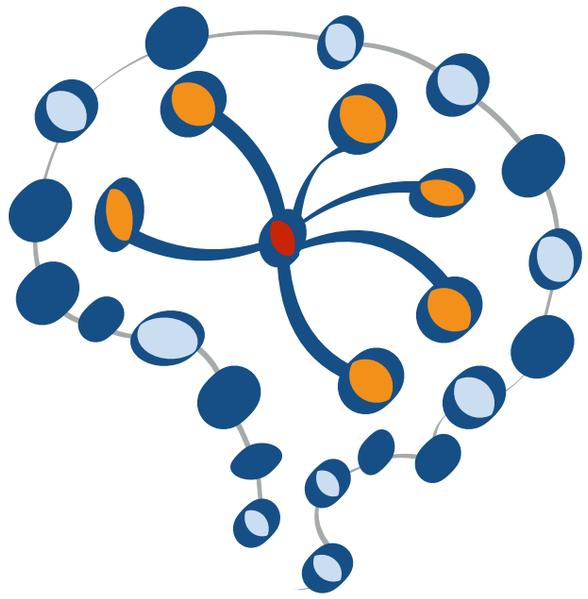
STAT 453: Introduction to Deep Learning and Generative Models

Sebastian Raschka

http://stat.wisc.edu/~sraschka/teaching



Deep Learning & AI News #4

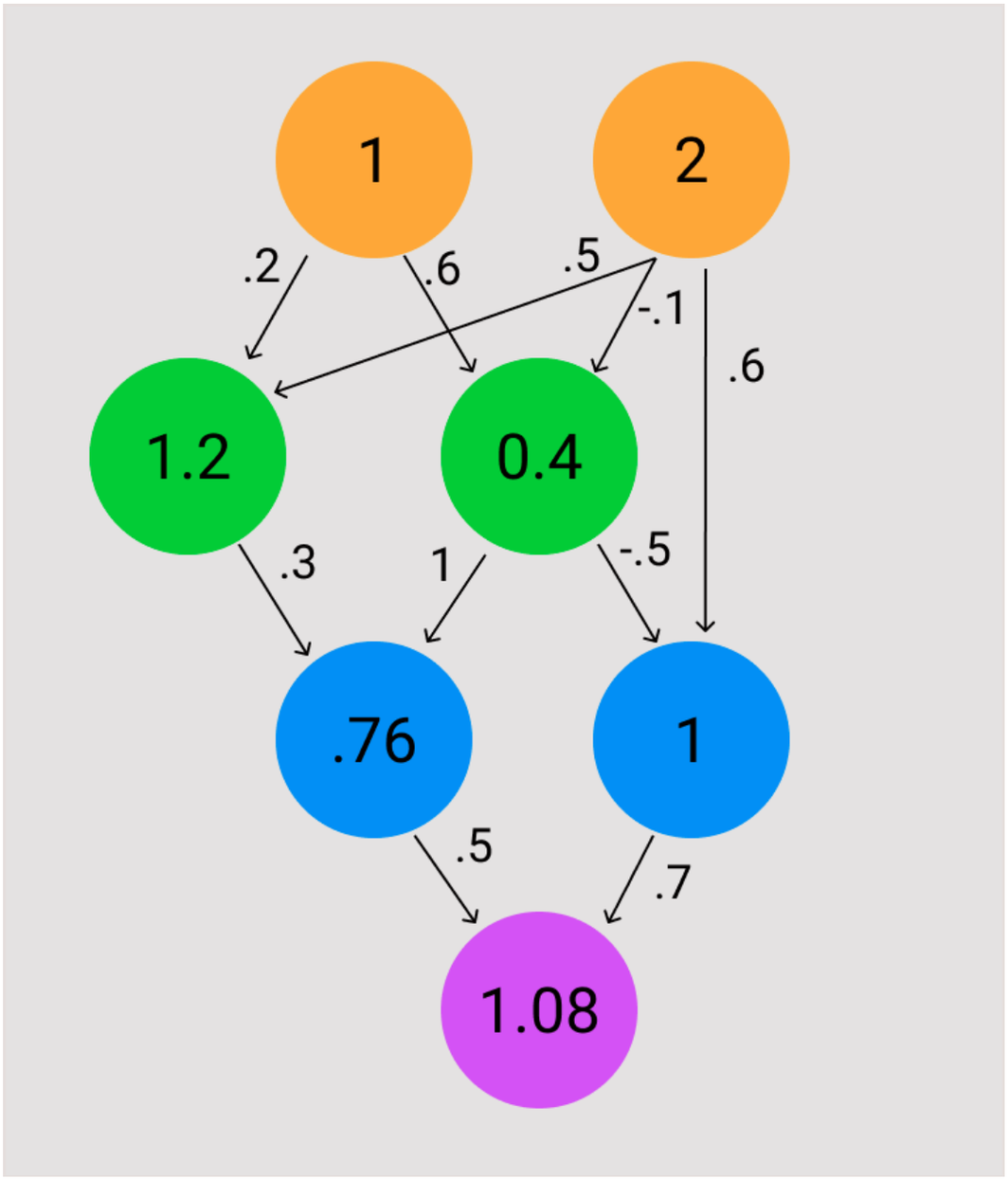
Interesting Things Related to Deep Learning

