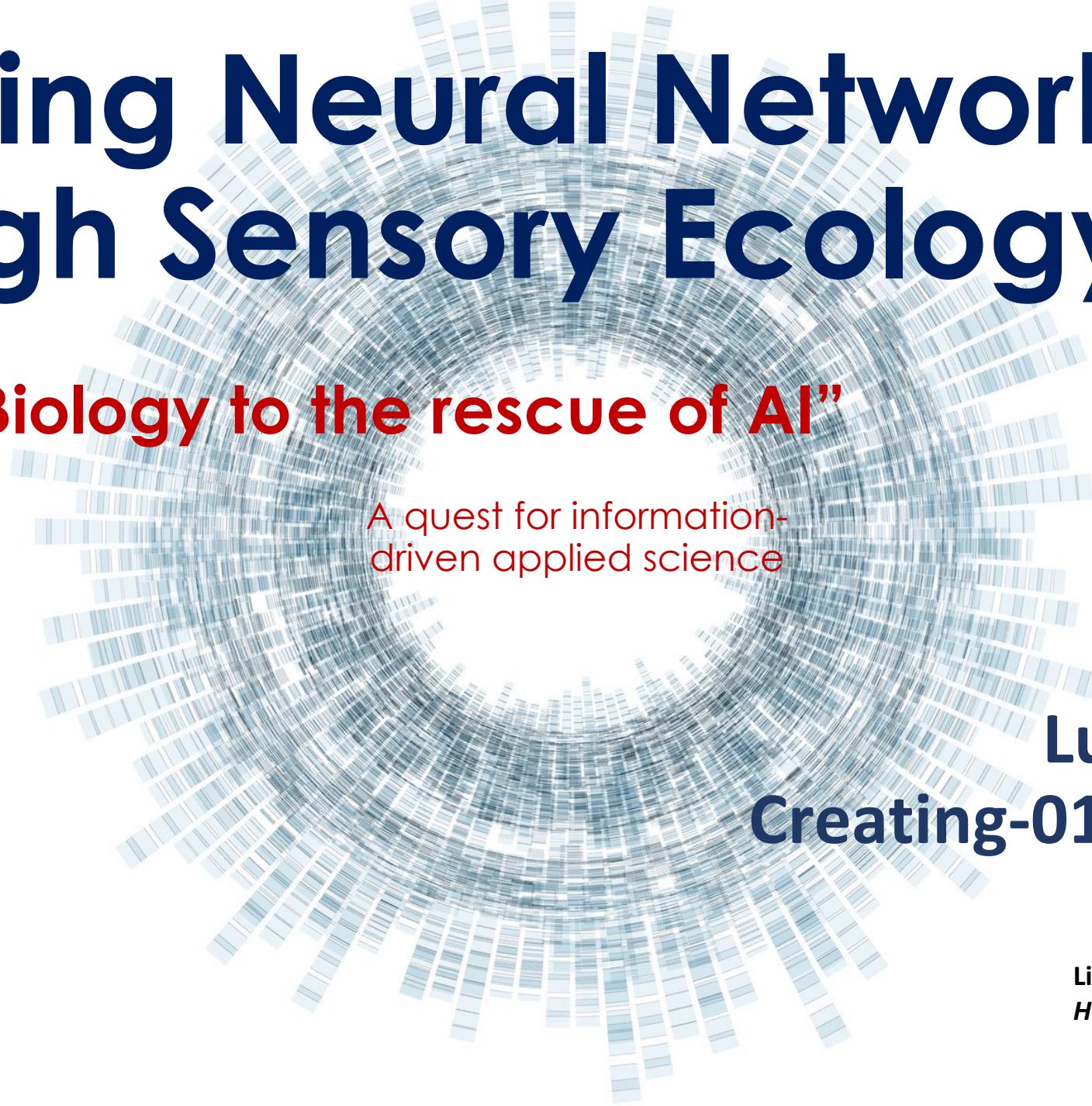


# Designing Neural Networks Through Sensory Ecology



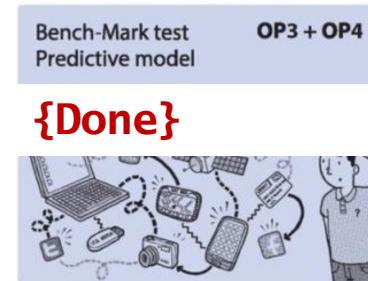
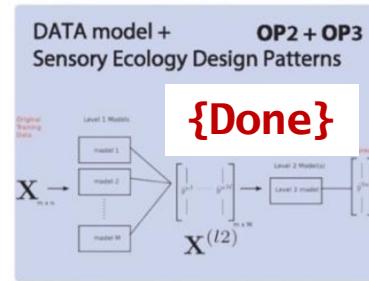
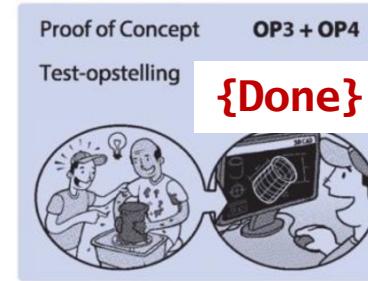
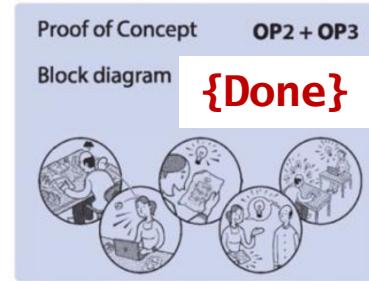
**“Biology to the rescue of AI”**

A quest for information-  
driven applied science

Lunch-Lezing  
Creating-010 FEB 2022

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

# {Envisioning a hands-on Data- Lab}



Experiment Findings File available

Making Human-Centered Sense of Functional Brain Imaging Data in a Smart Brave World: Designing artificial neural networks through sensory ecology.

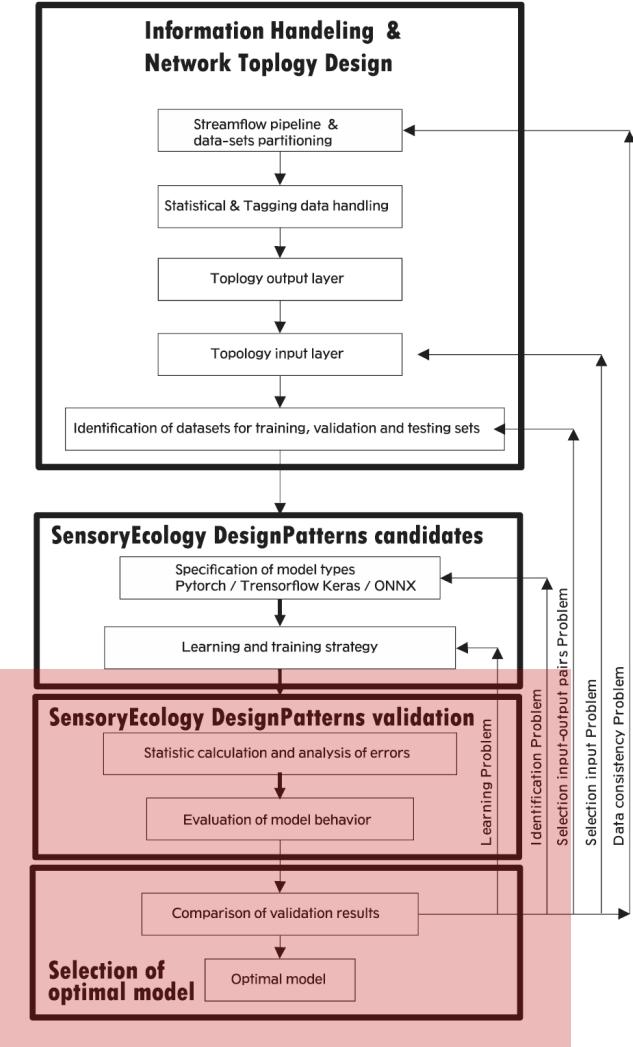
June 2020

DOI: [10.13140/RG.2.2.28861.18408](https://doi.org/10.13140/RG.2.2.28861.18408)

Project: [Making Human-Centered Sense of Functional Brain Imaging Data in a Smart Brave World](#)

Robert Frans van der Willigen

Research Interest 2.1  
Citations 0  
Recommendations 0 new 0  
Reads 103  
[See details](#)



# {Research Output}

## PART I

Literature review & taxonomy of {SOTA} Machine Learning

## PART II

Proof of concept of Sensory Ecology-based Sparse Coding PCA design patterns, allowing neural networks to learn complex cognitive tasks by reducing the nonlinear dimensionality of sensory information data-sets

## PART III

Description and implementation of a Data Fabric methodology by means of a high-performance computing “Data Lab” that allows for the implementation of the reuse of pre-trained deep neural networks {transfer-learning} on small-sized, custom data-sets

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

# {A working hypothesis}

## Barn owls: why give a hoot?

by Kara Rosania

<b>SCIENTIFIC NAME</b>
<i>Tyto alba</i>
<b>TAXONOMY</b>
PHYLUM: Chordata
CLASS: Aves
ORDER: Strigiformes
FAMILY: Tytonidae

© 2014 Nature America, Inc. All rights reserved.



### 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<sup>1</sup>. 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 shier, 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<sup>2</sup>. 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<sup>3</sup>. 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<sup>4</sup>. In the late 1970s, Knudsen and Konishi localized this auditory space map to a region of the owl's brain called the inferior colliculus<sup>5</sup>. 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<sup>6</sup> and that it shifts adaptively as the locations of sound sources move through space<sup>7</sup>.

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<sup>8</sup>. 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<sup>9</sup>.

1. Jaworski, J. & Peake, N. Vortex noise reductions from a flexible fiber model of owl down. *66<sup>th</sup> Annual Meeting of the American Physical Society's (APS) Division of Fluid Dynamics*, 24–26 November 2013, Pittsburgh, PA.
2. Knudsen, E.I. Auditory object localization learning in the auditory localization pathway of the barn owl. *Nature* **413**, 395–398 (2001).
3. Fischer, B.J., Christianson, G.B. & Perla, 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. Konishi, M. & Konishi, M. A neural map of auditory space in the owl. *Science* **200**, 795–797 (1978).
6. Efrat, A. & Gutfrund, 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 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's inferior colliculus. *J. Neurophysiol.* **101**, 2924–2933 (2009).

LAB ANIMAL



Karen Cetra/Nature Publishing Group

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”!

# **WHY DESMISTFYING AI?**

**“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 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 sword:**

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

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

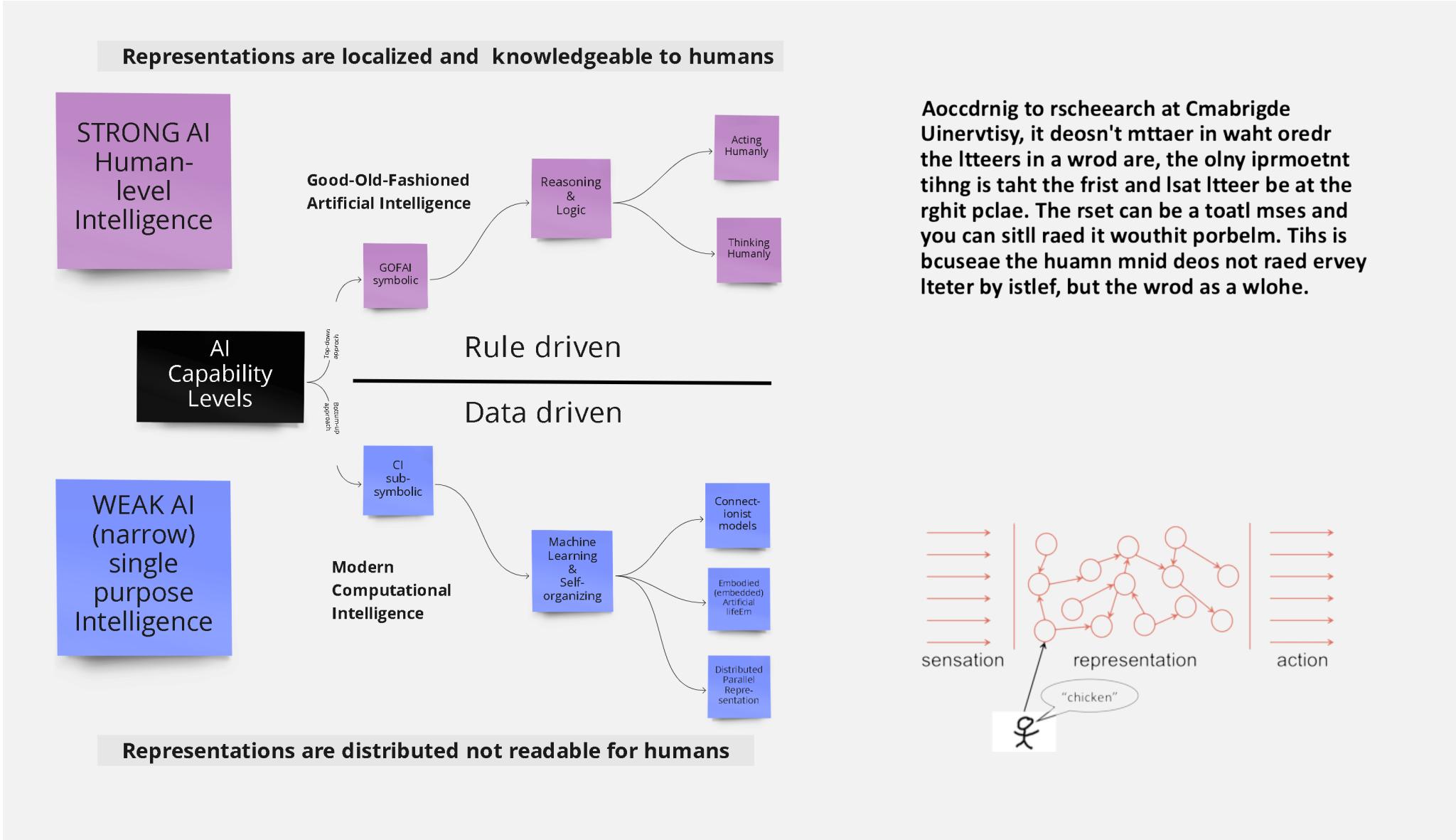
# {Brute Force}

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

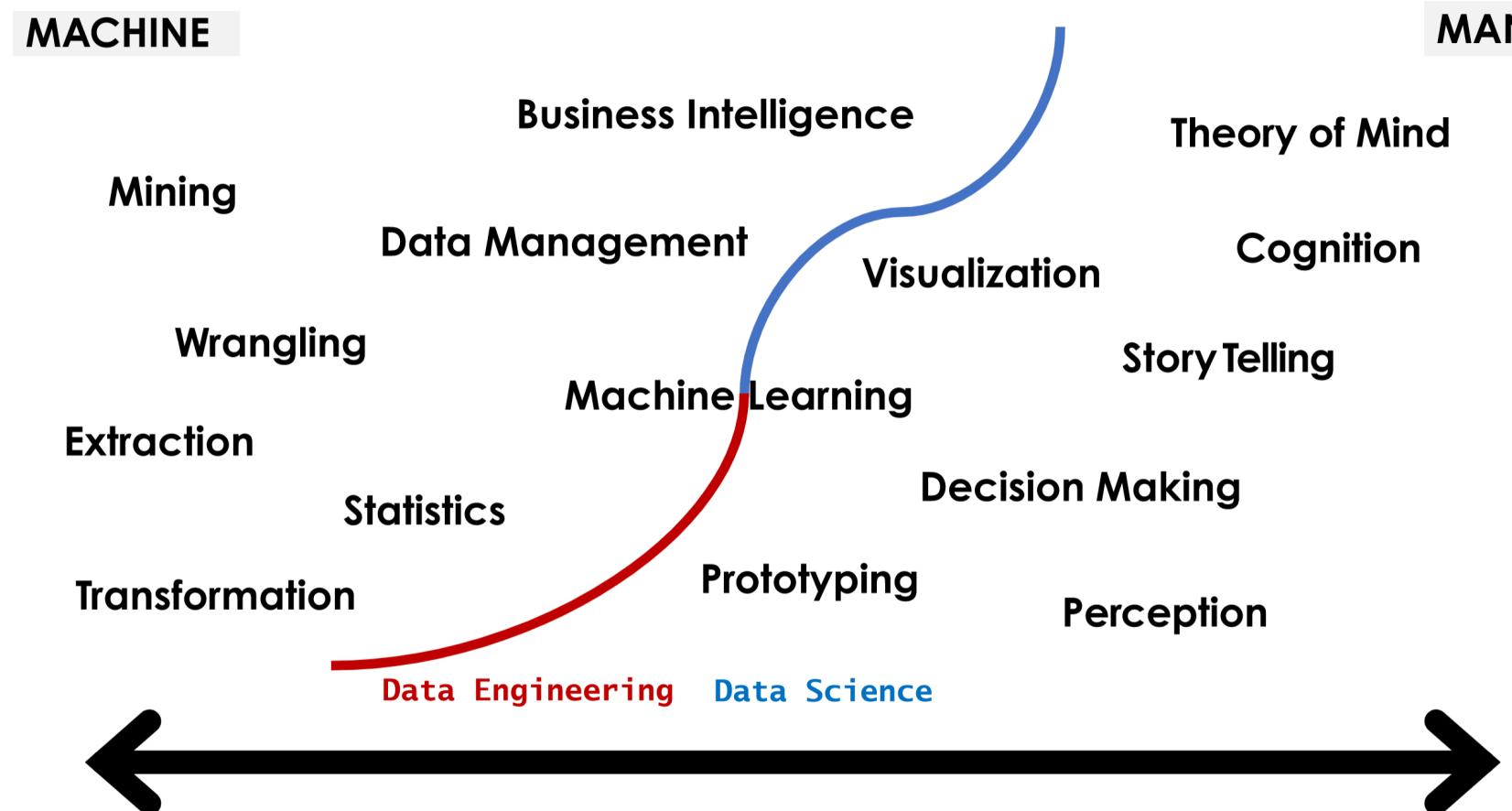
- High-performance GPU or TPU computing power
- Black-Box, poorly understood solutions
- Training with massive data-sets
- Deep Neural Network architecture
- Billions of parameters
- Lengthy Process & Huge budgets

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

# {AI taxonomy is complex}



# {AI integrates two Scientific Disciplines}



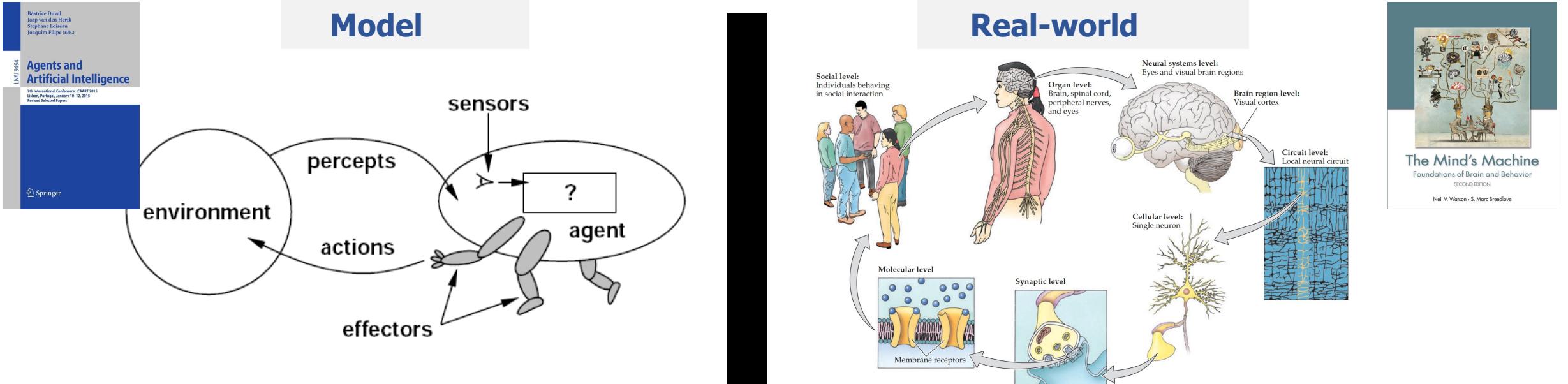
Inspired by Daniel Keim, "Visual Analytics: Definition, Process, and Challenges"

# {AI favors Agent-Based Models}

## 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



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

# {Chat-Bot Meena}

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

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

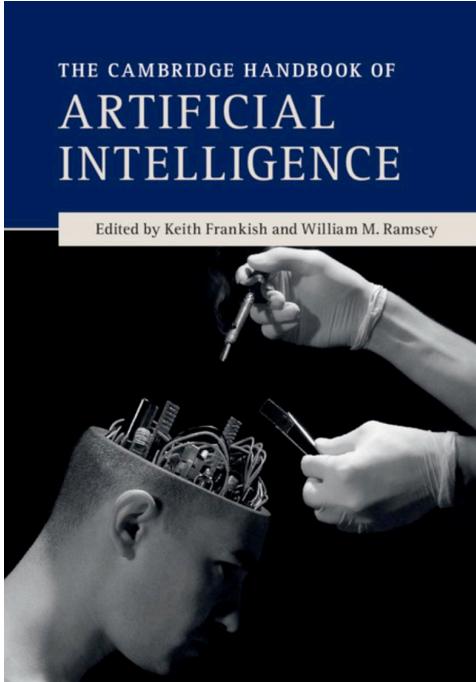
30 days of training, 4[Million] tokens per second

