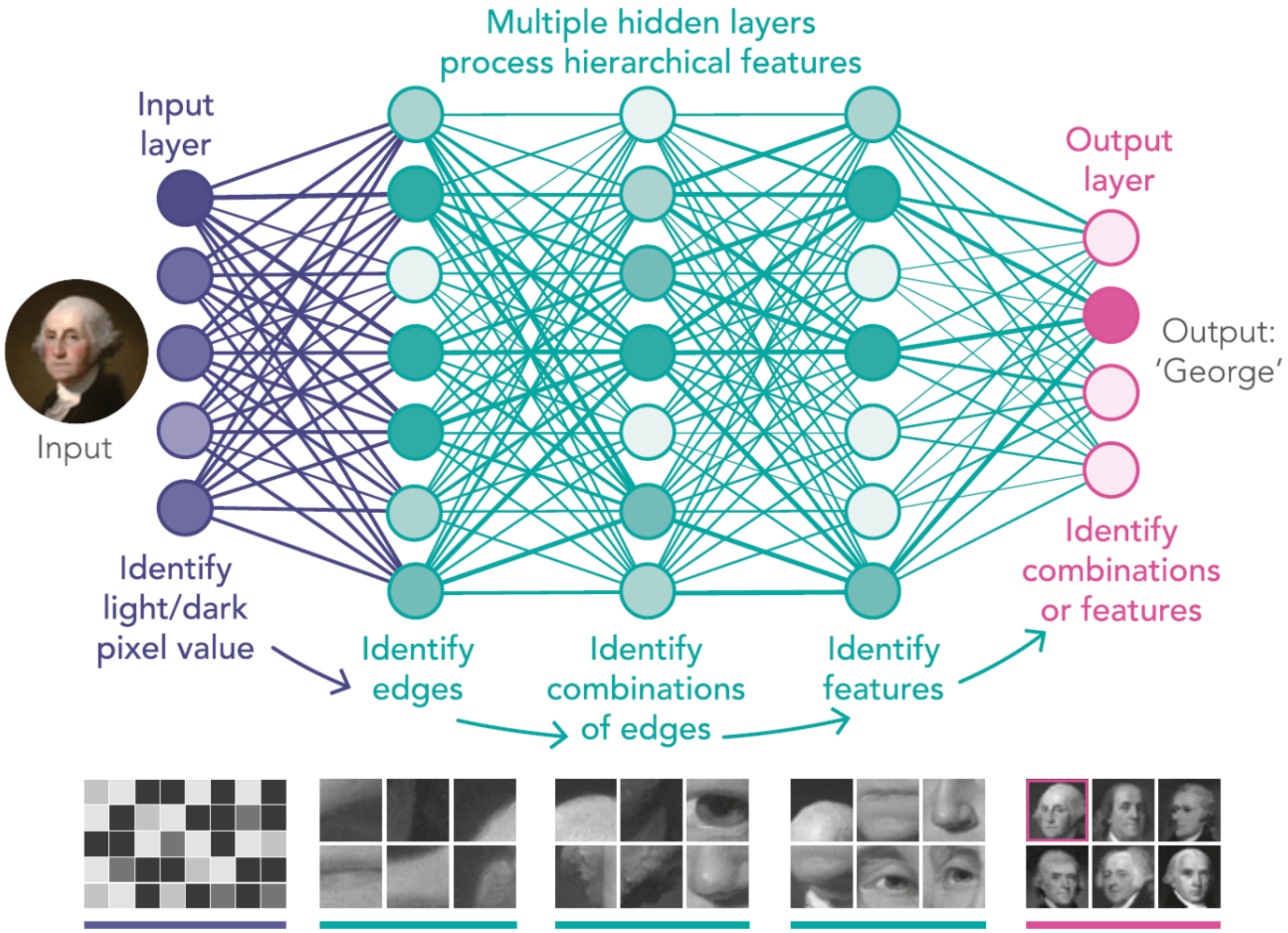
This example of what deep-learning researchers call an "adversarial attack," discovered by the Google Brain team in Mountain View, CA (1), highlights just how far AI still has to go before it remotely approaches human capabilities. "I initially thought that adversarial examples were just an annoyance," says Geoffrey Hinton, a computer scientist at the University of Toronto and one of the pioneers of deep learning. "But I now think they're probably quite profound. They tell us that we're doing something wrong."

That's a widely shared sentiment among AI practitioners, any of whom can easily rattle off a long list of deep learning's drawbacks. In addition to its vulnerability

Apparent shortcomings in deep-learning approaches have raised concerns among researchers and the general public as technologies such as driverless cars, which use deep-learning techniques to navigate, get involved in well-publicized mishaps. Image credit: Shutterstock.com/MONOPOLY919.

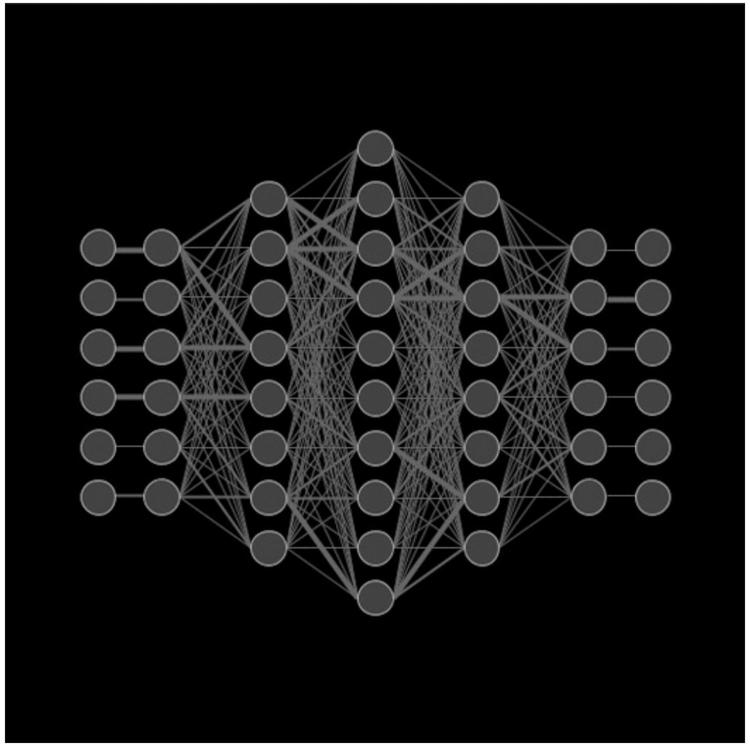
Published under the PNAS license.

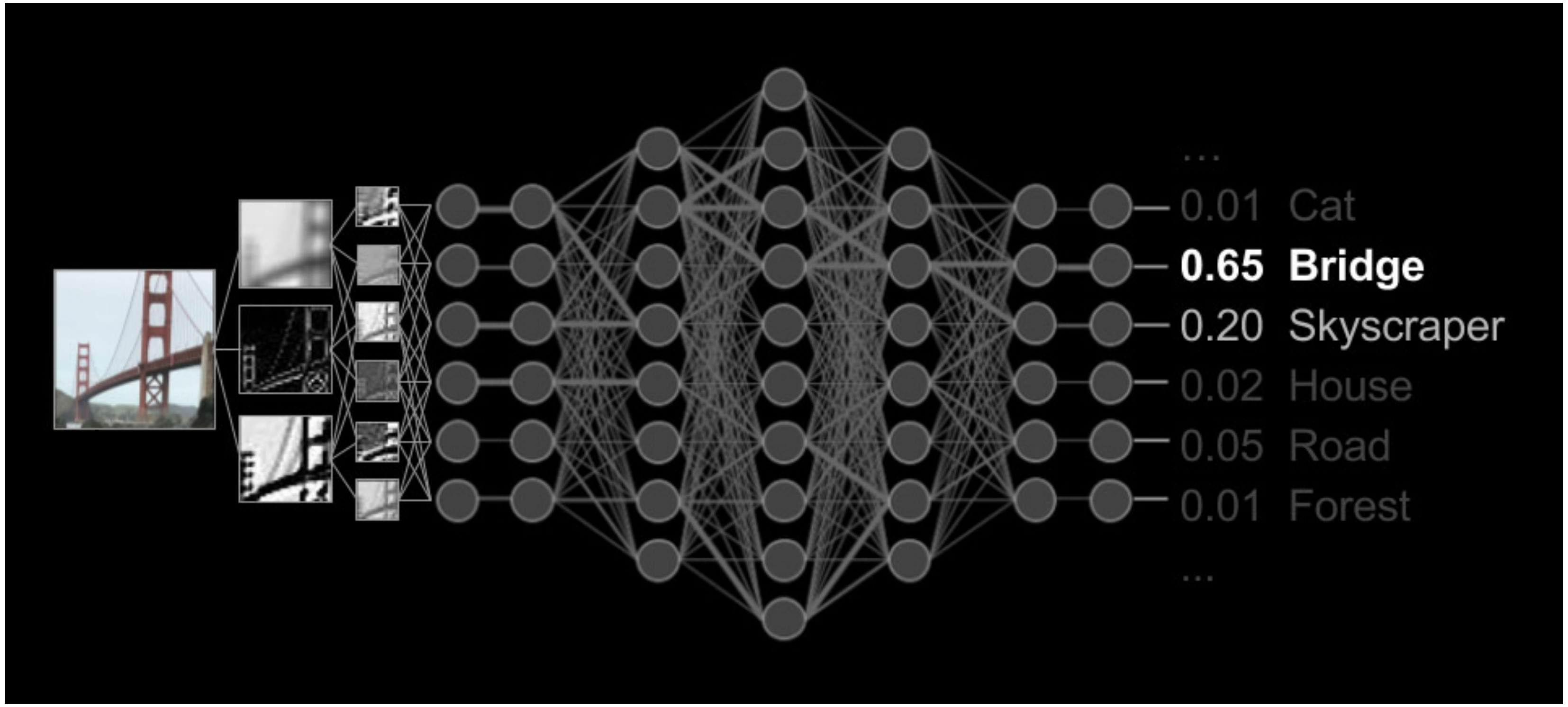
nuary 22, 2019 | vol. 116 | no. 4   www.pnas.org/cgi/doi/10.1073/pnas.1821594116



{Top-down} Top-down Encoding Capacity increases by adding hidden layers

HOGESCHOOL ROTTERDAM

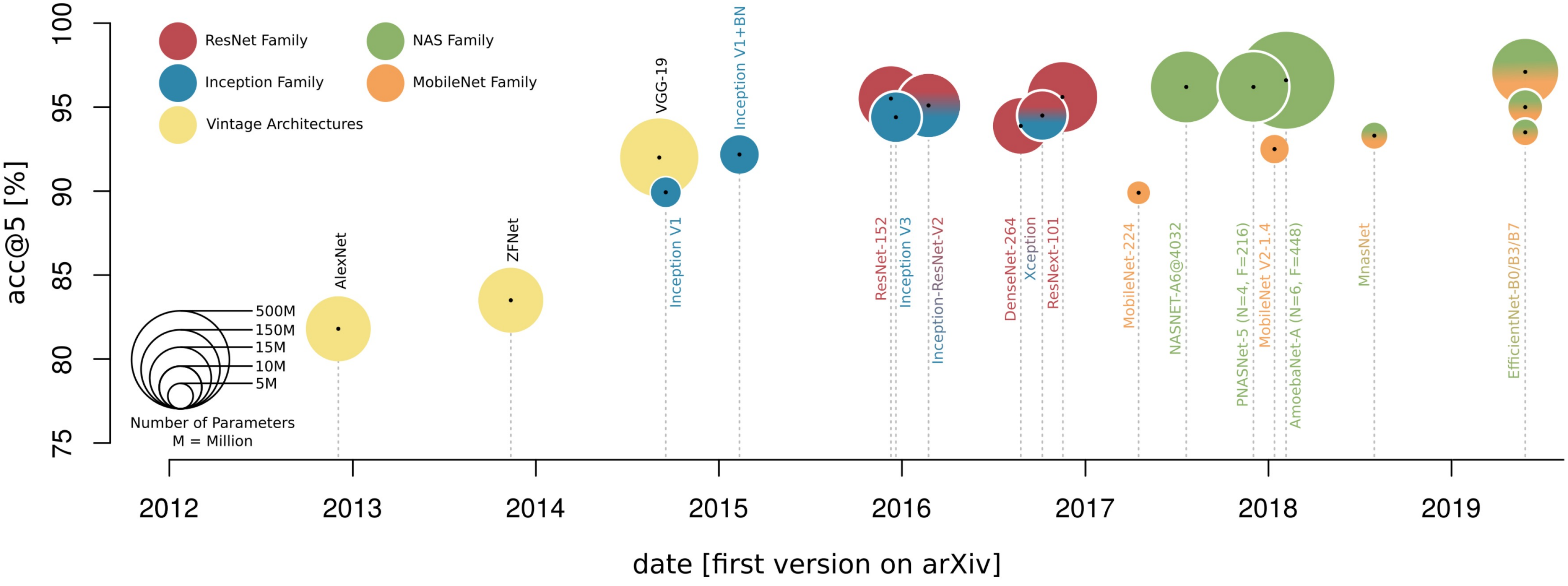




{Weights}

Performance increases by adding learnable parameters {weight's}

HOGESCHOOL ROTTERDAM



[How to calculate the number of learnable parameters]?

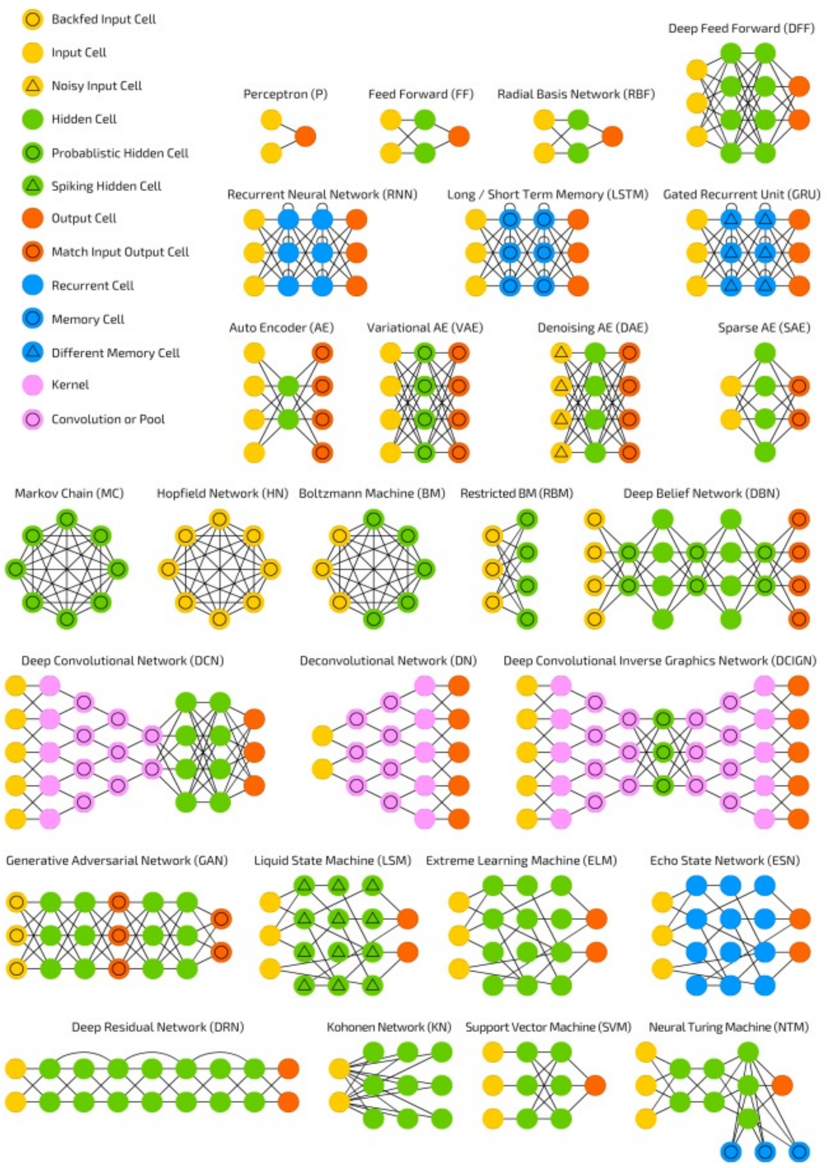
https://doi.org/10.3390/rs12101667

{Topology}

Topology of a neural network refers to the way
artificial neurons are connected to form a network.

Form follows function!
The topology of a network determines the degree of perplexity of the tasks it can perform.

https://pub.towardsai.net/main-types-of-neural-networks-and-its-applications-tutorial-734480d7ec8e

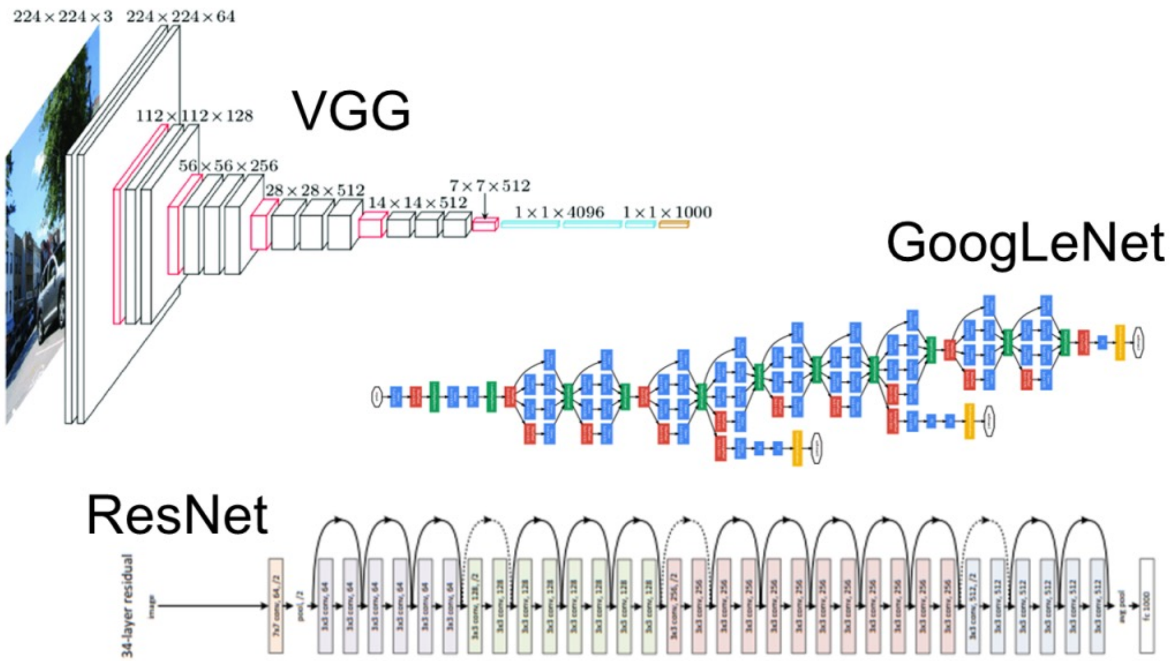


https://www.mdpi.com/2504-3900/47/1/9

HOGESCHOOL ROTTERDAM
CREATING

{Perplexity}

Toplogical complex Neural Networks Perform Better: have Low Perplexity



Why the simple strategy of scaling up neural networks has been so effective?

[2105.12806] A Universal Law of Robustness via Isoperimetry (arxiv.org)

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-
EfficientNetB0	29 MB	-	-	5,330,571	-
EfficientNetB1	31 MB	-	-	7,856,239	-
EfficientNetB2	36 MB	-	-	9,177,569	-
EfficientNetB3	48 MB	-	-	12,320,535	-
EfficientNetB4	75 MB	-	-	19,466,823	-
EfficientNetB5	118 MB	-	-	30,562,527	-
EfficientNetB6	166 MB	-	-	43,265,143	-
EfficientNetB7	256 MB	-	-	66,658,687	-

Keras https://keras.io/api/applications/

{Self-attention}

Transformer DNNs outperform GANs, CNN & RNNs by adding a stacked intermediate neural net topologies that attend to themselves called Transformers

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

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

1 Introduction

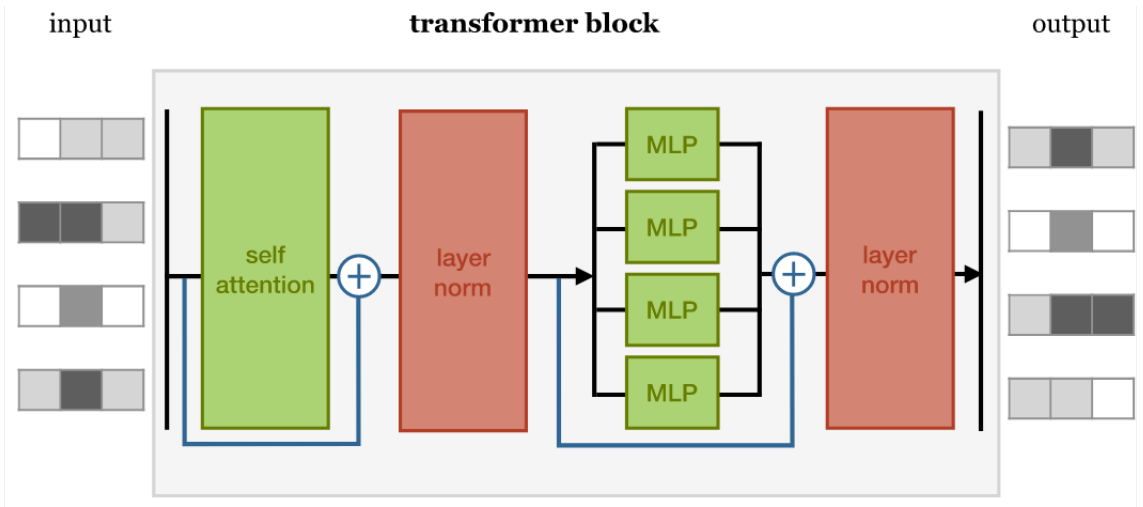
Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



[GitHub - pbloem/former: Simple transformer implementation from scratch in pytorch.](https://github.com/pbloem/former)
[Transformers from scratch | peterbloem.nl](http://peterbloem.nl)

Mukhamediev, R. I., Symagulov, A., Kuchin, Y., Yakunin, K., & Yelis, M. (2021). From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review. *Applied Sciences*, *11*(12), 5541. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app11125541

Why use biology in stead of Big-data ?