Feb 20th, 2021

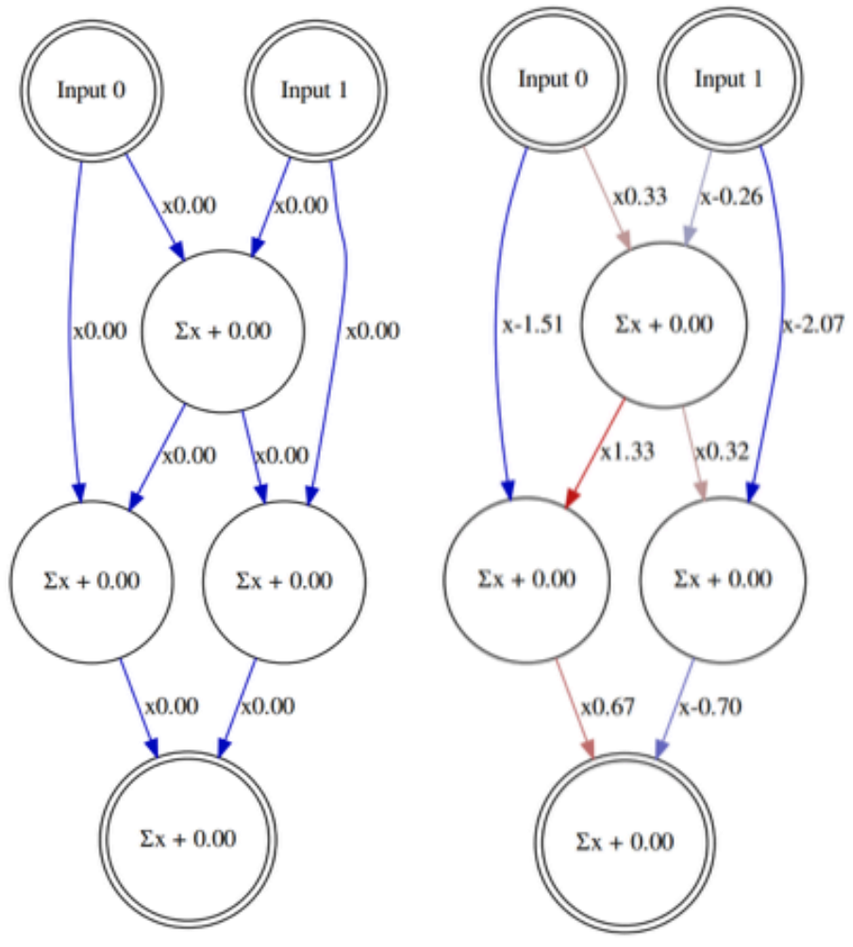
Freewire: Freely Wired Neural Networks

https://github.com/noahtren/Freewire

A PyTorch extension library for creating optimized freely wired neural networks to run on CUDA




```
m = Model(g, initialization="he")
visualize(g, title="architecture_and_weights")
```



arXiv.org > cs > arXiv:1904.01569

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 2 Apr 2019 (v1), last revised 8 Apr 2019 (this version, v2)]

Exploring Randomly Wired Neural Networks for Image Recognition

Saining Xie, Alexander Kirillov, Ross Girshick, Kaiming He

https://arxiv.org/abs/1904.01569

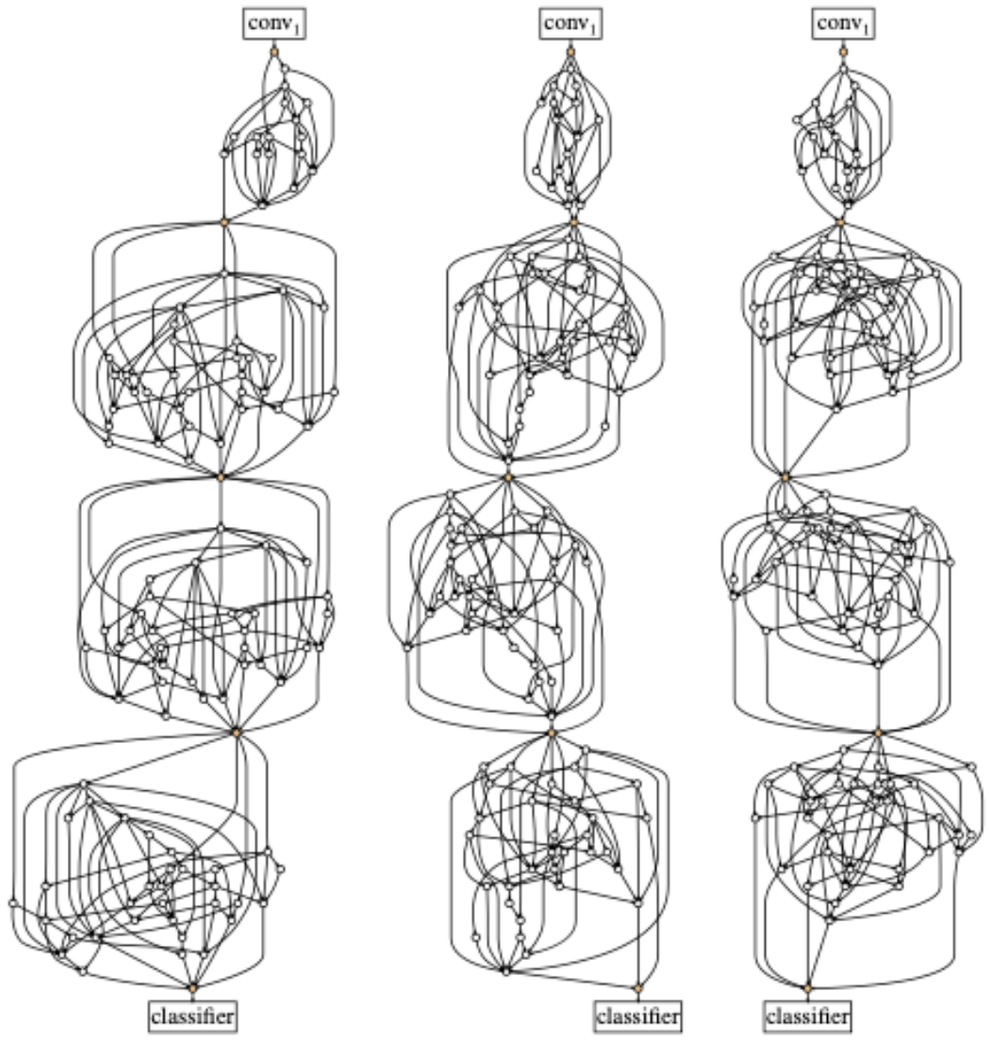


Figure 1. **Randomly wired neural networks** generated by the classical Watts-Strogatz (WS) [50] model: these three instances of random networks achieve (left-to-right) 79.1%, 79.1%, 79.0% classification accuracy on ImageNet under a similar computational budget to ResNet-50, which has 77.1% accuracy.