Training cost: 1,500,000\$

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

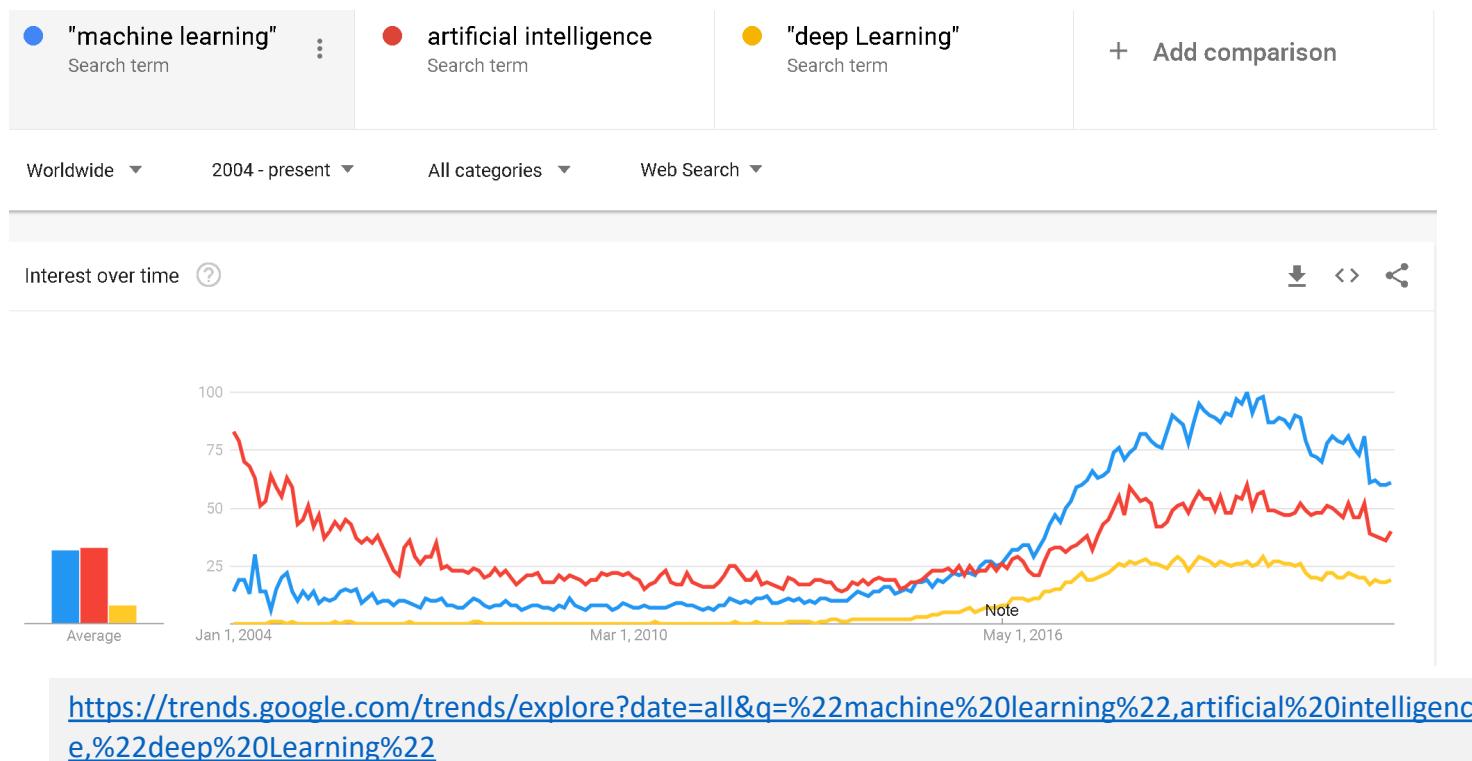
What constitutes a  
**{Deep}** Neural Net?

# {DL}



# (Pre-trained) Deep Learning {DL}

## Neural Networks {NNs} are the most advanced, successful & fastest growing Artificial Intelligence {AI} technology

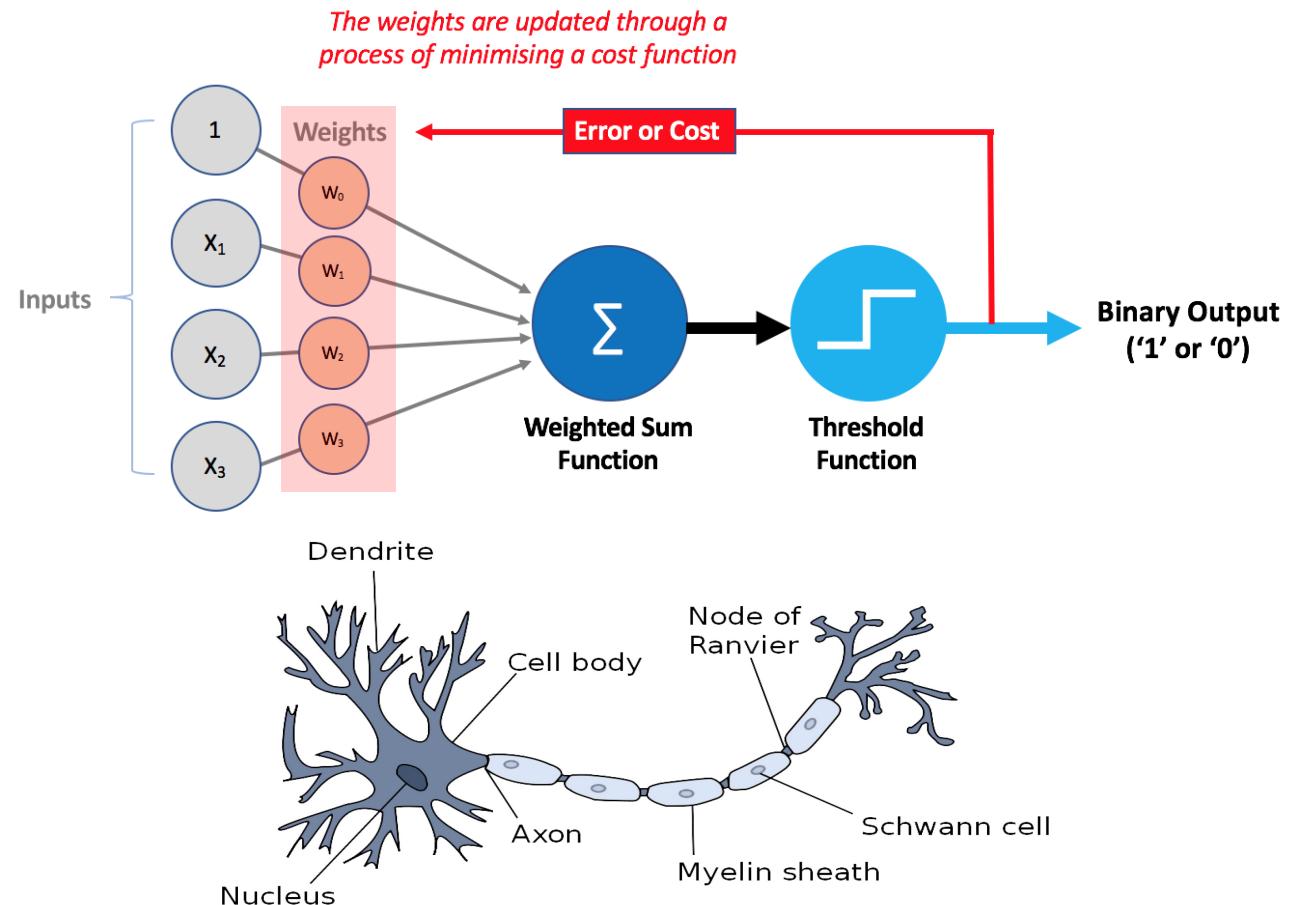


# {Artificial Neurons}

Deep Neural Nets {DNNs} harbor vast amounts of  
**“artificial neurons”** →smallest computational unit←

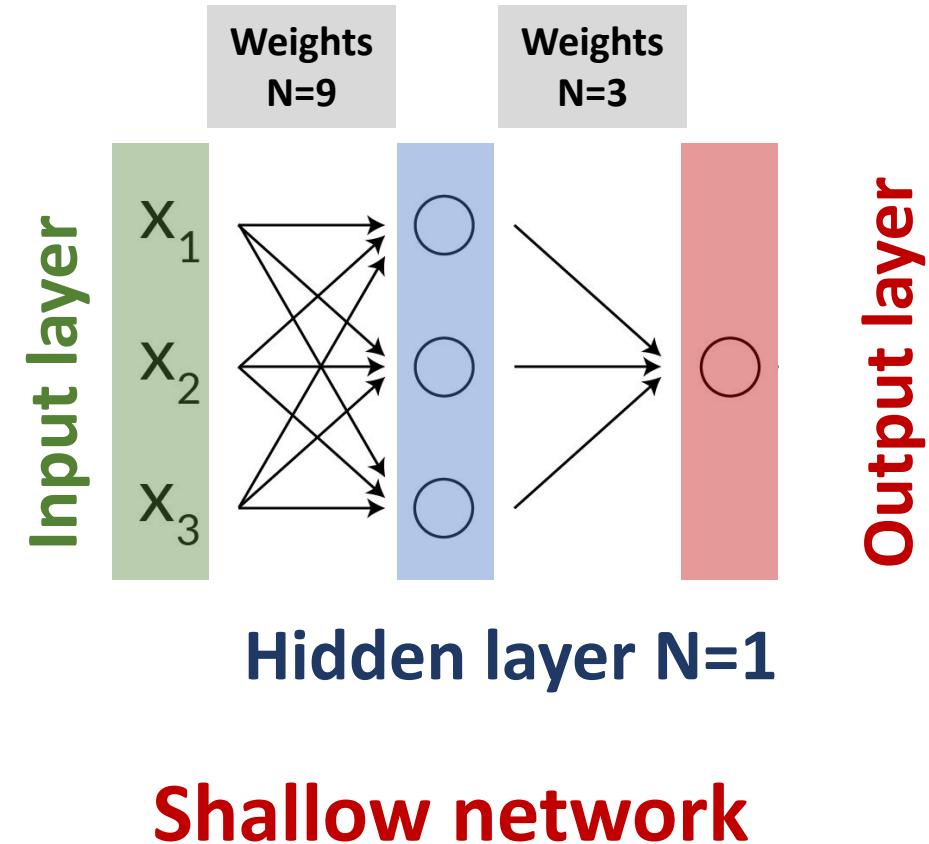
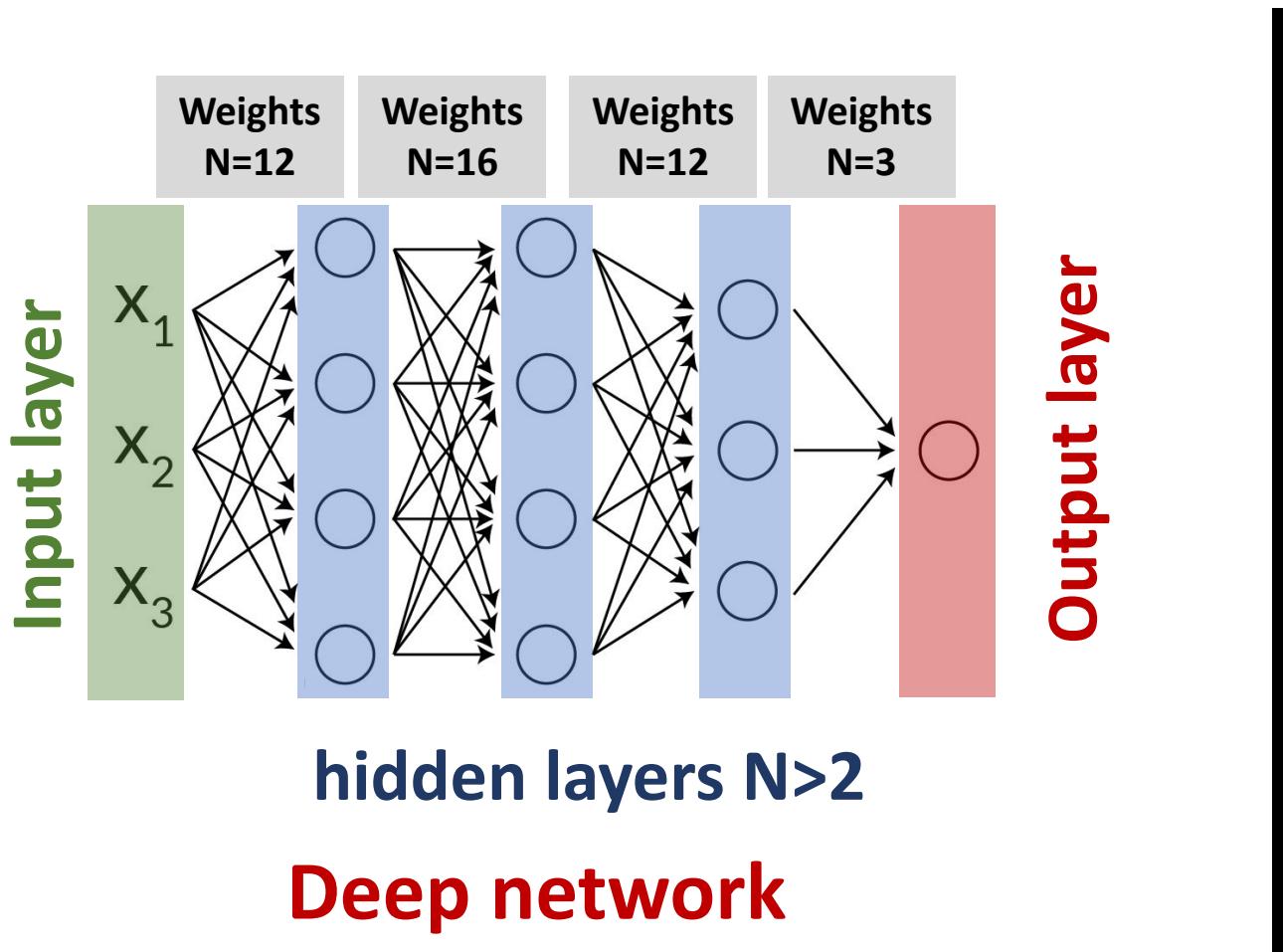
**Names for  
 Artificial Neurons**

**{unit}**  
**{cell}**  
**{node}**  
**{perceptron}**



# {NN Layers}

## Neural Network {NN} Layer Architecture



## 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>

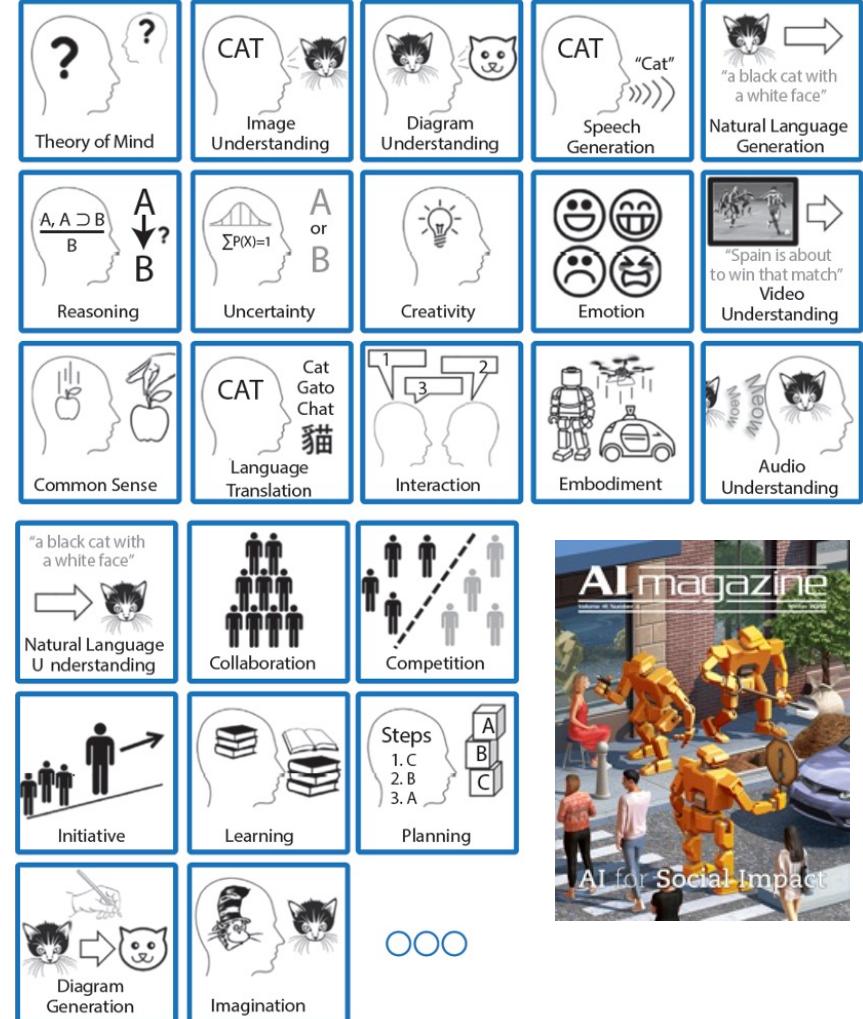
# {AI=ML=DL}

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

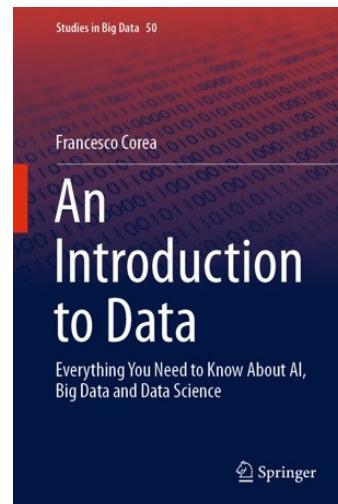
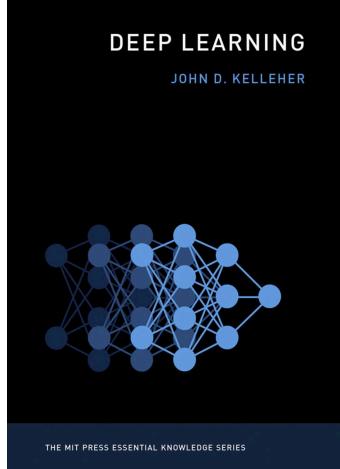
*human perception,  
cognition and/or  
motor skills*

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

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



# {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.

# {Human-level performance}

The screenshot shows the DeepL Translator interface. At the top, there's a navigation bar with the DeepL logo, "DeepL Translator", "DeepL Pro", "API", "Plans and pricing", "Apps" (with a "FREE" badge), "Contact Sales", "Start free trial", "Login", and a menu icon. Below the navigation bar are two main input fields: "Translate text" (26 languages) and "Translate files" (.pdf, .docx, .pptx). The left panel, set to "English", contains the following text:

{DL} must be understood as a major {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 between Languages at High Levels of Performance.

The right panel, set to "Dutch", contains the translated text:

{DL} moet worden opgevat als een belangrijk {ML} subdomein:  
Het creëren van Diepe Neurale Netwerken {DNNs} die menselijke prestaties kunnen bereiken op uitdagende cognitieve taken.

{DNNs} kunnen spraak of menselijke houdingen en gezichten herkennen; tekst vertalen tussen talen op hoog prestatieniveau.

At the bottom of each panel are interaction icons: a speaker icon, a thumbs up icon, a thumbs down icon, a share icon, and a copy icon. A URL at the bottom right provides the link to the translator: <https://www.deepl.com/translator>.