{Information Theory} vs {Brute-Force}

{Skewed}

Big-Data is Inherently Skewed

When artificial intelligence learns from humans, it's bad

James Crowder*

Learned bias

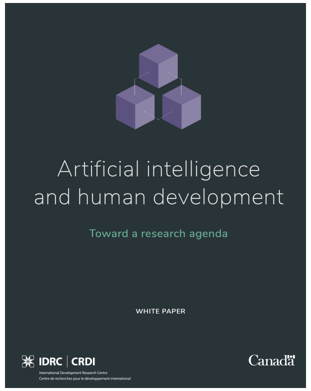
ML algorithms are not biased in and of themselves; they *learn* to be biased. This *algorithmic bias* (Danks and London 2017) has received a great deal of attention. It occurs when the learning algorithm is trained on biased datasets and subsequently "accurately" learns the patterns of bias inherent in the data (see, e.g., Caliskan et al. 2017). In some cases, the learned representations within ML algorithms can even exaggerate these biases (Zhao et al. 2017). Algorithmic bias has two sources: incomplete datasets and datasets that represent biased social phenomena.

Incomplete datasets are those that are not representative of the entire range of potential examples. Consequently, an algorithm trained on an incomplete dataset will perform poorly when given an example that falls outside the scope of the available data. For example, a facial recognition algorithm that is not trained on a wide variety of skin colours might not function accurately for faces of all skin tones. Indeed, one AI researcher with dark skin discovered that an otherwise functional facial recognition algorithm failed to recognize her face unless she put on a white mask[45] . Many similar examples have been reported.[46]

When critical decisions are made based on input from an algorithm trained on a database that is not representative of the entire user population, the results can impact health and well-being. Most clinical trials have highly selective criteria that exclude women (especially pregnant women), the elderly, and those with conditions beyond those being studied. Thus, participants tend to be white males.[47] In some cases, the findings of these studies do not generalize well across the broader population, with outcomes potentially compromised for individuals not represented in the research. As a result, AI algorithms trained on this data are

* bit.ly/2HFetvY

Photo by Marius Masalar



{Disparities}

Big Data causes racial & gender disparities

HOGESCHOOL ROTTERDAM

CREATING

The Race Gap in Speech Recognition Technology

Researchers from Stanford University found that state-of-the-art ASR systems exhibited significant racial and gender disparity—they misunderstand Black speakers twice as often as white speakers. In the paper, titled "Racial Disparities in Automated Speech Recognition," authors ran thousands of audio snippets of white and Black speakers, transcribed from interviews conducted with 42 white speakers and 73 Black speakers, through leading speech-to-text services by Amazon, Apple, Google, IBM, and Microsoft.

The results suggest that, on average, systems made 19 errors every hundred words for white speakers and 35 errors for Black speakers—nearly twice as many. Moreover, the systems performed particularly poorly for Black men, with more than 40 errors for every hundred words (Figure 2.5.3). The breakdown by ASR systems shows that gaps are similar across companies (Figure 2.5.4). This research emphasizes the importance of addressing the bias of AI technologies and ensuring equity as they become mature and deployed.

TESTINGS on LEADING SPEECH-to-TEXT SERVICES: WORD ERROR RATE by RACE and GENDER, 2019

Source: Koenecke et al., 2020 | Chart: 2021 AI Index Report

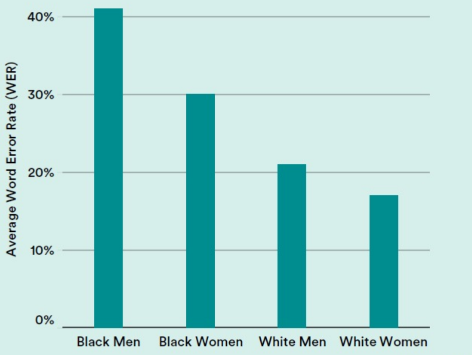


Figure 2.5.3

TESTINGS on LEADING SPEECH-to-TEXT SERVICES: WORD ERROR RATE by SERVICE and RACE, 2019

Source: Koenecke et al., 2020 | Chart: 2021 AI Index Report

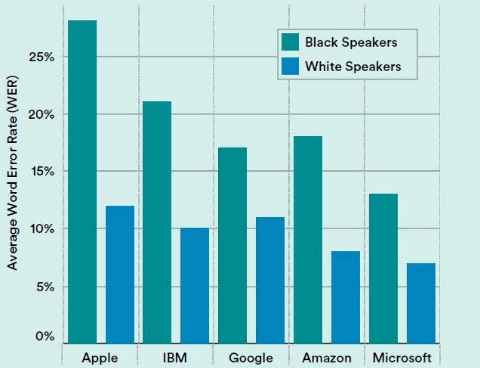


Figure 2.5.4

{Augmentation}

HOGESCHOOL ROTTERDAM

Big Data that is *not* augmented causes Overfitting

SURVEY PAPER Open Access

A survey on Image Data Augmentation for Deep Learning

Connor Shorten* and Taghi M. Khoshgoftaar

*Correspondence: cshorten2015@fau.edu
Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, USA

Abstract

Deep convolutional neural networks have performed remarkably well on many Computer Vision tasks. However, these networks are heavily reliant on big data to avoid overfitting. Overfitting refers to the phenomenon when a network learns a function with very high variance such as to perfectly model the training data. Unfortunately, many application domains do not have access to big data, such as medical image analysis. This survey focuses on Data Augmentation, a data-space solution to the problem of limited data. Data Augmentation encompasses a suite of techniques that enhance the size and quality of training datasets such that better Deep Learning models can be built using them. The image augmentation algorithms discussed in this survey include geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning. The application of augmentation methods based on GANs are heavily covered in this survey. In addition to augmentation techniques, this paper will briefly discuss other characteristics of Data Augmentation such as test-time augmentation, resolution impact, final dataset size, and curriculum learning. This survey will present existing methods for Data Augmentation, promising developments, and meta-level decisions for implementing Data Augmentation. Readers will understand how Data Augmentation can improve the performance of their models and expand limited datasets to take advantage of the capabilities of big data.

Keywords: Data Augmentation, Big data, Image data, Deep Learning, GANs

Introduction

Deep Learning models have made incredible progress in discriminative tasks. This has been fueled by the advancement of deep network architectures, powerful computation, and access to big data. Deep neural networks have been successfully applied to Computer Vision tasks such as image classification, object detection, and image segmentation thanks to the development of convolutional neural networks (CNNs). These neural networks utilize parameterized, sparsely connected kernels which preserve the spatial characteristics of images. Convolutional layers sequentially downsample the spatial resolution of images while expanding the depth of their feature maps. This series of convolutional transformations can create much lower-dimensional and more useful representations of images than what could possibly be hand-crafted. The success of CNNs has spiked interest and optimism in applying Deep Learning to Computer Vision tasks.

SpringerOpen

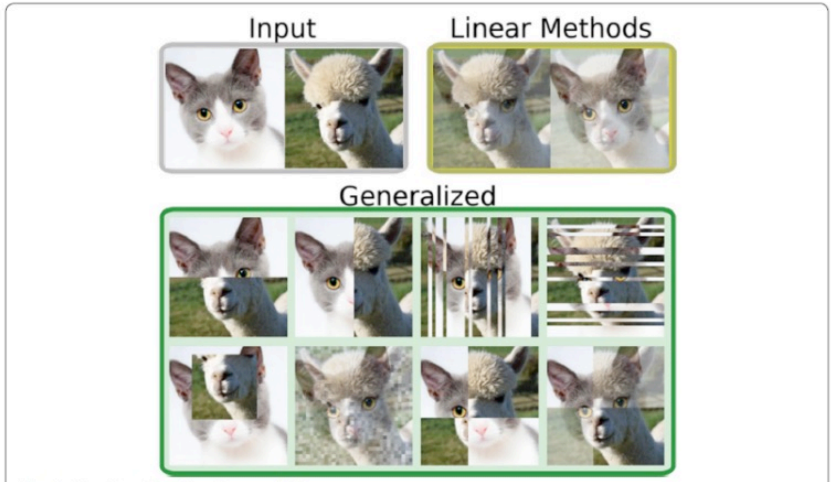
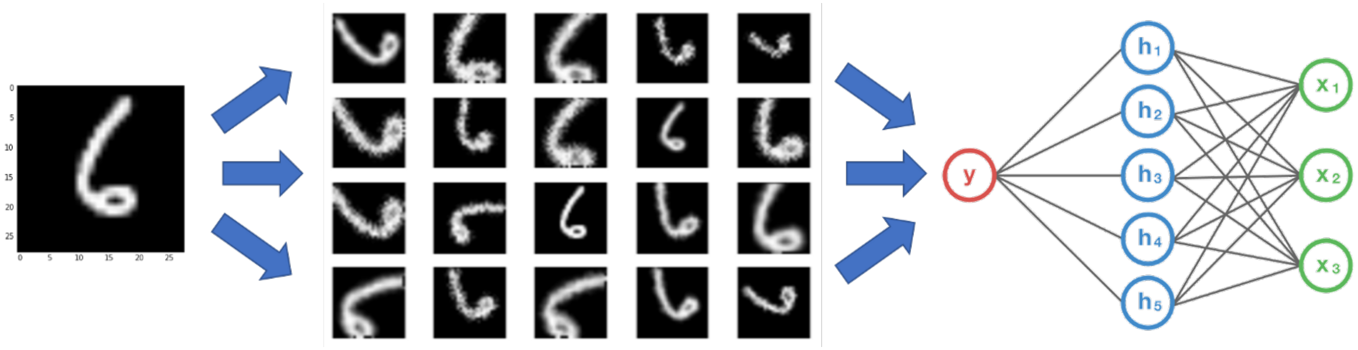


Fig. 9 Non-linearly mixing images [66]



{Regularization}

HOGESCHOOL ROTTERDAM

The process of adding information to adress **ill-posed problems:** How to increase sparsity of **Dense DNNs?**

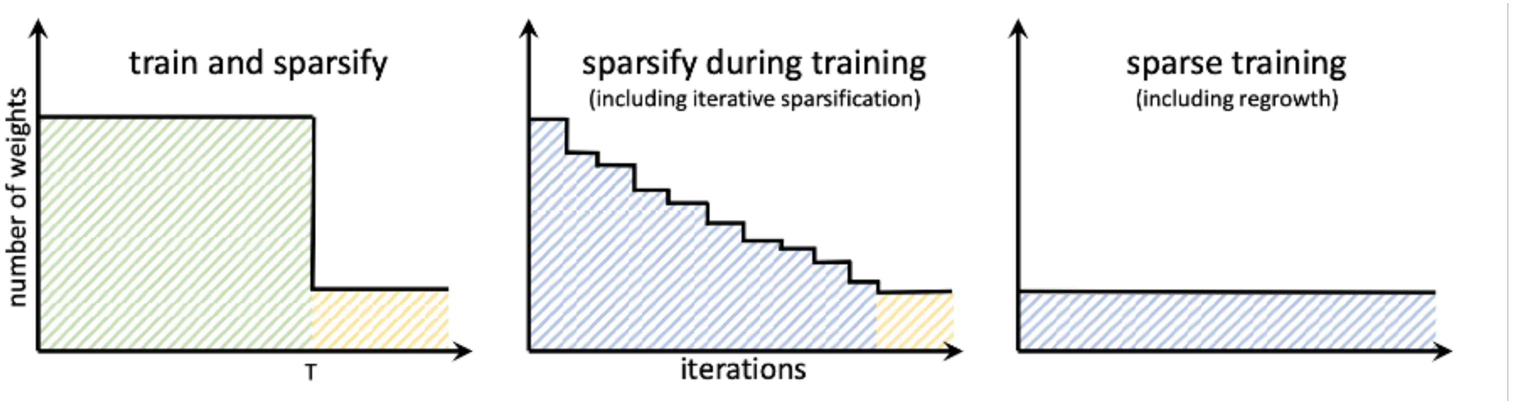
Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks

TORSTEN HOEFLER, ETH Zürich, Switzerland
DAN ALISTARH, IST Austria, Austria
TAL BEN-NUN, ETH Zürich, Switzerland
NIKOLI DRYDEN, ETH Zürich, Switzerland
ALEXANDRA PESTE, IST Austria, Austria

The growing energy and performance costs of deep learning have driven the community to reduce the size of neural networks by selectively pruning components. Similarly to their biological counterparts, sparse networks generalize just as well, if not better than, the original dense networks. Sparsity can reduce the memory footprint of regular networks to fit mobile devices, as well as shorten training time for ever growing networks. In this paper, we survey prior work on sparsity in deep learning and provide an extensive tutorial of sparsification for both inference and training. We describe approaches to remove and add elements of neural networks, different training strategies to achieve model sparsity, and mechanisms to exploit sparsity in practice. Our work distills ideas from more than 300 research papers and provides guidance to practitioners who wish to utilize sparsity today, as well as to researchers whose goal is to push the frontier forward. We include the necessary background on mathematical methods in sparsification, describe phenomena such as early structure adaptation, the intricate relations between sparsity and the training process, and show techniques for achieving acceleration on real hardware. We also define a metric of pruned parameter efficiency that could serve as a baseline for comparison of different sparse networks. We close by speculating on how sparsity can improve future workloads and outline major open problems in the field.

The supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience -

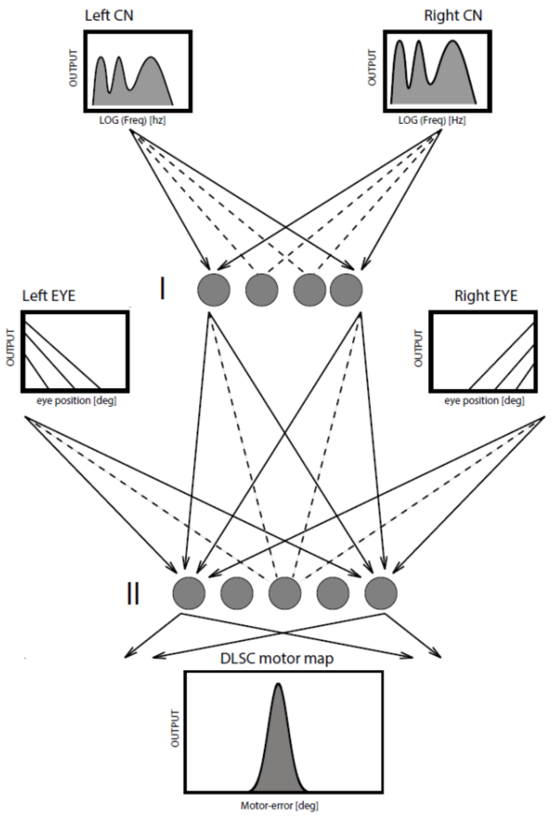
Albert Einstein, 1933



Hoefler, Torsten, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, and Alexandra Peste.
"Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks."
Journal of Machine Learning Research 22, no. 241 (2021): 1-124. https://doi.org/10.48550/arXiv.2102.00554

{Pruning}

DNNs are hard to down scale after the training phase, most DNNs are *not* sparse coded

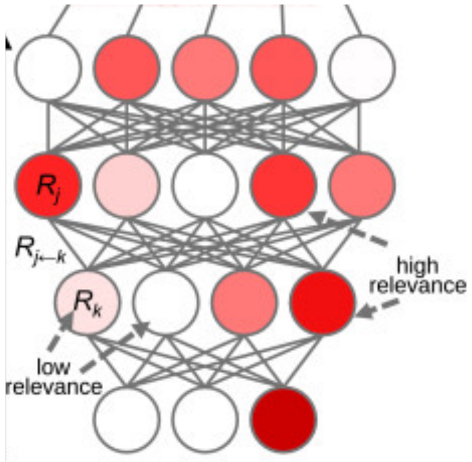


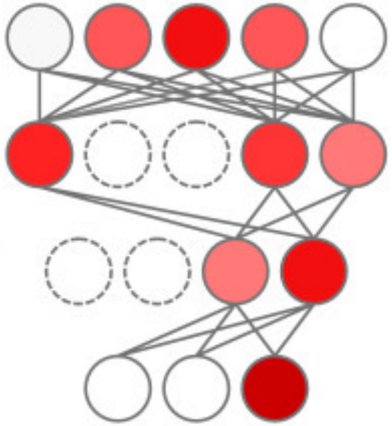
International Workshop on Biologically Motivated Computer Vision

Audio-Oculomotor Transformation (2002)

R.F. van der Willigen & Mark von Campenhausen

Part of the Lecture Notes in Computer Science book series (LNCS, volume 2525)

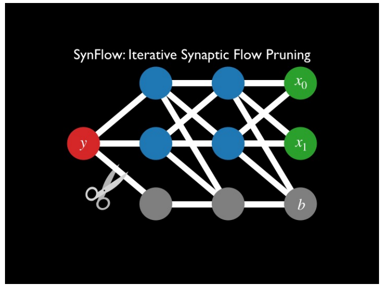




Yeom, S.K., Seegerer, P., Lapuschkin, S., Binder, A., Wiedemann, S., Müller, K.R., & Samek, W. (2021). Pruning by explaining: A novel criterion for DNN pruning. Pattern Recognition, 115, 107899.

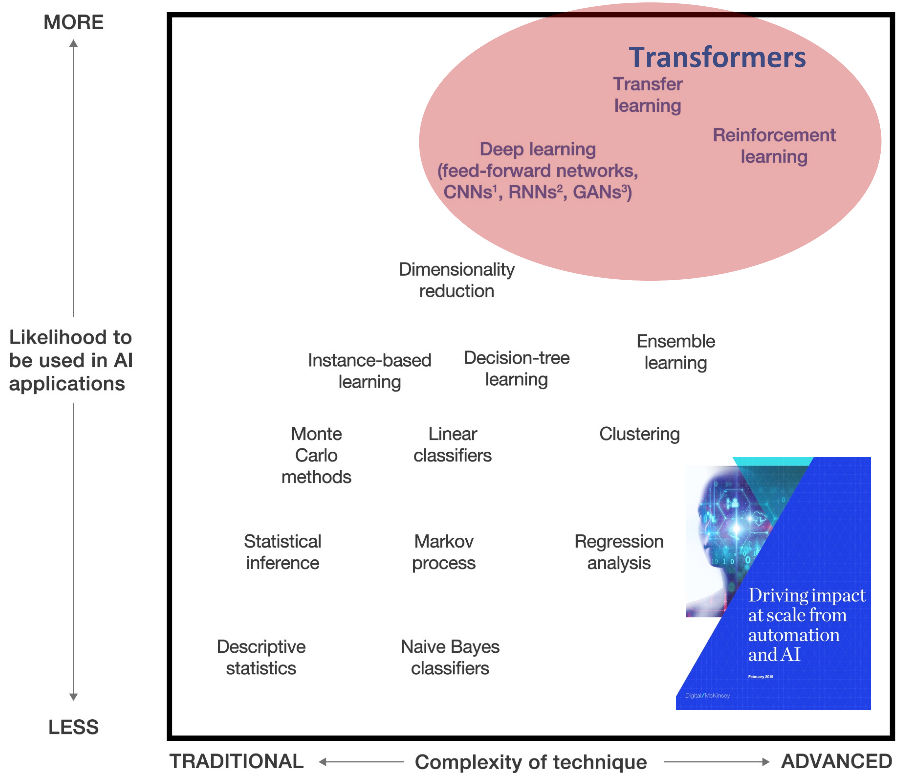
Blalock, D., Gonzalez Ortiz, J. J., Frankle, J., & Guttag, J. (2020). What is the state of neural network pruning? Proceedings of machine learning and systems, 2, 129-146.

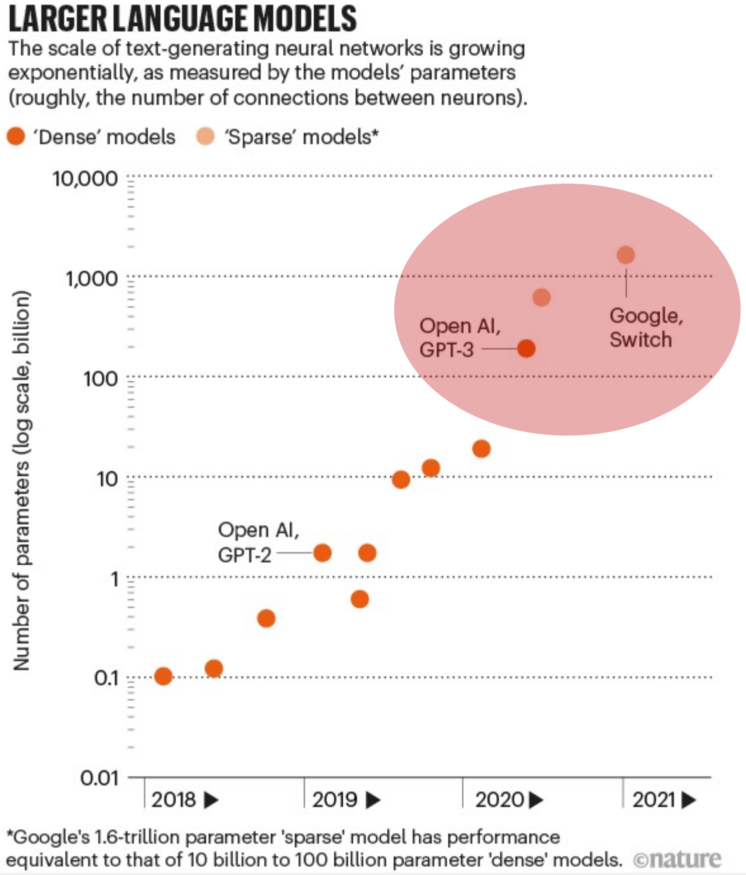
Frankle, J., & Carbin, M. (2018). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*.