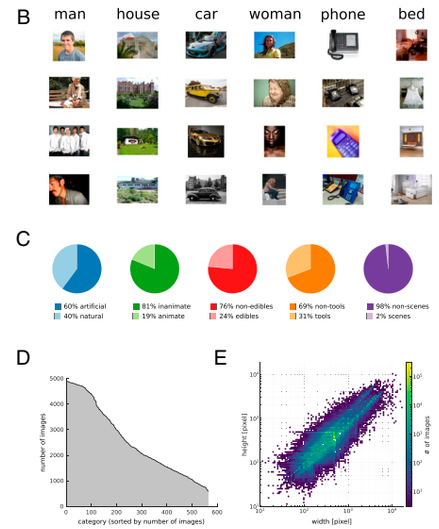
RESEARCH ARTICLE

An ecologically motivated image dataset for deep learning yields better models of human vision

Johannes Mehrer, Courtney J. Spoerer, Emer C. Jones, Nikolaus Kriegeskorte, and Tim C. Kietzmann

PNAS February 23, 2021 118 (8) e2011417118; https://doi.org/10.1073/pnas.2011417118

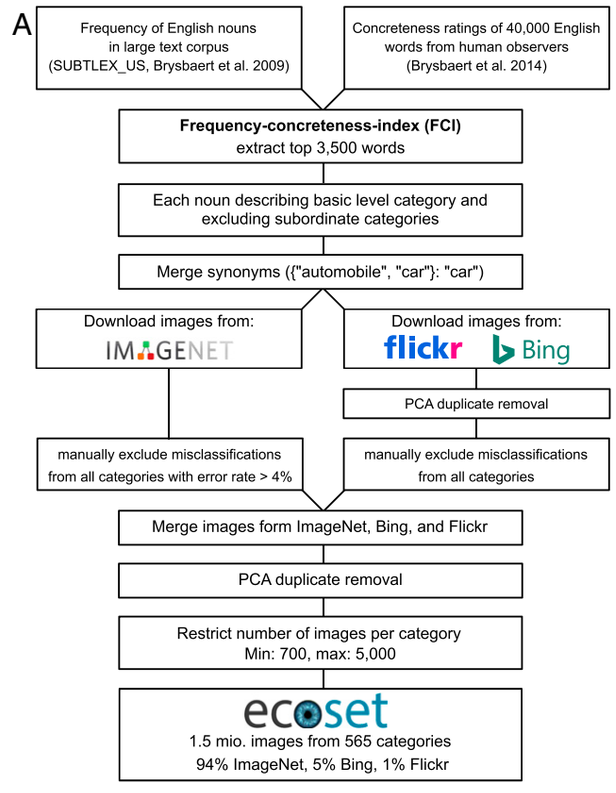
Edited by J. Anthony Movshon, New York University, New York, NY, and approved January 12, 2021 (received for review June 13, 2020)



- Object categories important to humans
- 1.5 million images
- 565 basic-level categories

Code (dataset and pre-trained CNNs):
https://codeocean.com/capsule/9570390/tree/v1

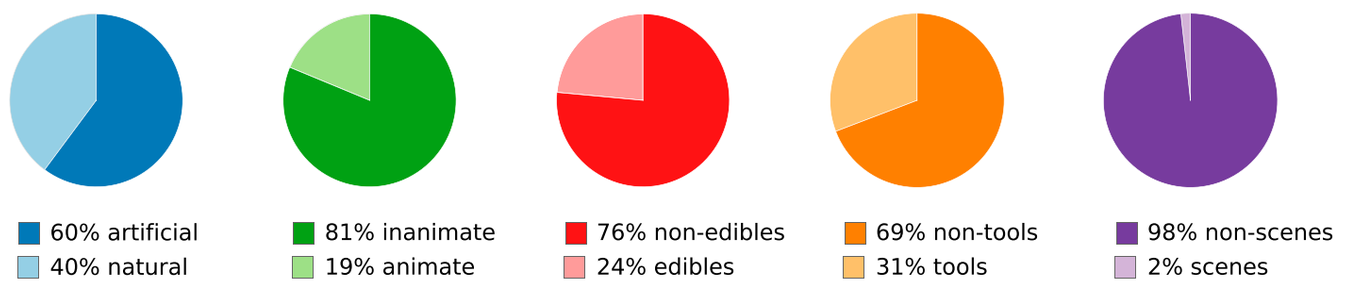
Paper: https://www.pnas.org/content/118/8/e2011417118



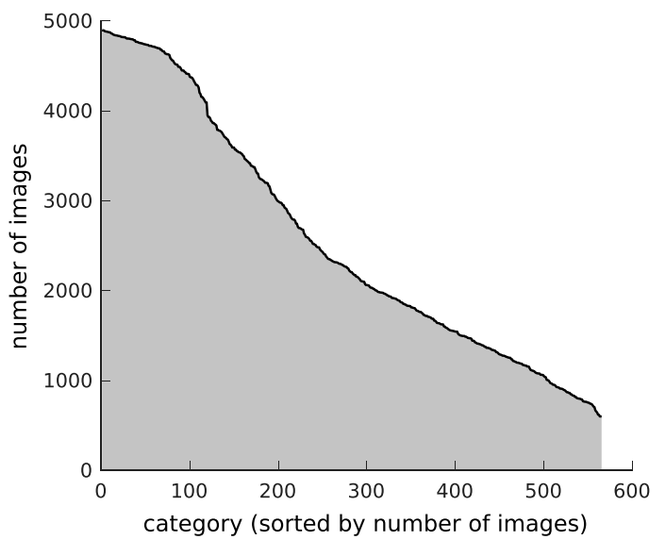
B


man
house
car
woman
phone
bed

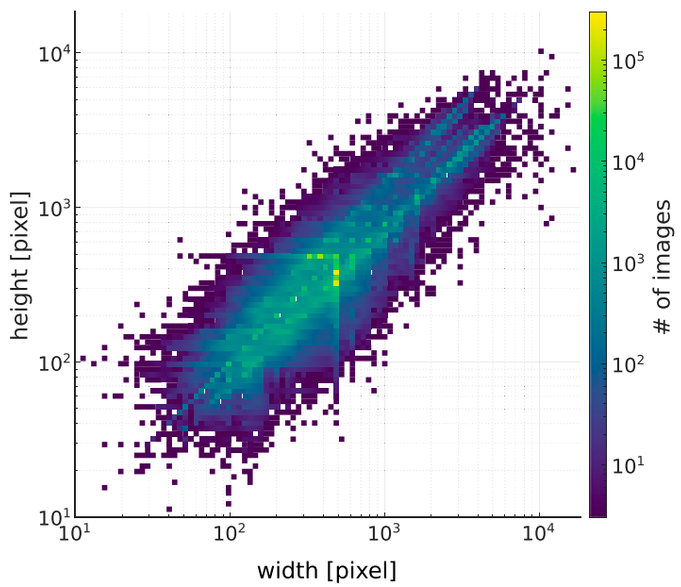
C


60% artificial
40% natural
81% inanimate
19% animate
76% non-edibles
24% edibles
69% non-tools
31% tools
98% non-scenes
2% scenes