<https://www.deepl.com/translator>

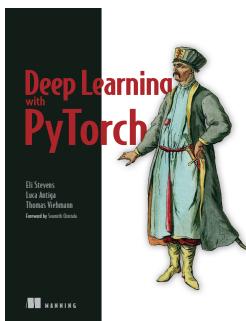
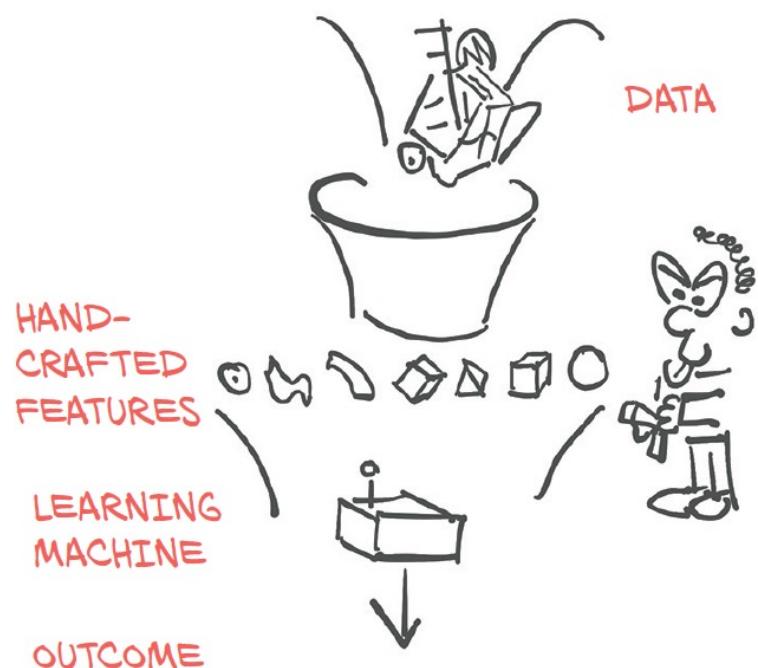
**{DL} represents an  
{AI} breakthrough**

**Paradigm-Shift**

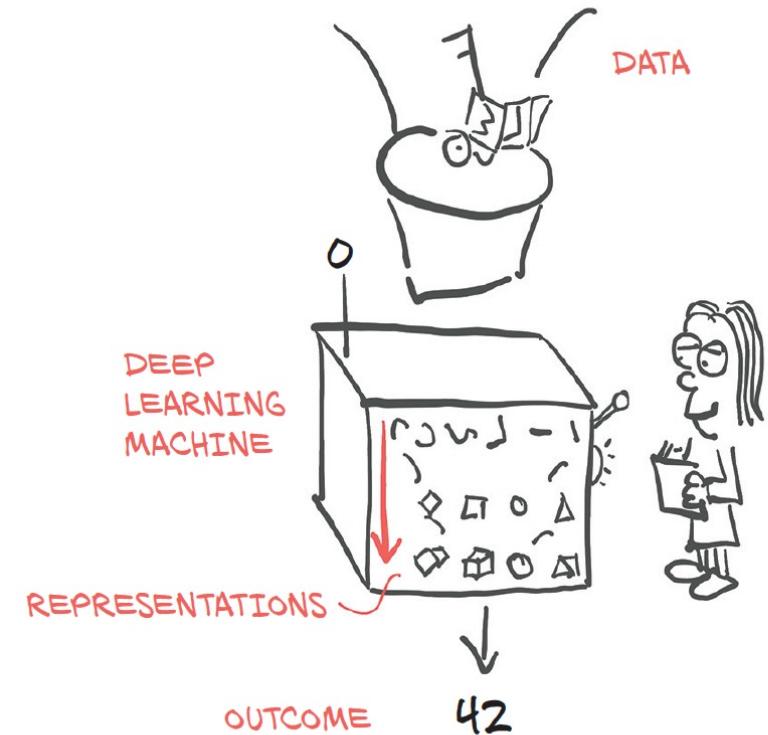
# {AI Paradigm-shift}

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

## A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg  
 Paul G. Allen School of Computer Science  
 University of Washington  
 Seattle, WA 98105  
 slundb@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 models is often achieved by methods that are not easily interpretable or explainable 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

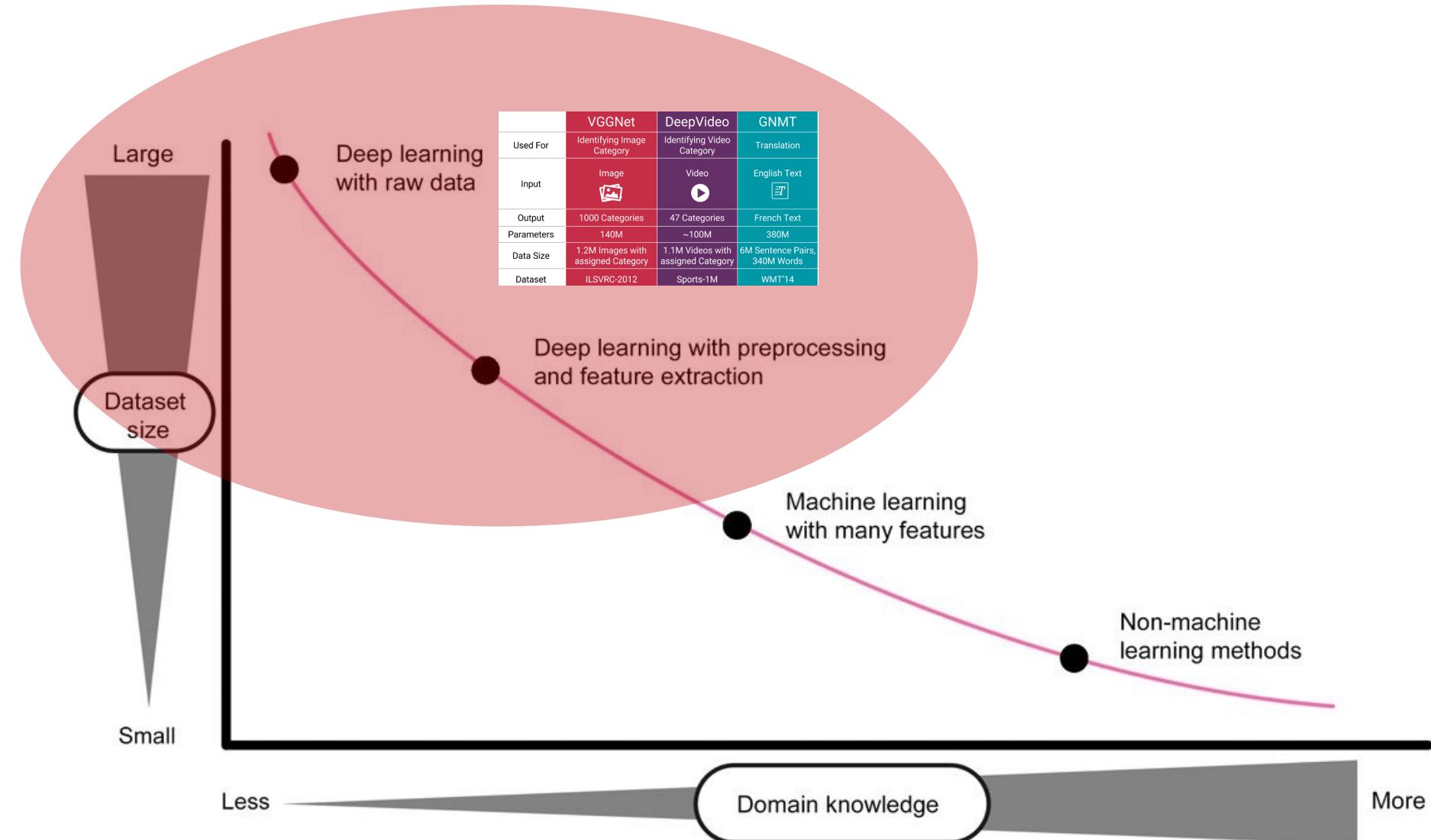
The ability to correctly interpret a prediction model's output is extremely important. It engenders appropriate user trust, provides insight into how a model may be improved, and supports understanding of the process being modeled. In some applications, simple models (e.g., linear models) are often preferred 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 benefit 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.<sup>1</sup> Our approach leads to three potentially surprising results that bring clarity to the growing space of methods:

- 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.

<sup>1</sup><https://github.com/slundberg/shap>

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



# {Top-down}

# Top-down Encoding Capacity increases by adding hidden layers

## 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

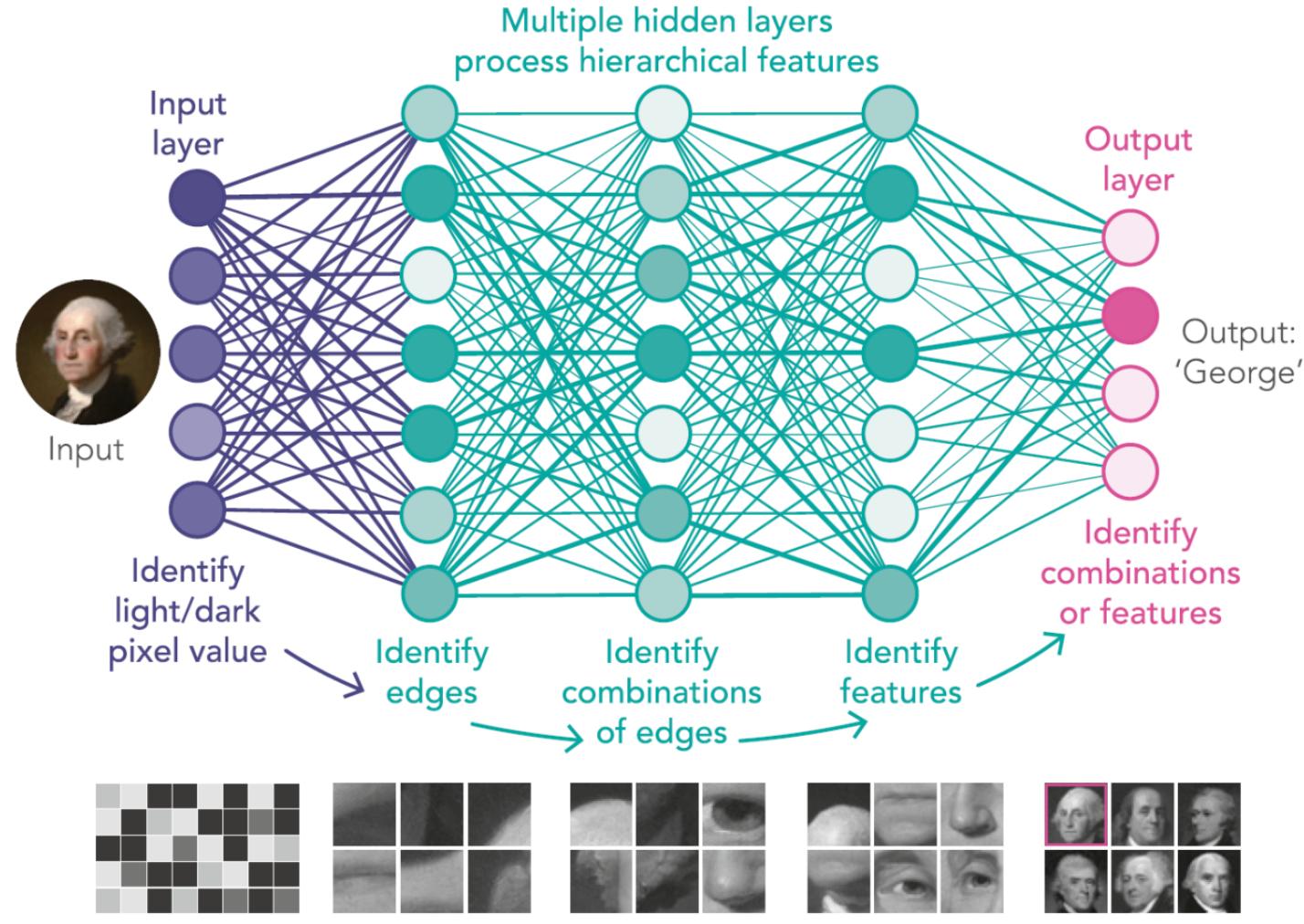


Apparent shortcomings in deep-learning approaches have raised concerns among researchers and the general public as technologies such as driverless cars, which use deep-learning techniques to navigate, get involved in well-publicized mishaps. Image credit: Shutterstock.com/MONOPOLY919.

Published under the PNAS license.

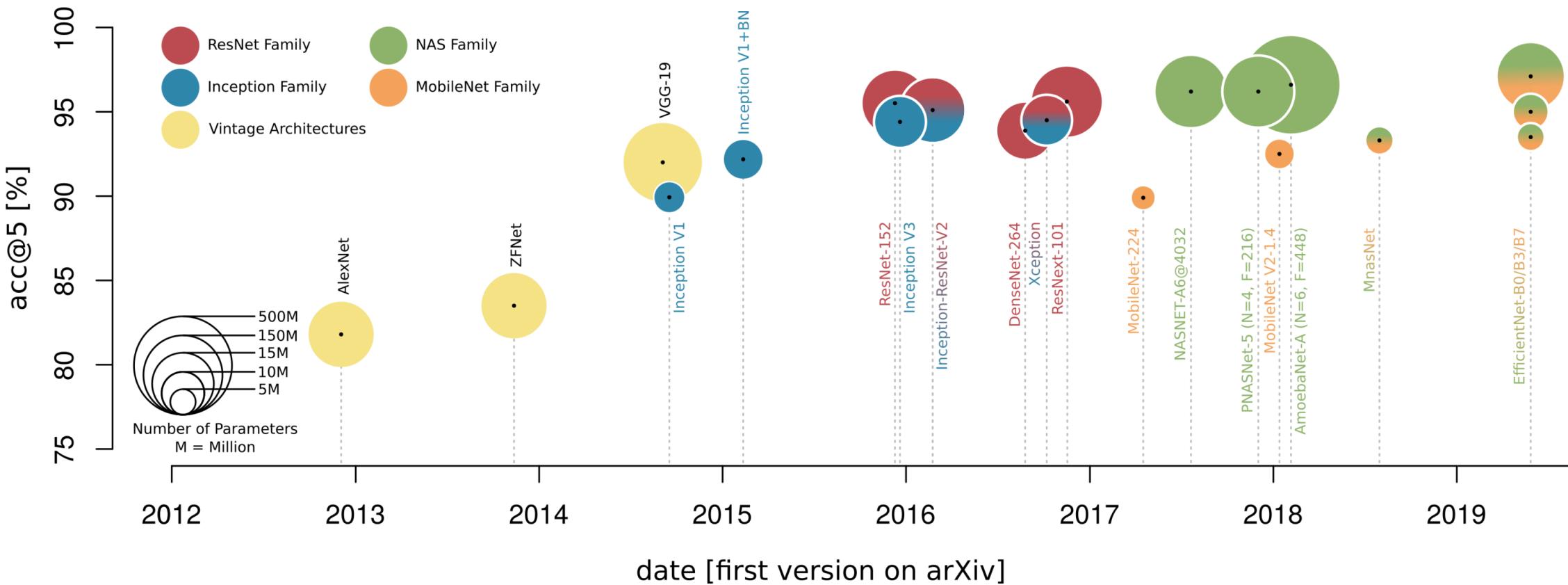
January 22, 2019 | vol. 116 | no. 4

[www.pnas.org/cgi/doi/10.1073/pnas.1821594116](http://www.pnas.org/cgi/doi/10.1073/pnas.1821594116)



# {Weights}

## Performance increases by adding learnable parameters {weight's}



How to calculate the number of learnable parameters?

<https://doi.org/10.3390/rs12101667>

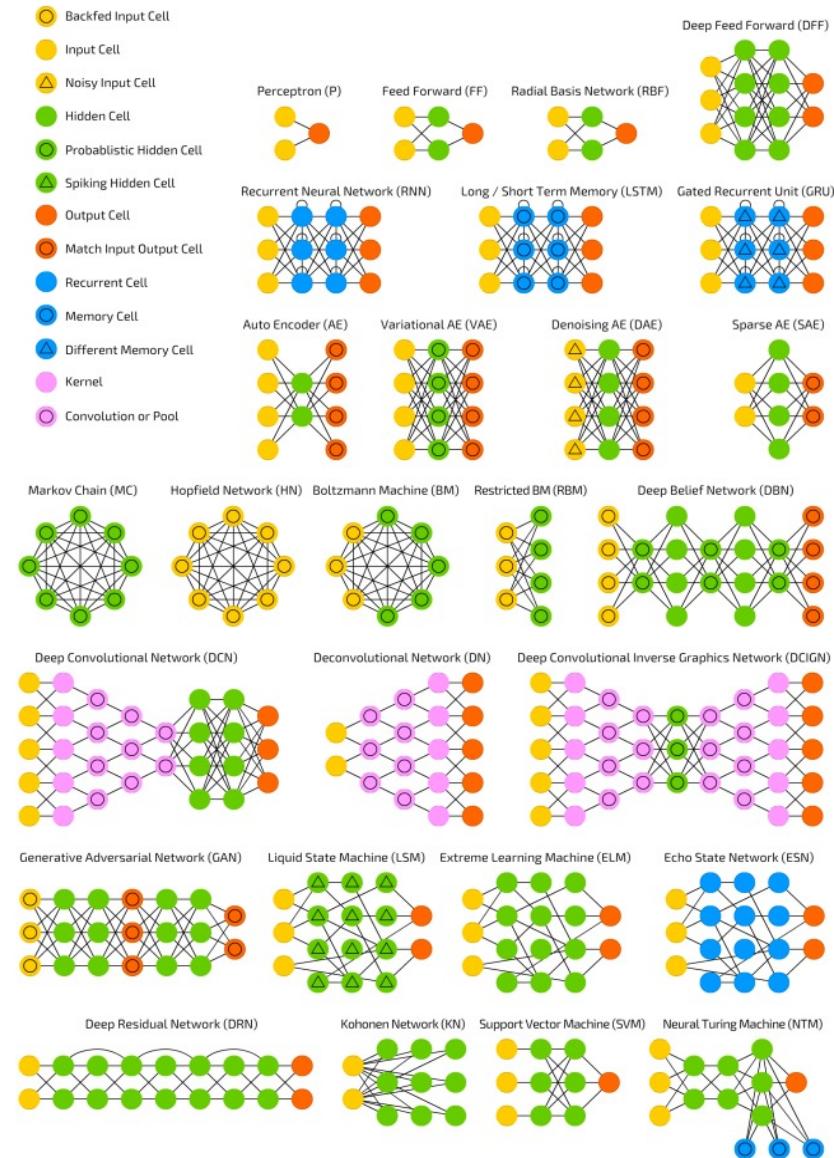
# {Topology}

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

**Form follows function!**

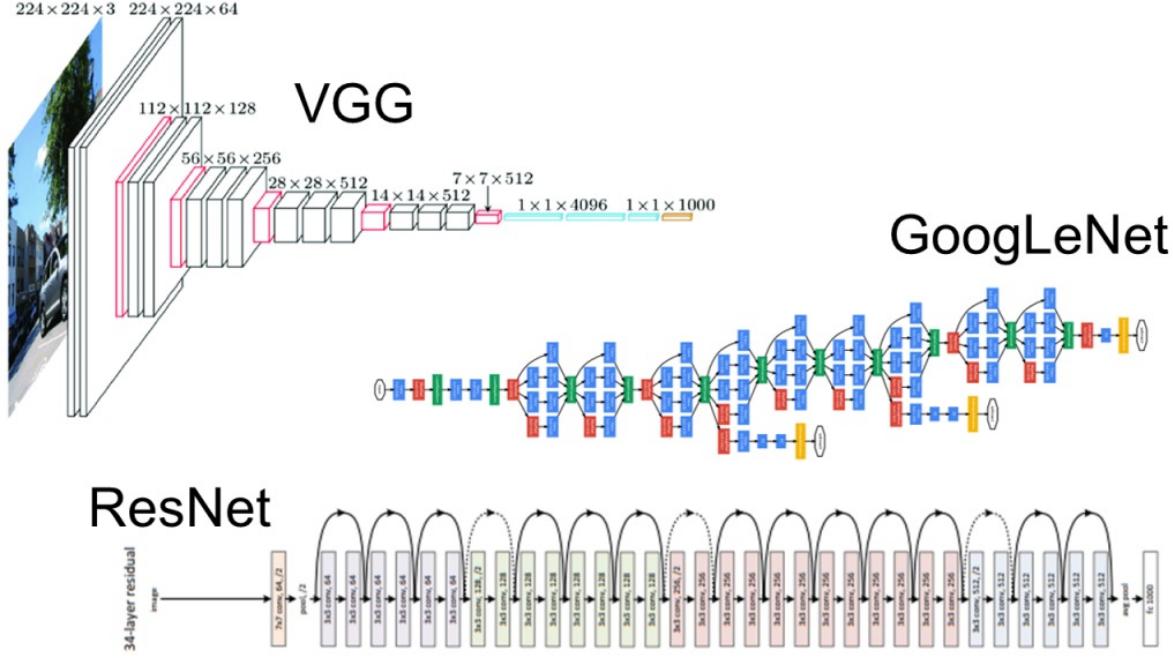
The topology of a network determines the degree of **perplexity** of the tasks it can perform.

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



# {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	-

# {Self-attention}

## 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

Lukasz Kaiser\*  
 Google Brain  
 lukasz.kaiser@google.com

Ilia 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

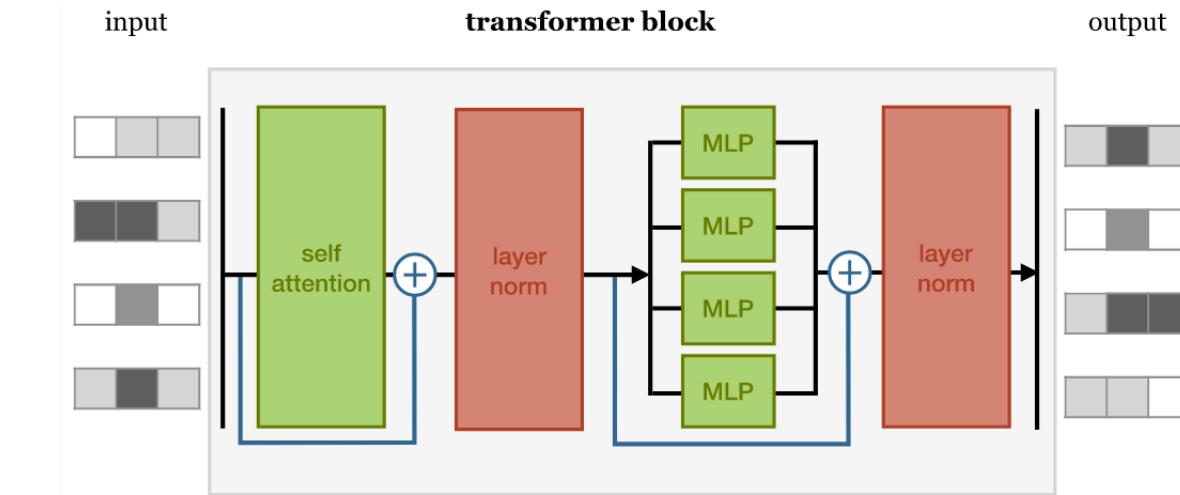
Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

\*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.