{large scale}

Only very large scale {DNNs} are useful [can compete with human performance]





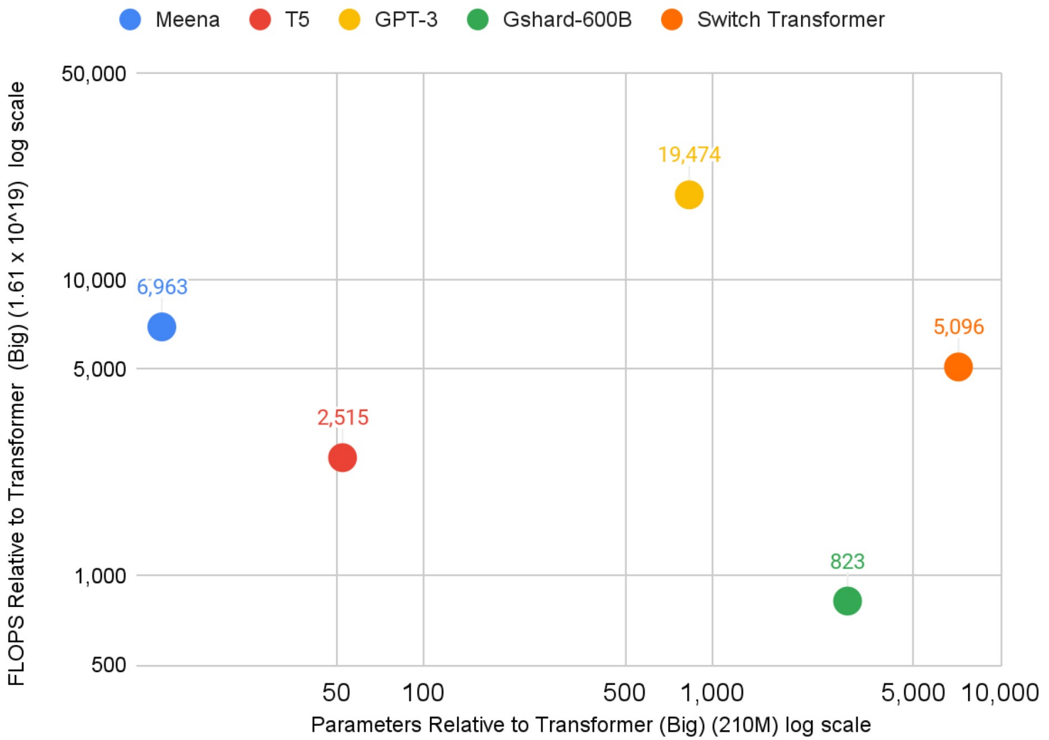
https://www.nature.com/articles/d41586-021-00530-0

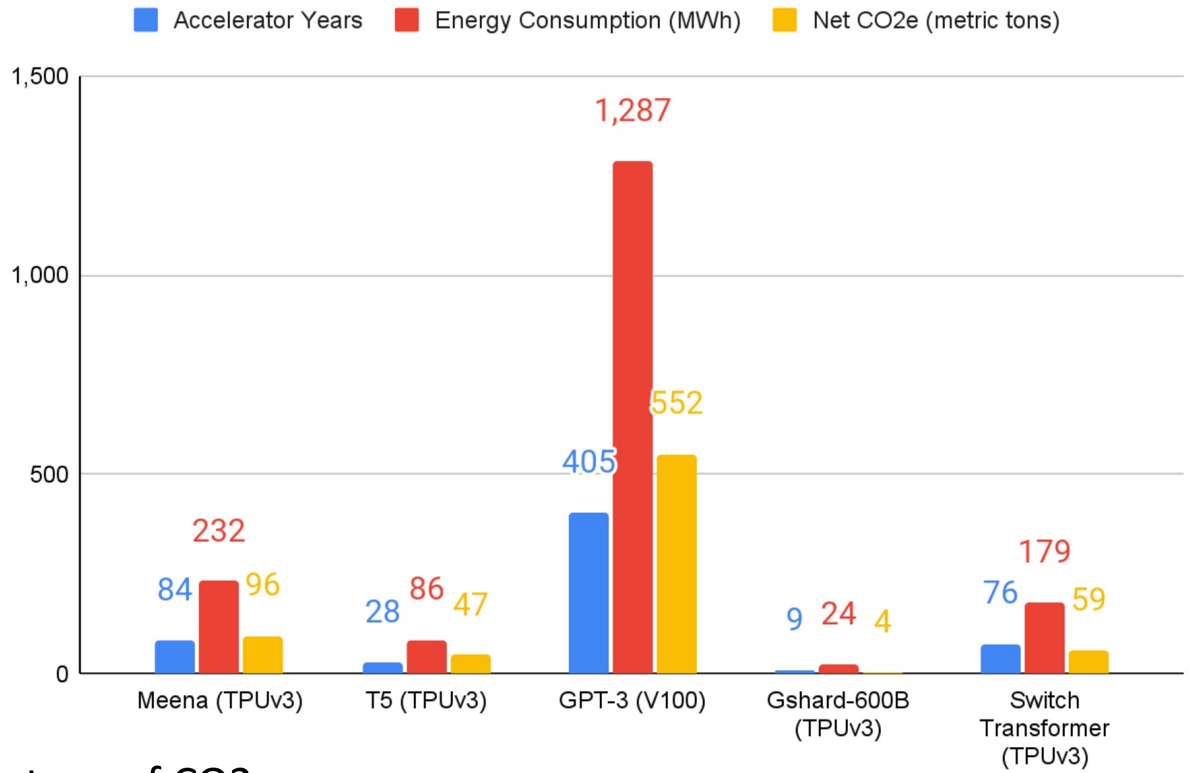
HOGESCHOOL ROTTERDAM
CREATING

{CO2 foot-print}

HOGESCHOOL ROTTERDAM
CREATING

Training large scale transformer {DNNs} produce massive Carbon Emissions





As of 2007, the average U.S. household emits 20 metric tons of CO2 per year. In comparison to a world average of 4 tons.

https://arxiv.org/ftp/arxiv/papers/2104/2104.10350.pdf

Carbon Footprint_CSS09-05_e2021.pdf (umich.edu)

{computational unsustainability}

The scale of state-of-the-art {SOTA} –near human level– DNNs

– combined with a blind Brute-Force implementation + post-hoc analysis –

is becoming more and more

computationally unsustainable,

even to the point

that hypernetworks are employed

to help humans to make DDNs work.

[2110.13100v1.pdf (arxiv.org)](https://arxiv.org/pdf/2110.13100v1.pdf)

https://paperswithcode.com/sota/

AI research needs to prioritize computationally sustainable paradigms

HOW?

Can “lessons learned” from Sensory Ecology make the application of {DNNs} less problematic?

PART I

The Science

SCIENTIFIC REPORTS
natureresearch

HOGESCHOOL ROTTERDAM

OPEN

Biological learning curves outperform existing ones in artificial intelligence algorithms

Article | Open Access | Published: 09 August 2019

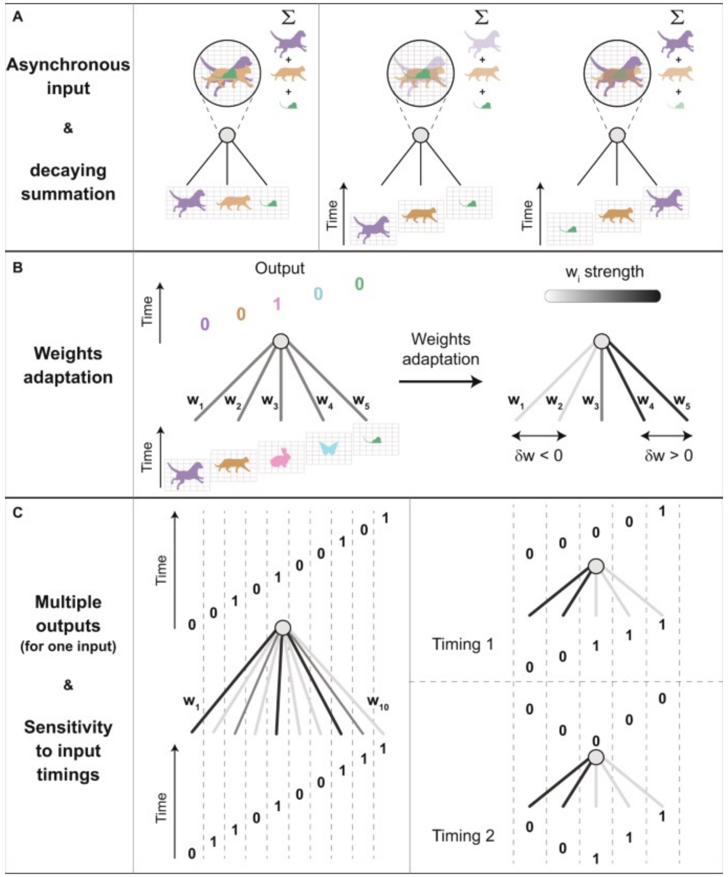
Biological learning curves outperform existing ones in artificial intelligence algorithms

Herut Uzan, Shira Sardi, Amir Goldental, Roni Vardi & Ido Kanter

Scientific Reports 9, Article number: 11558 (2019) | Cite this article

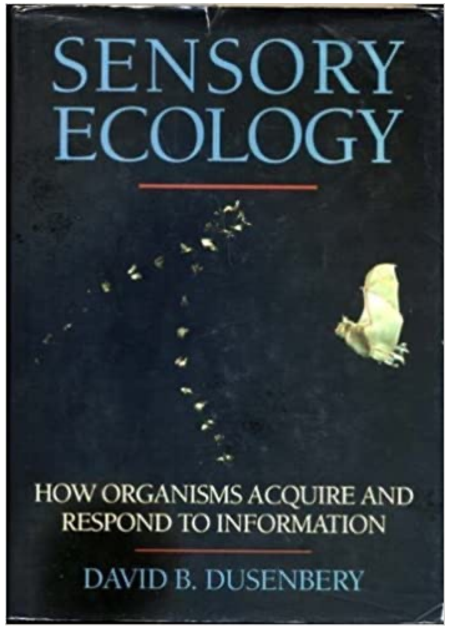
Abstract

Recently, deep learning algorithms have outperformed human experts in various tasks across several domains; however, their characteristics are distant from current knowledge of neuroscience. The simulation results of biological learning algorithms presented herein outperform state-of-the-art optimal learning curves in supervised learning of feedforward networks. The biological learning algorithms comprise asynchronous input signals with decaying input summation, weights adaptation, and multiple outputs for an input signal. In particular, the generalization error for such biological perceptrons decreases rapidly with increasing number of examples, and it is independent of the size of the input. This is achieved using either synaptic learning, or solely through dendritic adaptation with a mechanism of swinging between reflecting boundaries, without learning steps. The proposed biological learning algorithms outperform the optimal scaling of the learning curve in a traditional perceptron. It also results in a considerable robustness to disparity between weights of two networks with very similar outputs in biological supervised learning scenarios. The simulation results indicate the potency of neurobiological mechanisms and open opportunities for developing a superior class of deep learning algorithms.



https://doi.org/10.1038/s41598-019-48016-4

{Sensory Ecology}



Sensory Ecology
studies how
Sensory Specialists
acquire and respond to
Information

Received: 29 November 2019 | Revised: 9 February 2020 | Accepted: 17 February 2020
DOI: 10.1002/cne.24896

Check for updates

JCN WILEY

REVIEW

Sensory systems in birds: What we have learned from studying sensory specialists

Andrew N. Iwaniuk[1] | Douglas R. Wylie[2]

[1]Department of Neuroscience, University of Lethbridge, Lethbridge, Alberta, Canada
[2]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Correspondence
Andrew N. Iwaniuk, Department of Neuroscience, University of Lethbridge, Lethbridge, AB T1K3M4, Canada.
Email: andrew.iwaniuk@uleth.ca

Funding information
Canada Research Chairs; Natural Sciences and Engineering Research Council of Canada

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/cne.24896.

Abstract

"Diversity" is an apt descriptor of the research career of Jack Pettigrew as it ranged from the study of trees, to clinical conditions, to sensory neuroscience. Within sensory neuroscience, he was fascinated by the evolution of sensory systems across species. Here, we review some of his work on avian sensory specialists and research that he inspired in others. We begin with an overview of the importance of the Wulst in stereopsis and the need for further study of the Wulst in relation to binocularity across avian species. Next, we summarize recent anatomical, behavioral, and physiological studies on optic flow specializations in hummingbirds. Beyond vision, we discuss the first evidence of a tactile "fovea" in birds and how this led to detailed studies of tactile specializations in waterfowl and sensorimotor systems in parrots. We then describe preliminary studies by Pettigrew of two endemic Australian species, the plains-wanderer (*Pedionomus torquatus*) and letter-winged kite (*Elanus scriptus*), that suggest the evolution of some unique auditory and visual specializations in relation to their unique behavior and ecology. Finally, we conclude by emphasizing the importance of a comparative and integrative approach to understanding avian sensory systems and provide an example of one system that has yet to be properly examined: tactile facial bristles in birds. Through reviewing this research and offering future avenues for discovery, we hope that others also embrace the comparative approach to understanding sensory system evolution in birds and other vertebrates.

1 | INTRODUCTION

Birds are an exceptionally diverse clade. Not only have they speciated into 10,000 species, they inhabit nearly every terrestrial habitat, occupy a highly diverse range of ecological niches and express an immense range of behavioral variation. This diversity among (and within) bird species has been critical for the development of evolutionary theory (Darwin, 1859, 1888), key concepts in behavioral evolution (Lorenz, 1971), and our understanding of evolutionary processes (Grant & Grant, 2014). The behavioral and ecological diversity of birds is based on an equally impressive diversity in the brain.

The relative size and composition of avian brains varies greatly across orders, families and even genera, a fact that was evident in some of the earlier descriptions of avian brain anatomy by Craigie (Craigie, 1928, 1930, 1940) and others (Stingelin, 1958, 1961). Even in these early anatomical studies, it was clear that some aspects of avian brain anatomy reflected sensory specializations. That is, in accordance with the principle of proper mass (Jerison, 1973), the size of sensory regions in the avian brain varies with the importance of the respective sensory modality. It was this neuroanatomical variation that led to speculation over the importance of olfaction to kiwi, New World vultures, and seabirds (Bang & Cobb, 1968). Our more recent studies have provided further insights into the relationship between sensory ecology and brain anatomy (D. R. Wylie, Gutierrez-Ibanez, & Iwaniuk,

Andrew N. Iwaniuk and Douglas R. Wylie authors contributed equally to this work.

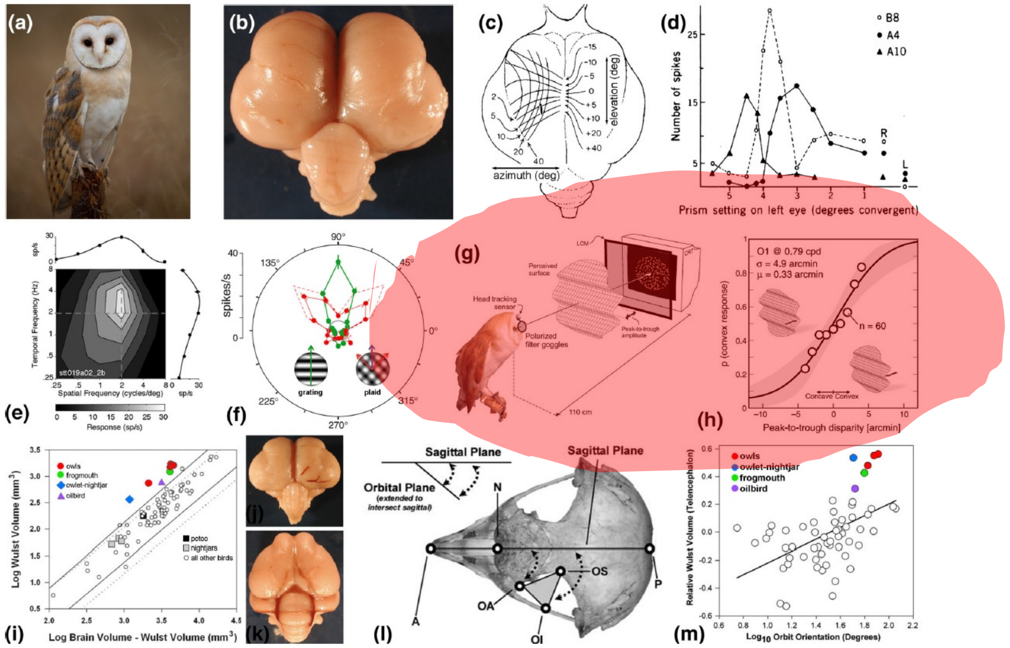


FIGURE 1 The visual Wulst and binocular vision. (a) A barn owl (*Tyto alba*) (photo by Miroslav Zitek). (b) The brain of a greater sooty owl (*Tyto tenebricosa*). (c) The retinotopic map in the Wulst of a barn owl. (adapted from Pettigrew (1979)). (d) Disparity tuning of binocular neurons in the Wulst of the barn owl. For three neurons, firing rate is plotted as a function of retinal disparity (deg). The responses to monocular stimulation of the right and left eyes is shown on the right-hand side of the panel (R, L). (from Pettigrew & Konishi, 1976a, 1976b). (e) Spatio-temporal tuning of a motion sensitive neuron in the Wulst of a burrowing owl (*Athene cunicularia*). (from Pinto and Baron (2009)). (f) Responses of a neuron in the Wulst of a burrowing owl to drifting gratings (green) and "plaids" (red) composed of two gratings 90° apart. Firing rate is plotted as a function of direction in polar coordinates. The dashed red line indicates the predicted response to the plaid patterns if the neuron was responding to the sum of the two component gratings. (adapted from Baron, Pinto, Dias, Lima, and Neuenschwander (2007)). (g) The psycophysical method used to show that barn owls have stereopsis. Random dot stereograms were displayed on the screen and appeared as either a concave or convex corrugated surface of varying depth (LCM = liquid crystal modulator) (copyright ARVO). (h) A psychometric function where the proportion (p) of convex responses is plotted as a function of disparity. Theta, the standard deviation of the Gaussian fit to the function, represents the stereo acuity threshold. Mu, is the mean position of the function and is a measure of response bias toward either convex or concave corrugations (copyright ARVO). (g) and (h) are from van der Willigen, Harmening, Vossen, and Wagner (2010). (i) Wulst volume as a function of brain-Wulst volume for owls, caprimulgiformes and other birds. The pairs of lines represent 95% confidence interval of least-squares linear regressions calculated for "all other birds" (open circles) using conventional statistics (solid lines) and statistics that include phylogenetic information (dashed lines). Adapted from Iwaniuk and Wylie (2006). (j) The brain of a tawny frogmouth (*Podargus strigoides*). (k) The brain of a spotted nightjar (*Eurostopodus argus*). (l) Measuring orbit orientation in birds. Orbit orientation was measured as the dihedral angle formed by the intersection of the orbital and sagittal planes. The morphometric points used to define orbital and sagittal planes are shown on the skull of a snowy owl (*Nyctea scandiaca*). The sagittal plane is defined by points a (anterior), N (nasal-maxillary junction), and P (posterior).The orbital plane is defined by the superior, inferior and anterior points on the orbit (OS, OI, and OA). From Iwaniuk, Heesy, Hall, and Wylie (2008). (m) Relative Wulst volume as a function of orbit orientation. The black line shows the least-squares linear regression line. Relative Wulst volumes were residuals derived from a least-squares linear regression of Wulst volume vs. brain-Wulst volume. Adapted from Iwaniuk et al. (2008)

2015), but much of our interpretation was dependent on neurophysiological and behavioral data that was unavailable to early avian neuroanatomists.

In spite of the variation in avian brain anatomy documented by earlier researchers, it was not until Pettigrew and others started using modern neuroscience techniques to probe the function of sensory regions that we started to understand the extent to which sensory ecology shapes brain anatomy and physiology in birds. More importantly, Pettigrew not only provided ground-breaking insights into how birds see, touch, and hear the world around them, his studies

https://doi.org/10.1002/cne.24896

HOGESCHOOL ROTTERDAM

{Sensory Ecology vs DNNs}

HOGESCHOOL ROTTERDAM

Deep neural networks {DNNs} have attained human-level performances on challenging cognitive tasks.

https://doi.org/10.3390/rs12101667

Yet, it remains unclear how information is represented in {DNNs}.

https://doi.org/10.1016/j.inffus.2019.12.012

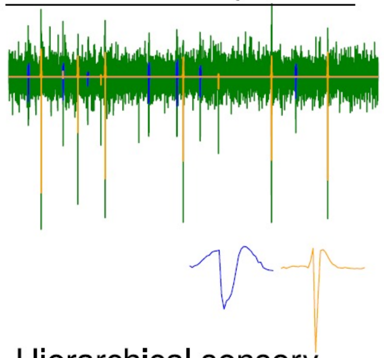
In parallel, the same cognitive tasks have been extensively studied in the brains of owls (visual, auditory specialist), humans & monkeys (cognitive task specialist).

https://doi.org/10.3389/fncom.2020.578158

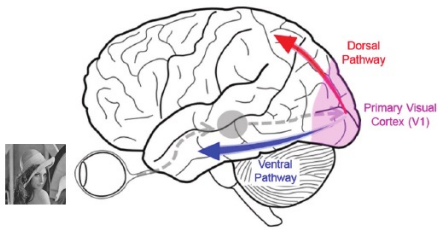
{Sensory Ecology vs DNN}

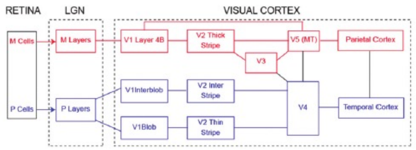
Unsupervised learning to decipher neural code

Blind source separation

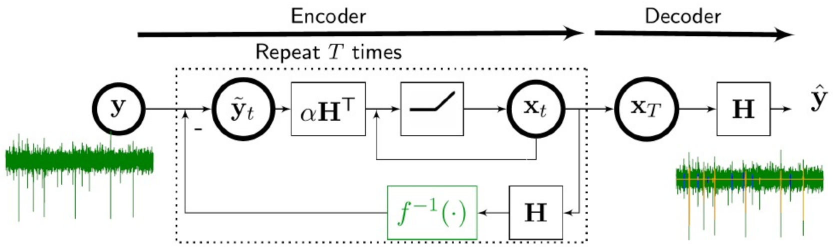


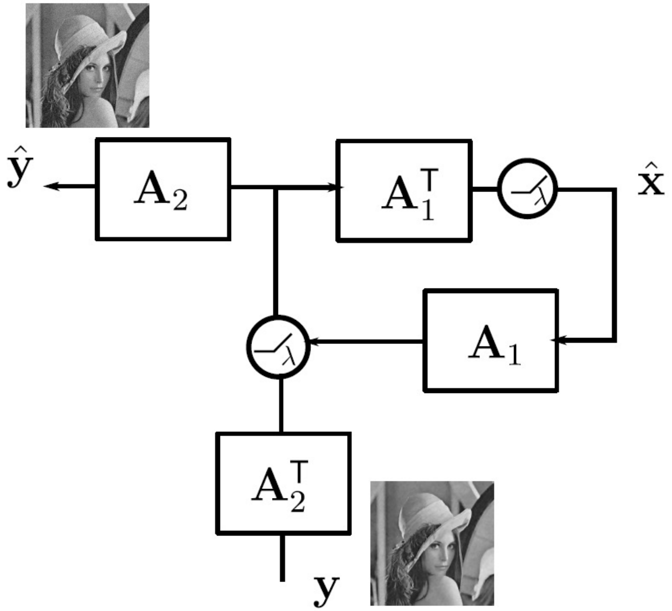
Hierarchical sensory processing principles





How to design interpretable deep nets?





https://iaifi.org/talks/2021_04_15_IAIFI_Colloquium_Ba.pdf

{Information Theory}

Information theory comes handy here. We have defined entropy, surprisal, and cross-entropy when we introduced the softmax regression (Section 3.4.7) and more of information theory is discussed in the online appendix on information theory. If we want to compress text, we can ask about predicting the next token given the current set of tokens. A better language model should allow us to predict the next token more accurately. Thus, it should allow us to spend fewer bits in compressing the sequence. So we can measure it by the cross-entropy loss averaged over all the n tokens of a sequence:

$$\frac{1}{n}\sum_{t=1}^{n} -\log P(x_t \mid x_{t-1}, \ldots, x_1), \tag{8.4.7}$$

where P is given by a language model and x_t is the actual token observed at time step t from the sequence. This makes the performance on documents of different lengths comparable. For historical reasons, scientists in natural language processing prefer to use a quantity called *perplexity*. In a nutshell, it is the exponential of (8.4.7):

$$\exp\left(-\frac{1}{n}\sum_{t=1}^{n} \log P(x_t \mid x_{t-1}, \ldots, x_1)\right). \tag{8.4.8}$$

Perplexity can be best understood as the harmonic mean of the number of real choices that we have when deciding which token to pick next. Let

This formula is a measure of the **thermodynamic entropy** of a system when k is equal to the Boltzmann's constant ($1.38064852 \times 10^{-23}$ J/K)

In order for Boltzmann's thermodynamic entropy definition to apply, the natural logarithm must be used (i.e., $b=e$). In the context of Shannon entropy, the scale of the logarithm can be chosen for convenience. For state spaces defined by binary decision trees, $\log_2$ is typically used so that one unit of information (one **bit**) reduces uncertainty by a half. If one uses $\log_{10}$, one unit of information reduces uncertainty 10 fold (a quantity termed a **Hartley** of information).