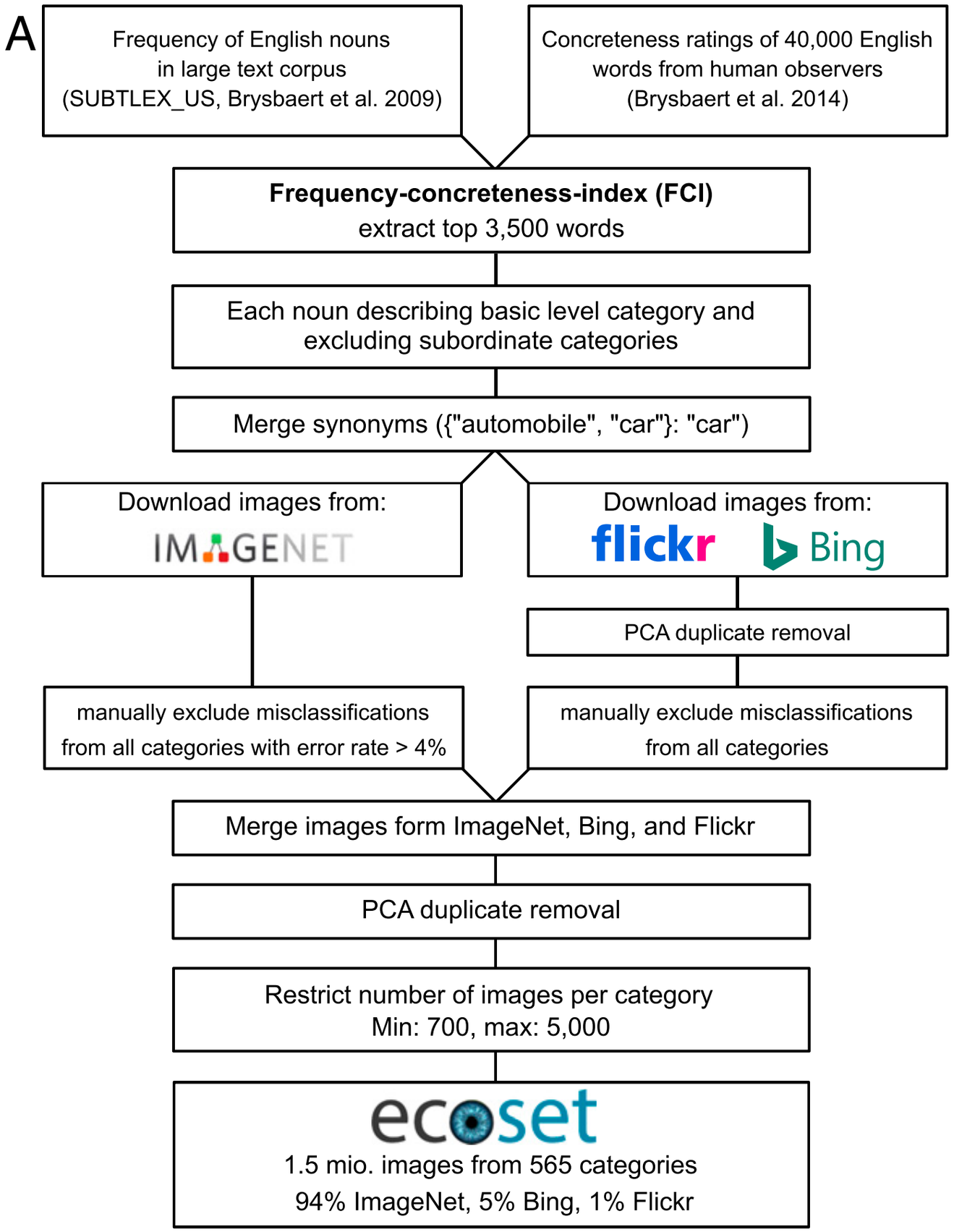
D

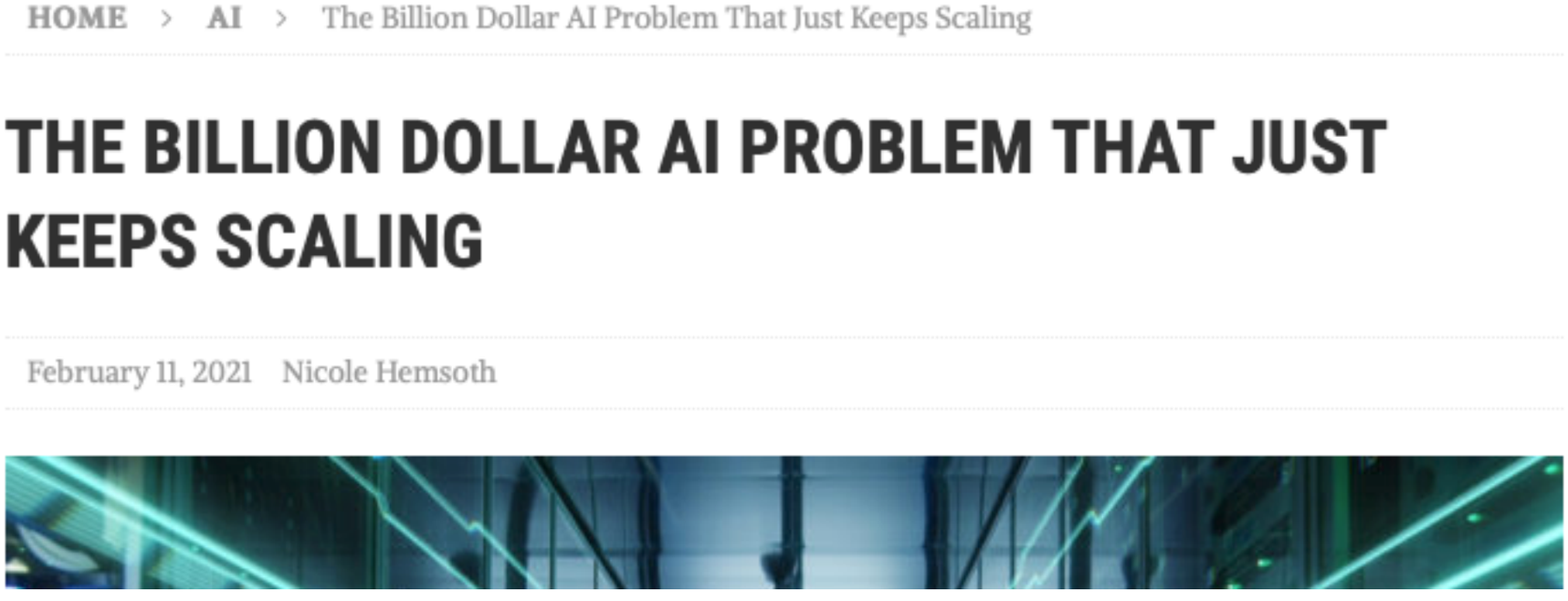

number of images
category (sorted by number of images)
0
1000
2000
3000
4000
5000
0
100
200
300
400
500
600