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



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

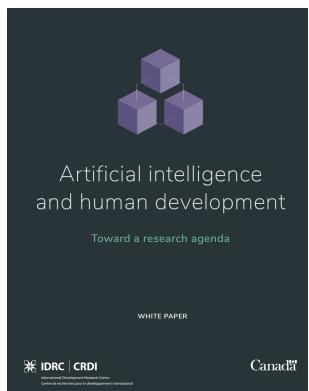
Mukhamediev, R. I., Symagulov, A., Kuchin, Y., Yakunin, K., & Yelis, M. (2021). From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review. *Applied Sciences*, 11(12), 5541. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/app11125541>

# Why use biology instead 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<sup>45</sup>. Many similar examples have been reported.<sup>46</sup>

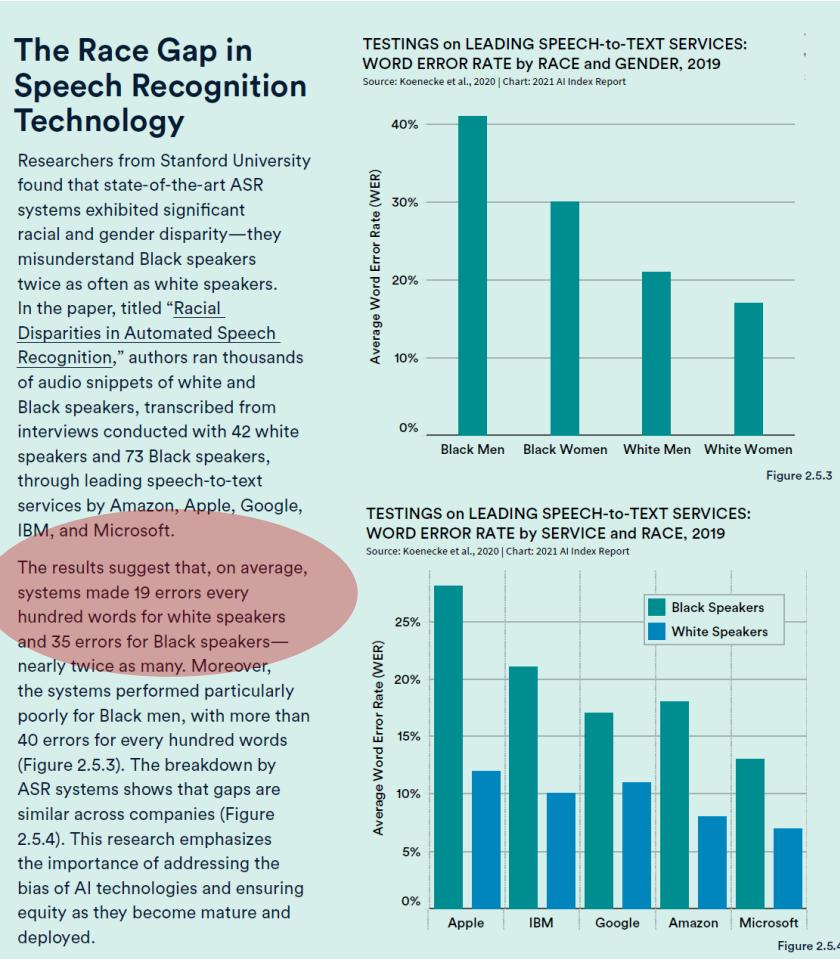
When critical decisions are made based on input from an algorithm trained on a database that is not representative of the entire user population, the results can impact health and well-being. Most clinical trials have highly selective criteria that exclude women (especially pregnant women), the elderly, and those with conditions beyond those being studied. Thus, participants tend to be white males.<sup>47</sup> 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](https://bit.ly/2HFetvY)

Photo by Marius Masala

# {Disparities}

## Big Data causes racial & gender disparities



# {Augmentation}

## 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

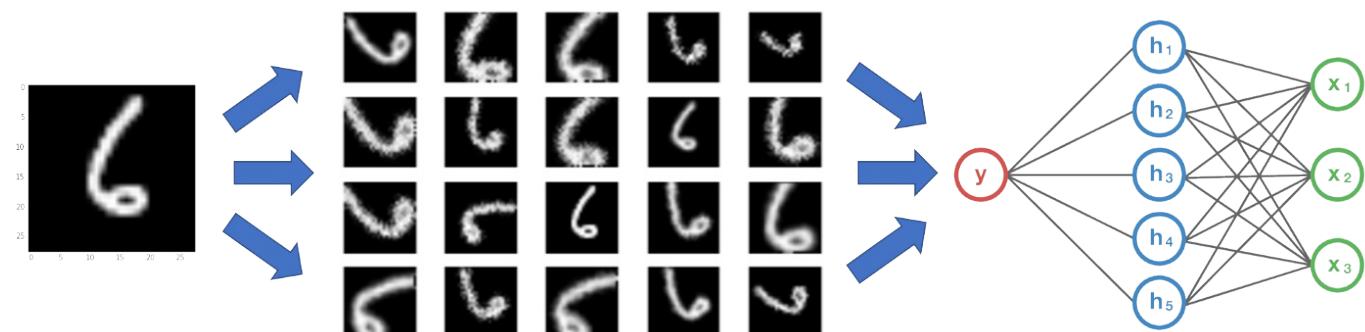
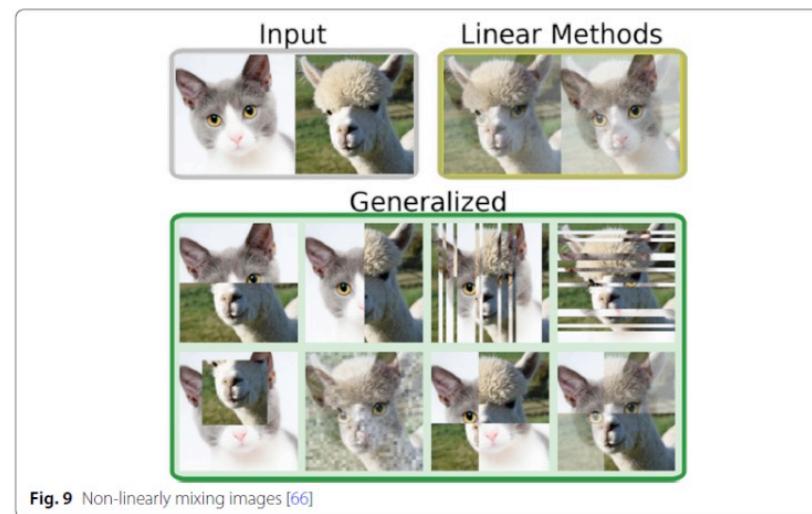
### Abstract

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 sparked interest and optimism in applying Deep Learning to Computer Vision tasks.



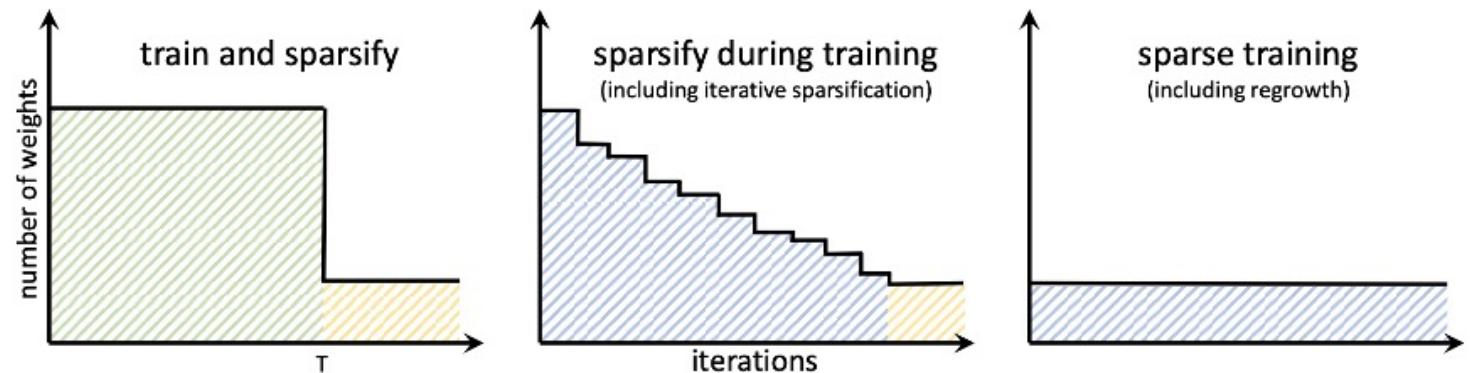
# {Regularization}

## The process of adding information to address 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.

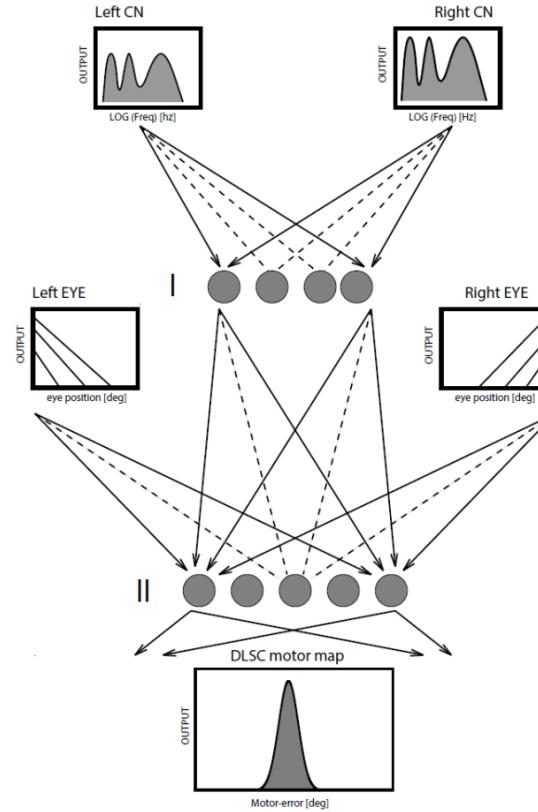


*The supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience -*

Albert Einstein, 1933

Hoefer, 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}



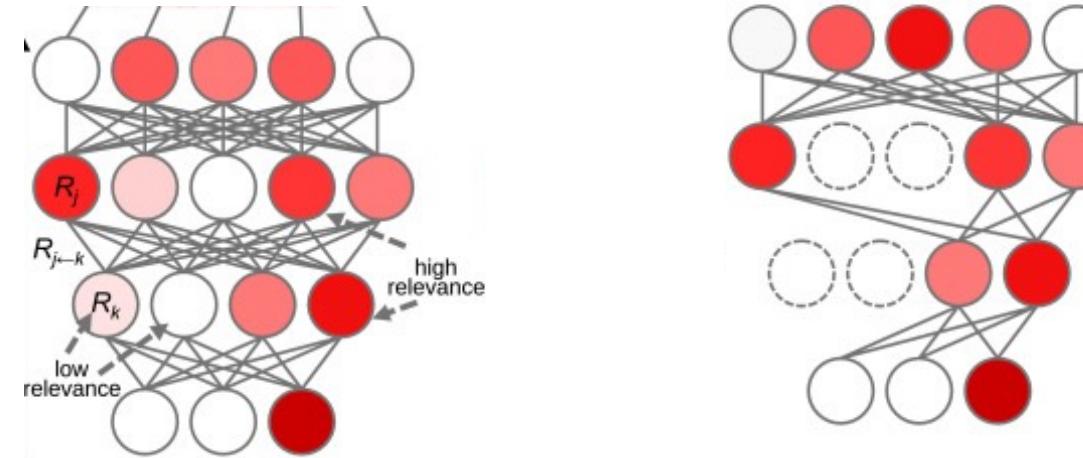
[International Workshop on Biologically Motivated Computer Vision](#)

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)

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



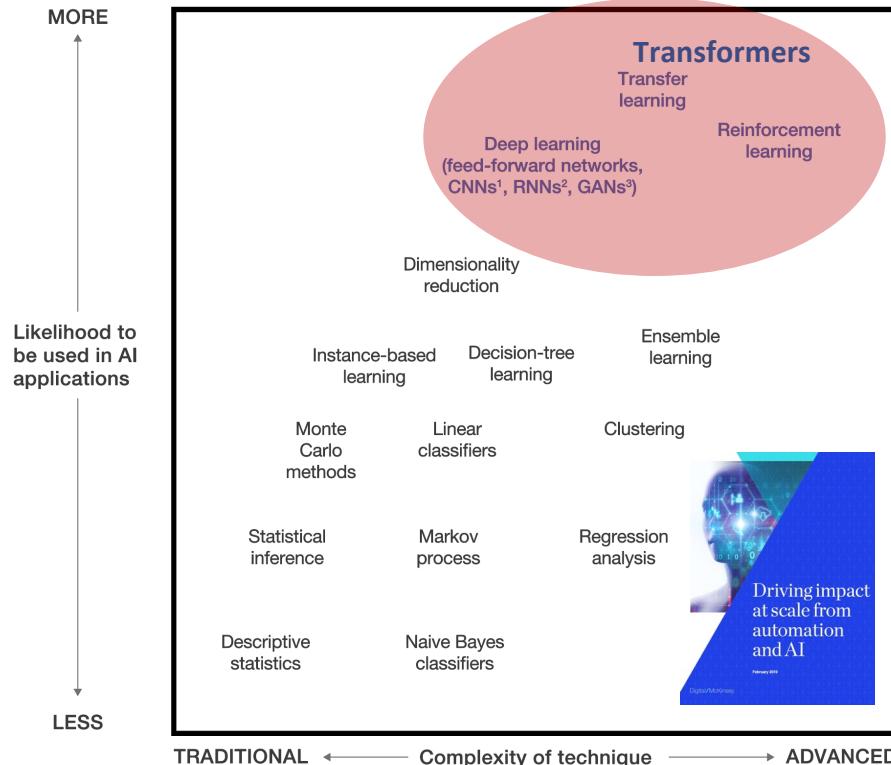
Yeom, S.K., Seegerer, P., Lapuschkin, S., Binder, A., Wiedemann, S., Müller, K.R., & Samek, W. (2021). Pruning by explaining: A novel criterion for DNN pruning. *Pattern Recognition*, 115, 107899.

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}

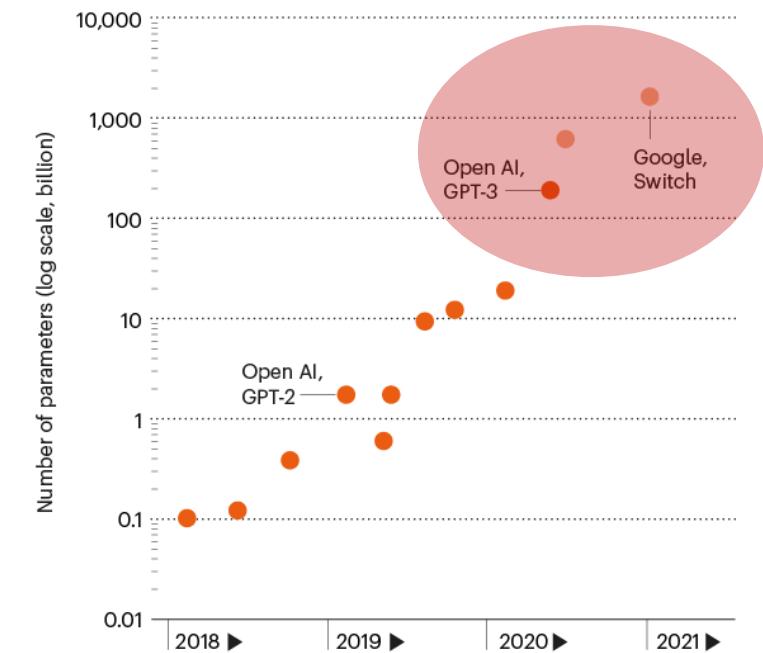
Only very large scale {DNNs} are useful  
[can compete with human performance]



## LARGER LANGUAGE MODELS

The scale of text-generating neural networks is growing exponentially, as measured by the models' parameters (roughly, the number of connections between neurons).

● 'Dense' models ● 'Sparse' models\*

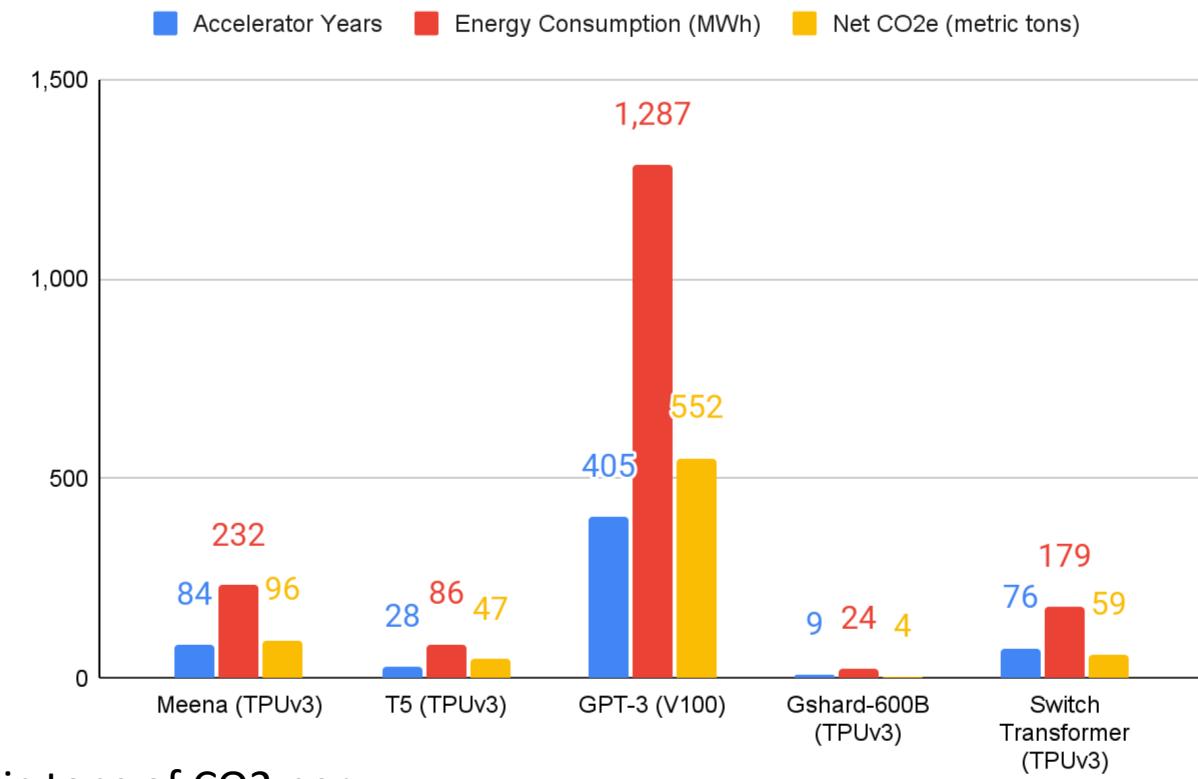
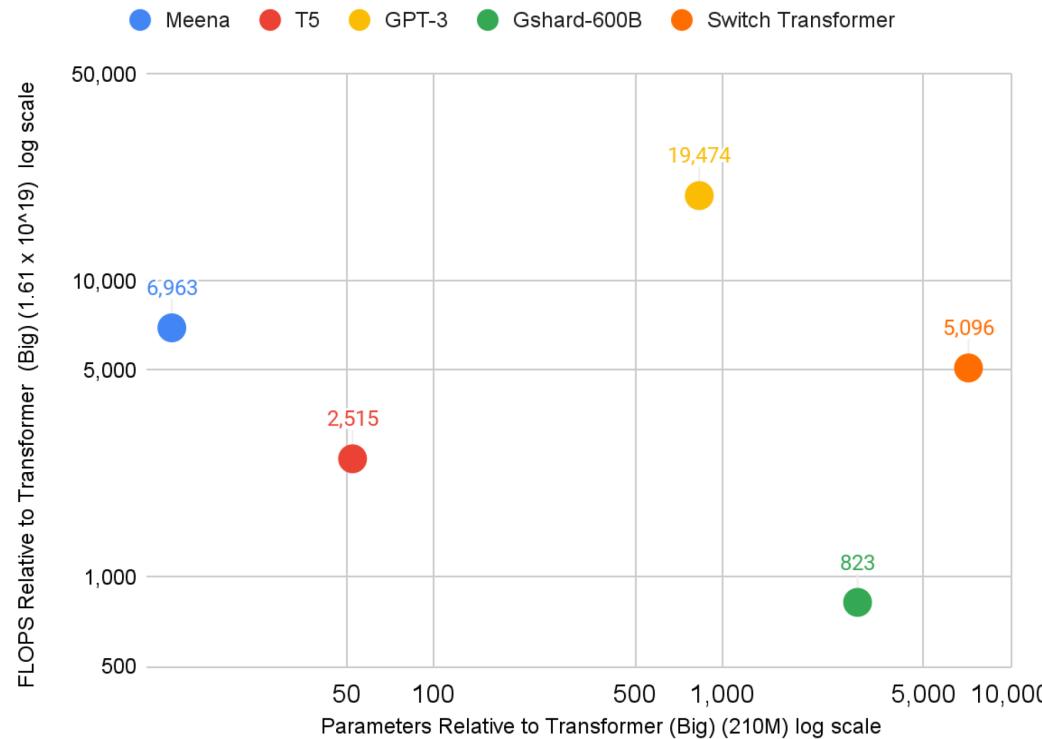


\*Google's 1.6-trillion parameter 'sparse' model has performance equivalent to that of 10 billion to 100 billion parameter 'dense' models. ©nature

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

# {CO<sub>2</sub> foot-print}

## Training large scale transformer {DNNs} produce massive Carbon Emissions



As of 2007, the average U.S. household emits 20 metric tons of CO<sub>2</sub> per year. In comparison to a world average of 4 tons.