Shannon generalized information theory to non-thermodynamic systems by demonstrating that this formula can be used to quantify the carrying capacity of any channel of communication. In this general case, H is referred to as **Shannon entropy**. Here the value of k is arbitrary and unit-less and often set to 1 for convenience.

Given a probability distribution **p** $p(x_i)$ defines the probability that the system is in a specific state

If all m states of a system are equally probable, the uncertainty is maximized. This is often a useful reference point for comparison (see Box 3). In this case, the formula reduces to $H = -k \log_b(m)$ (or, $S = k \ln(W)$ using Boltzmann's famous notation).

$$H = -k\sum_{i=1}^{m} p(x_i) \log_b p(x_i)$$

As discussed in Box 2, the **Kullbeck-Leibler divergence** D_{KL} measures the "distance" between two probability distributions (**p** and **p***). It is called a divergence and not a distance because it asymmetrical and depends on what is chosen as the reference (or prior, in Bayesian contexts) distribution. If **p** is the reference, then:

$$D_{KL}(\mathbf{p}^*, \mathbf{p}) = -\sum_{i=1}^{m} p^*(x_i) \log\left(\frac{p^*(x_i)}{p(x_i)}\right)$$

If, and only if, **p** is a uniform distribution (such that all states are equally likely), then

$$D_{KL}(\mathbf{p}^*, \mathbf{p}) = H(\mathbf{p}) - H(\mathbf{p}^*)$$

If **p*** and **p** are marginal posterior and marginal prior distributions respectively, the quantity $H(\mathbf{p}) - H(\mathbf{p}^*)$ is referred to as **Lindley information**. More generally, the sum of $D_{KL}(\mathbf{p}^*, \mathbf{p}) + D_{KL}(\mathbf{p}, \mathbf{p}^*)$ is called **Jeffrey's divergence**, which is proportional to **Fisher's information**, a widely used statistical measure of uncertainty. Frank (2009, 2012) has demonstrated that natural selection maximizes Fisher's Information.

Knowledge of the state of one variable can also provide information (reduce uncertainty) about the state of another variable. This is called **Mutual Information**. Mutual information (which we denote $I(X,Y)$) is simply the difference between the **marginal** $H(X)$ and the **conditional** $H(X|Y)$ **entropies** (or because unlike the Kullbeck-Leibler divergence [right], it is symmetrical, $H(Y) - H(Y|X)$). We can compute this as

$$I(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i, y_j) \log\left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right)$$

If the variables are independent, then knowledge of the state of one provides no information on the state of the other and the $I(X,Y) = 0$.

For derivations and details, see the excellent treatments by Lindley (1956), Cover and Thomas (2001), Amari and Nagaoka (2000), Burnham and Anderson (2001), and Frank (2009, 2012).

HOGESCHOOL ROTTERDAM

CREATING

Information {en}coding can be measured objectively, whereas Big-data cannot!

frontiers in Ecology and Evolution

HYPOTHESIS AND THEORY
published: 18 June 2019
doi: 10.3389/fevo.2019.00219

Check for updates

Principles of Ecology Revisited: Integrating Information and Ecological Theories for a More Unified Science

Mary I. O'Connor[1], Matthew W. Pennell[1], Florian Altermatt[2,3], Blake Matthews[4,5], Carlos J. Melián[4,5] and Andrew Gonzalez[6]*

[1] Biodiversity Research Centre and Department of Zoology, University of British Columbia, Vancouver, BC, Canada, [2] Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, [3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [4] Department of Fish Ecology and Evolution, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland, [5] Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland, [6] Department of Biology, Quebec Centre for Biodiversity Science, McGill University, Montreal, QC, Canada

OPEN ACCESS

Edited by:
Luciano Bosso, Department of Agriculture, University of Naples Federico II, Italy

Reviewed by:
Melanie E. Moses, University of New Mexico, United States
Brian Gregory Eddy, Natural Resources Canada, Canada

***Correspondence:**
Mary I. O'Connor
oconnor@zoology.ubc.ca

Specialty section:
This article was submitted to Biogeography and Macroecology, a section of the journal Frontiers in Ecology and Evolution

Received: *31 December 2018*
Accepted: *24 May 2019*
Published: *18 June 2019*

Citation:
O'Connor MI, Pennell MW, Altermatt F, Matthews B, Melián CJ and Gonzalez A (2019) Principles of Ecology Revisited: Integrating Information and Ecological Theories for a More Unified Science. Front. Ecol. Evol. 7:219. doi: 10.3389/fevo.2019.00219

The persistence of ecological systems in changing environments requires energy, materials, and information. Although the importance of information to ecological function has been widely recognized, the fundamental principles of ecological science as commonly expressed do not reflect this central role of information processing. We articulate five fundamental principles of ecology that integrate information with energy and material constraints across scales of organization in living systems. We show how these principles outline new theoretical and empirical research challenges, and offer one novel attempt to incorporate them in a theoretical model. To provide adequate background for the principles, we review major concepts and identify common themes and key differences in information theories spanning physics, biology and semiotics. We structured our review around a series of questions about the role information may play in ecological systems: (i) what is information? (ii) how is information related to uncertainty? (iii) what is information processing? (iv) does information processing link ecological systems across scales? We highlight two aspects of information that capture its dual roles: *syntactic information* defining the processes that encode, filter and process information stored in biological structure and *semiotic information* associated with structures and their context. We argue that the principles of information in living systems promote a unified approach to understanding living systems in terms of first principles of biology and physics, and promote much needed theoretical and empirical advances in ecological research to unify understanding across disciplines and scales.

Keywords: information theory, semiotic, entropy, organization, first principles, ecology, evolution

https://d2l.ai/chapter_appendix-mathematics-for-deep-learning/information-theory.html

https://doi.org/10.3389/fevo.2019.00219

{Sensory Ecology Design Patterns}

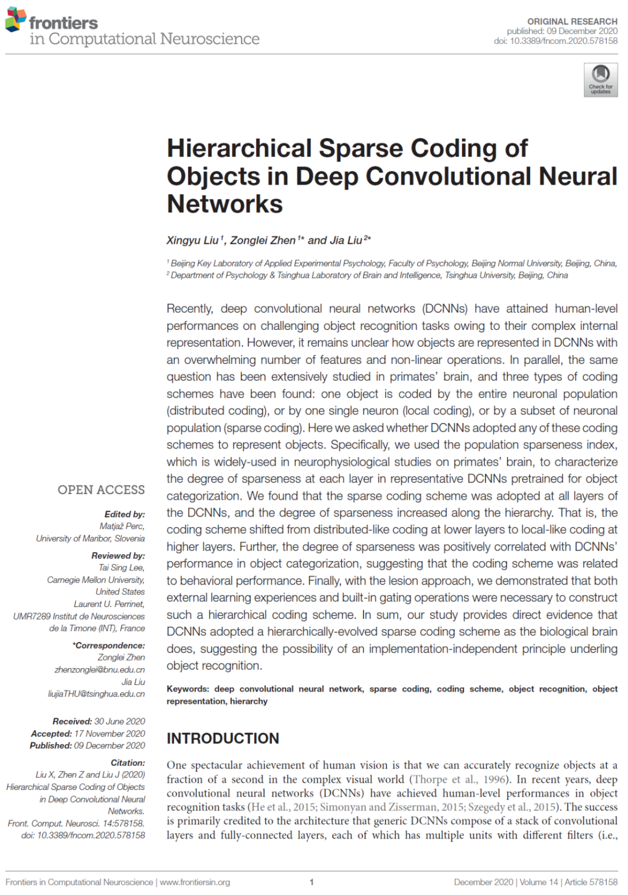
HOGESCHOOL ROTTERDAM

Sensory ecology has uncovered 3 types **information coding schemes** in the brain of owls, humans & monkeys:

Distributed {en}Coding

Local {en} Coding

Sparse {en} Coding


frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 09 December 2020
doi: 10.3389/fncom.2020.578158

Check for updates

Hierarchical Sparse Coding of Objects in Deep Convolutional Neural Networks

Xingyu Liu[1], *Zonglei Zhen*[1*] *and Jia Liu*[2*]

[1] Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China,
[2] Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China

Recently, deep convolutional neural networks (DCNNs) have attained human-level performances on challenging object recognition tasks owing to their complex internal representation. However, it remains unclear how objects are represented in DCNNs with an overwhelming number of features and non-linear operations. In parallel, the same question has been extensively studied in primates' brain, and three types of coding schemes have been found: one object is coded by the entire neuronal population (distributed coding), or by one single neuron (local coding), or by a subset of neuronal population (sparse coding). Here we asked whether DCNNs adopted any of these coding schemes to represent objects. Specifically, we used the population sparseness index, which is widely-used in neurophysiological studies on primates' brain, to characterize the degree of sparseness at each layer in representative DCNNs pretrained for object categorization. We found that the sparse coding scheme was adopted at all layers of the DCNNs, and the degree of sparseness increased along the hierarchy. That is, the coding scheme shifted from distributed-like coding at lower layers to local-like coding at higher layers. Further, the degree of sparseness was positively correlated with DCNNs' performance in object categorization, suggesting that the coding scheme was related to behavioral performance. Finally, with the lesion approach, we demonstrated that both external learning experiences and built-in gating operations were necessary to construct such a hierarchical coding scheme. In sum, our study provides direct evidence that DCNNs adopted a hierarchically-evolved sparse coding scheme as the biological brain does, suggesting the possibility of an implementation-independent principle underling object recognition.

Keywords: deep convolutional neural network, sparse coding, coding scheme, object recognition, object representation, hierarchy

OPEN ACCESS

Edited by:
Matjaž Perc, University of Maribor, Slovenia

Reviewed by:
Tai Sing Lee, Carnegie Mellon University, United States
Laurent U. Perrinet, UMR7289 Institut de Neurosciences de la Timone (INT), France

***Correspondence:**
Zonglei Zhen
zhenzonglei@bnu.edu.cn
Jia Liu
liujiaTHU@tsinghua.edu.cn

Received: *30 June 2020*
Accepted: *17 November 2020*
Published: *09 December 2020*

Citation:
Liu X, Zhen Z and Liu J (2020) Hierarchical Sparse Coding of Objects in Deep Convolutional Neural Networks. Front. Comput. Neurosci. 14:578158. doi: 10.3389/fncom.2020.578158

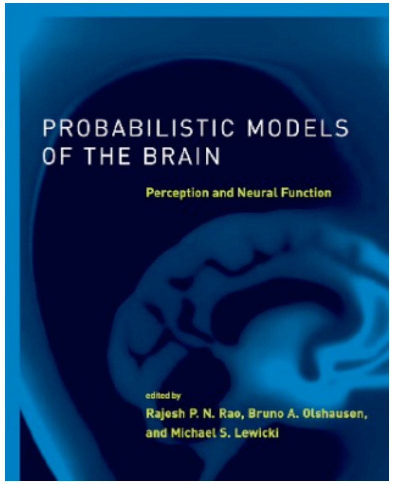
INTRODUCTION

One spectacular achievement of human vision is that we can accurately recognize objects at a fraction of a second in the complex visual world (Thorpe et al., 1996). In recent years, deep convolutional neural networks (DCNNs) have achieved human-level performances in object recognition tasks (He et al., 2015; Simonyan and Zisserman, 2015; Szegedy et al., 2015). The success is primarily credited to the architecture that generic DCNNs compose of a stack of convolutional layers and fully-connected layers, each of which has multiple units with different filters (i.e.,

https://doi.org/10.3389/fncom.2020.578158

{Sparse Coding}

Sensory Ecology has shown
that in order to enhance survival-rate:



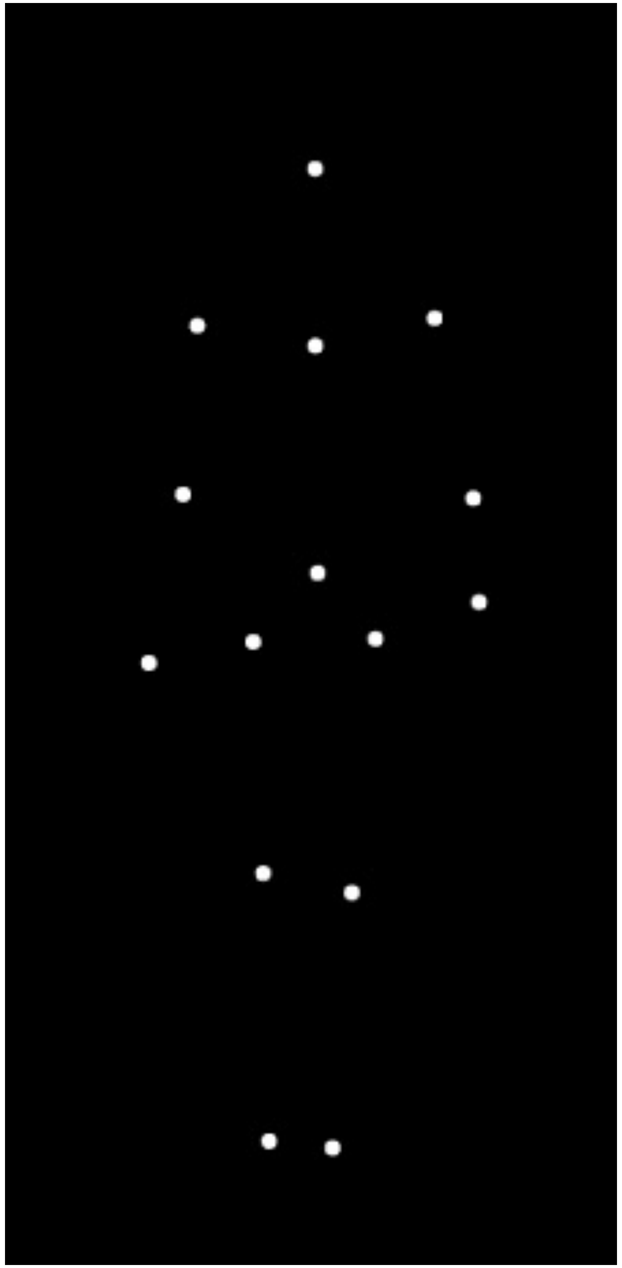
"A sensing organism should attempt to
represent stimuli
– *encoding of sensory information* –
as a combination of
as few putative causes
as possible"

{Biological Motion}

Biological motion is a prime example of sparse coding as performed by the human brain.

Only 12 moving-dots are needed to determine direction of motion.

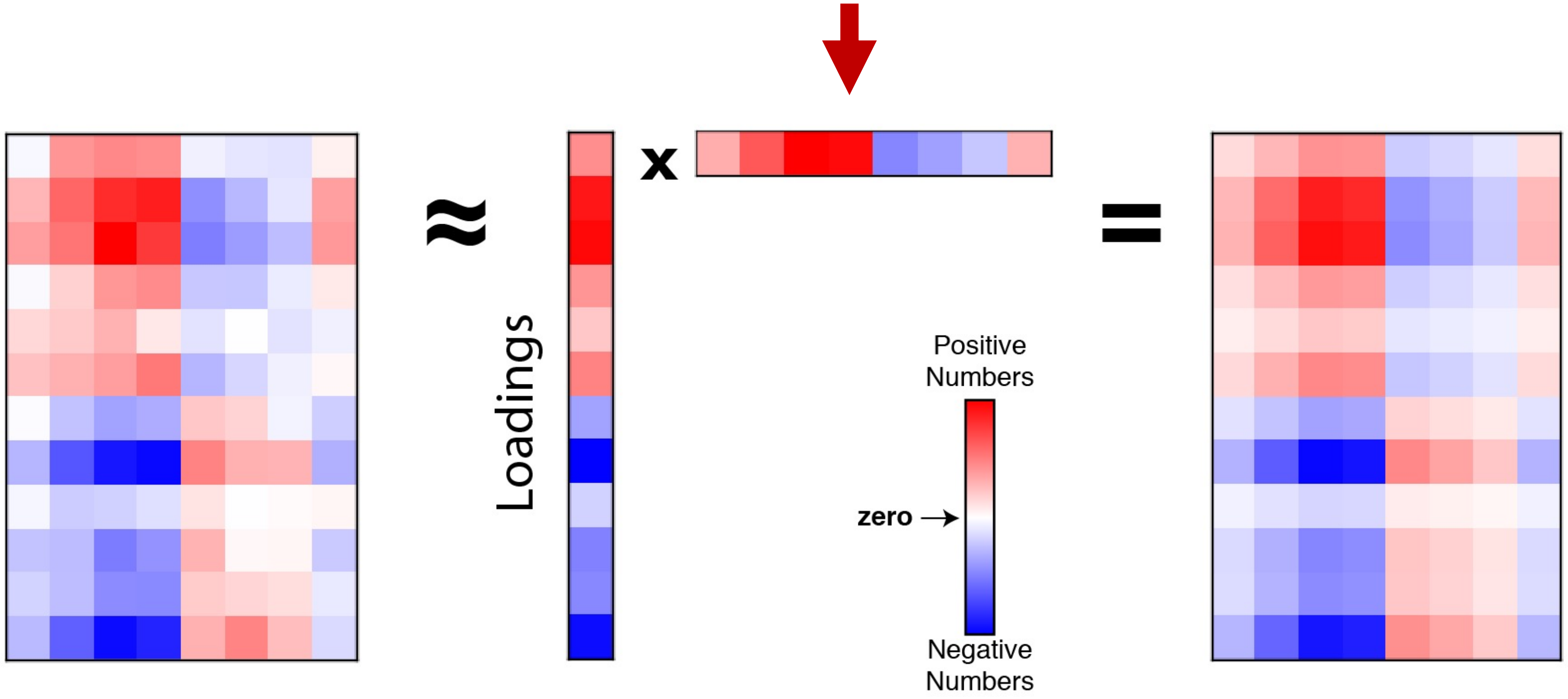
Bradshaw, M. & van der Willigen, R. F., (1999).
The walker's direction affects the perception of biological motion.
In M. A. Grealy & J. A. Thomson (Eds.),
Studies in perception and action V (pp. 3–6). London: Academic Press.



HTML5 BML Walker – BioMotionLab

{Maths: Sparse PCA}

Reconstruction of sensory input with 1 Principal Component (PC)



http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

PART II

Proof of Concept

PRIMATE Language Perception

{Speech Sensitivity hypothesis}

HOGESCHOOL ROTTERDAM
CREATING

Hallmark neurophysiological research focusing on macaque vocalizations implicates an evolutionary ancient cortical system to represent spectrotemporal modulations needed for speech recognition.