E


height [pixel]
width [pixel]
# of images

A


Frequency of English nouns
in large text corpus
(SUBTLEX_US, Brysbaert et al. 2009)
Concreteness ratings of 40,000 English
words from human observers
(Brysbaert et al. 2014)
Frequency-concreteness-index (FCI)
extract top 3,500 words
Each noun describing basic level category and
excluding subordinate categories
Merge synonyms ({"automobile", "car"}: "car")
Download images from:
IMAGENET
Download images from:
flickr
Bing
PCA duplicate removal
manually exclude misclassifications
from all categories with error rate > 4%
manually exclude misclassifications
from all categories
Merge images form ImageNet, Bing, and Flickr
PCA duplicate removal
Restrict number of images per category
Min: 700, max: 5,000
ecoset
1.5 mio. images from 565 categories
94% ImageNet, 5% Bing, 1% Flickr



https://www.nextplatform.com/2021/02/11/the-billion-dollar-ai-problem-that-just-keeps-scaling/

Two main methods to parallelize across GPUs:

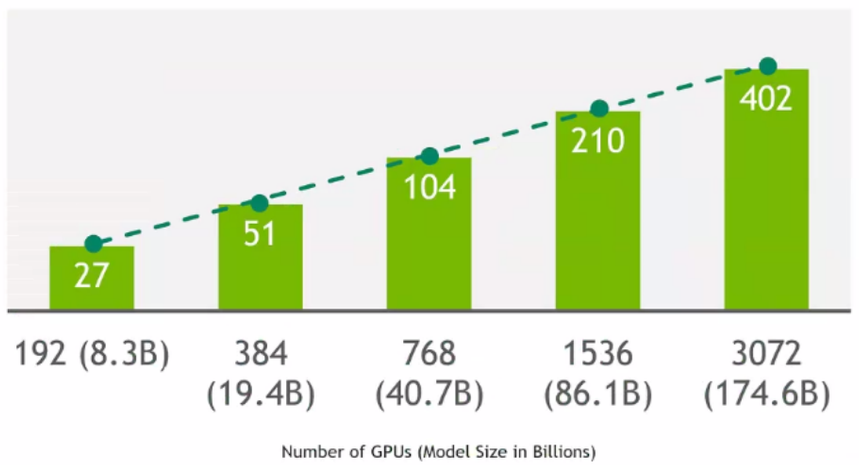
- Data parallelism
- Tensor parallelism

MEGATRON SCALING ON NVIDIA'S DGX-A100 CLUSTER

SELENE

- Batch size: 3072
- 2048 tokens sequences
- 48-way data parallel
- Vocabulary size: 51200

Case	Hidden Size	Number of Layers	Model Parallel Size	Number of GPUs
174.6B (GPT-3)	12288	96	64	3072
86.1B	10240	68	32	1536
40.7B	8192	50	16	768
19.4B	6144	42	8	384
8.3B	4096	40	4	192



280 DGX-A100 systems, which cost $199,000 each
+15% networking cost of the total cost
+20% storage

List price: 75 million (electricity not included)

https://www.newsbreak.com/news/2164940176428/study-shows-that-federated-learning-can-lead-to-reduced-carbon-emissions

"models are routinely trained for thousands of hours on specialized hardware accelerators in datacenters estimated to use 200 terawatt-hours per year. (The average U.S. home consumes about 10,000 kilowatt-hours per year, a fraction of that total.)"

Federated learning has an environmental advantage partly due to the cooling needs of datacenters,

1 terawatt = 1,000,000,000 kilowatts

However, federated learning can also be less efficient:

- prolonged training times due to distributed data bases
- lower hardware efficiency
- data transfer via wifi