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

# {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 **DNNs** 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

**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

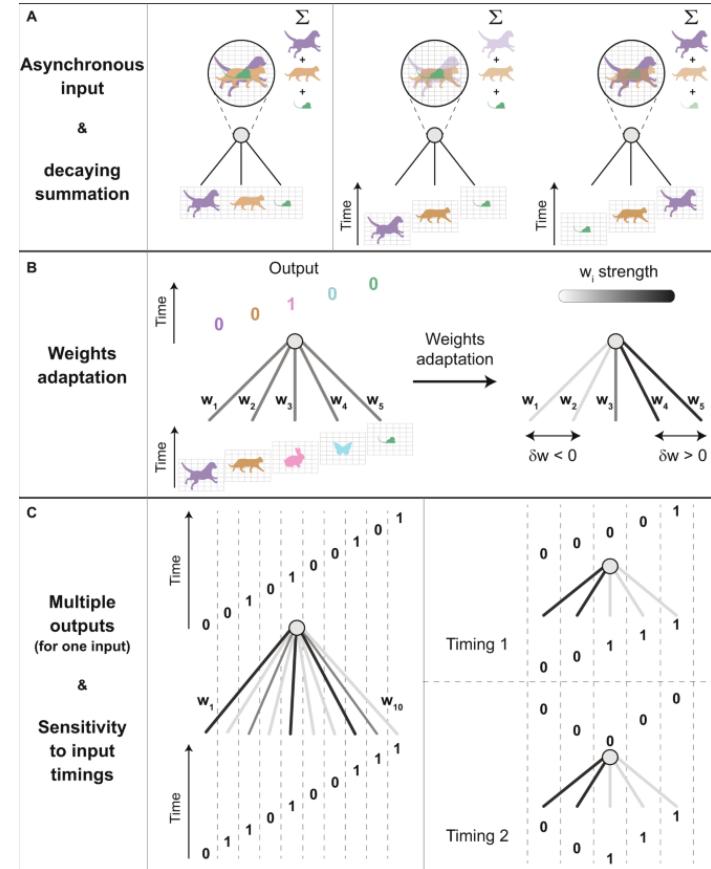
Herut Uzan, Shira Sardi, Amir Goldental, Roni Vardi & Ido Kanter 

*Scientific Reports* 9, Article number: 11558 (2019) | [Cite this article](#)

19k Accesses | 5 Citations | 223 Altmetric | [Metrics](#)

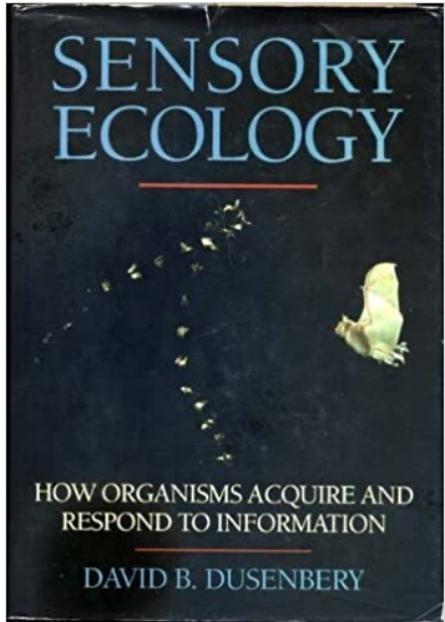
#### Abstract

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**

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

Andrew N. Iwaniuk<sup>1</sup>  | Douglas R. Wylie<sup>2</sup> 

<sup>1</sup>Department of Neuroscience, University of Lethbridge, Lethbridge, Alberta, Canada

<sup>2</sup>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>.

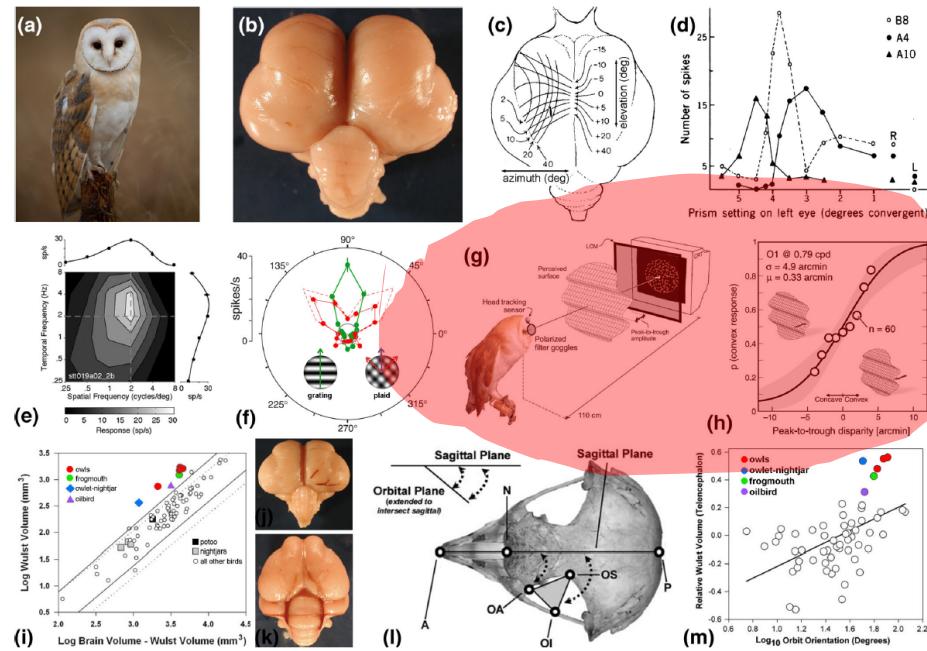
### Abstract

"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 binocular vision 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,



**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, Diaz, Lima, and Neuenchwander (2007)). (g) The psychophysical 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, Harmering, Vossen, and Wagner (2010). (i) Wulst volume as a function of brain-Wulst volume for owls, caprimulgiforms 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 (*Ninox 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

# {Sensory Ecology vs DNNs}

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>

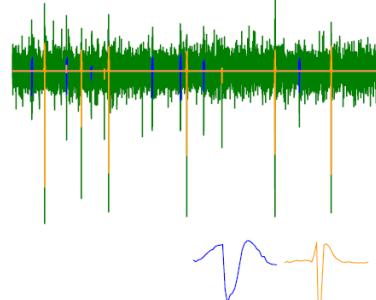
In parallel, the same cognitive tasks have been extensively studied in the brains of owls (visual, auditory specialist), humans & monkeys (cognitive task specialist).

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

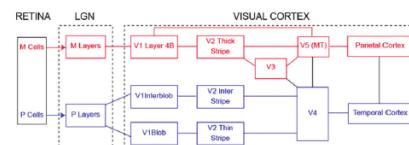
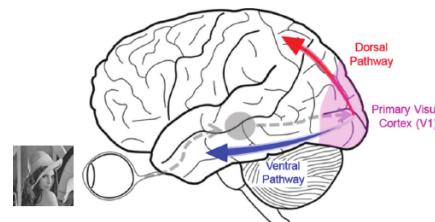
# {Sensory Ecology vs DNN}

Unsupervised learning  
to decipher neural code

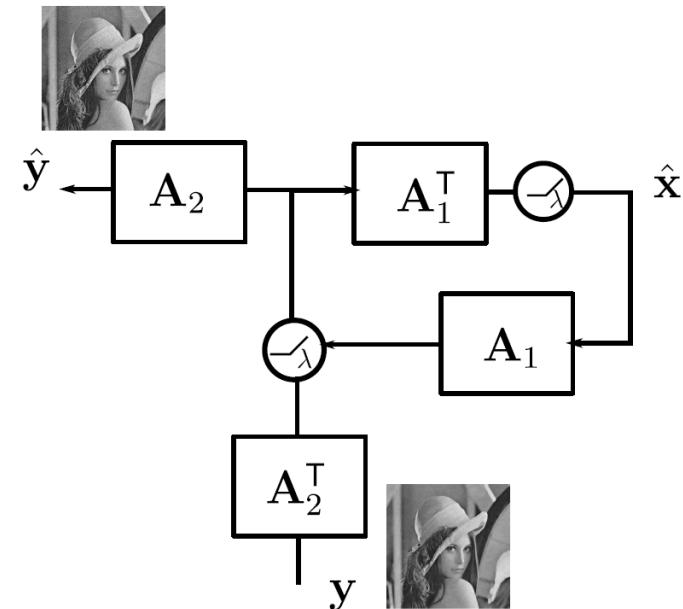
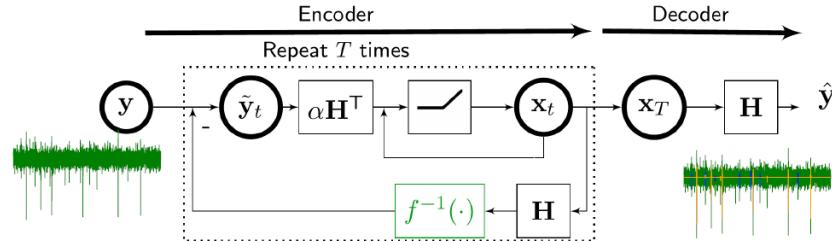
Blind source separation



Hierarchical sensory  
processing principles



How to design  
interpretable deep nets?



# {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:

$$\frac{1}{n} \sum_{t=1}^n -\log P(x_t | x_{t-1}, \dots, x_1), \quad (8.4.7)$$

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):

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

**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)

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

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 unitless and often set to 1 for convenience.

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).

Given a probability distribution  $p$  ( $p(x)$ ) 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  $H = -k \log(m)$  (or,  $S=k \ln(W)$  using Boltzmann's famous notation).

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

[https://d2l.ai/chapter\\_appendix-mathematics-for-deep-learning/information-theory.html](https://d2l.ai/chapter_appendix-mathematics-for-deep-learning/information-theory.html)

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

# 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



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

Many J. O'Connor<sup>1\*</sup>, Matthew W. Pennell<sup>1,2</sup>, Florian Altermatt<sup>2,3</sup>, Blake Matthews<sup>4,5</sup>, Carlos J. Mellán<sup>4,6</sup> and Andrew Gonzalez<sup>6</sup>

<sup>1</sup>Ecology Research Centre and Department of Zoology, University of British Columbia, Vancouver, BC, Canada

<sup>2</sup>Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland; <sup>3</sup>Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

<sup>4</sup>Department of Ecology and Evolution, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland; <sup>5</sup>Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland; <sup>6</sup>Department of Biology, Quebec Centre for Biodiversity Science, McGill University, Montreal, QC, Canada

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.

**Citation:** O'Connor M, Pennell MW, Altermatt F, Mellán 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

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

# {Sensory Ecology Design Patterns}

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

## Distributed {en}Coding

Local

Sparse

{en} Coding

{en} Coding



### Hierarchical Sparse Coding of Objects in Deep Convolutional Neural Networks

Xingyu Liu<sup>1</sup>, Zonglei Zhen<sup>1\*</sup> and Jia Liu<sup>2\*</sup>

<sup>1</sup> Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China

<sup>2</sup> 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 underlying object recognition.

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

#### OPEN ACCESS

**Edited by:**  
Matej Perc,  
University of Maribor, Slovenia

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

**Edited by:**

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

**Citation:**  
Liu X, Zhen Z and Liu J (2020)

Hierarchical Sparse Coding of Objects

in Deep Convolutional Neural

Networks.

Front. Comput. Neurosci. 14:578158.

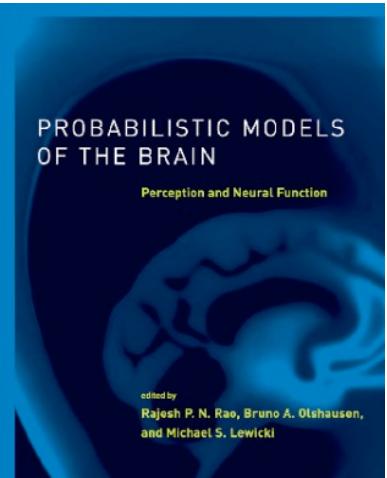
doi: 10.3389/fncom.2020.578158

#### 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., 1998). 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.,

# {Sparse Coding}

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



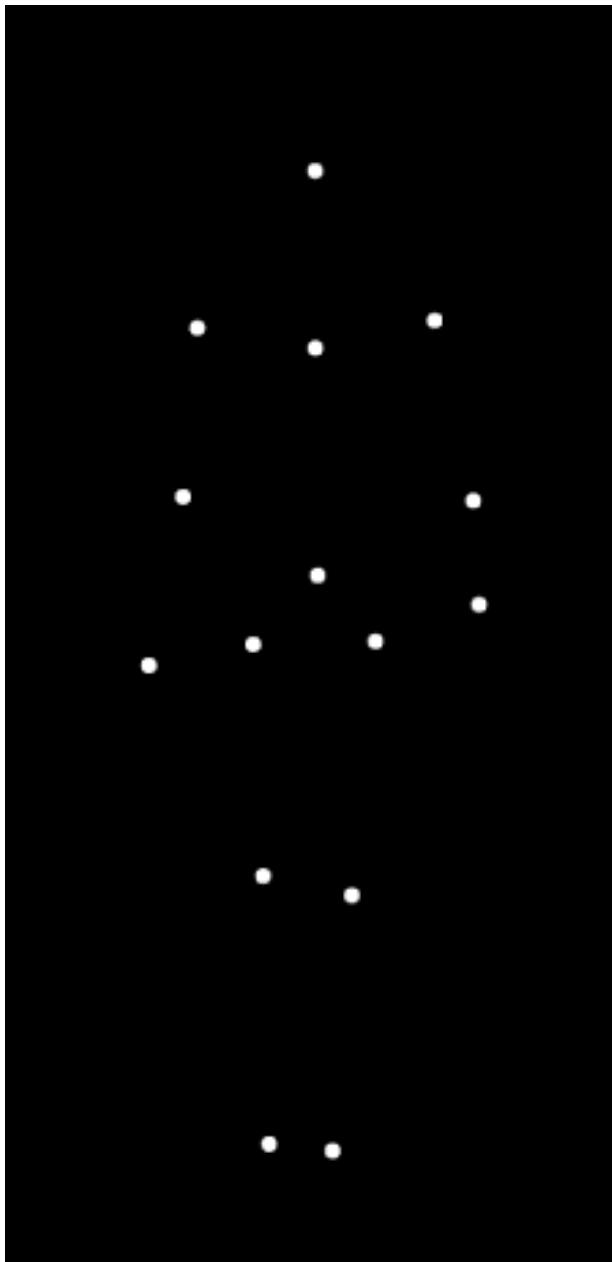
"A **sensing organism** should attempt to  
**represent stimuli**  
– *encoding of sensory information* –  
as a combination of  
**as few putative causes**  
**as possible**"

# {Biological Motion}

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

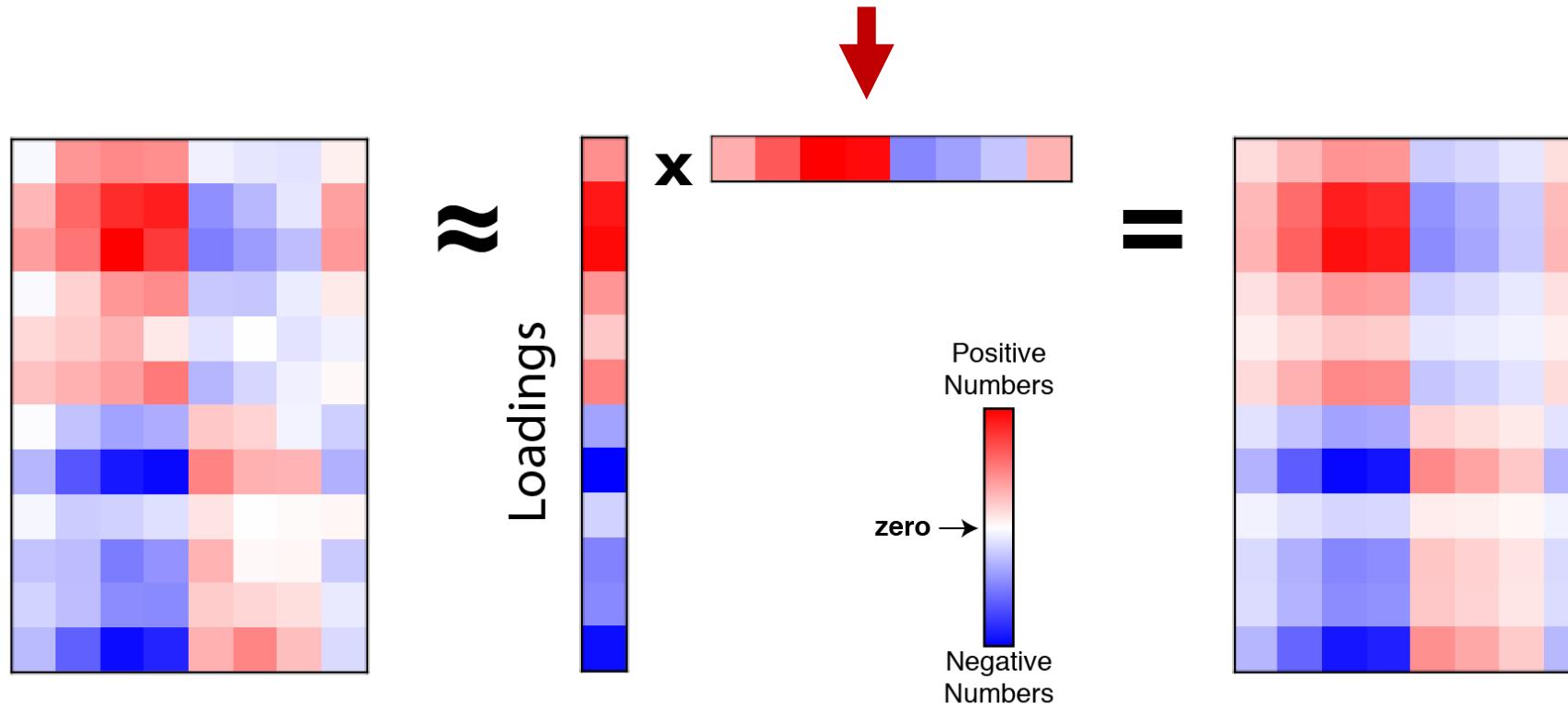
**Only 12 moving-dots are needed to determine direction of motion.**

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.