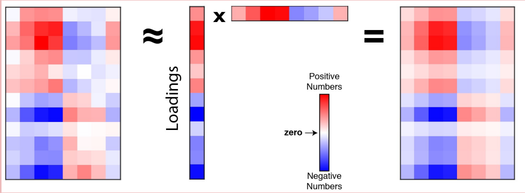
If the mechanism by which non-human primates process vocalizations extends to humans and occurs via **spectrotemporal separable** representation of Naturalistic sounds **{Rippled-noises}**

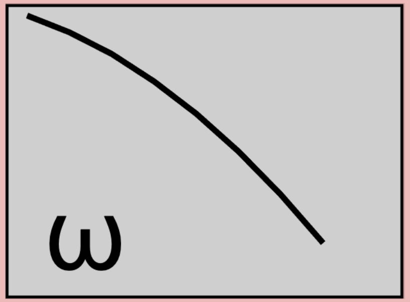
then the sensitivity to vocalizations in the primate brain is likely to be represented through a sparse coding scheme.

http://dx.doi.org/10.13140/RG.2.2.28232.03844
(Source van der Willigen et al. 2022)

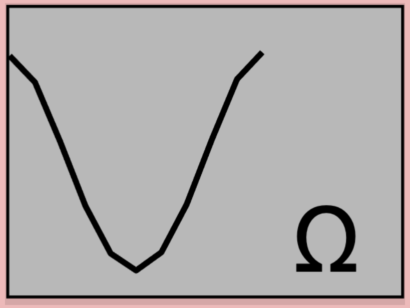
{Sparce coding Prediction}

Sparse coded

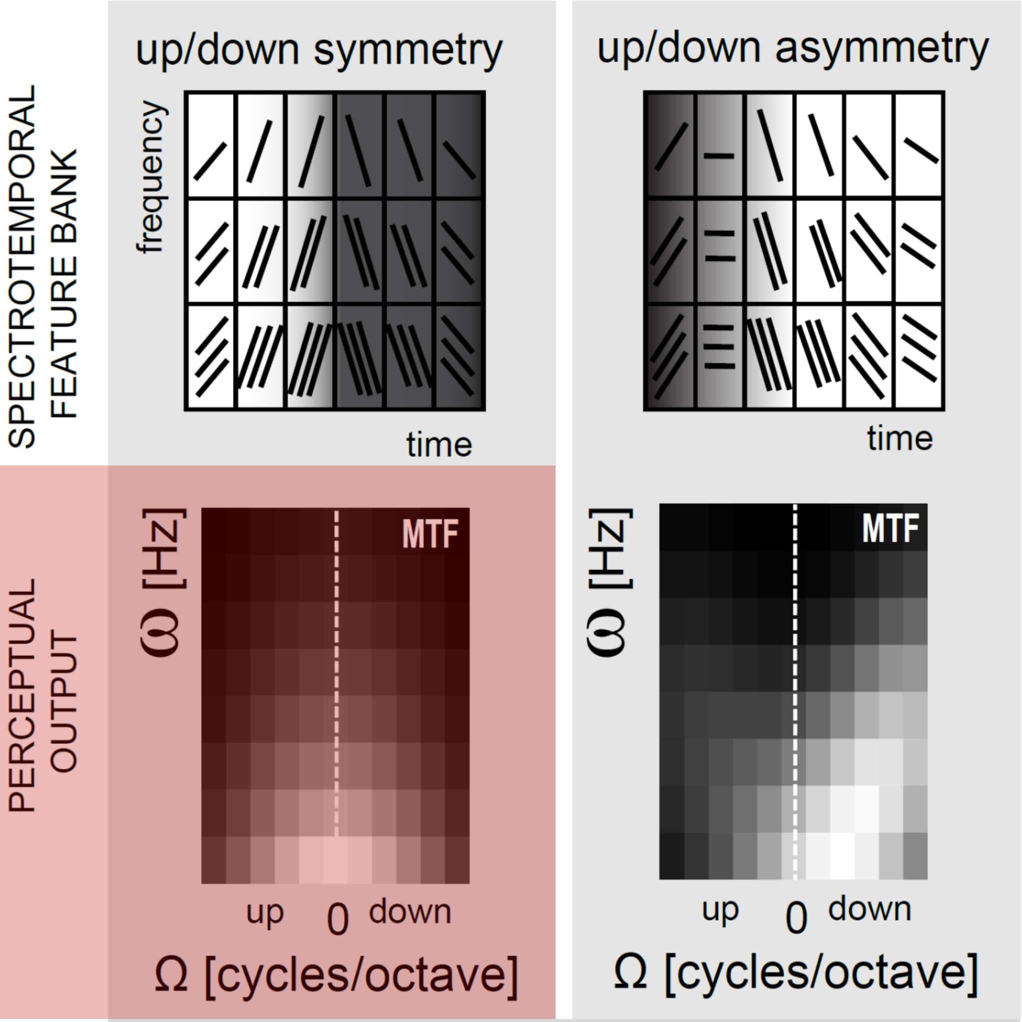




X



=



http://dx.doi.org/10.13140/RG.2.2.28232.03844

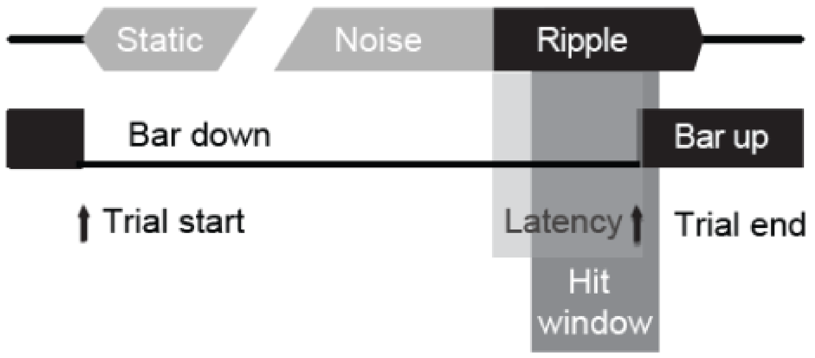
(Source van der Willigen et al. 2022)

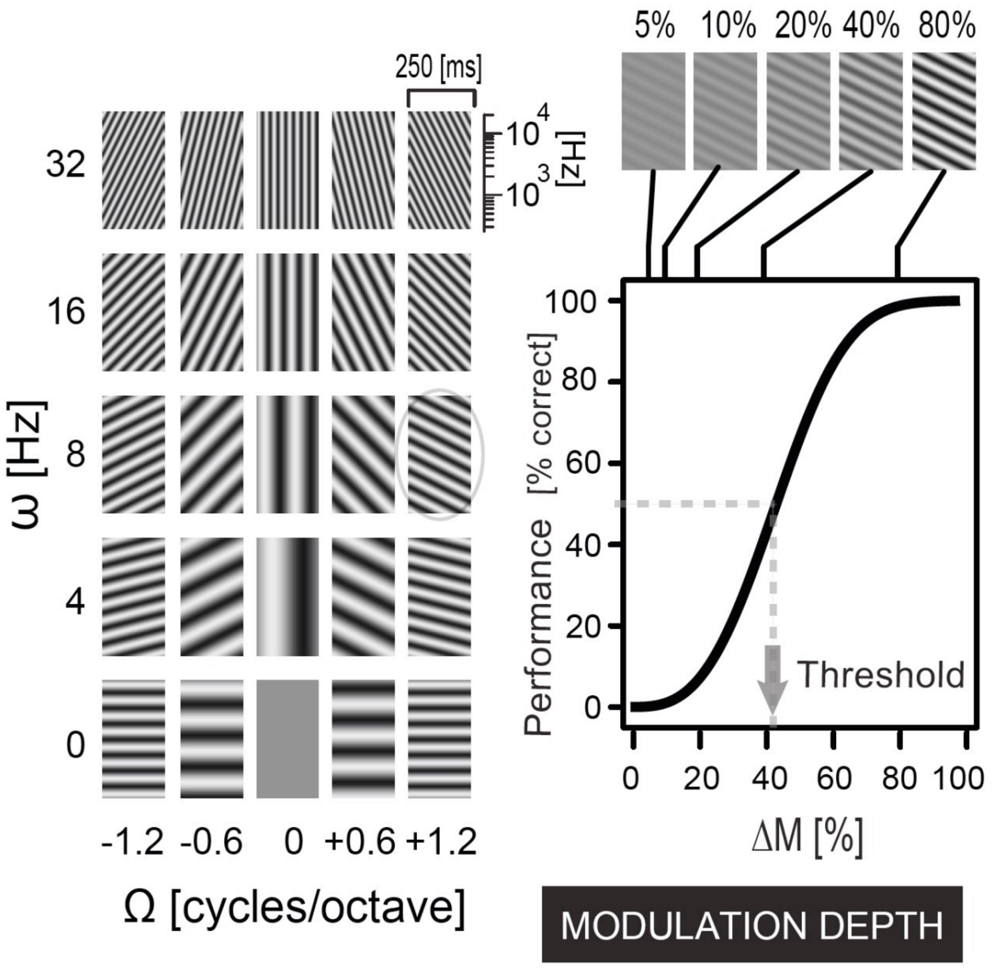
{Naturalistic Sounds}

HOGESCHOOL ROTTERDAM

CREATING

With the separable spectrotempral primate speech sensitivity hypothesis in mind; (N=5) humans and (N=5) monkeys were exposed to a wide range of naturalistic rippled-noises to characterize their perceptual abilities to process acoustic spectrotemporal modulations.





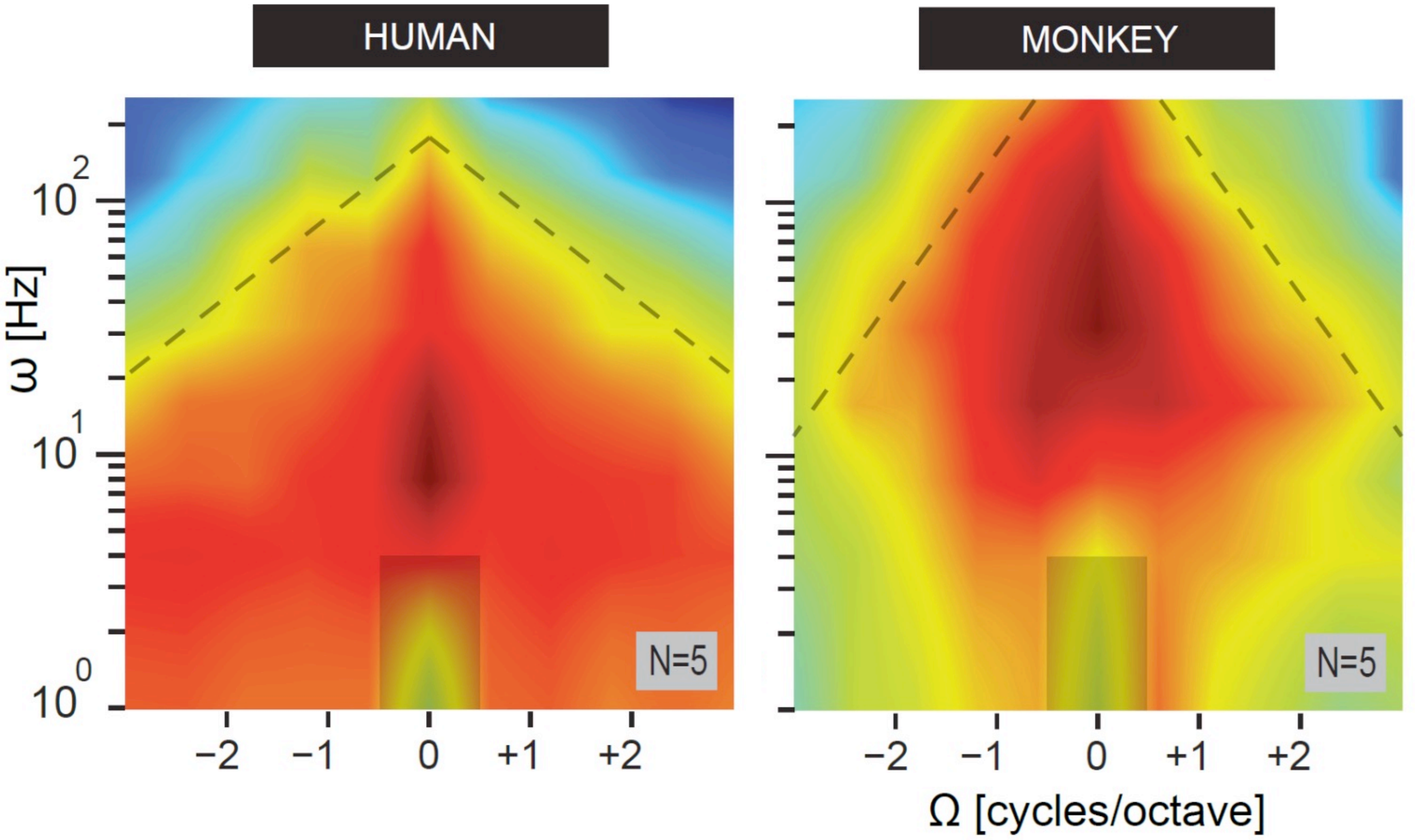
http://dx.doi.org/10.13140/RG.2.2.28232.03844

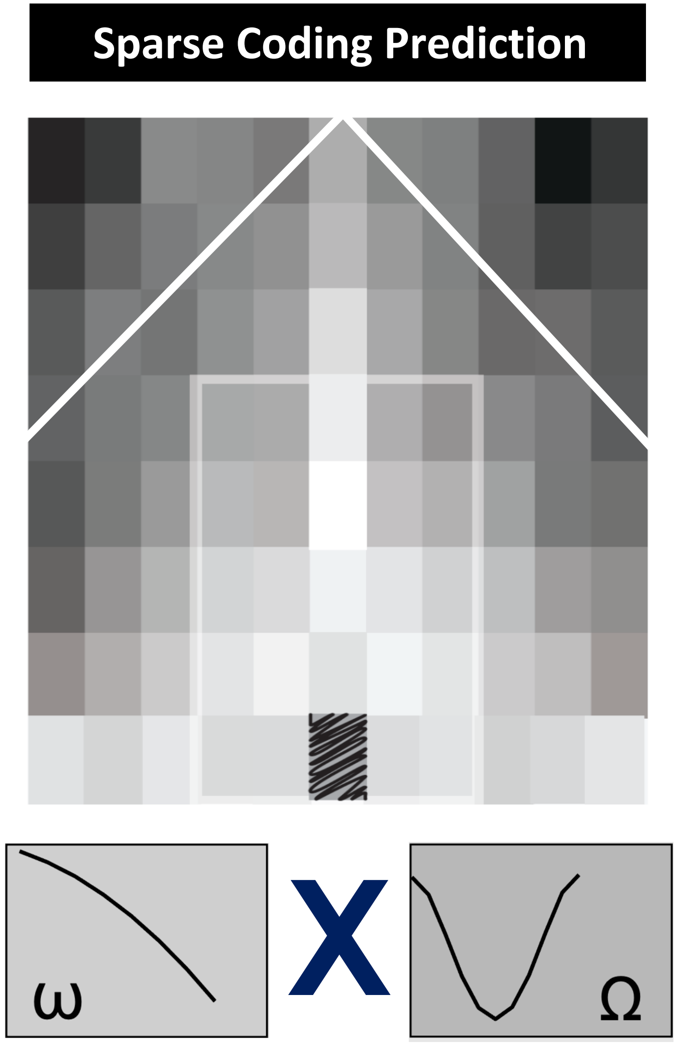
(Source van der Willigen et al. 2022)

{Empirical results confirm Sparse Coding}

HOGESCHOOL ROTTERDAM

CREATING





http://dx.doi.org/10.13140/RG.2.2.28232.03844
(Source van der Willigen et al. 2022)

OWL
Depth
vision

{Stereo Motion-Parallax hypothesis}

Stereopsis constitutes a passive viewing strategy, available to adult owls, for judging an object's three-dimensional shape based on the slight differences in the position of that object as viewed by the two eyes.

Yet, stereopsis alone cannot deliver metrically accurate information about depth relationships.

One possibility, then, is that:

Owls learn to overcome this limitation of stereopsis through Motion Parallax by making head movements while viewing 3D-objects stereoscopically.

(van der Willigen et al. 2022, submitted)

{Sparce Coding 3D-shape}

HOGESCHOOL ROTTERDAM

CREATING

When having access to disparity information at the same time as motion parallax , then the different ways in which judging object distance, D, affects perceived depth, d, from either stereopsis or motion parallax make it possible to obtain a metrically correct estimate of 3D shape.

Sparse coded

$$d = \eta \frac{D^2}{I} \quad \cap \quad D = \frac{D_1 + D_2}{2}$$

Stereopsis

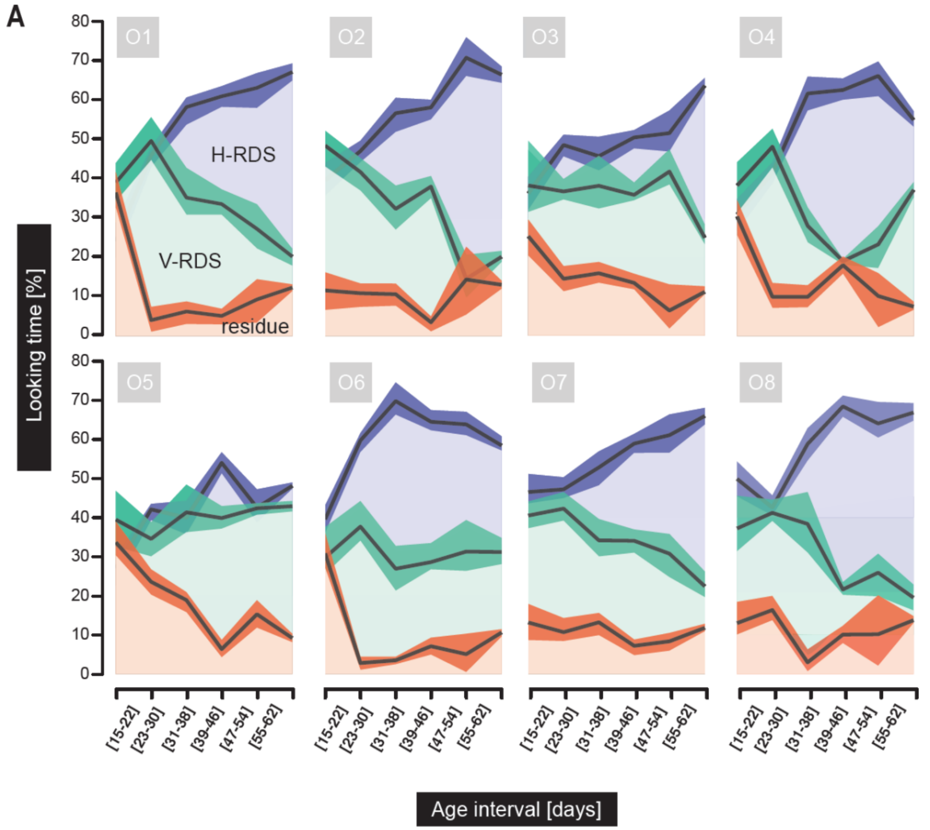
$$d = D_1 - D_2 = \frac{A}{2\tan(\theta_1/2)} - \frac{A}{2\tan(\theta_2/2)}$$

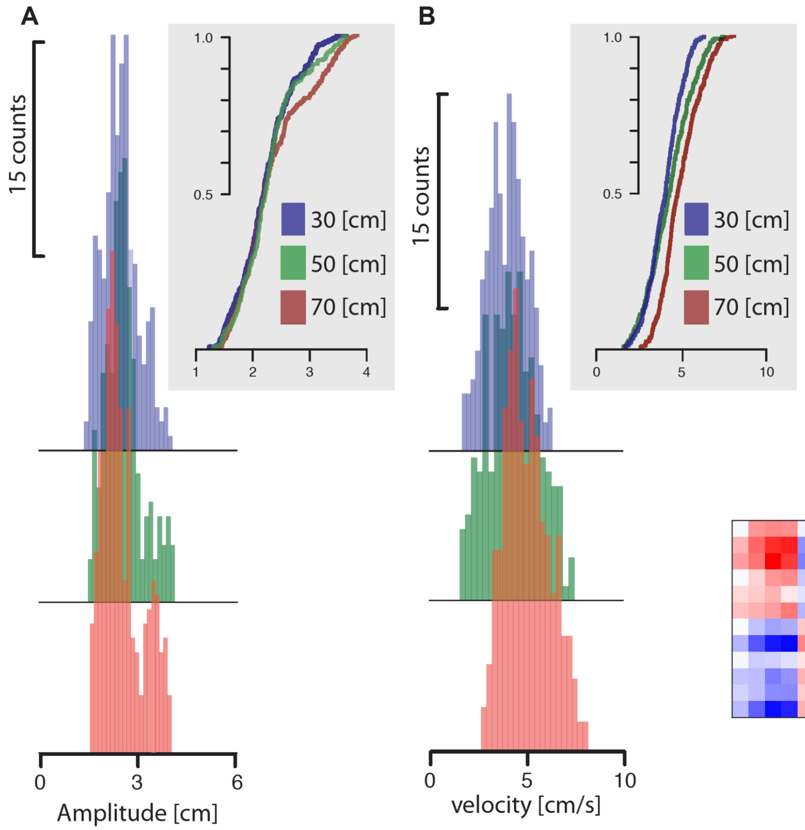
motion parallax

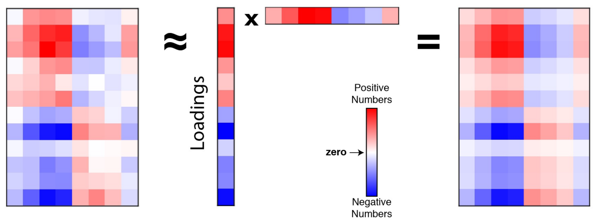
If the same observer monitors the difference in motion parallax of the depth interval between the back, D1, and front, D2, of the object, it follows then that there is only one solution for D whereby stereopsis and motion parallax specify the same amount of depth.