https://science.sciencemag.org/content/367/6481/984

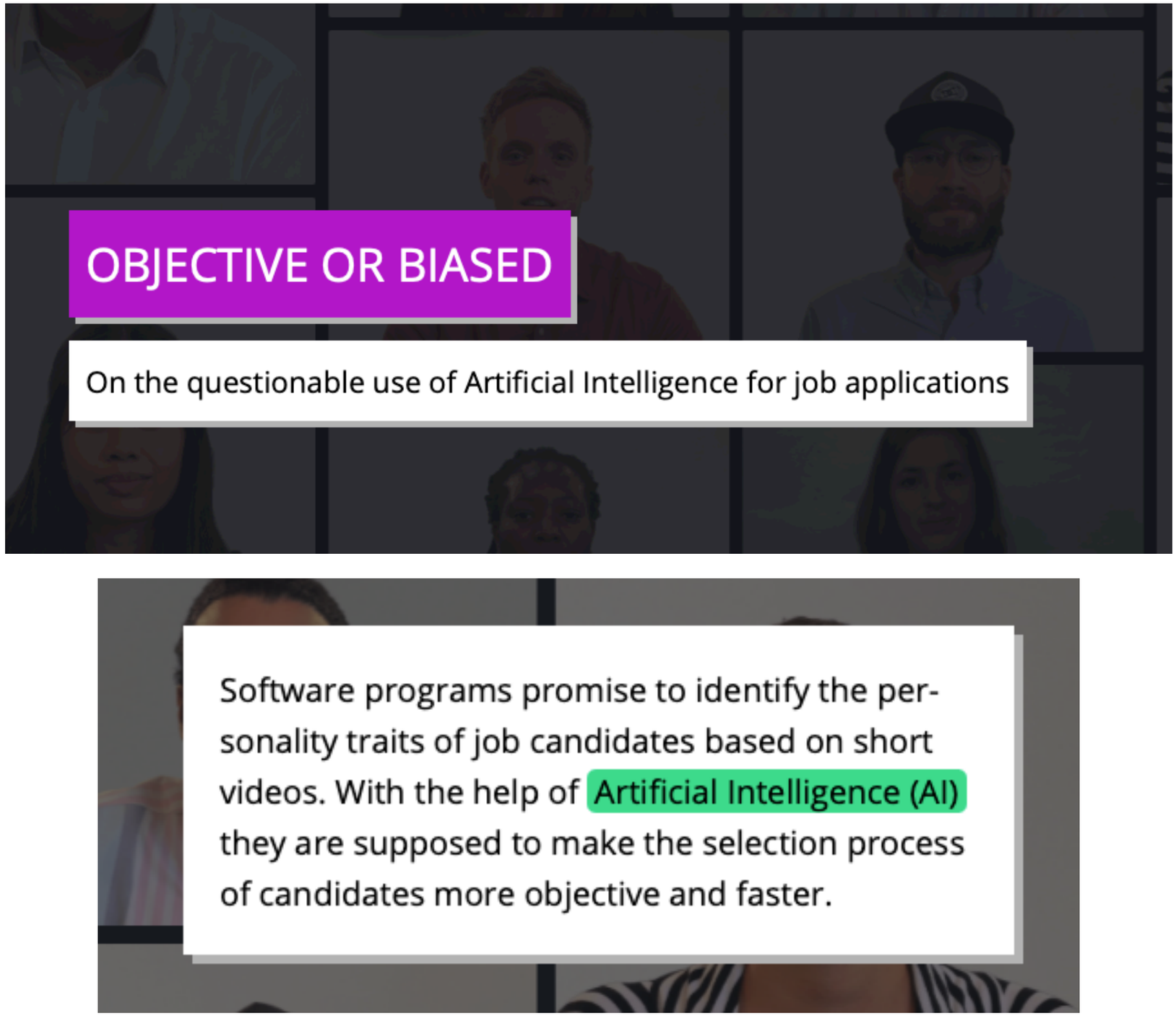
Chip startup NeuReality launches from stealth to make AI inference more efficient

BY MARIA DEUTSCHER

https://siliconangle.com/2021/02/10/chip-startup-neureality-launches-stealth-make-ai-inference-efficient/

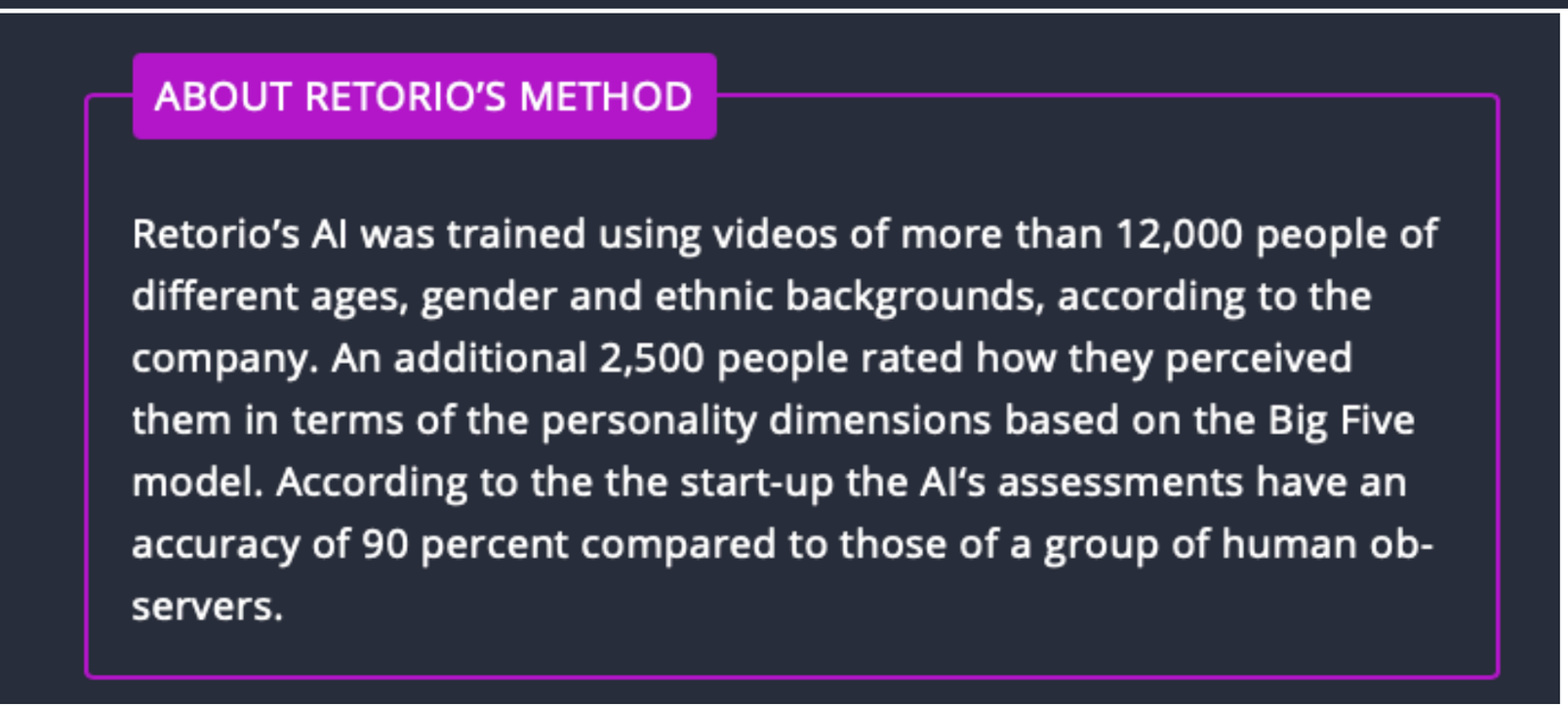
NeuReality Ltd., a startup working to develop more efficient artificial intelligence chips, today exited stealth mode and disclosed on the occasion that it has raised $8 million in seed funding.