# {Maths: Sparse PCA}

Reconstruction of sensory input  
with 1 Principal Component (PC)



## PART II

# Proof of Concept

# PRIMATE Language Perception

# {Speech Sensitivity hypothesis}

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**}

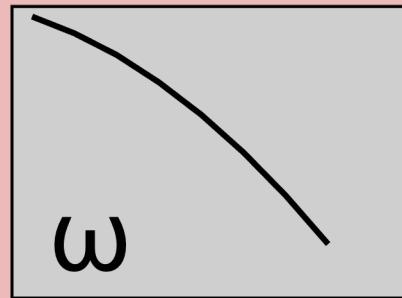
*then the sensitivity to vocalizations in the primate brain is likely to be represented through a sparse coding scheme.*

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

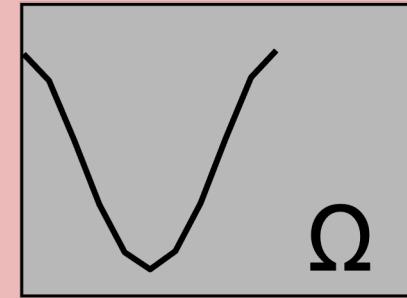
(Source van der Willigen et al. 2022)

# {Sparse coding Prediction}

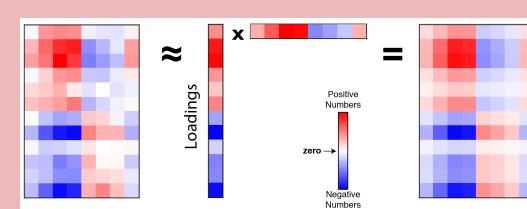
## Sparse coded



$\times$

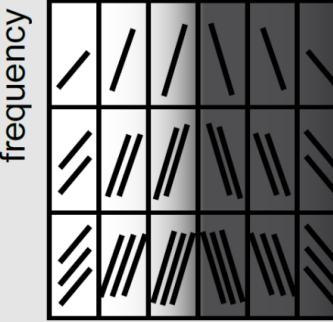


$=$

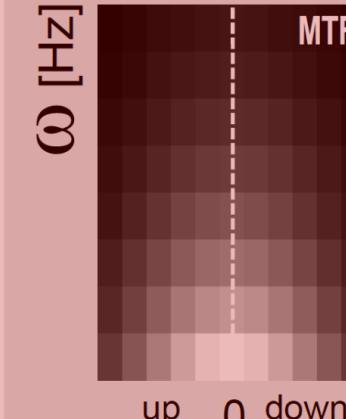


SPECTROTEMPORAL  
FEATURE BANK

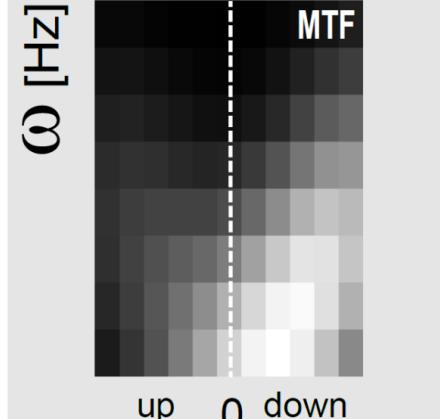
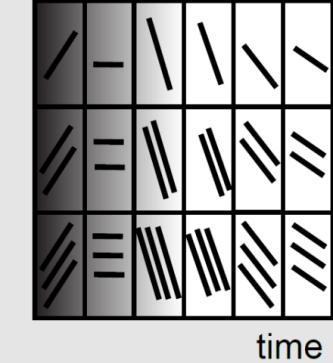
up/down symmetry



PERCEPTUAL  
OUTPUT

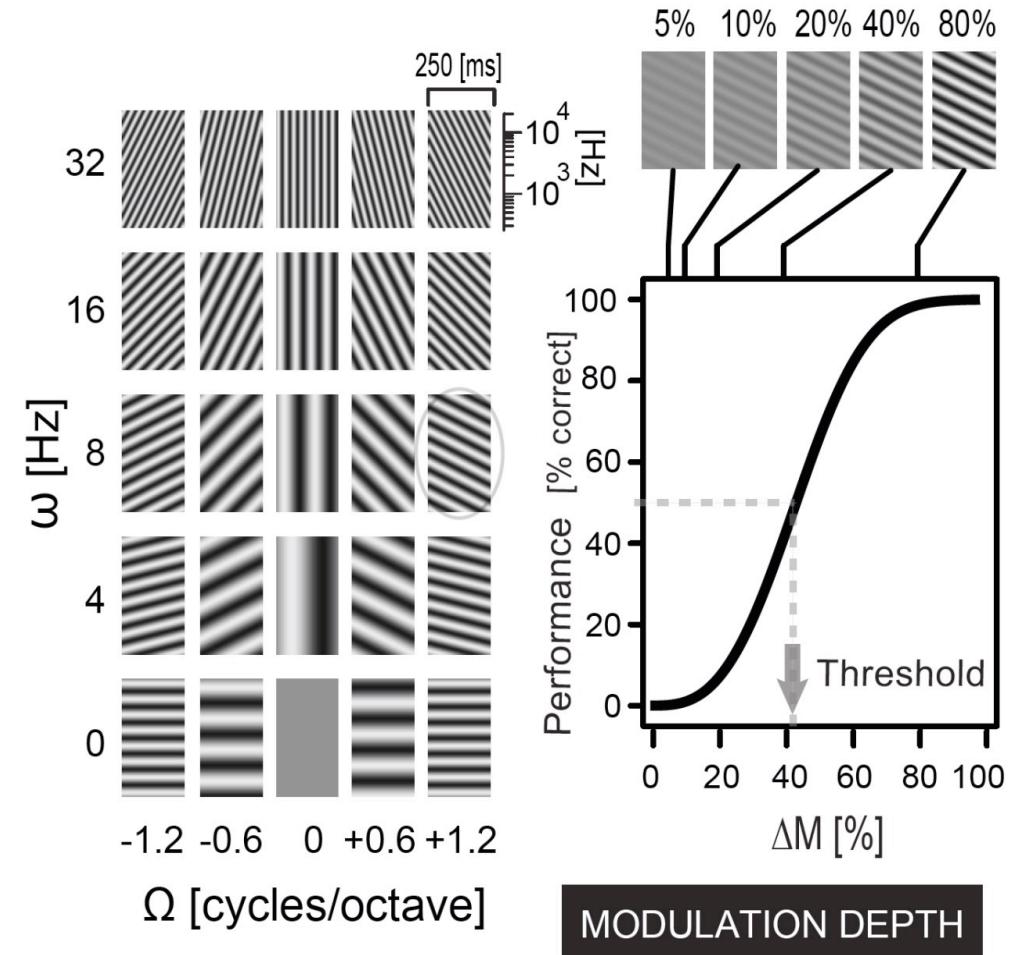
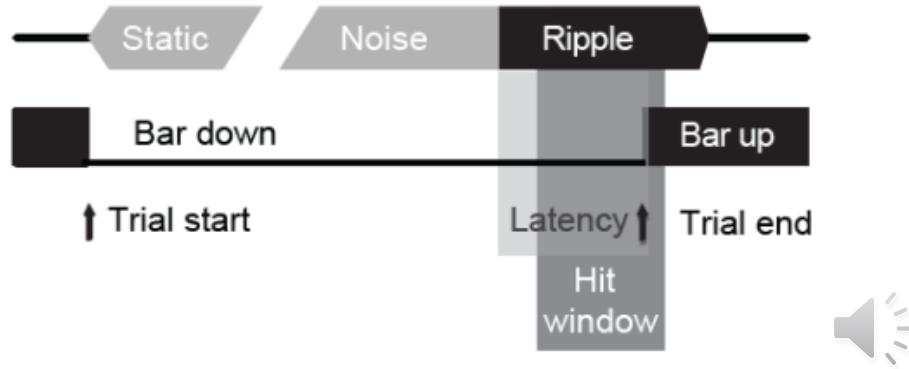


up/down asymmetry



# {Naturalistic Sounds}

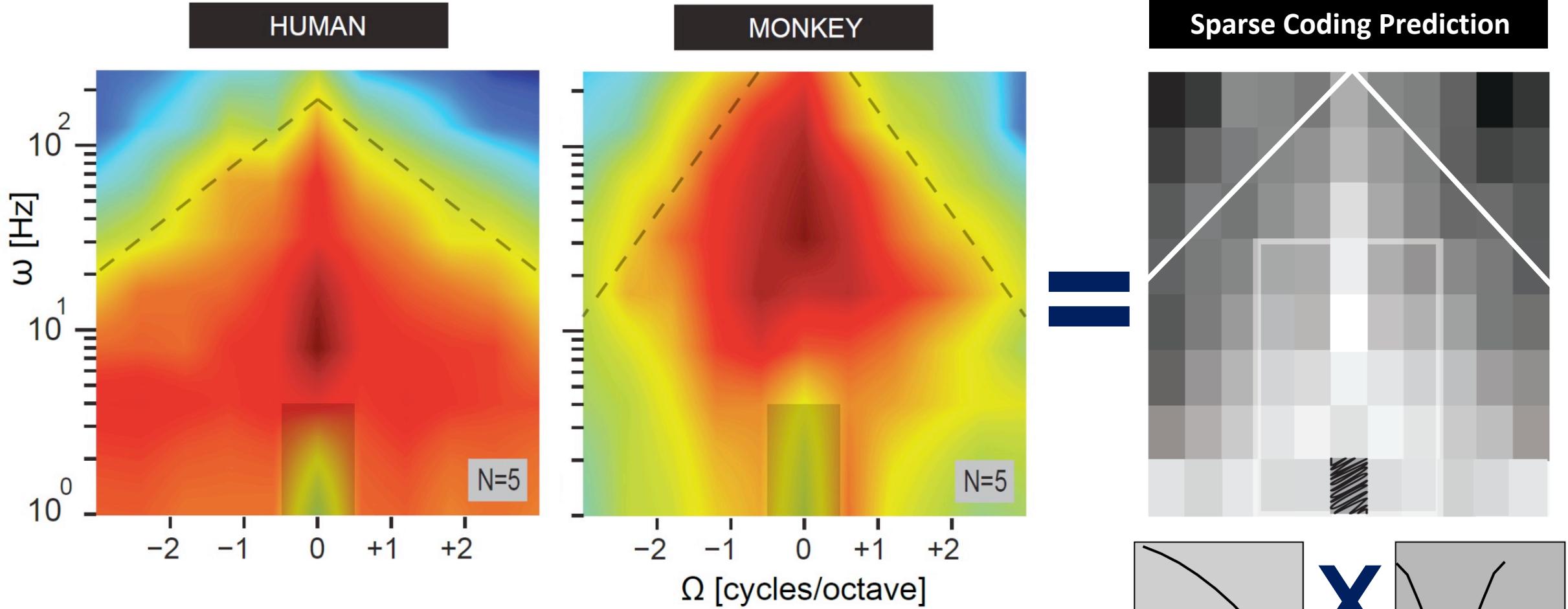
With the **separable spectrotemporal primate speech sensitivity hypothesis** in mind; (N=5) humans and (N=5) monkeys were exposed to a wide range of **naturalistic rippled-noises** to characterize their **perceptual abilities** to process acoustic **spectrotemporal modulations**.



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

(Source van der Willigen et al. 2022)

# {Empirical results confirm Sparse Coding}



<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.

# {Sparse Coding 3D-shape}

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

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

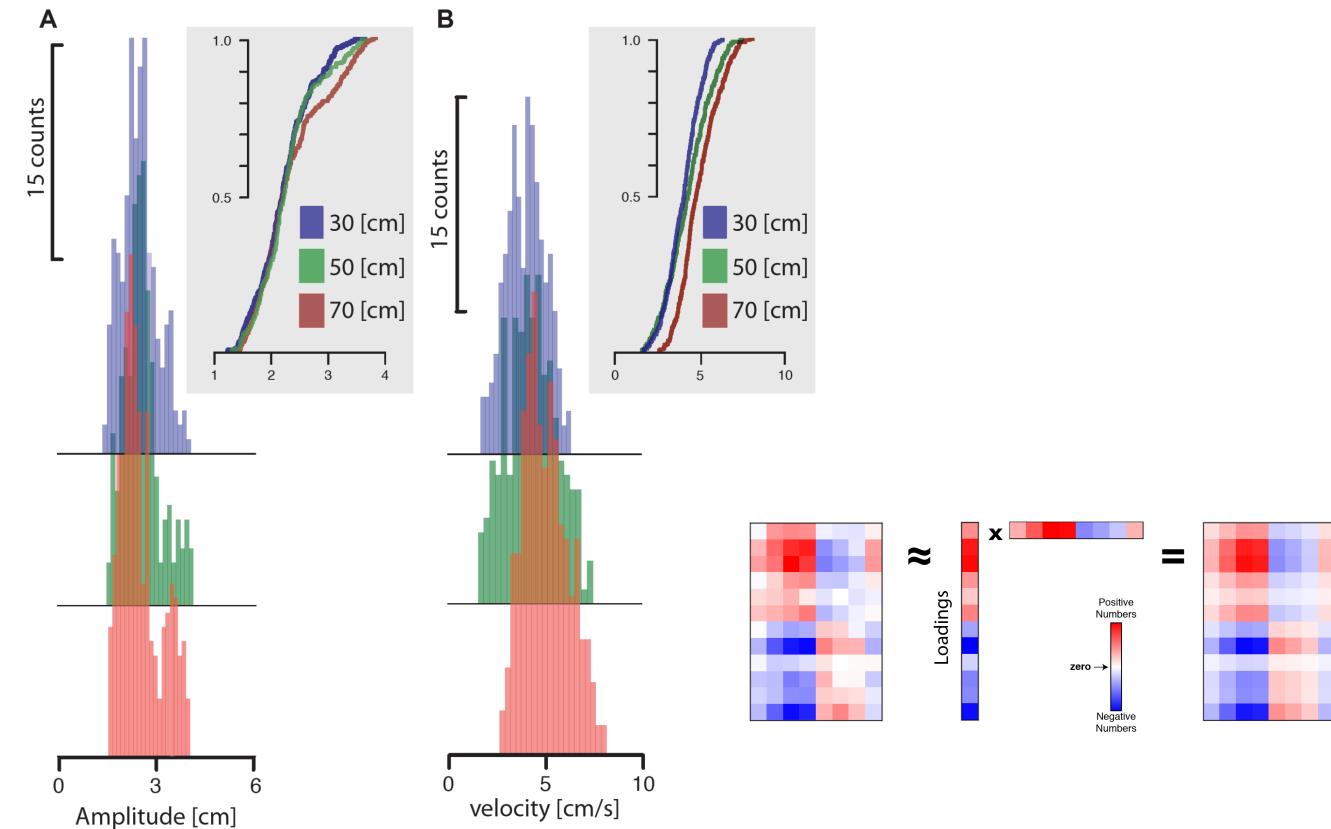
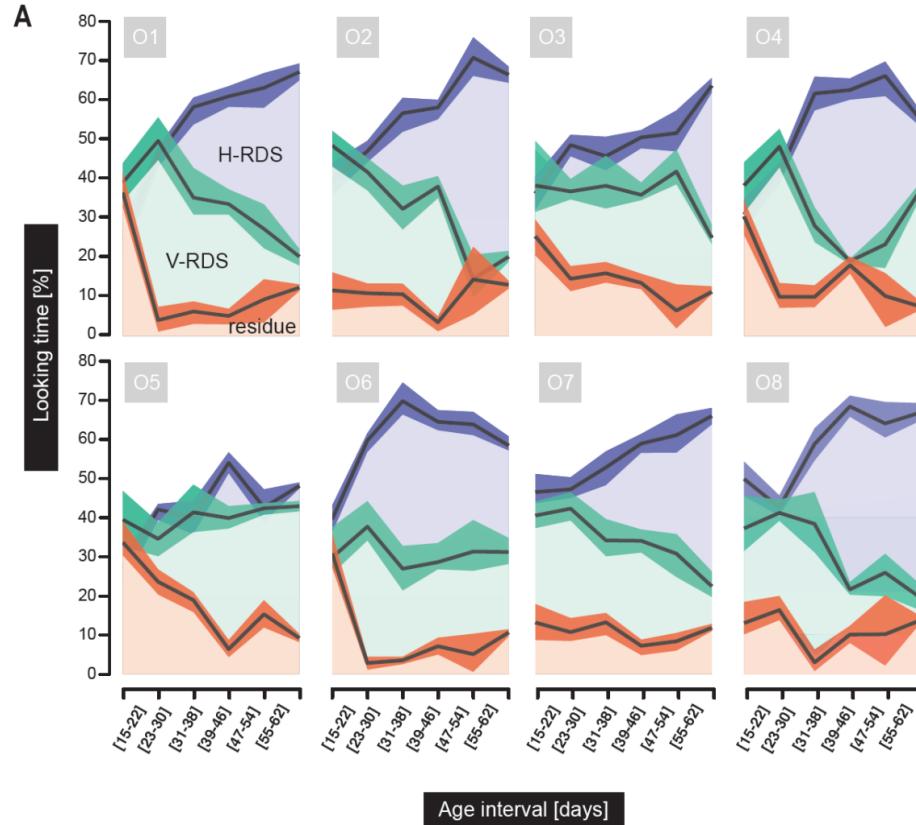
**Stereopsis**

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

**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**.

# {Empirical results confirm Sparse Coding}



Stereopsis

X

motion  
parallax

=

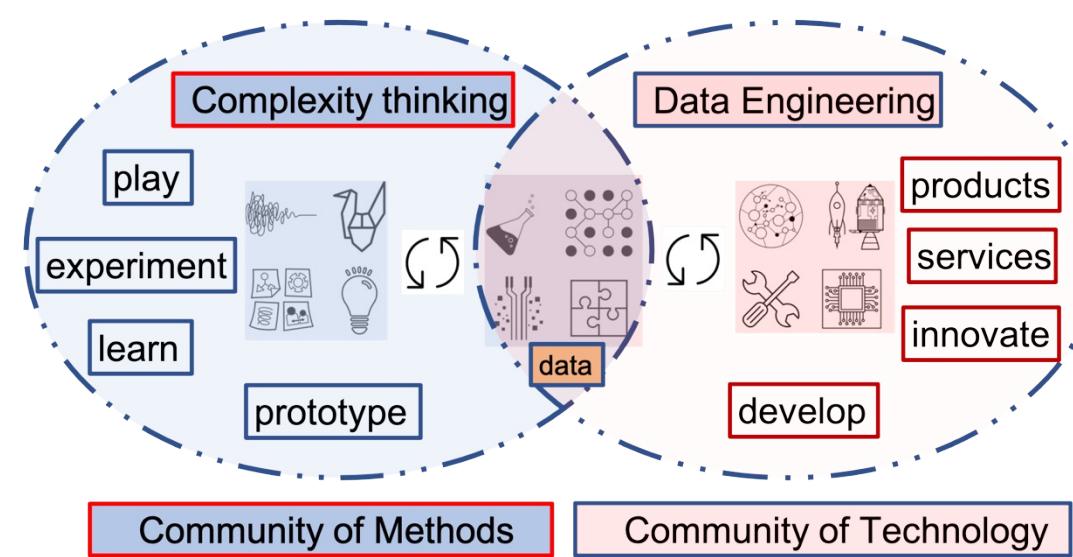
3D  
shape

# PART III

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

# PoP-UP AiRA-Lab: PROMETHEUS

Opbouwen van kennis & expertise  
door hands-on seminarie van datatechnologie



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

# Creating DATA FABRIC by building a *data infrastructure*



Inzetten van **data**technologie & AI**** om te komen tot **Bildung von 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\\_LivingLab\\_AI\\_ETHICS#fullTextFileContent](https://www.researchgate.net/publication/357933768_NO_MORE_SECRETS_AIRA_LivingLab_AI_ETHICS#fullTextFileContent)

# *Data infrastructure drivers*

People

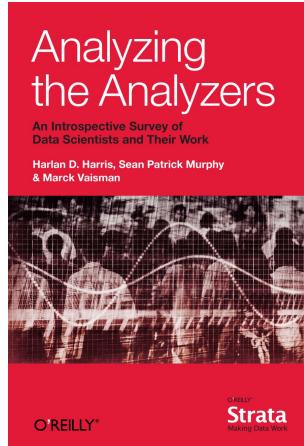
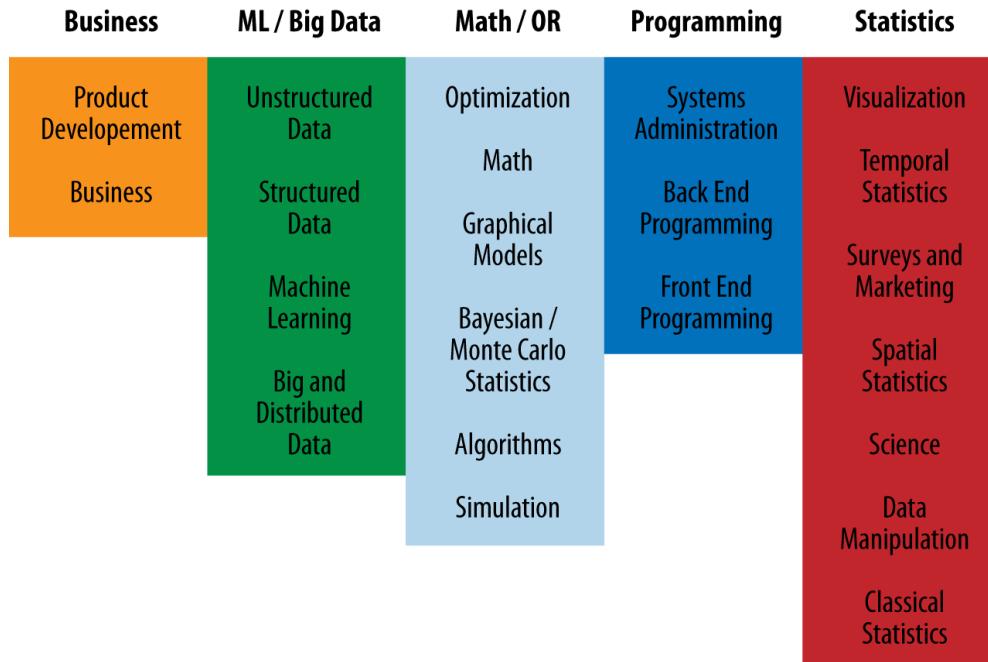
Data Access & Tooling

Knowledge Dissemination & Data Curation

Interoperability & Standards

Cloud Computing Capabilities & Accessibility

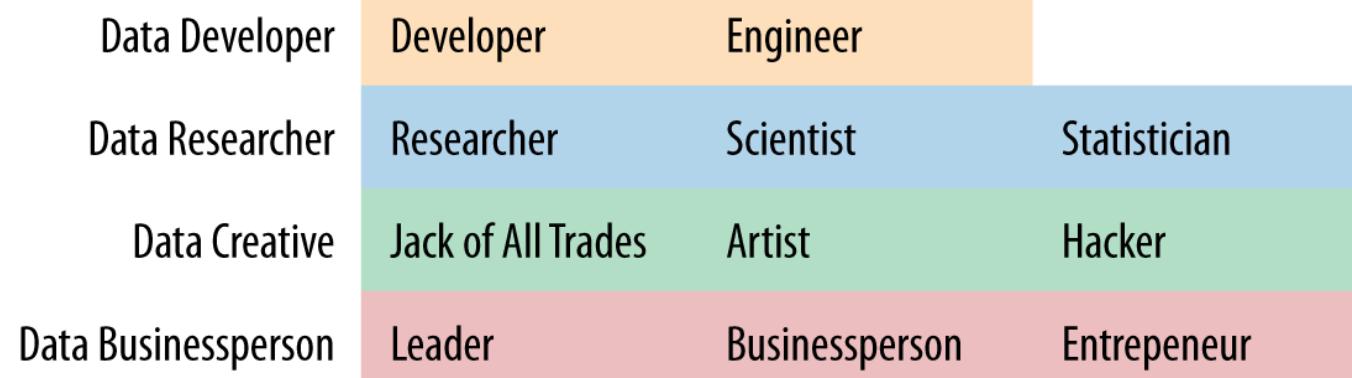
# {People}



The size of your organization often determines role overlap



[https://mleo.github.io/ml/GCP\\_day1/](https://mleo.github.io/ml/GCP_day1/)



# {Data Access & Tooling}

**Doing the right things**

**Doing things right**

**Discover**  
**Research Phase**

- [0] Collection
- [1] Access + Retrieval

Big Data (Acquisition/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  
Encryption