(van der Willigen et al. 2022, submitted)

{Empirical results confirm Sparse Coding}







Stereopsis X **motion parallax** = **3D shape**

(van der Willigen et al. 2022, submitted)

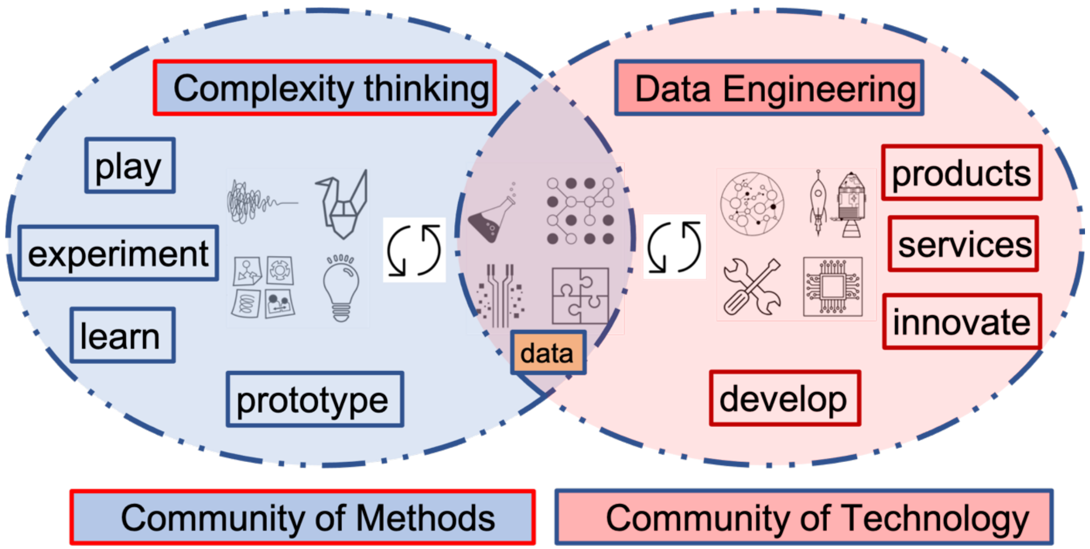
PART III

Living-Lab: AiRA, Hub for Data & Responsible {AI}

PoP-UP AiRA-Lab: PROMETHEUS

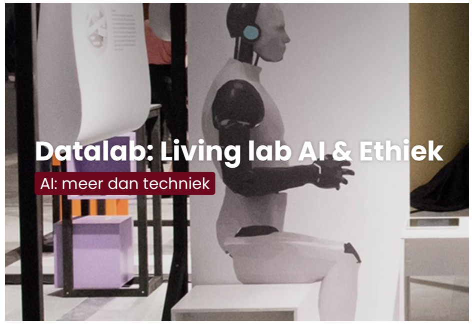
HOGESCHOOL ROTTERDAM

Opbouwen van kennis & expertise
door *hands-on* seminatie van datatechnologie



https://robfvdw.medium.com/a-generic-approach-to-data-driven-activities-d85ad558b5fa

Creating **DATA FABRIC** by building a ***data infrastructure***



Inzetten van ***datatechnologie & AI*** *om te komen tot* ***Bildung van complexiteitsdenken****, enerzijds, en* ***het laagdrempelig, verantwoord delen van inzichten voor maatschappelijke vraagstukken die voortvloeien uit het op grote schaal automatisch verzamelen van data****, anderzijds.*

https://www.researchgate.net/publication/357933768_NO_MORE_SECRETS_AIRA_LivinG-Lab_AI_ETHICS#fullTextFileContent

Data infrastructure drivers

People

Data Access & Tooling

Knowledge Dissemination & Data Curation

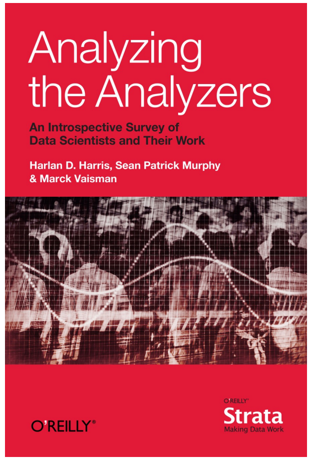
Interoperability & Standards

Cloud Computing Capabilities & Accessibility

{People}

Business	ML / Big Data	Math / OR	Programming	Statistics
Product Developement	Unstructured Data	Optimization	Systems Administration	Visualization
Business	Structured Data	Math	Back End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian / Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

https://mlleo.github.io/ml/GCP_day1/



Data Developer	Developer	Engineer	
Data Researcher	Researcher	Scientist	Statistician
Data Creative	Jack of All Trades	Artist	Hacker
Data Businessperson	Leader	Businessperson	Entrepeneur

{Data Access & Tooling}

HOGESCHOOL ROTTERDAM

Doing the right things

Doing things right

Discover
Research Phase

[0] Collection

[1] Access
+ Retrieval

Big Data (Acuisition/Aggregation) Gathering
Empirical(Sensor/IoT Measuring/Sampling)

Ownership (Open/Closed)
Storage (Cloud/Database)

Define
Synthesis Phase

[2] Preparation
+ Wrangling (Munging)

Loading
Feature Extraction/Reduction
Normalization
Transformation
Conversion

Develop
Ideation Phase

[3] Exploration

[4] Analysis
+ Machine-Learning

[5] Abstraction

Graphical (spatial)
Ontological (Language)
Semantic (text)
Rule-based/Algorithmic
Quantitative/Qualitative
Numerical/Categorical/Symbolic

Mining (Heuristics/Statistics/Descriptive/Prescriptive)
Construct Useful Insights/Trends/Patterns/Diagnosis(Information)

Parameter Selection + Representation
Summarization
Problem Solving
Diagnostic
Prediction
Encription

Deliver
Implementation Phase

[6] Organization
+ Managing

[7] Automation
+ Reporting

Visualization
Virtualization

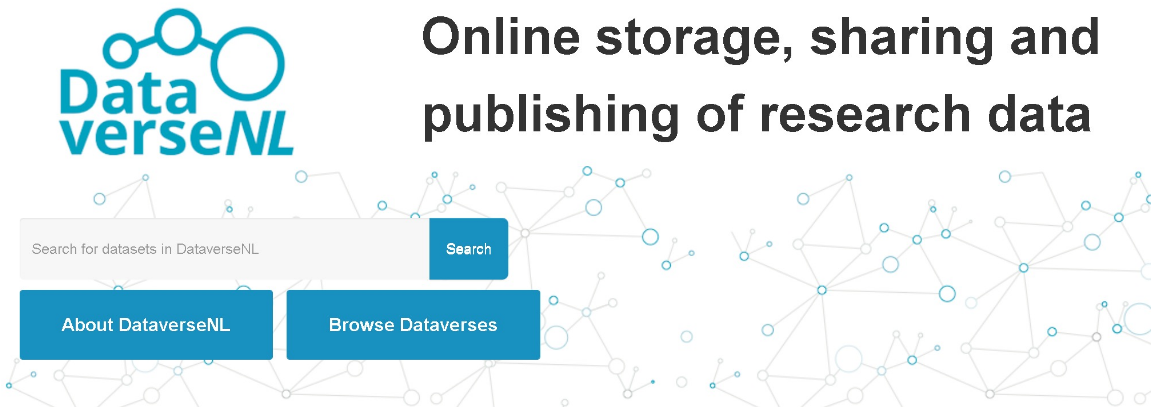
Performance (Measure/Monitor)
Evaluation & Review
Decision & Advise or Prescription
(Interactive/Passive) Story Telling

**Data
Gathering & Ingesting**

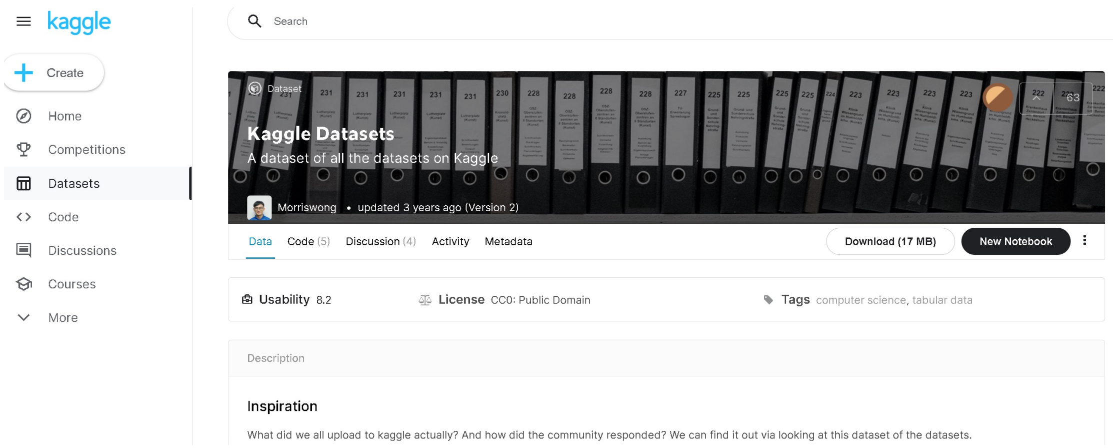
**Data
Dissemination & Curation**

{Data Access & Tooling}

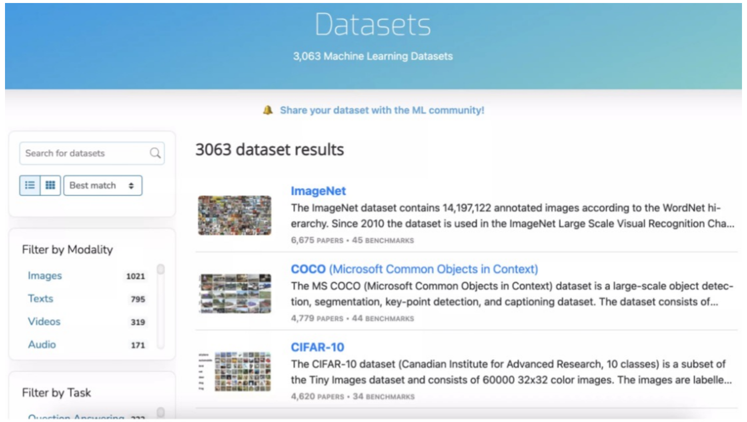
HOGESCHOOL ROTTERDAM



https://dataverse.nl/



https://www.kaggle.com/morriswongch/kaggle-datasets



paperswithcode.com

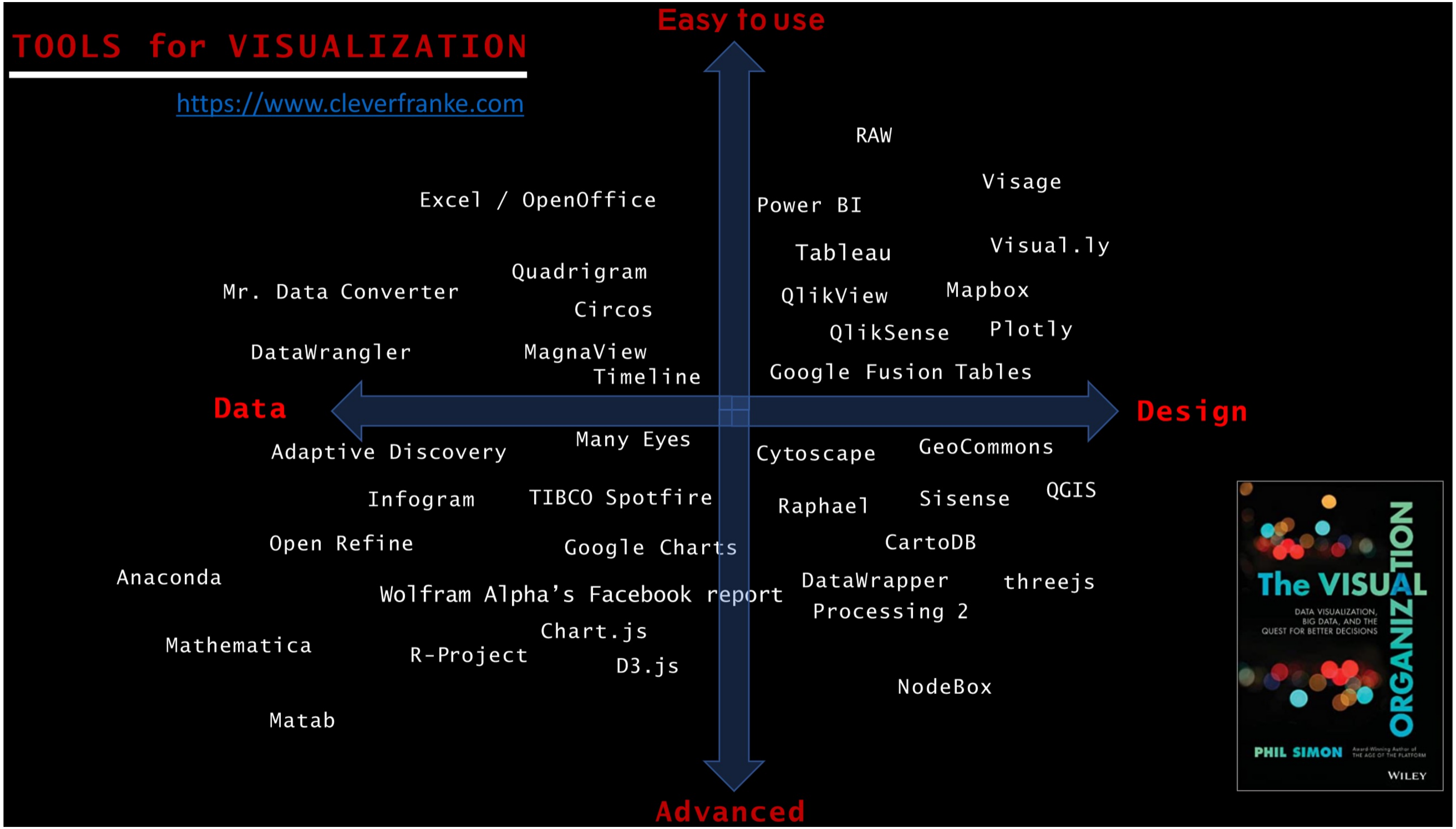
Papers with Code - Machine Learning Datasets

5577 datasets * 65018 papers with code.

https://paperswithcode.com/datasets

{Tooling}

HOGESCHOOL ROTTERDAM



{Knowledge Dissemination & Curation}

HOGESCHOOL ROTTERDAM

Employ decision-three flowcharts to apply (D)NNs appropriately

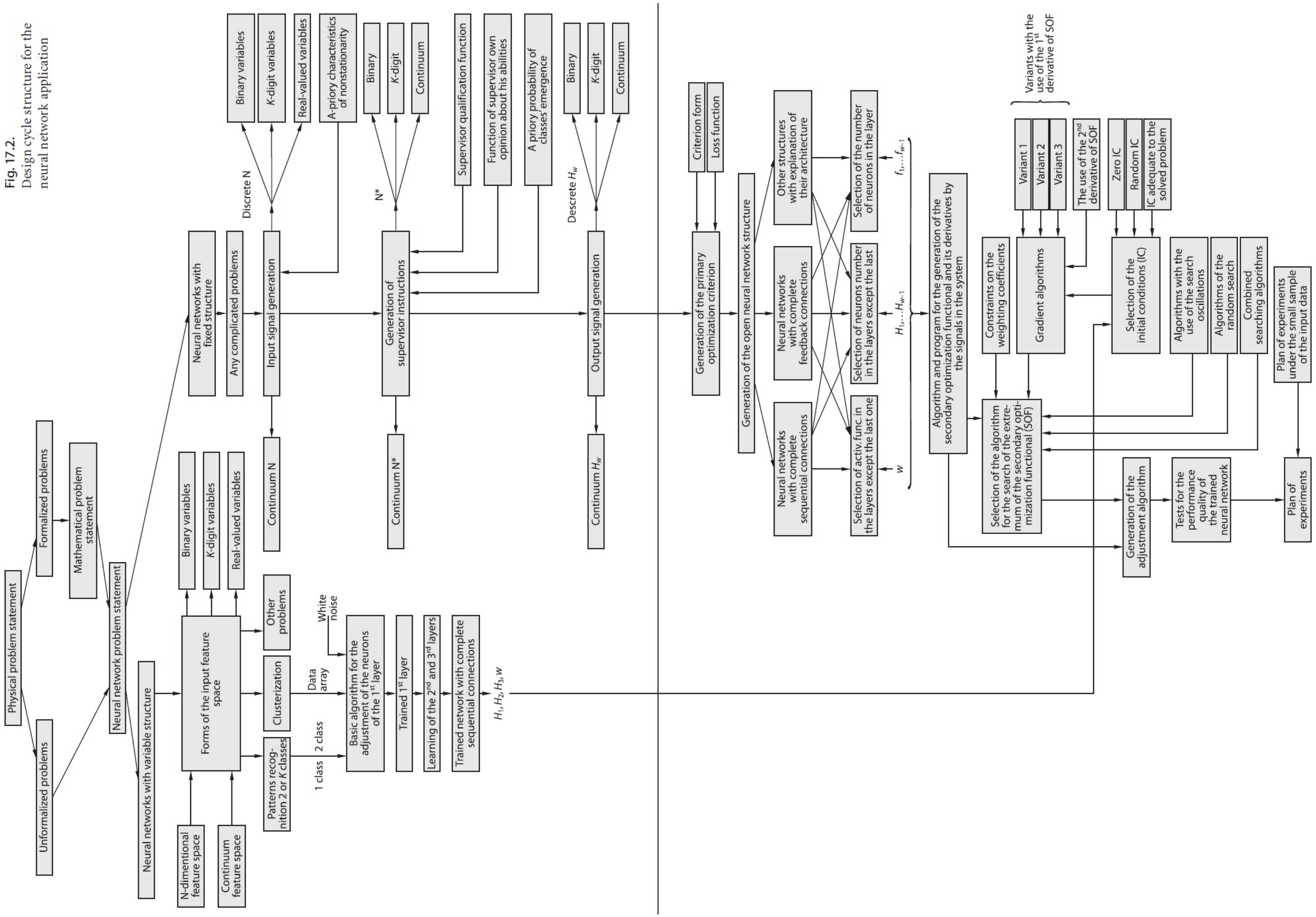
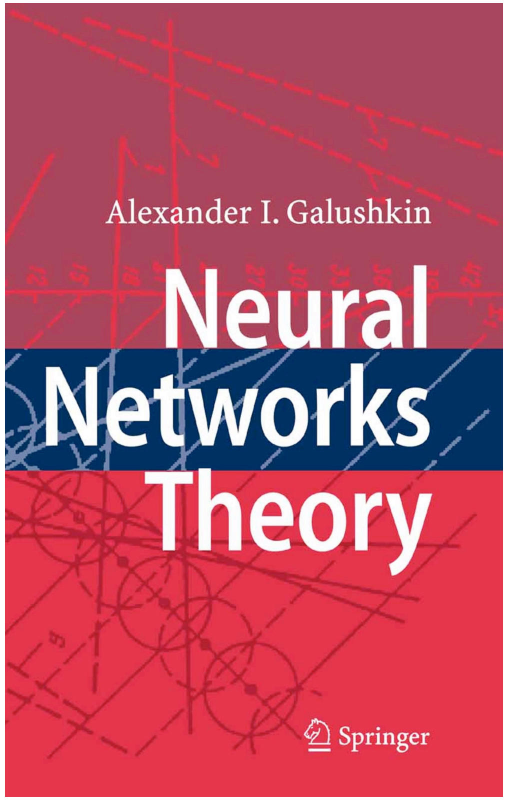


Fig. 17.2. Design cycle structure for the neural network application



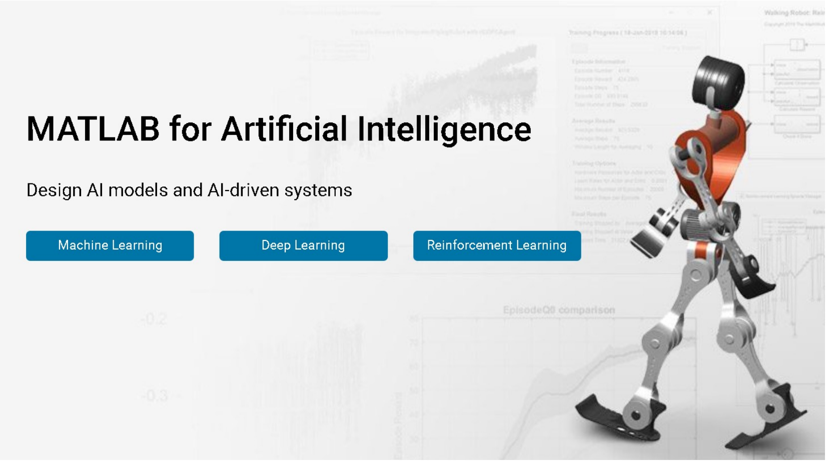
{Knowledge Dissemination & Curation}

HOGESCHOOL ROTTERDAM

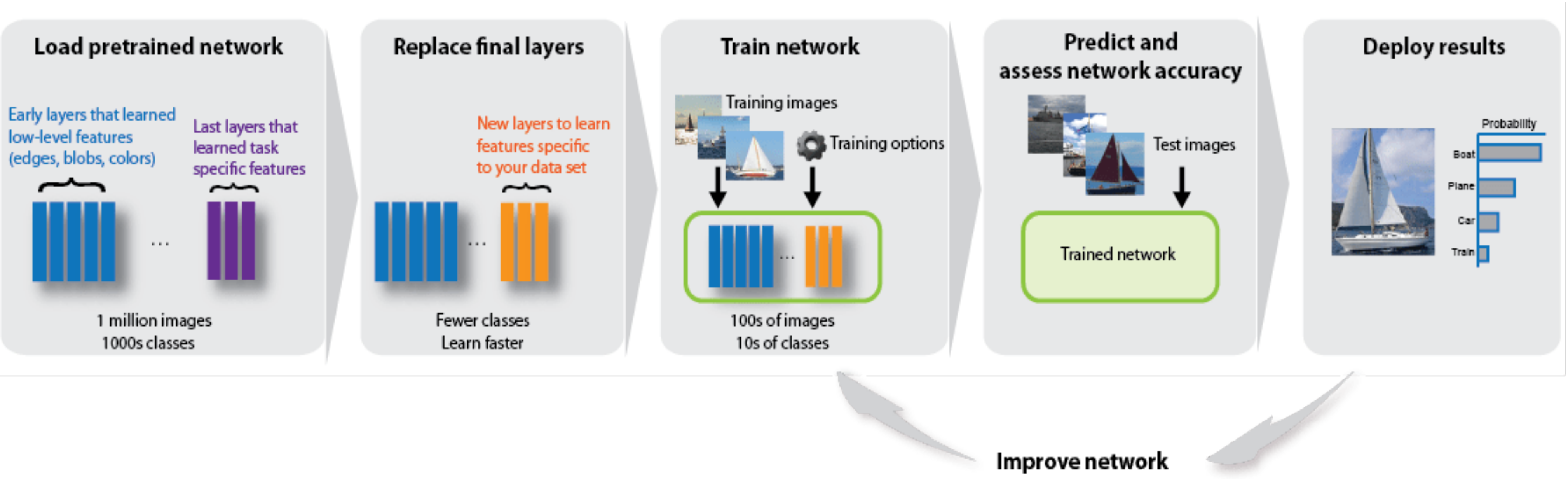
Hight quality Dutch reviews on state-of-the-art AI

{Knowledge Dissemination & Curation}

HOGESCHOOL ROTTERDAM



Hands-on demonstrations of how to reuse pre-trained DNNs on custom datasets by means of High-End Gaming-PCs