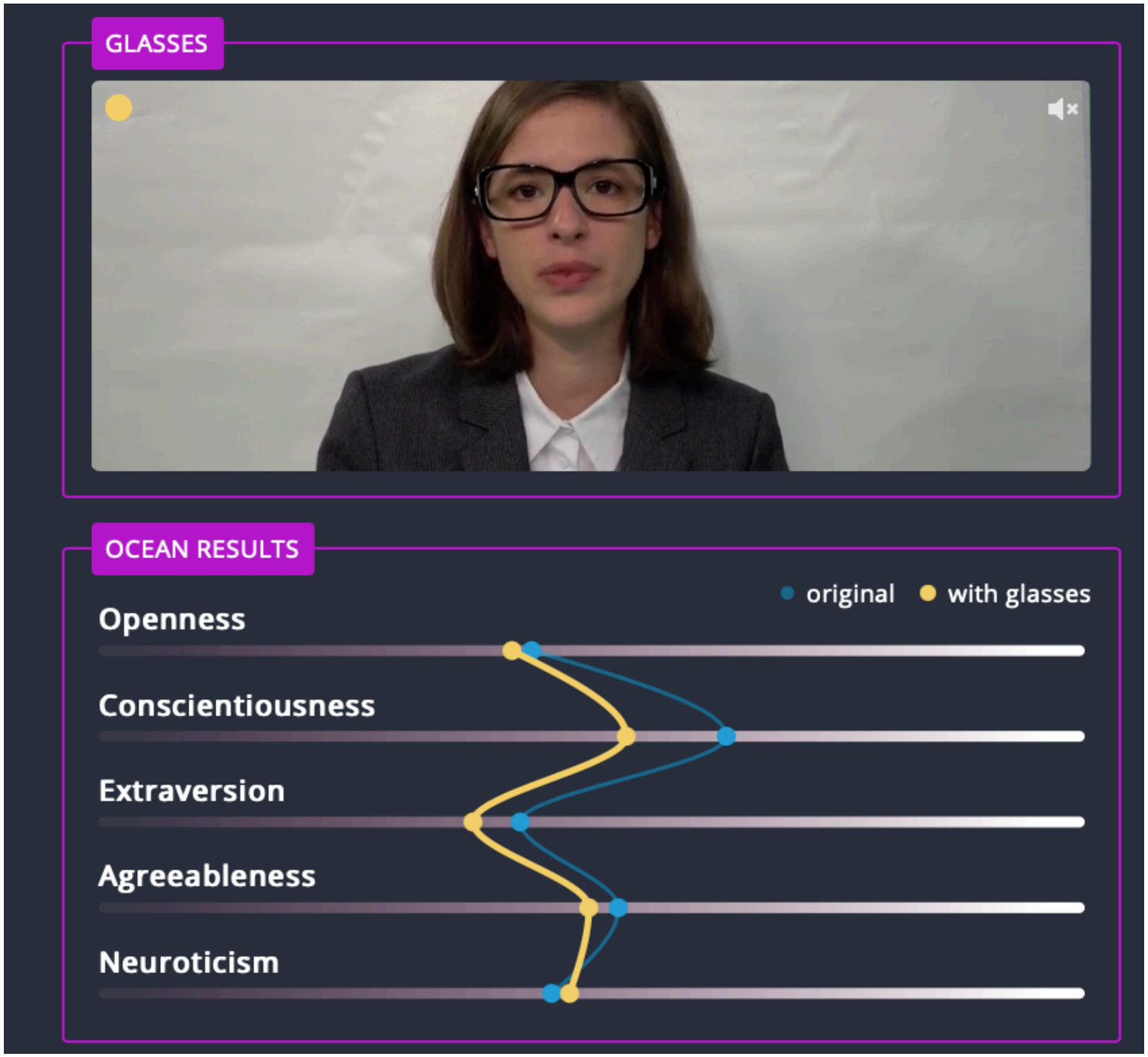
The startup says its chip can perform inference with 15 times higher performance per dollar than the competition. That efficiency will come from lower hardware costs and decreased power consumption, as well as a reduction in data center space requirements, NeuReality claims.



https://web.br.de/interaktiv/ki-bewerbung/en/

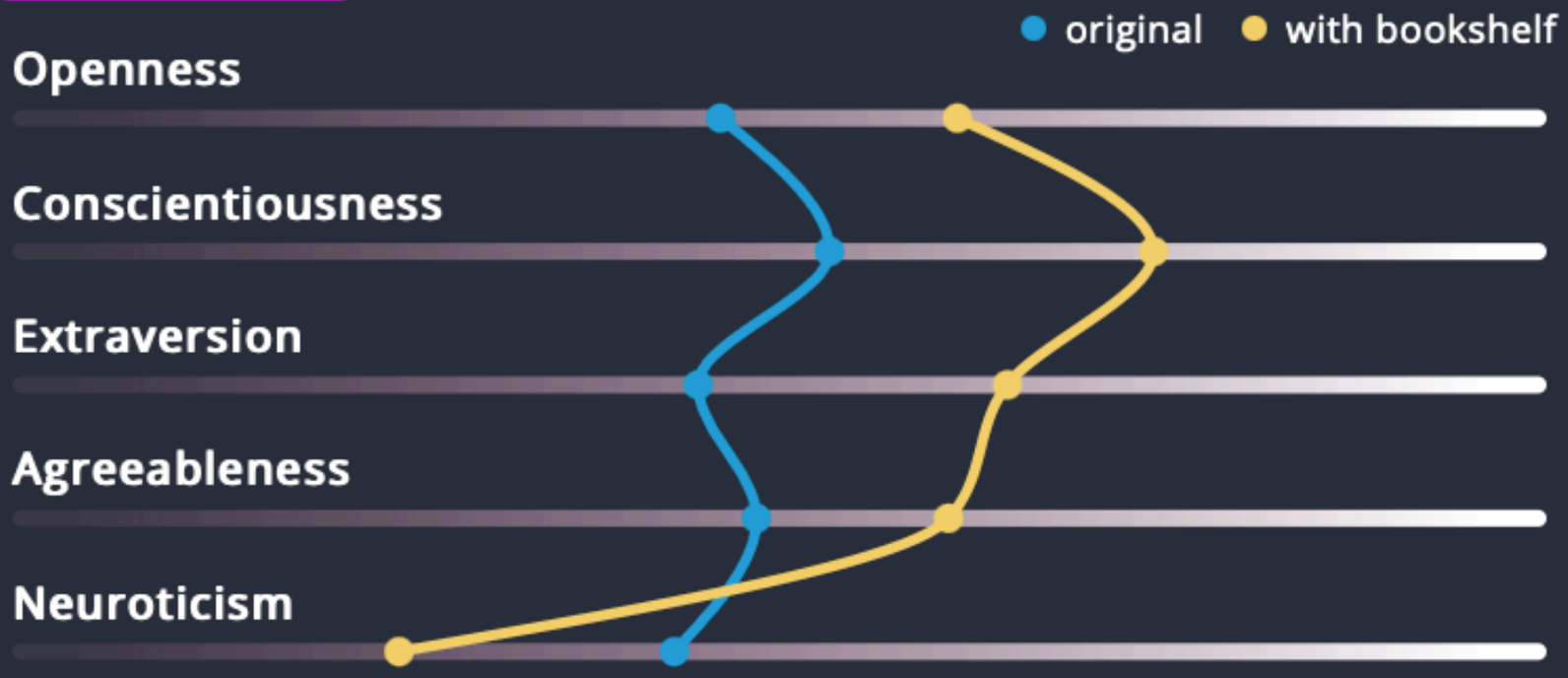
The software refers to the so-called OCEAN model for personality traits. According to this model, personality can be assessed in five dimensions: Openness, conscientiousness, extraversion, agreeableness, and neuroticism.