**Deliver**  
**Implementation Phase**

- [6] Organization + Managing
- [7] Automation + Reporting

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

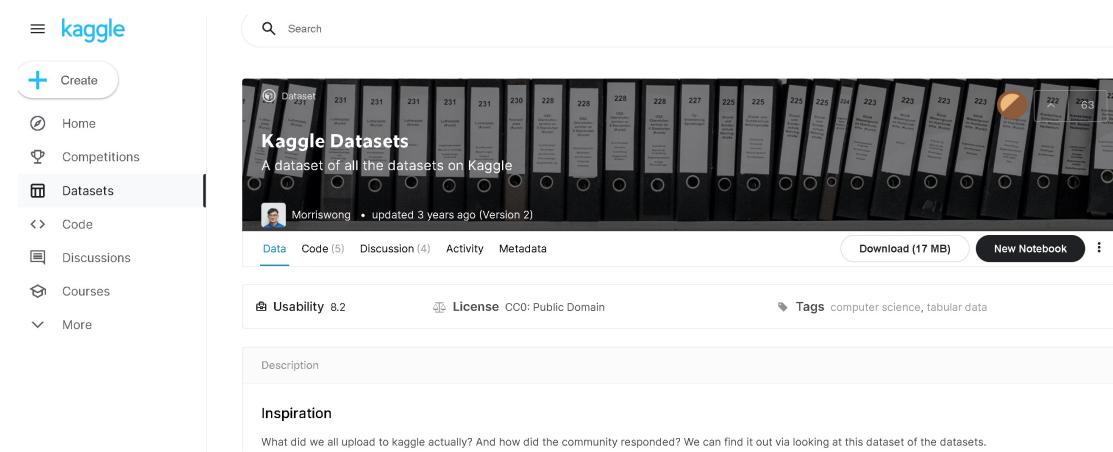
**Data**  
**Gathering & Ingesting**

**Data**  
**Dissemination & Curation**

# {Data Access & Tooling}



<https://dataverse.nl/>



Kaggle Datasets

A dataset of all the datasets on Kaggle

Morriswong • updated 3 years ago (Version 2)

Data Code (5) Discussion (4) Activity Metadata

Download (17 MB) New Notebook

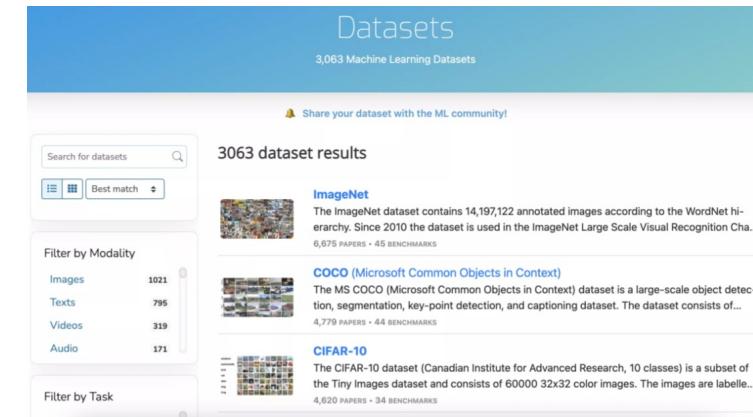
Usability 8.2 License CC0: Public Domain Tags computer science, tabular data

Description

Inspiration

What did we all upload to kaggle actually? And how did the community responded? We can find it out via looking at this dataset of the datasets.

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



Datasets

3,063 Machine Learning Datasets

Share your dataset with the ML community!

3063 dataset results

**ImageNet**  
The ImageNet dataset contains 14,197,122 annotated images according to the WordNet hierarchy. Since 2010 the dataset is used in the ImageNet Large Scale Visual Recognition Cha...  
6,675 PAPERS • 45 BENCHMARKS

**COCO (Microsoft Common Objects in Context)**  
The MS COCO (Microsoft Common Objects in Context) dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of...  
4,779 PAPERS • 44 BENCHMARKS

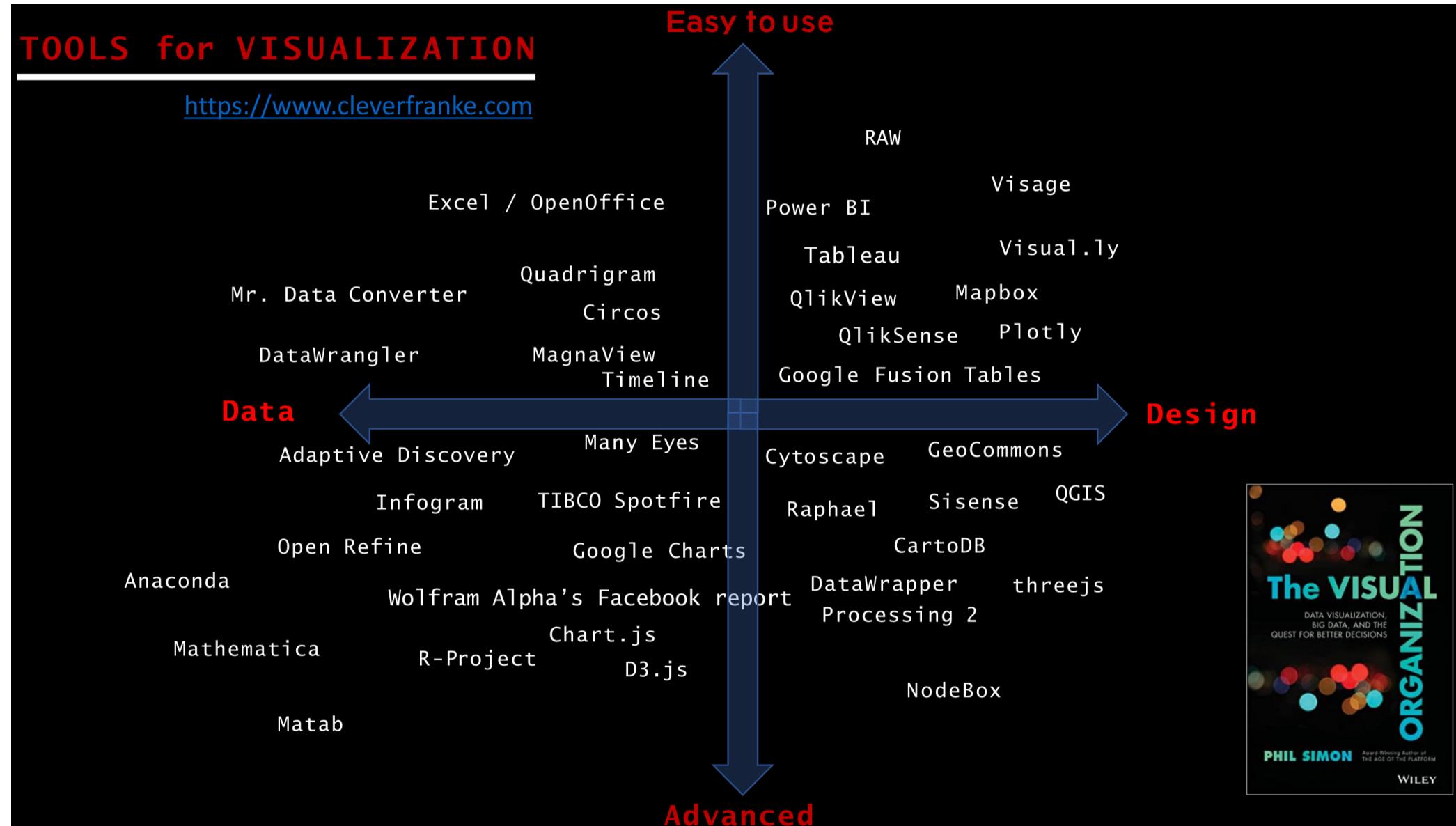
**CIFAR-10**  
The CIFAR-10 dataset (Canadian Institute for Advanced Research, 10 classes) is a subset of the Tiny Images dataset and consists of 60000 32x32 color images. The images are labelle...  
4,620 PAPERS • 34 BENCHMARKS

 [paperswithcode.com](https://paperswithcode.com)

## Papers with Code - Machine Learning Datasets

5577 datasets \* 65018 papers with code.

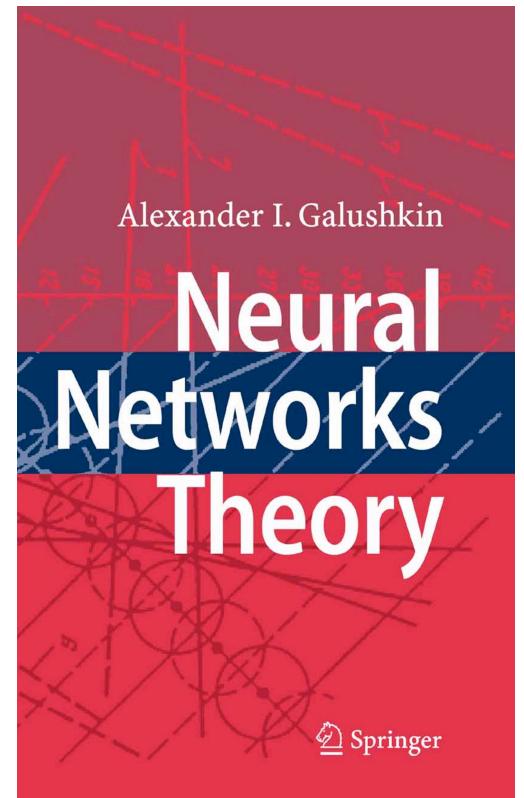
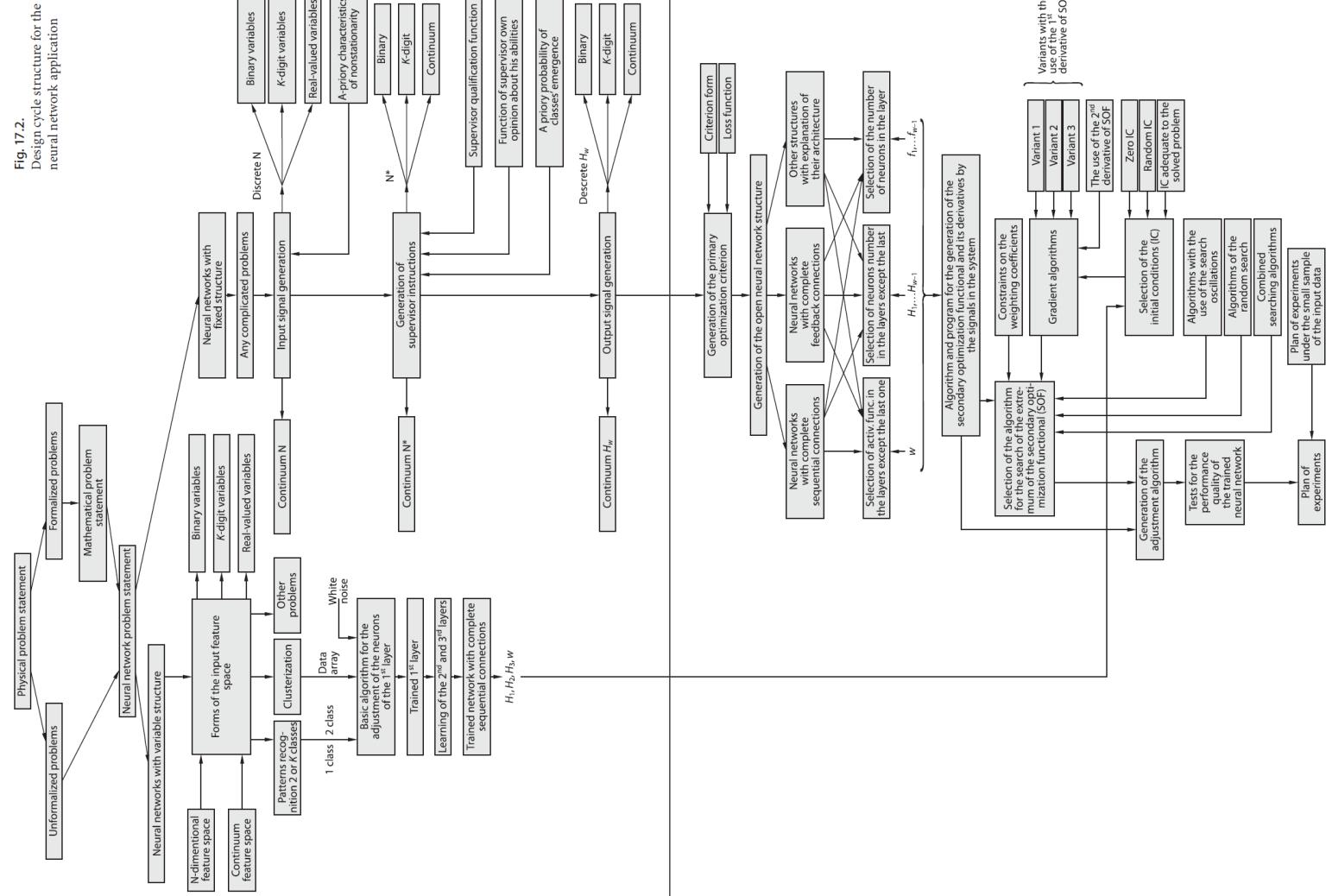
<https://paperswithcode.com/datasets>



# {Knowledge Dissemination & Curation}



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



# {Knowledge Dissemination & Curation}

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



De (on)mogelijkheden van kunstmatige intelligentie in het onderwijs

In opdracht van:  
Ministerie van Onderwijs, Cultuur & Wetenschap

Project:  
2018.068

Publicatienummer:  
2018.068.1828 v1.0.116

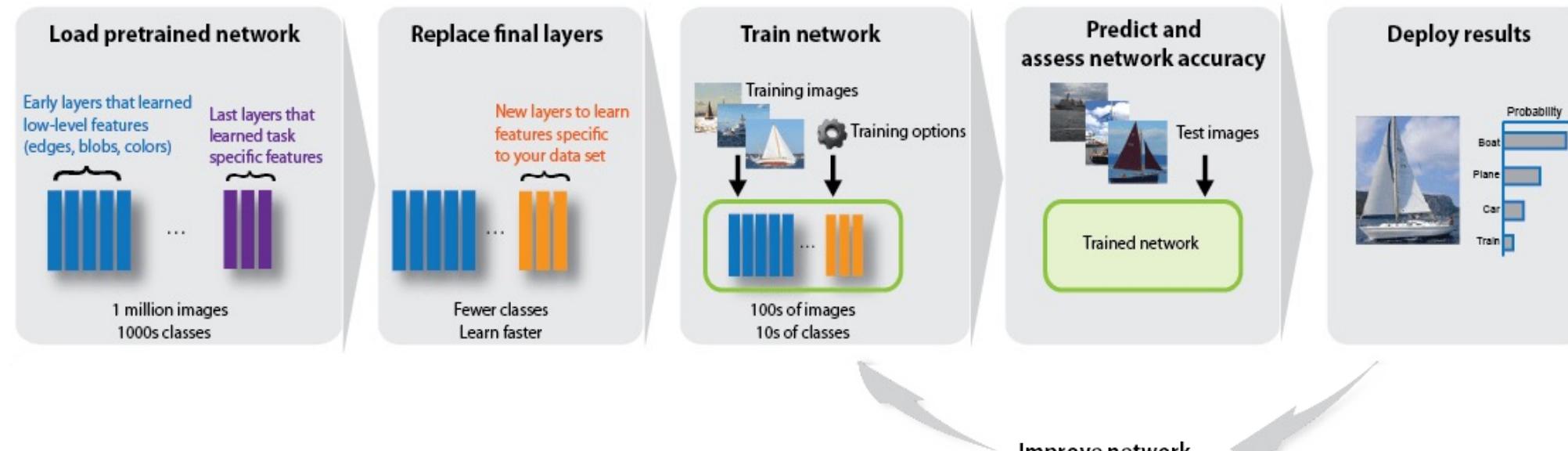
Datum:  
Utrecht, 21 januari 2019

Auteurs:  
ir. Tommy van der Vorst  
ir. Nick Jelicic  
mr. Marc de Vries  
Julie Albers

# {Knowledge Dissemination & Curation}



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



# {Knowledge Dissemination & Curation}

## Learn from prime examples

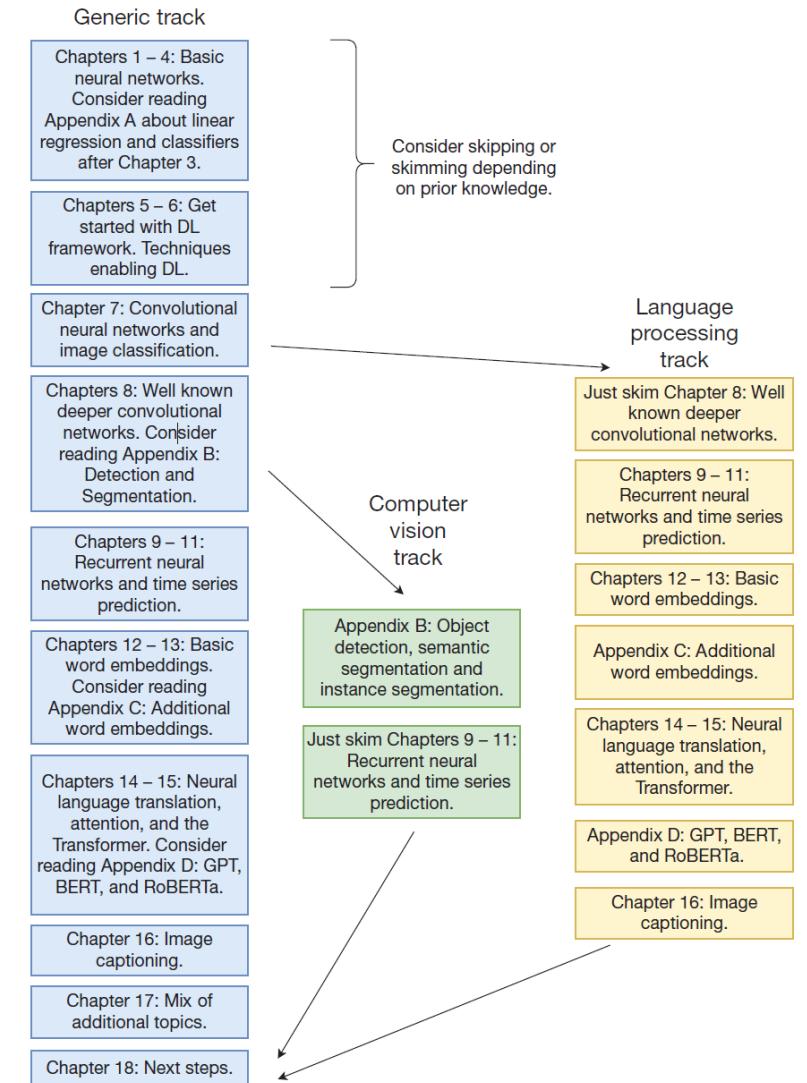
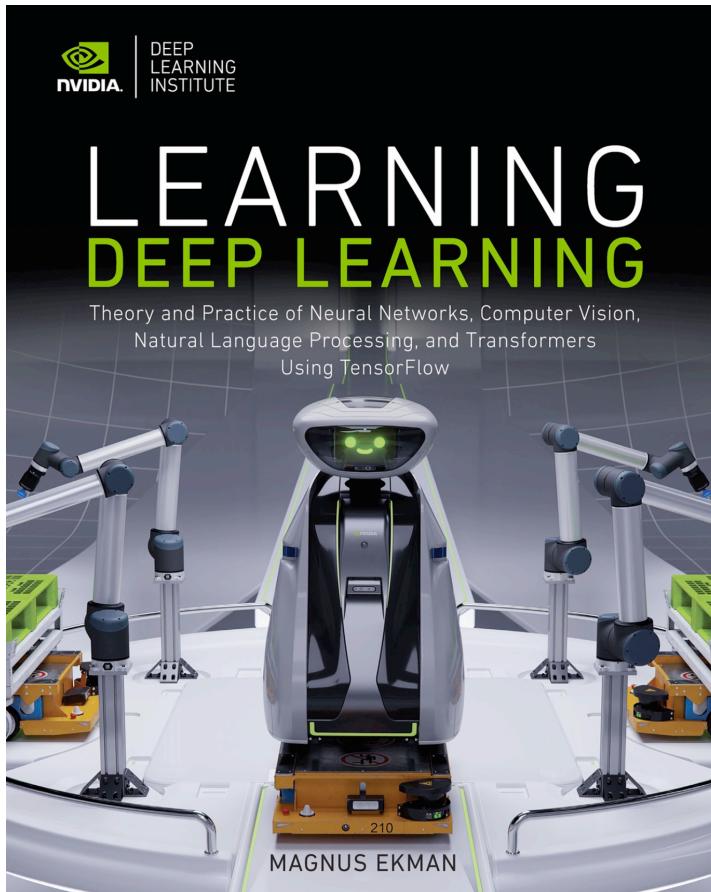


Figure P-5 Three different tracks to follow when reading this book

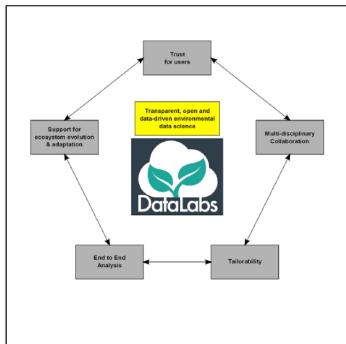
# {Knowledge Dissemination & Curation}

## Learn from prime examples

### Patterns

#### Tackling the Challenges of 21<sup>st</sup>-Century Open Science and Beyond: A Data Science Lab Approach

##### Graphical Abstract



##### 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.

##### 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

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

CellPress

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

### 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 Konvalinov 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

[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

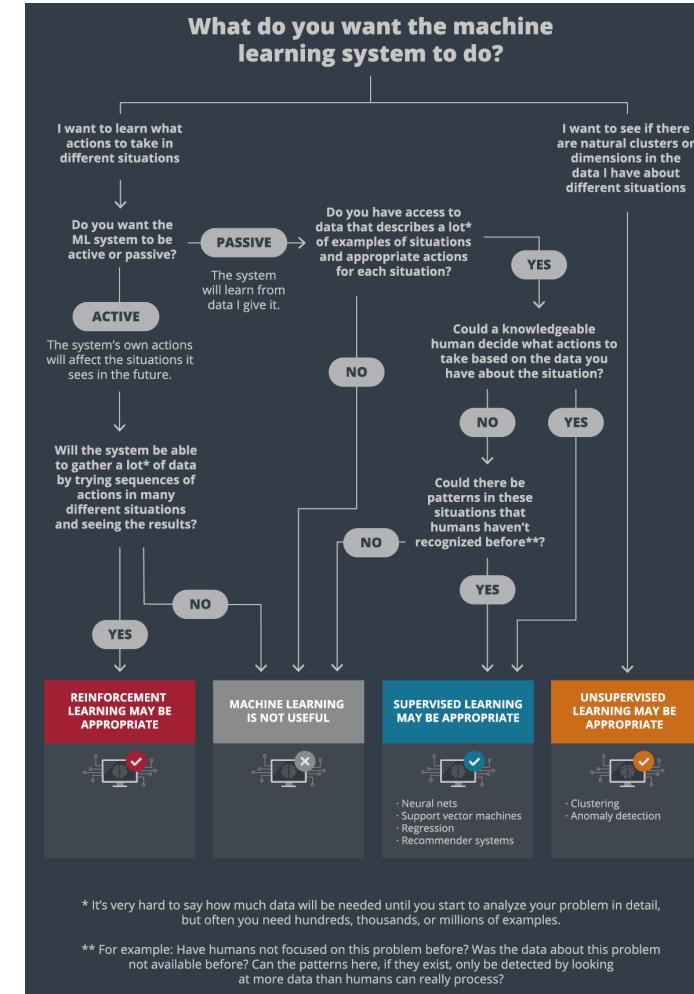
RB17-2020



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

 newsmap.js.herokuapp.com

## NewsMap.js

A squarified treemap visualization of Google News based on the original newsmap.jp

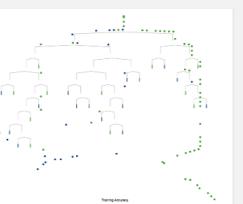




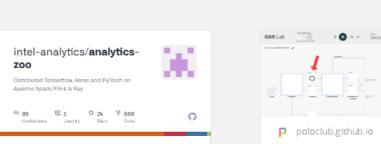


**Making predictions**

The newly-trained decision tree model determines whether a home is in San Francisco or New York by running each point through the branches.