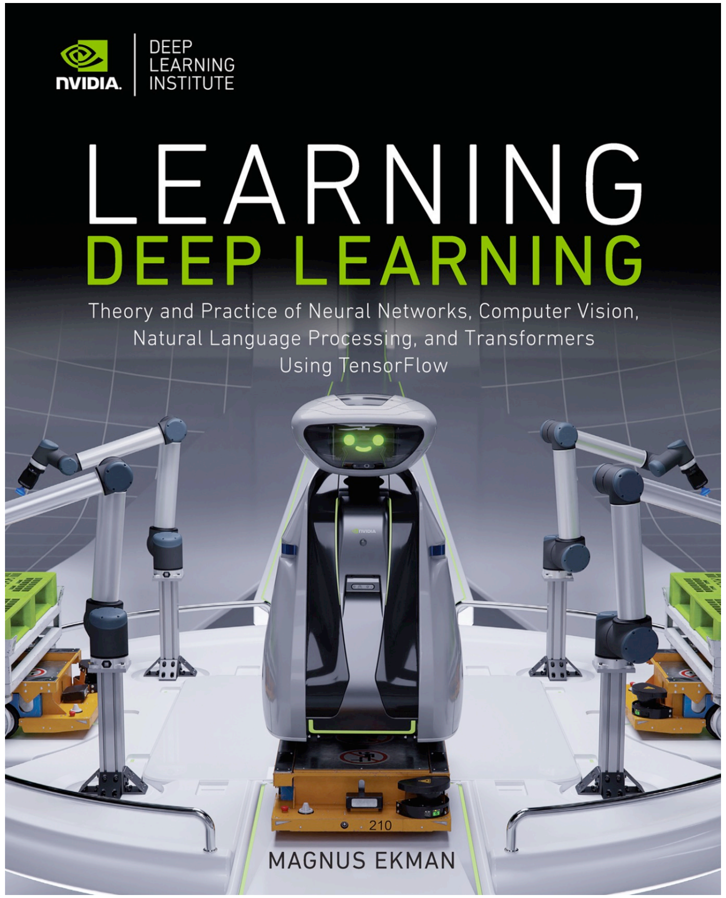


[Train Deep Learning Network to Classify New Images - MATLAB & Simulink - MathWorks Benelux]()

{Knowledge Dissemination & Curation}

HOGESCHOOL ROTTERDAM

Learn from prime examples



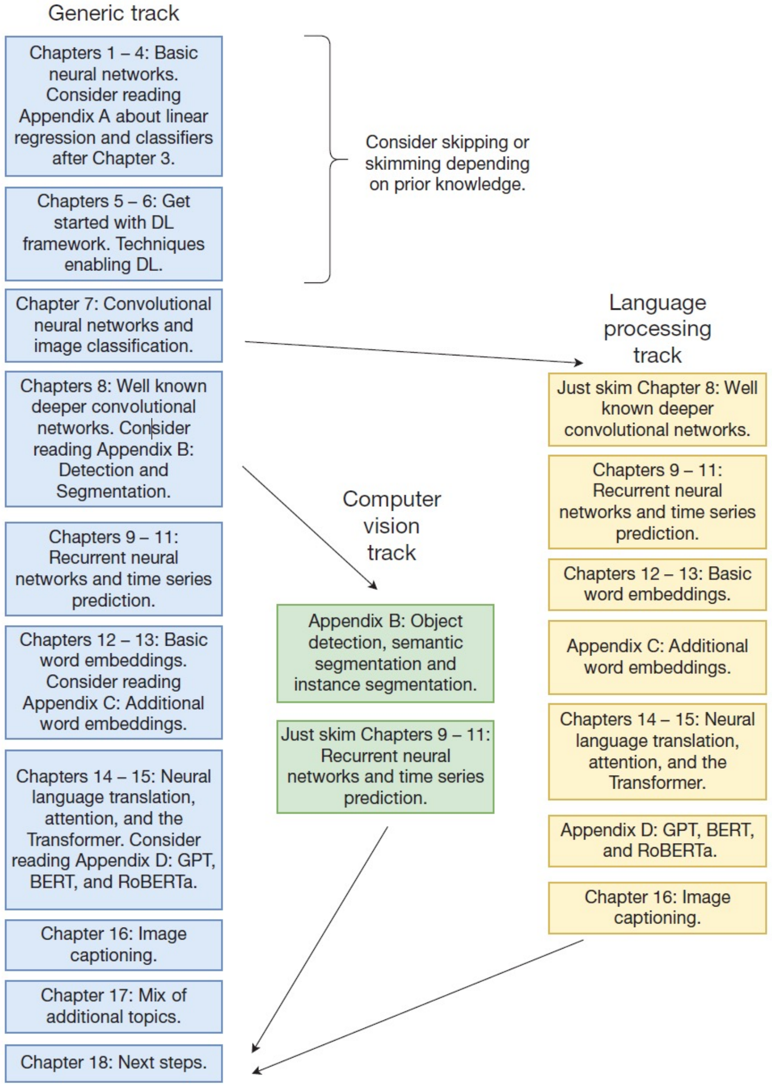


Figure P-5 Three different tracks to follow when reading this book

{Knowledge Dissemination & Curation}

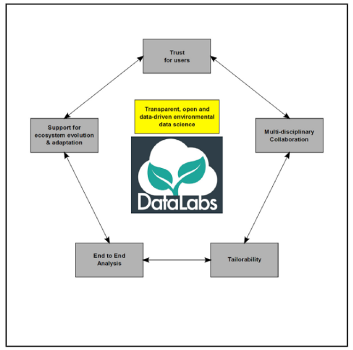
HOGESCHOOL ROTTERDAM

Learn from prime examples

Patterns

Tackling the Challenges of 21st-Century Open Science and Beyond: A Data Science Lab Approach

Graphical Abstract



Highlights

- Offer a vision of data science labs as open, collaborative platforms in the cloud
- Discussion of how data science labs support open and transparent science
- Discussion of experiences around implementing data labs in practice
- The definition of a roadmap of research challenges around virtual data labs

Authors

Michael J. Hollaway, Graham Dean, Gordon S. Blair, Mike Brown, Peter A. Henrys, John Watkins

Correspondence

mhollaway@ceh.ac.uk

In Brief

We present the "data science lab" concept as a potential solution to champion cross-disciplinary and open science. Data science labs are cloud-based, collaborative, and tailorable platforms enabling users with different requirements and expertise to find data-driven solutions to a wide range of environmental challenges. We present examples of methodological and infrastructural developments using data science labs along with a detailed research roadmap to serve as a focal point for developing a more data-driven and transparent approach to environmental data science.

Hollaway et al., 2020, Patterns 1, 100103
October 9, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.patter.2020.100103

CellPress

THEME ARTICLE: JUPYTER IN COMPUTATIONAL SCIENCE

Using Jupyter for Reproducible Scientific Workflows

Marijan Beg, *Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ, Southampton, U.K.*

Juliette Taka, *Logilab, 75013, Paris, France*

Thomas Kluyver, *European XFEL, 22869, Schenefeld, Germany*

Alexander Konovalov, *School of Computer Science, University of St. Andrews, KY16 9SX, St. Andrews, U.K.*

Min Ragan-Kelley, *Simula Research Laboratory, 1364, Fornebu, Norway*

Nicolas M. Thiéry, *Laboratoire de Recherche en Informatique, Université Paris-Saclay, CNRS, 91405, Orsay, France*

Hans Fangohr, *Max Planck Institute for the Structure and Dynamics of Matter, 22761, Hamburg, Germany*

Literate computing has emerged as an important tool for computational studies and open science, with growing folklore of best practices. In this work, we report two case studies—one in computational magnetism and another in computational mathematics—where domain-specific software was exposed to the Jupyter environment. This enables high level control of simulations and computation, interactive exploration of computational results, batch processing on HPC resources, and reproducible workflow documentation in Jupyter notebooks. In the first study, Ubermag drives existing computational micromagnetics software through a domain-specific language embedded in Python. In the second study, a dedicated Jupyter kernel interfaces with the GAP system for computational discrete algebra and its dedicated programming language. In light of these case studies, we discuss the benefits of this approach, including progress toward more reproducible and reusable research results and outputs, notably through the use of infrastructure such as JupyterHub and Binder.

Research usually results in a publication that presents and shares the obtained findings and conclusions. For a publication to be scientifically valid, it must present the methodology rigorously, so that readers can follow the "recipe" and reproduce the results. If this criterion is met, the publication is considered reproducible. *Reproducible* publications are more easily *reusable* and, thus, provide a significant opportunity to make (often tax-payer funded) research more impactful. However, the reproducibility of computational work is usually hindered not only by a lack of data or metadata but also by a lack of details on the procedure and tools used.

1) The source code of the software used is not available.
2) Information on the computing environment, such as the hardware, operating system, supporting libraries, and (if required) code compilation details is not revealed.
3) The exact procedure, which led to the results reported in the publication, is not shared. This should include the set of parameters used, the simulation and data analysis procedure, and any additional data cleaning, processing, and visualization. Ideally, these are shared as open-source code and analysis scripts used to perform the simulation and to read, analyze, and visualize the resulting data. This way, the entire process can be repeated by rerunning simulation and/or analysis scripts. A human-readable document detailing the computational steps taken, despite being "better-than-

1521-9615 © 2021 IEEE
Digital Object Identifier 10.1109/MCSE.2021.3052101
Date of publication 15 January 2021; date of current version 25 March 2021.

[Tackling the Challenges of 21st-Century Open Science and Beyond: A Data Science Lab Approach - ScienceDirect]

[Using Jupyter for Reproducible Scientific Workflows | IEEE Journals & Magazine | IEEE Xplore]

{Knowledge Dissemination & Curation}

HOGESCHOOL
ROTTERDAM

Learn from prime examples

RESEARCH BRIEFS | DECEMBER 17, 2020

Artificial Intelligence and the Future of Work

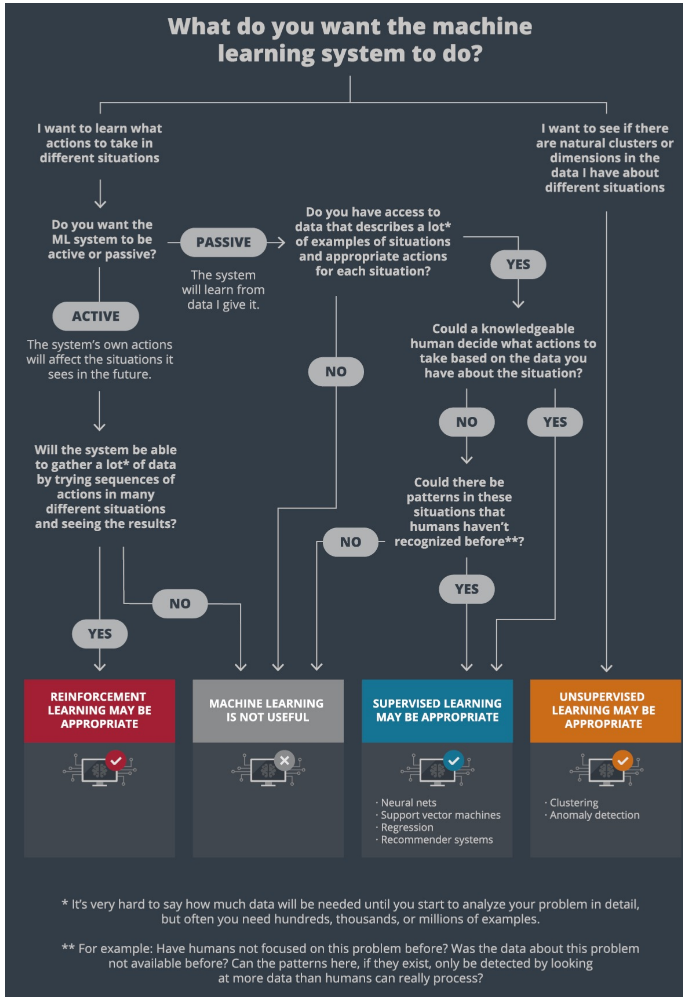
THOMAS MALONE, DANIELA RUS, ROBERT LAUBACHER

RB17-2020

SHARE

DOWNLOAD RESEARCH BRIEF

https://workofthefuture.mit.edu/wp-content/uploads/2020/12/2020-Research-Brief-Malone-Rus-Laubacher2.pdf

