Course evaluations

https://edstem.org/us/courses/70815/discussion/6340135

Please do them

Lecture 20

CV/AI Frontiers

Administrative

A5 is out

- Due May 14th

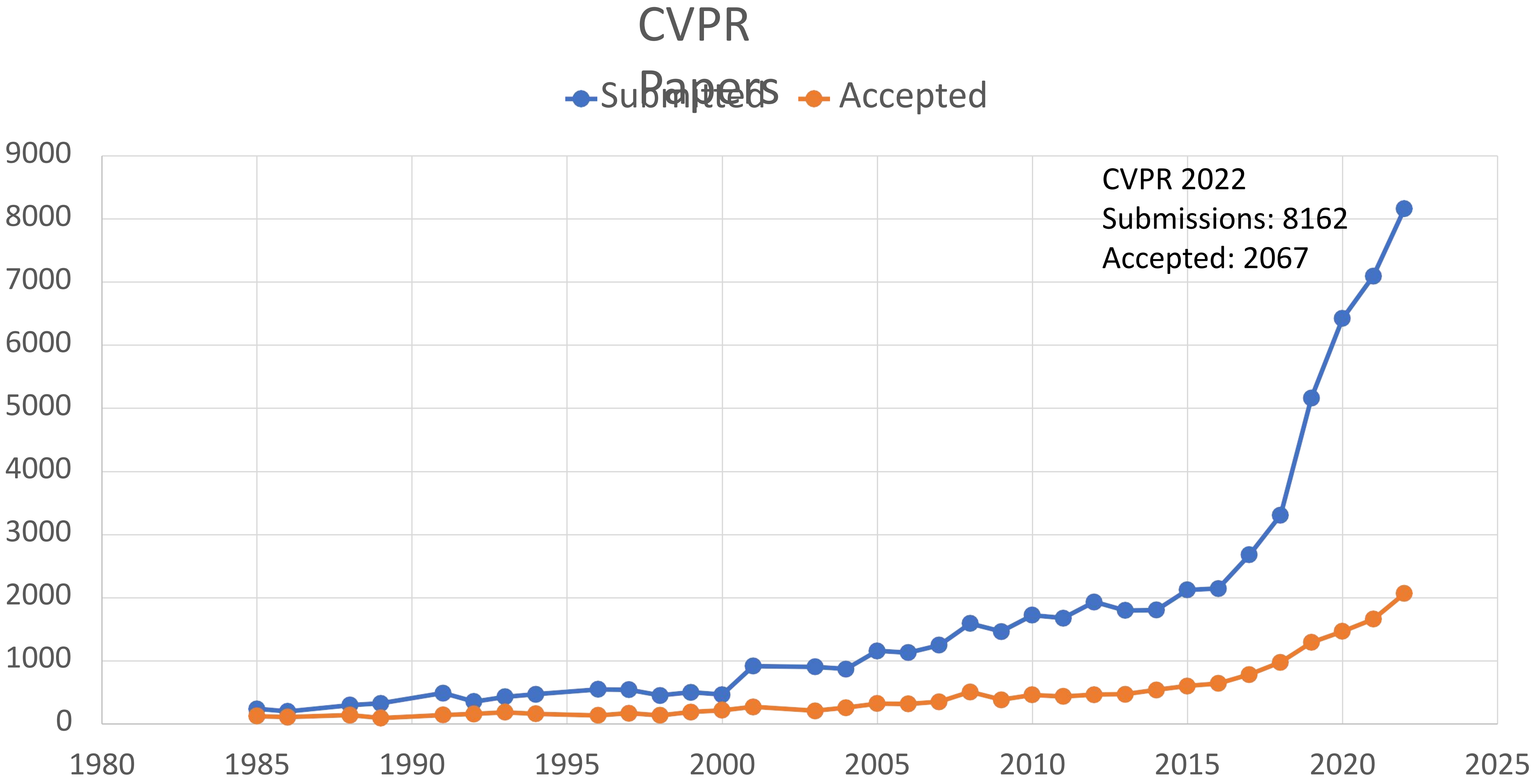
Exam

- Mon, Mar 17 10:30-12:20
- Same room as lecture: G20
- List of topics posted on EdStem

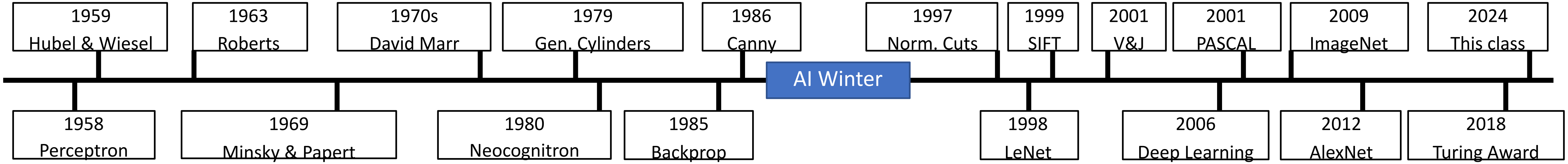
Makeup exam

- Friday 14th (emails have been sent out with location and time)

We are in the era of deep learning


CVPR Papers
Submitted
Accepted
CVPR 2022
Submissions: 8162
Accepted: 2067
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025

Deep Learning wasn’t invented overnight!