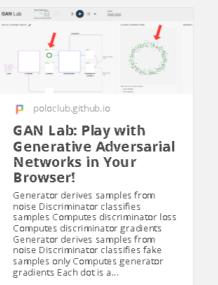
**CodePen**

An online code editor, learning environment, and community for front-end web development using HTML, CSS and JavaScript code snippets, projects, and web applications.



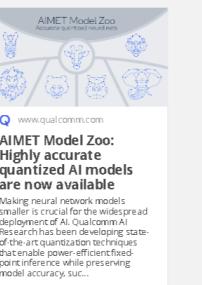
**intel-analytics/analytics-zoo**

Distributed TensorFlow, Keras and PyTorch, Keras and BigDL, on Apache Spark & Ray



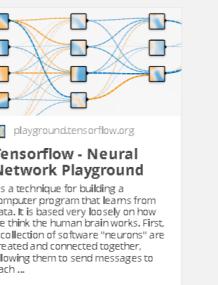
**GAN Lab: Play with Generative Adversarial Networks in Your Browser!**

Generator derives samples from noise Discriminator classifies samples Computer computes discriminator loss



**AIMET Model Zoo:**  
Highly accurate quantized AI models are available

Making neural network models smaller and faster for the widest pre-deployment AI Quantization Research has been developing state-of-the-art quantization techniques that enable power-efficient/fingerprint inference while preserving model accuracy such as...



**Tensorflow - Neural Network Playground**

It's a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a number of small components are created and connected together, allowing them to send messages to each other.



**A MATLAB implementation of the TensorFlow Neural Network Playground**

In Big Data containers like there is no way that you will ever work with TensorFlow directly in your favorite frameworks, like Hadoop or Spark. At some point, I had to learn how to use TensorFlow and as Apache spark seems to be more flexible and faster, I decided to...



**playground.info - Wolfram|Alpha**

Uh oh! Wolfram|Alpha doesn't run without JavaScript. Please enable JavaScript. If you don't know how, you can find instructions. Once you've done that, refresh this page to start using Wolfram|Alpha.



**Wolfram|Alpha:**  
Making the world's knowledge computable

Uh oh! Wolfram|Alpha doesn't run without JavaScript. Please enable JavaScript. If you don't know how, you can find instructions. Once you've done that, refresh this page to start using Wolfram|Alpha.



**SURFara Knowledge Base**

Welcome to the SURFara Wiki! For users! In this Wiki you can find all information about our services.



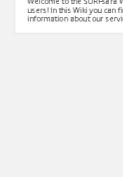
**neo4j**

The #1 Database for Connected Data



**Processing Foundation**

We create platforms to provide creative tools for artists, designers, educators, students, and hobbyists to make art, code, and share their work.



**TensorFlow**

TensorFlow is a platform for machine learning and deep learning research and development.



**Explaining Predictions of Machine Learning Models with LIME - Miro Data Science Meetup**

Slides from Miro Data Science Meetup at the Miro Data Science Meetup on December 12th, 2017.

<https://miro.com>

# Learn from prime examples

## DATA Science Playgrounds

[https://miro.com/app/board/uXjVOCnGVhU/#?invite\\_link\\_id=716932584109](https://miro.com/app/board/uXjVOCnGVhU/#?invite_link_id=716932584109)

# {Knowledge Dissemination & Curation}



HOGESCHOOL  
ROTTERDAM

## Post-Doc Voucher Milestones

**HUMAN\_IN\_THE\_LOOP**  
How to design deep nets that are interpretable by humans

**SENSORY ECOLOGY**  
A Data Science Quest On the Origin of Natural Language Processing

**Milestones**  
Post-Doc Voucher & PopUP AI-LAB Prometheus (jan 2022)

**Planning**  
JAAR1 2020-2021 Engineering Phase  
JAAR2 2021-2022 Kritisch Thinking Phase

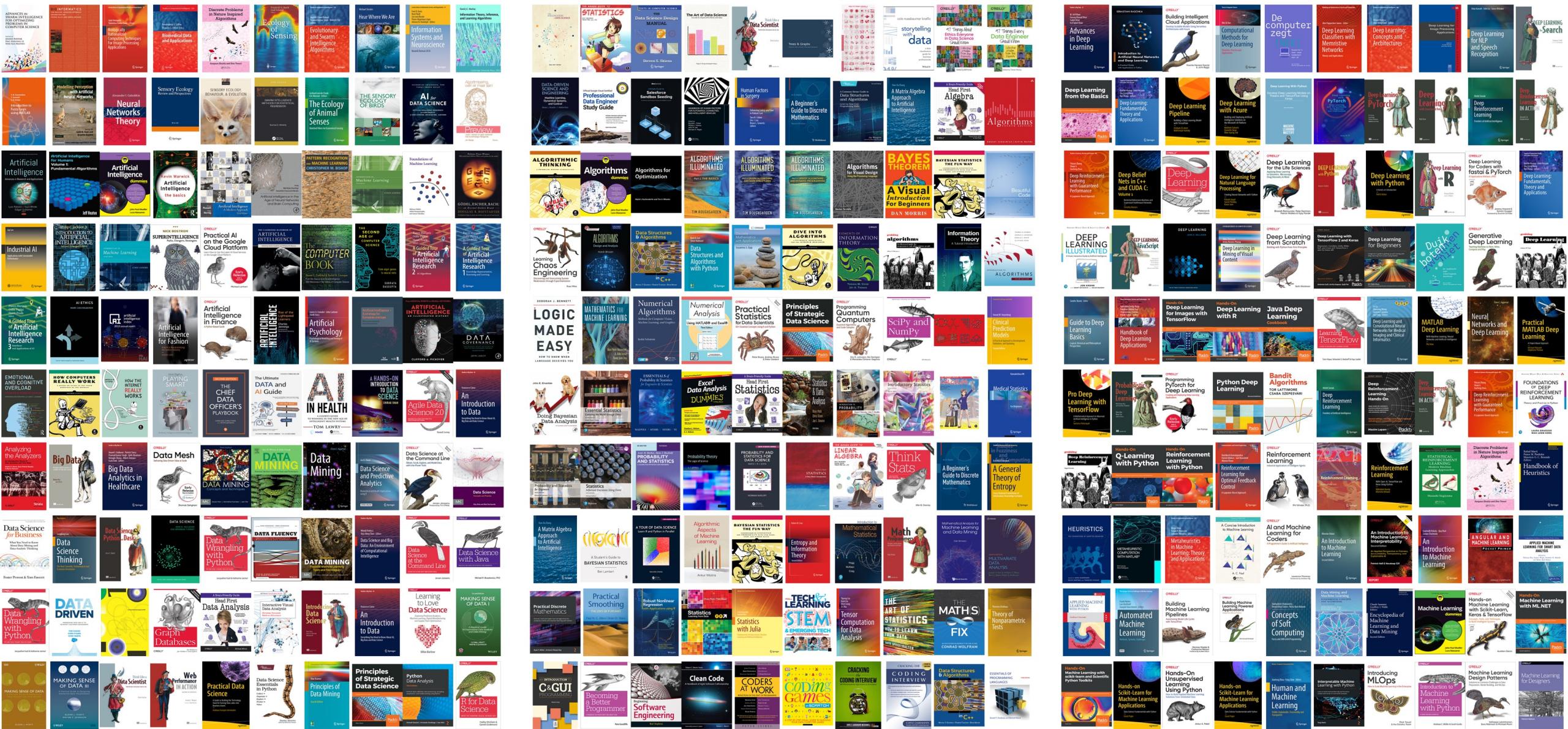
**DO IT NOW**

**TOOLS FOR VISUALIZATION**

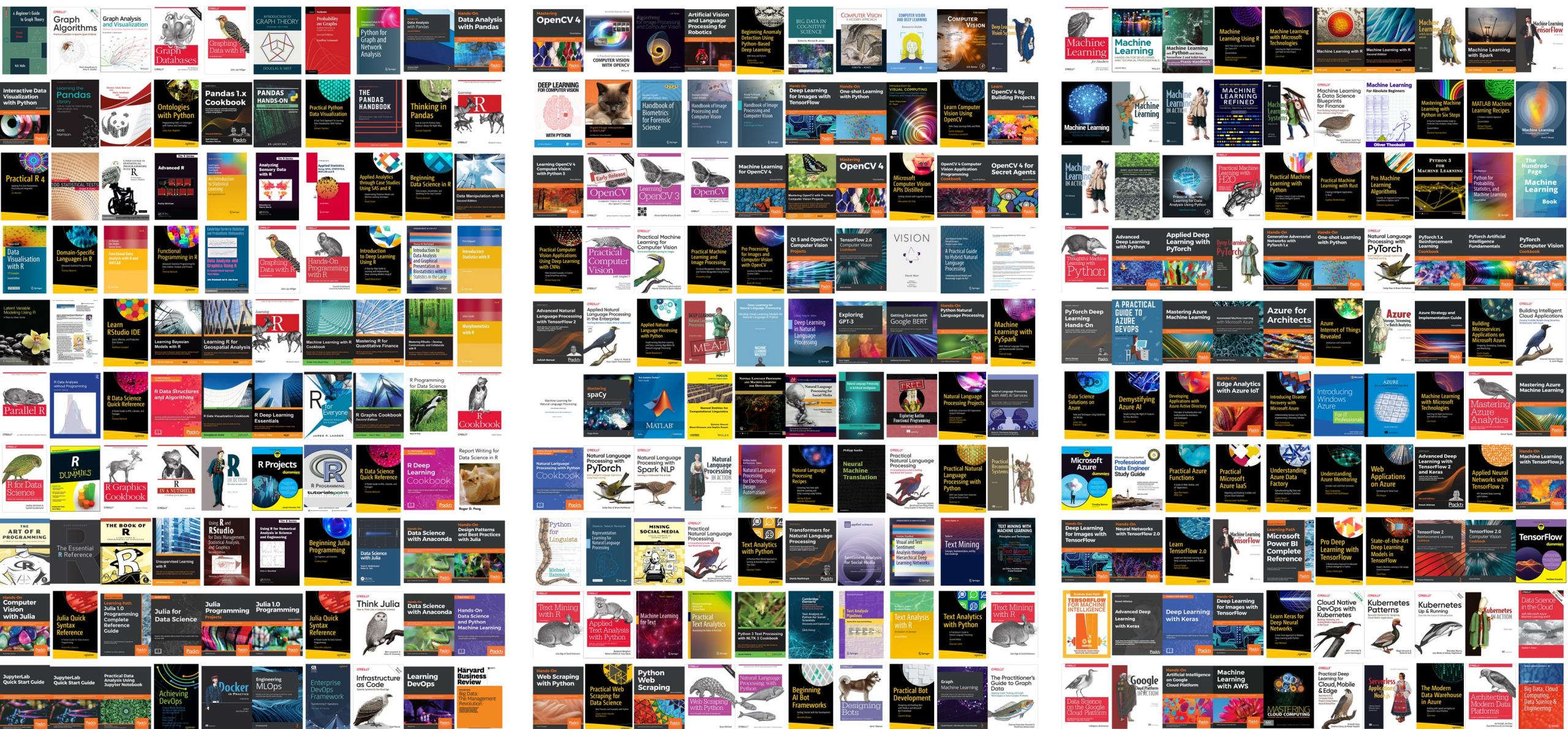
**DATA SCIENCE / ENGINEERING PROCESS**

[https://miro.com/app/board/o9J\\_l-CrLzg=/?invite\\_link\\_id=427711164595](https://miro.com/app/board/o9J_l-CrLzg=/?invite_link_id=427711164595)

# {Studied Materials: books}

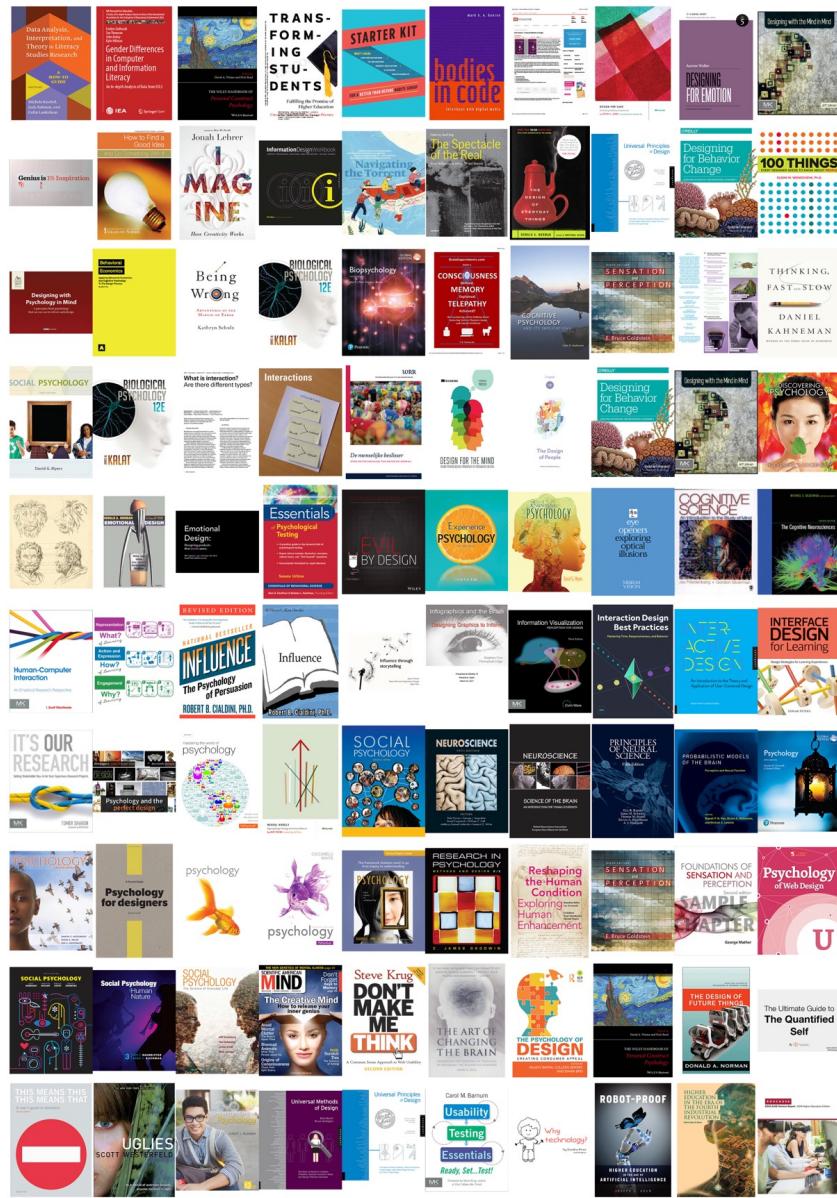


# {Studied Materials: books}



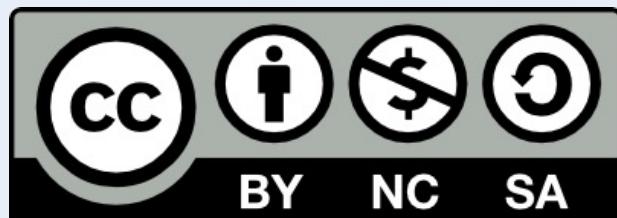
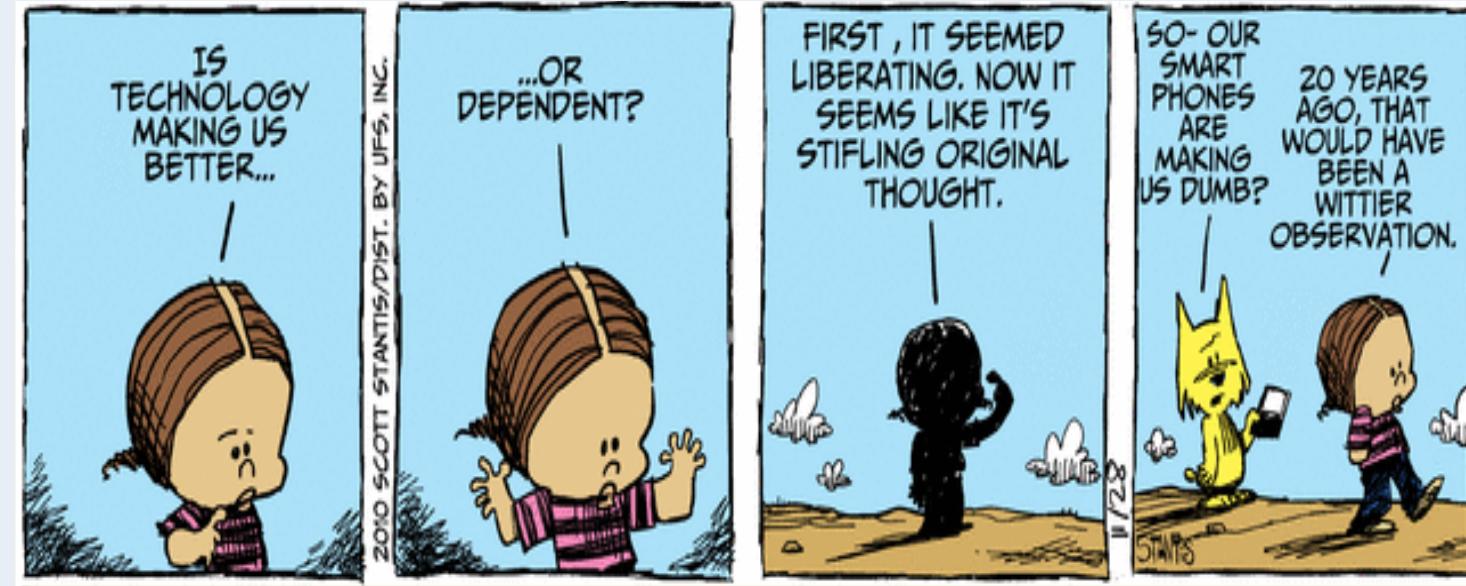
# {Studied Materials: books}

# {Studied Materials: books}



<http://creativecommons.org/licenses/by-nc-sa/3.0/>

These materials are licensed under a Creative Commons Attribution-Share-Alike license. You can change it, transmit it, show it to other people. Just always give credit to RFvdW.



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

Creative Commons License Types		
	Can someone use it commercially?	Can someone create new versions of it?
Attribution	①	
Share Alike	①②	 Yup, AND they must license the new work under a Share Alike license.
No Derivatives	①③	
Non-Commercial	①④	 Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
Non-Commercial Share Alike	①②③	 Yup, AND they must license the new work under a Non-Commercial Share Alike license.
Non-Commercial No Derivatives	①③④	

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