GLASSES
OCEAN RESULTS
original
with glasses
Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism

BACKGROUND

OCEAN RESULTS



A bookshelf alters the results even more than the picture frame. The result calculated by the AI differs significantly from that of the original version.

For computer science professor Katharina Zweig, who heads the Algorithm Accountability Lab at Technical University Kaiserslautern, these results indicate a difficulty well known: “The fundamental problem with face recognition by machine learning is that we never know exactly which pattern in an image these machines are responding to.”

https://arxiv.org/abs/2102.06604

When engineers train deep learning models, they are very much "flying blind".

- collection of instruments to look into the inner workings closer look into the inner workings of a learning machine
- useful for trouble-shooting during training
- open-source, for PyTorch: https://github.com/f-dangel/cockpit

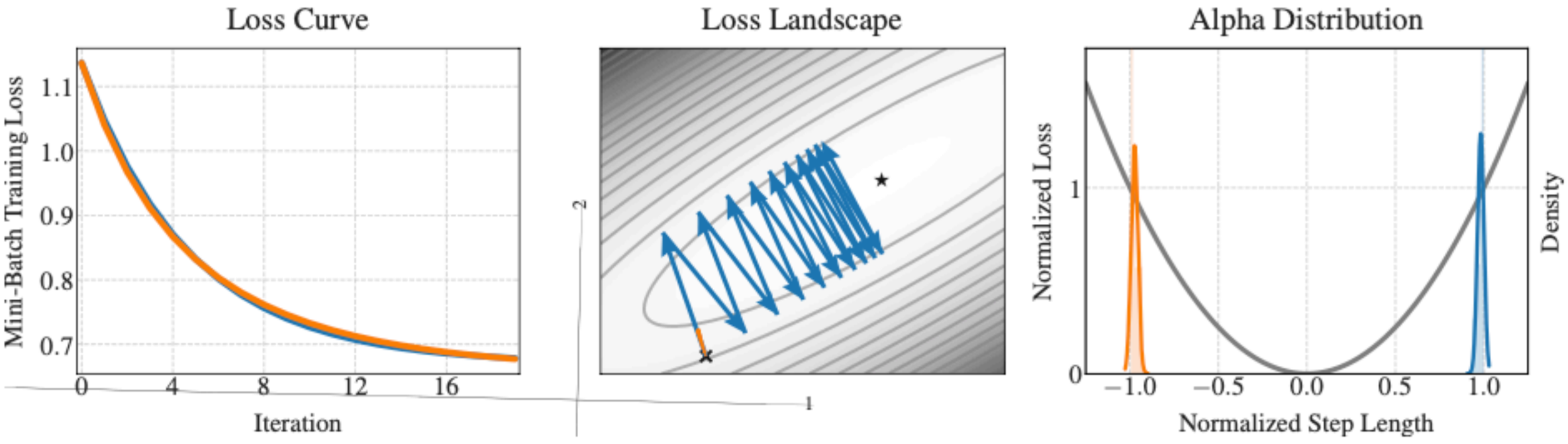


Figure 1. **Illustrative example: Learning curves do not tell the whole story**. Two different optimization runs (—/—) can lead to virtually the same loss curve (*left*). However, the actual optimization trajectories (*middle*), exhibit vastly different behaviors. In practice, they are intractably large and cannot be visualized directly. Recommendable actions for both scenarios (increase/decrease the learning rate) cannot be inferred from the loss curve. The α- distribution, one COCKPIT instrument (*right*), not only clearly distinguishes the two scenarios, but also allows for taking decisions regarding how the learning rate should be adapted. See Section 3.3 for further details.

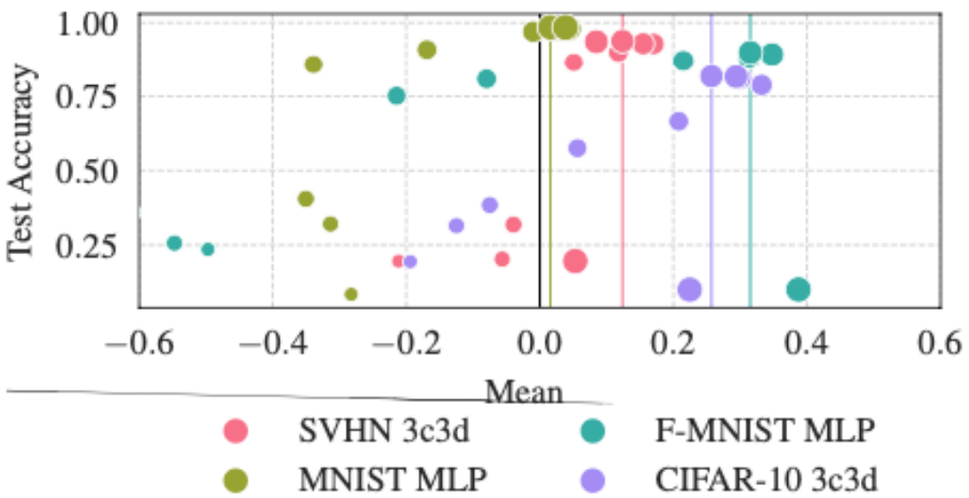


Figure 5. **Final test accuracy as a function of standardized step-size** α. For multiple test problems, final test accuracy is shown against average α-value over the whole training period. Marker size indicates the magnitude of the raw learning rate used, marker color identifies tasks (legend below). For each problem, the best-performing setting is highlighted by a vertical colored line.