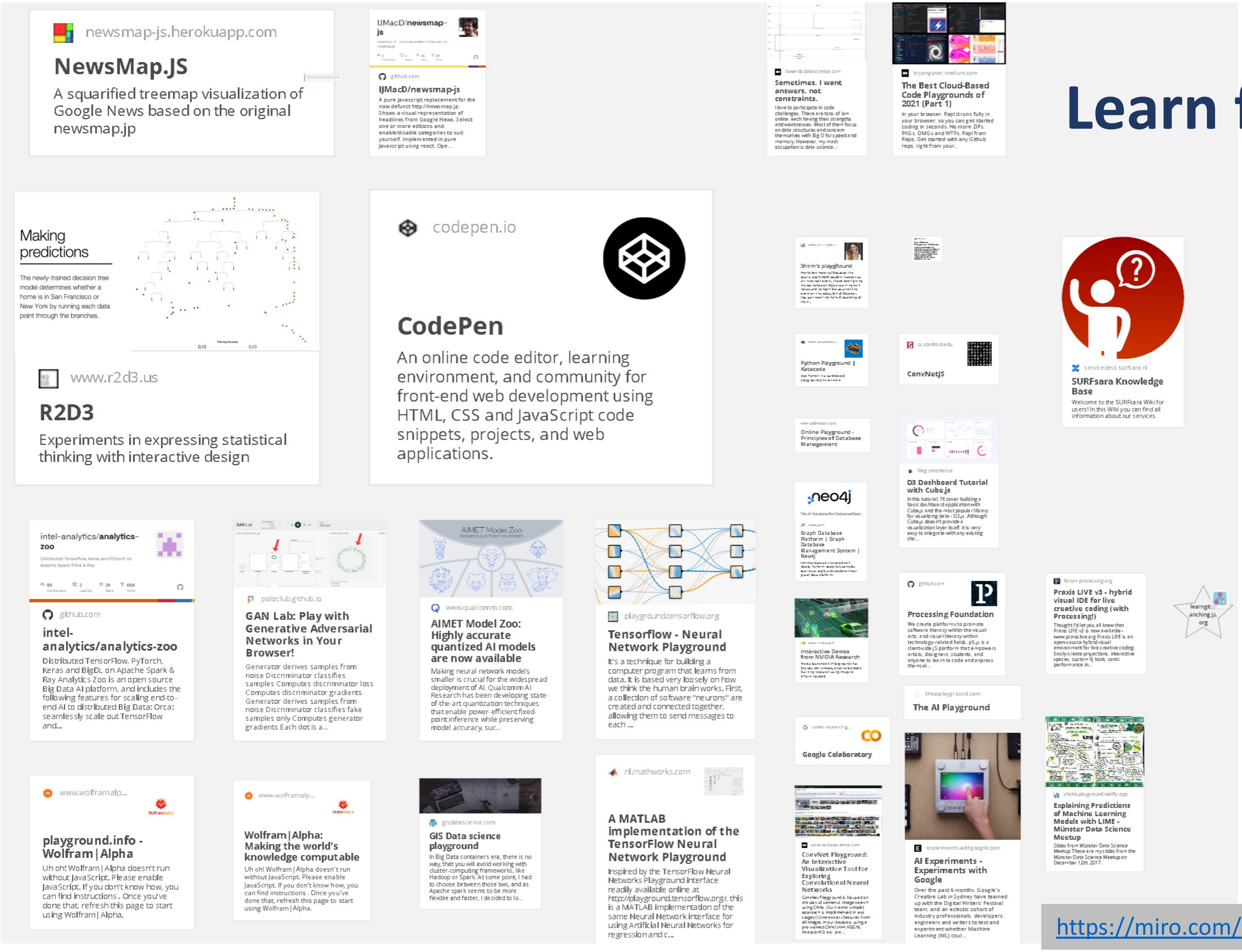


{Knowledge Dissemination & Curation}

HOGESCHOOL ROTTERDAM

Learn from prime examples

DATA Science Playgrounds

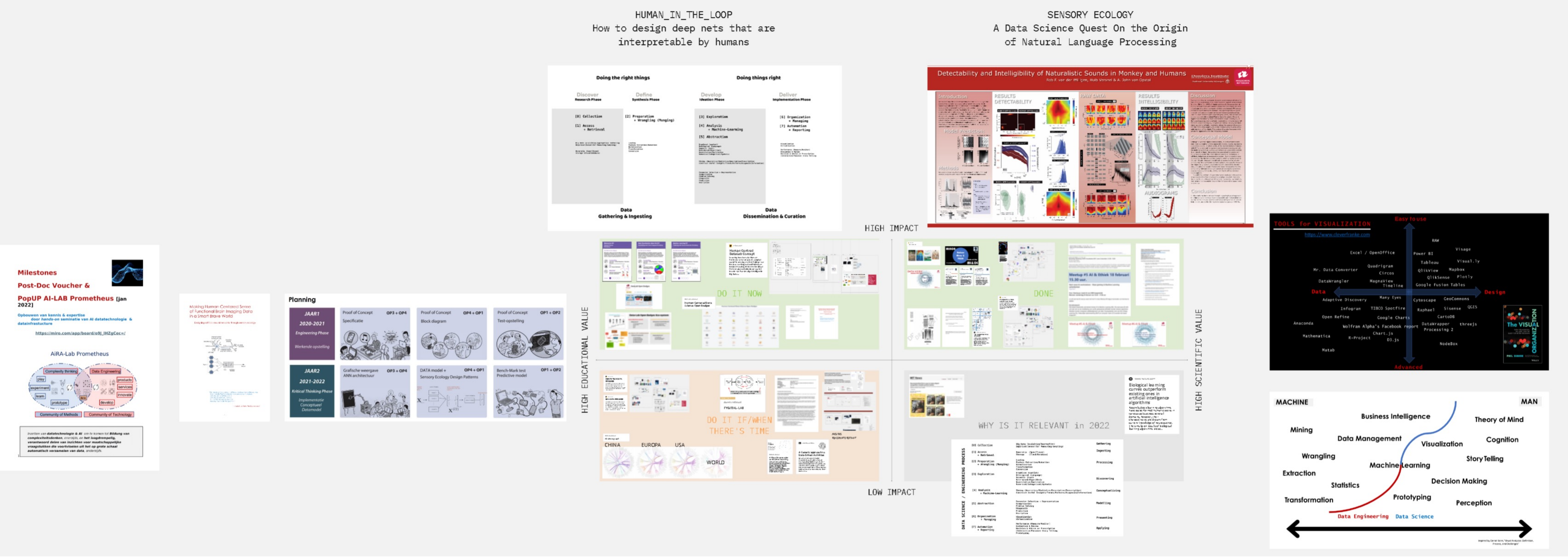


https://miro.com/app/board/uXjVOCnGVbU=/?invite_link_id=716932584109

{Knowledge Dissemination & Curation}

HOGESCHOOL
ROTTERDAM

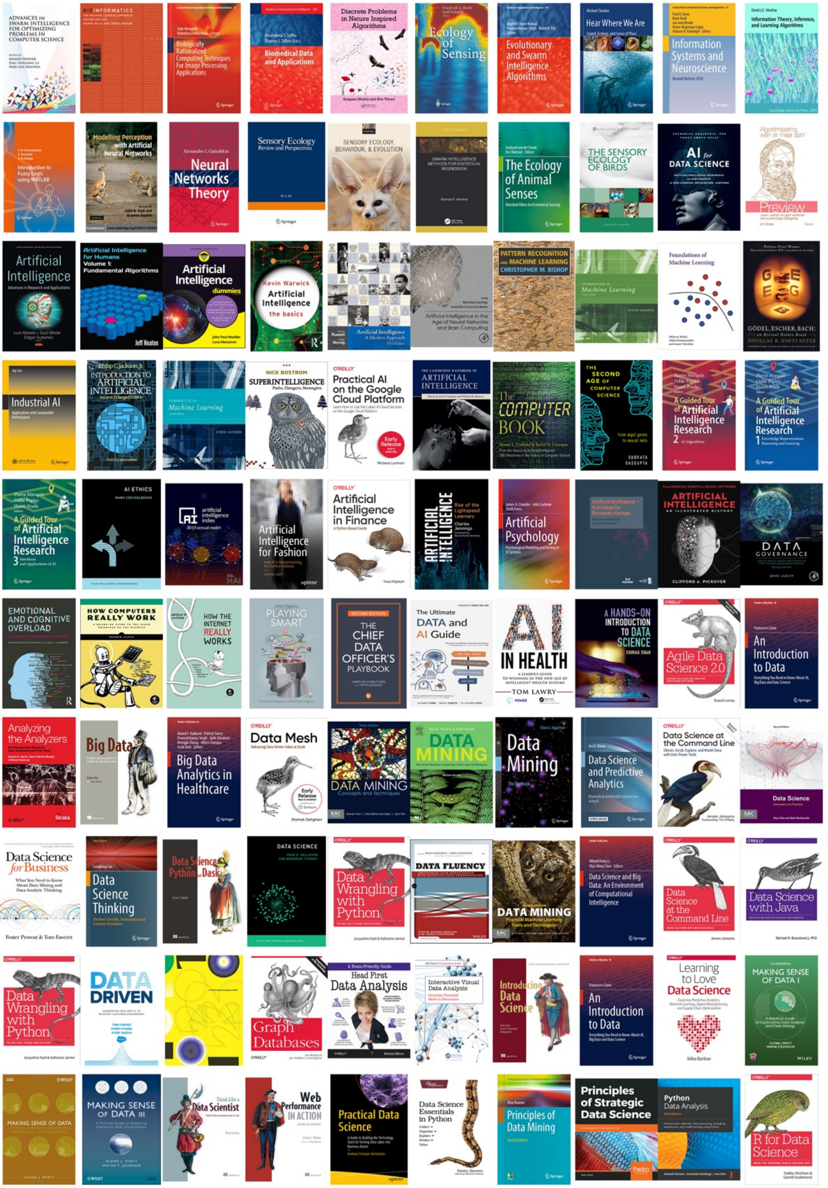
Post-Doc Voucher Milestones

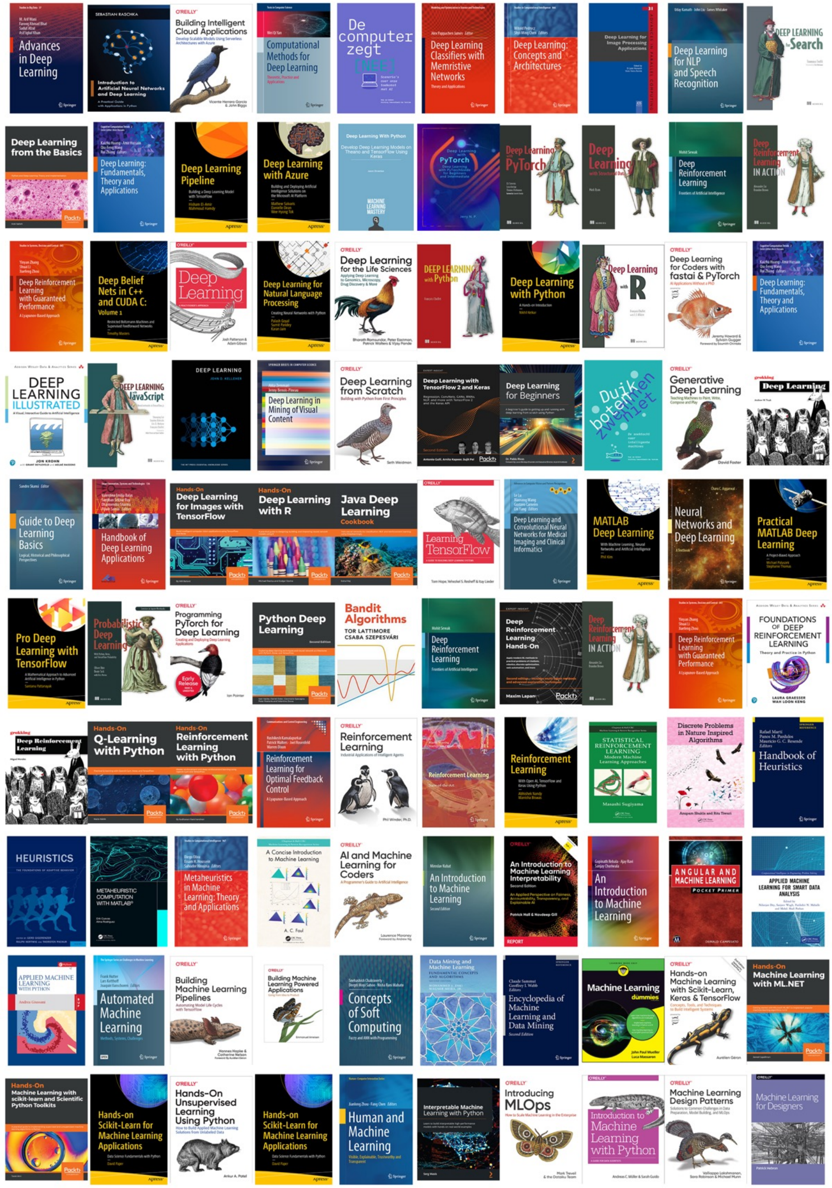


https://miro.com/app/board/o9J_l-CrLzg=/?invite_link_id=427711164595

{Studied Materials: books}

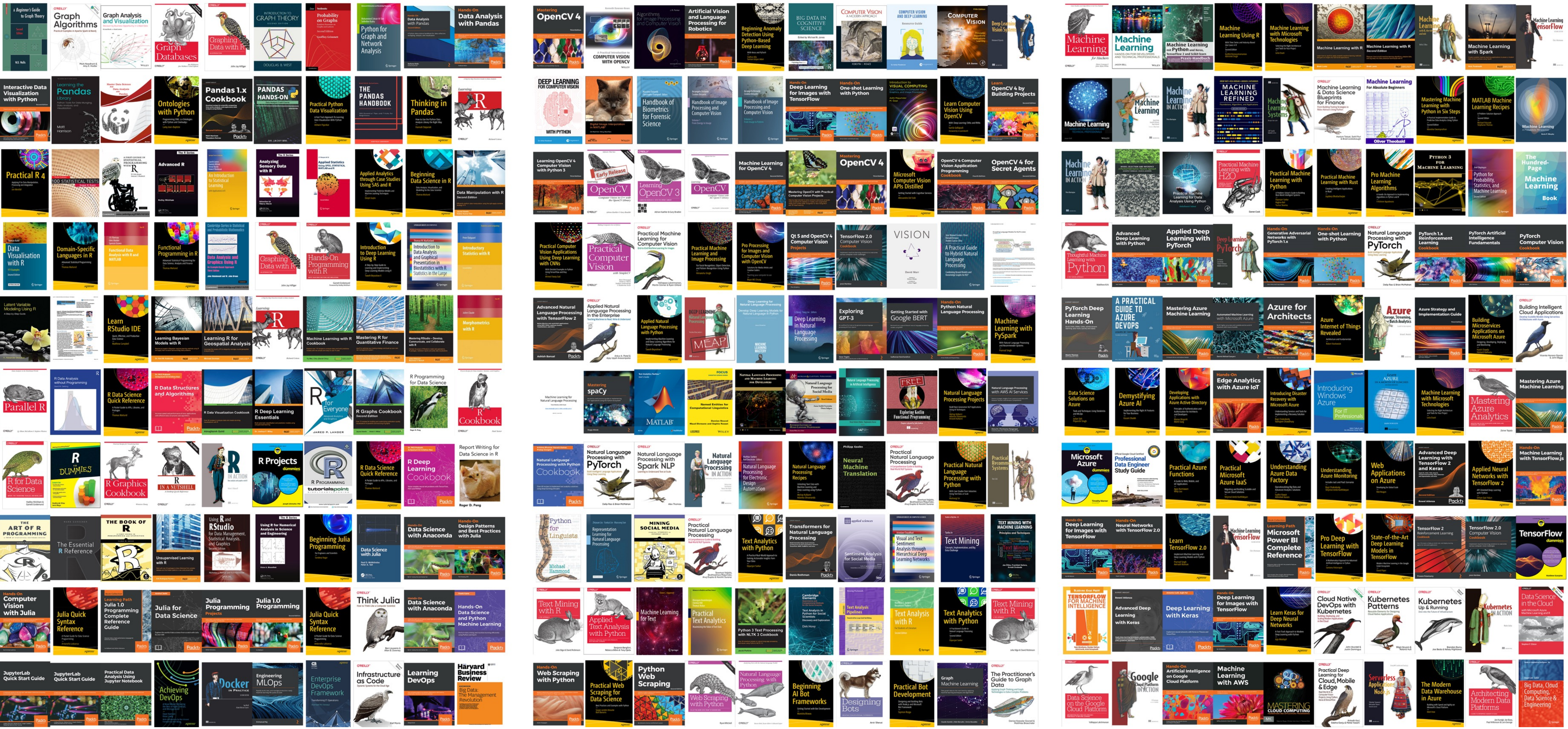
HOGESCHOOL ROTTERDAM





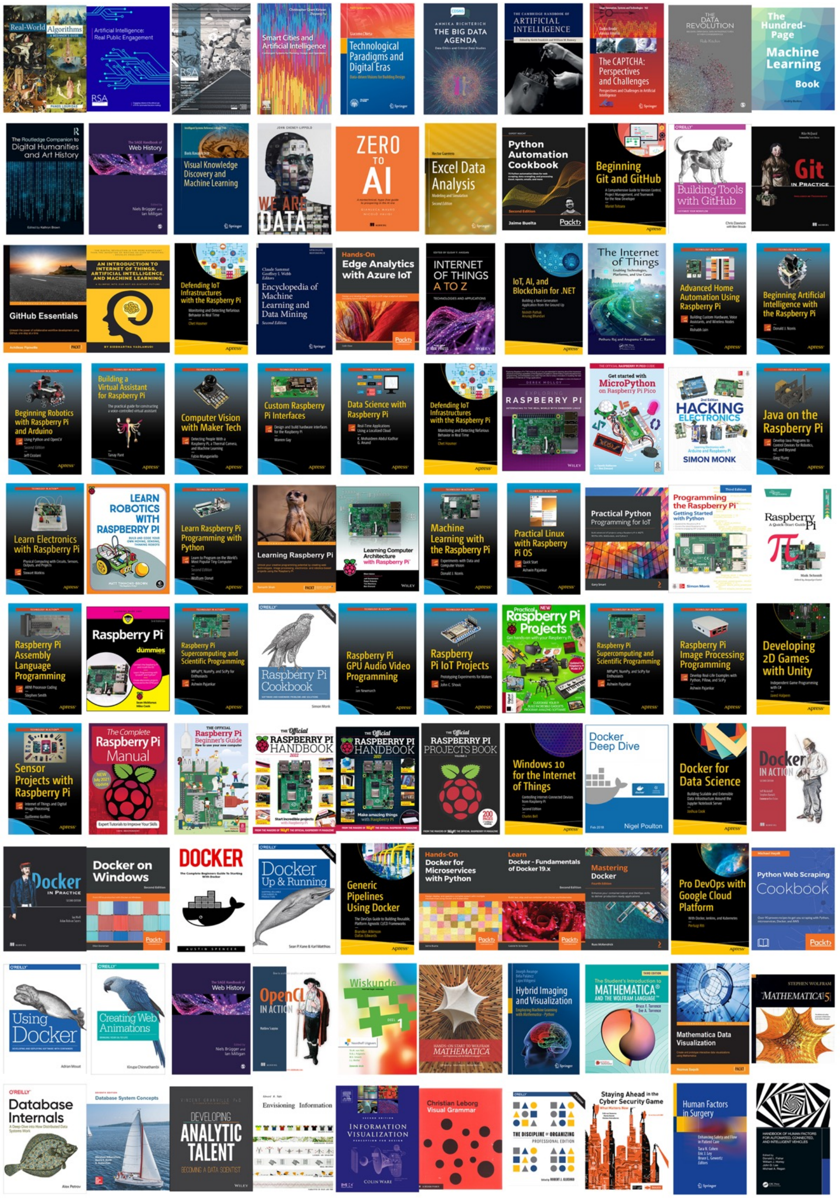
{Studied Materials: books}

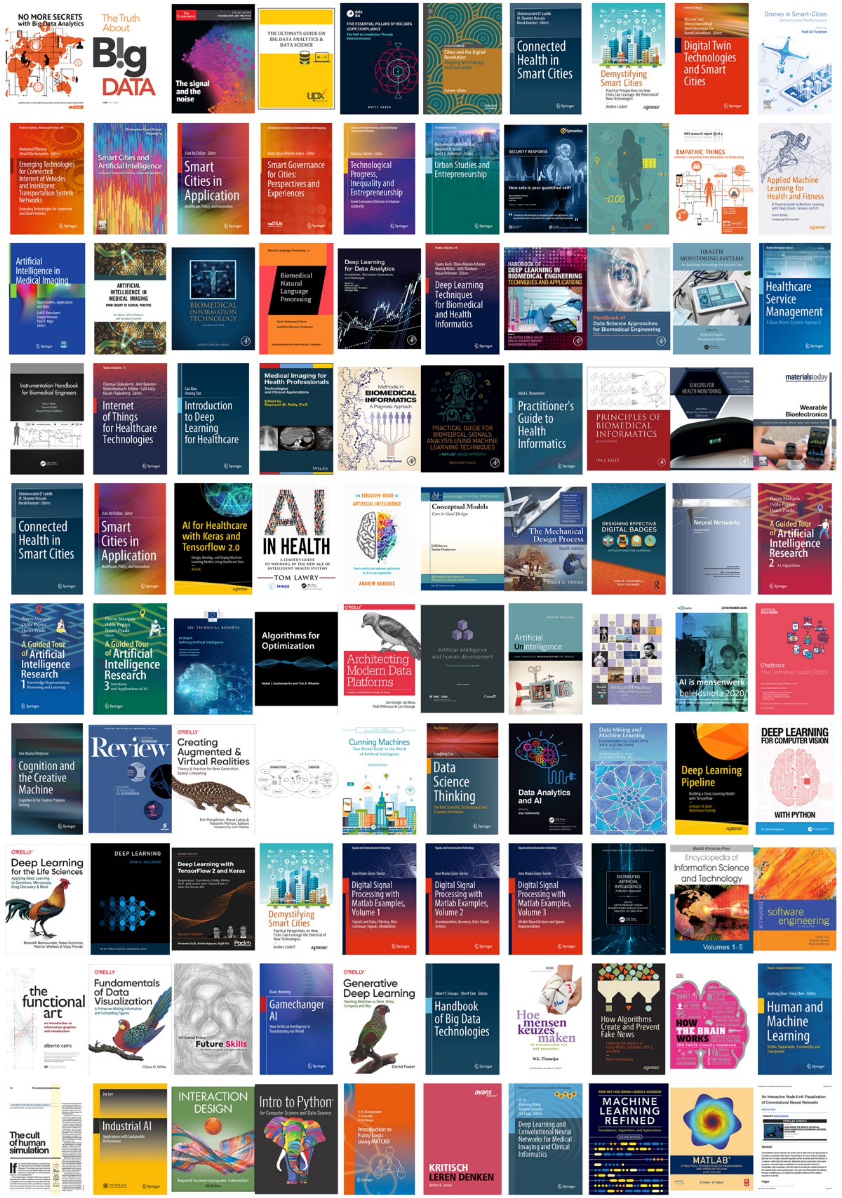
HOGESCHOOL ROTTERDAM

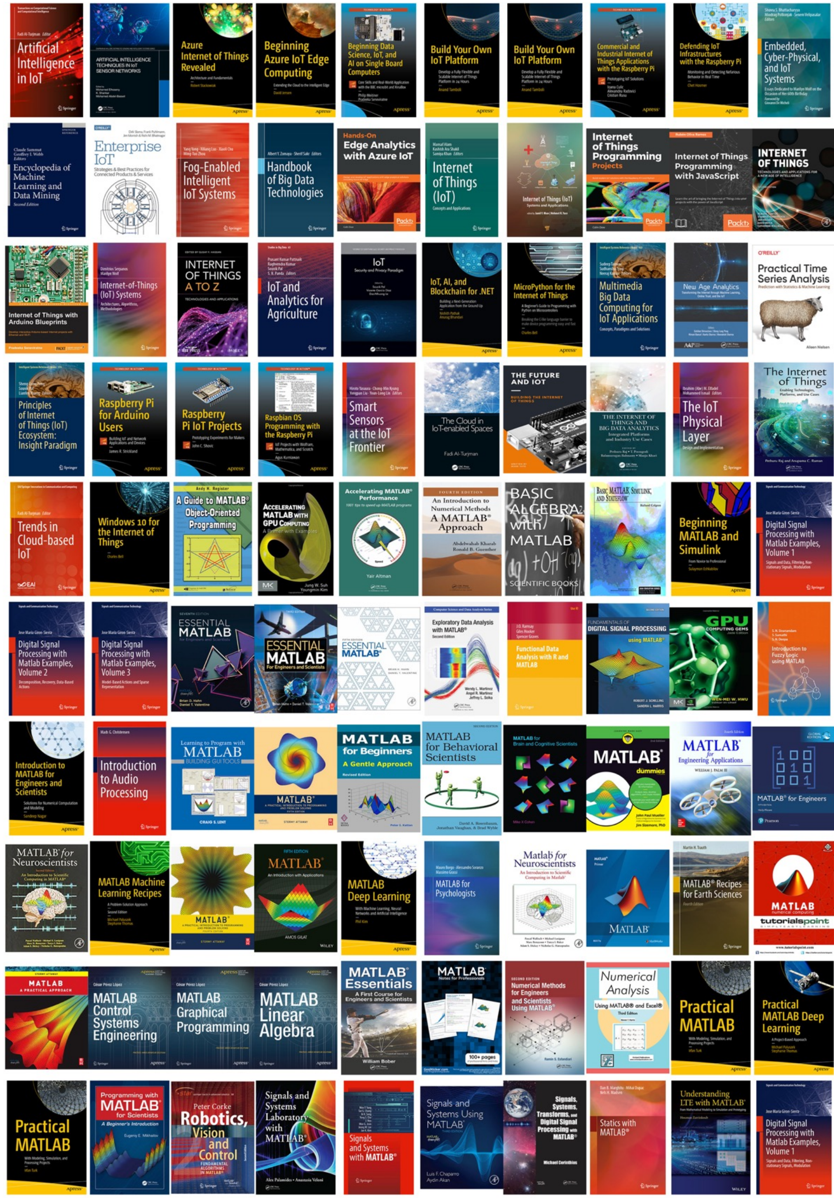


{Studied Materials: books}

HOGESCHOOL ROTTERDAM

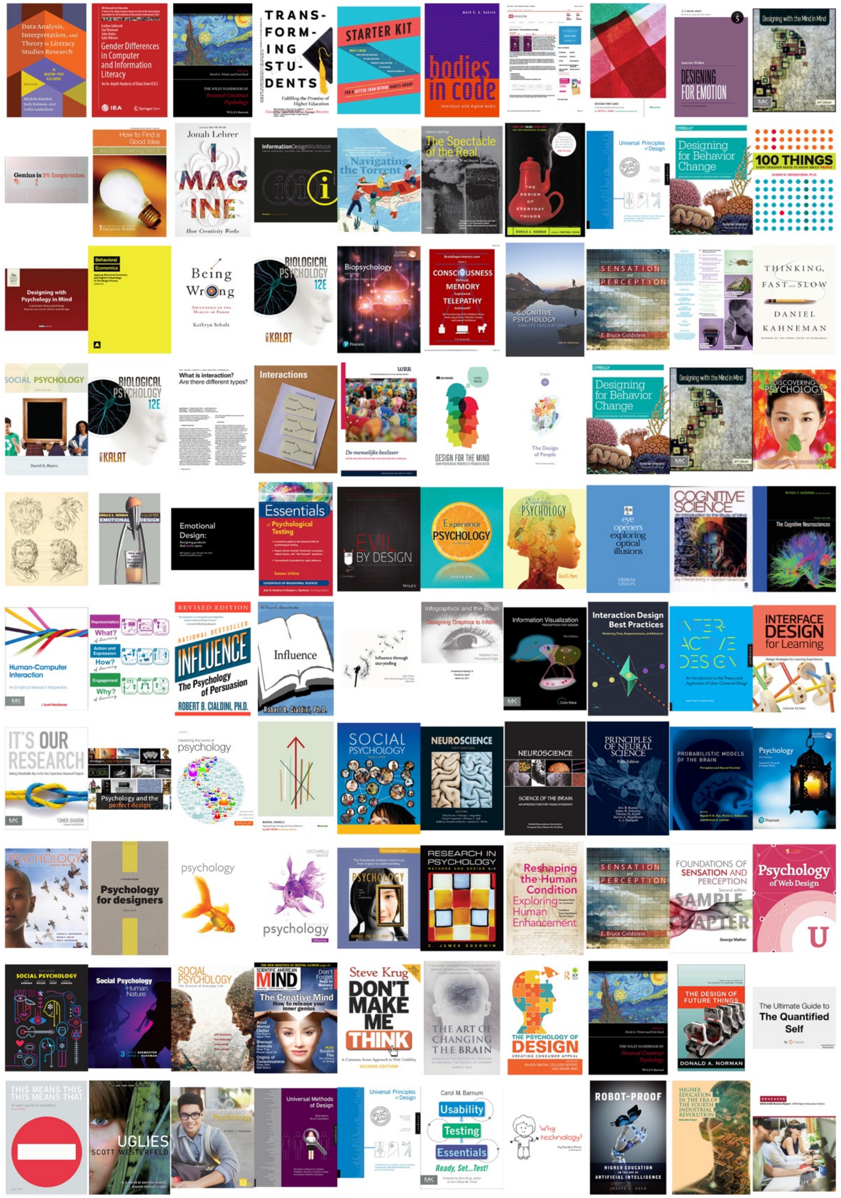


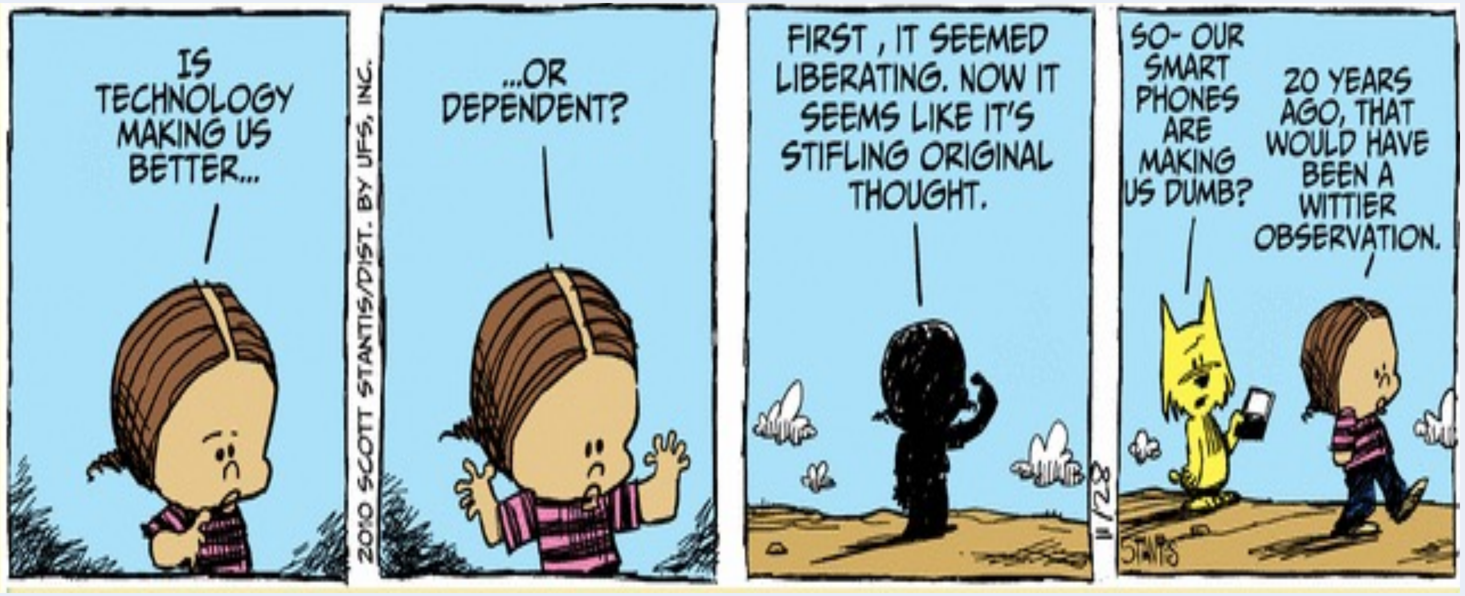




{Studied Materials: books}

HOGESCHOOL ROTTERDAM





Creative Commons License Types

	Can someone use it commercially?	Can someone create new versions of it?
Attribution	👍	👍
Share Alike	👍	Yup, AND they must license the new work under a Share Alike license.
No Derivatives	👍	👎
Non-Commercial	👎	Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
Non-Commercial Share Alike	👎	Yup, AND they must license the new work under a Non-Commercial Share Alike license.
Non-Commercial No Derivatives	👎	👎

SOURCE
http://www.masternewmedia.org/how-to-publish-a-book-under-a-creative-commons-license/

This seminar was developed by:
Living-Lab: *AiRA,*
Hub voor Data & Responsible AI
Rob van der Willigen
FEB 2022