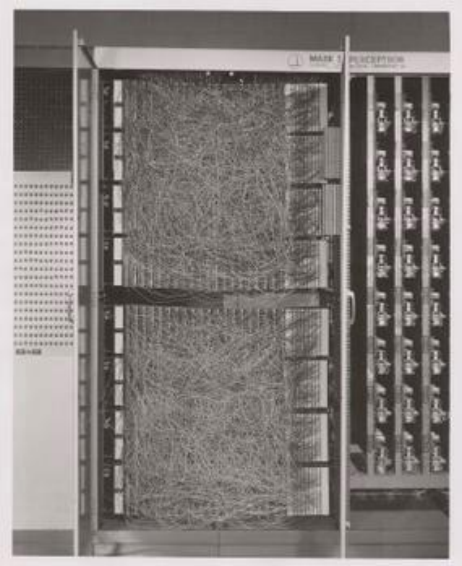
Deep Learning wasn't invented overnight!

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1997 Norm. Cuts
1999 SIFT
2001 V&J
2001 PASCAL
2009 ImageNet
2024 This class

AI Winter

1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
1985 Backprop
1998 LeNet
2006 Deep Learning
2012 AlexNet
2018 Turing Award

Perceptron



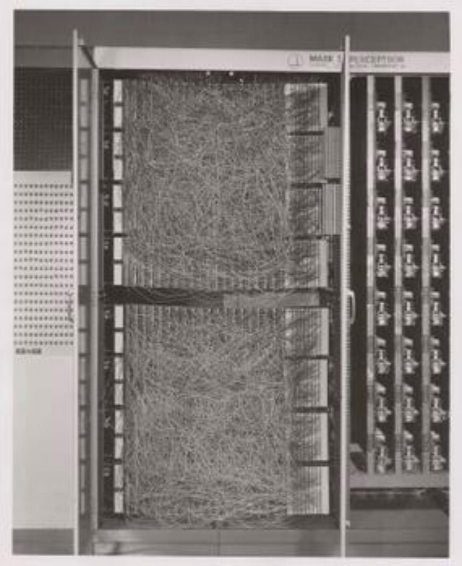
Frank Rosenblatt, ~1957

Deep Learning wasn't invented overnight!

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2001 PASCAL | 2009 ImageNet | 2024 This class

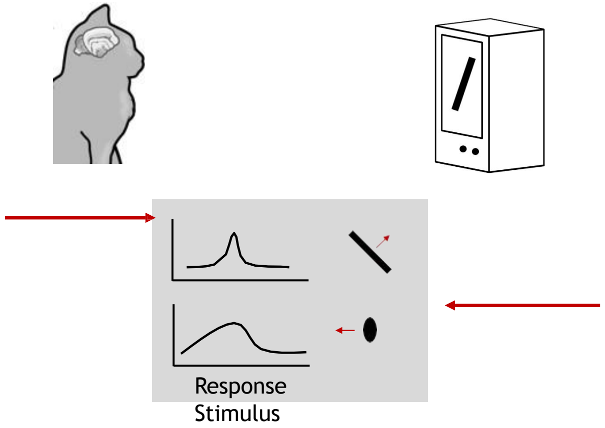
1958 Perceptron | 1969 Minsky & Papert | 1980 Neocognitron | 1985 Backprop | AI Winter | 1998 LeNet | 2006 Deep Learning | 2012 AlexNet | 2018 Turing Award

Perceptron



Frank Rosenblatt, ~1957

Simple and Complex cells



Hubel and Wiesel, 1959

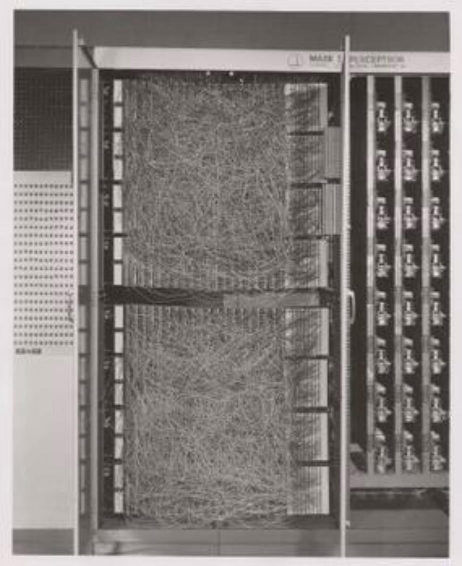
Deep Learning wasn't invented overnight!

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2001 PASCAL | 2009 ImageNet | 2024 This class

AI Winter

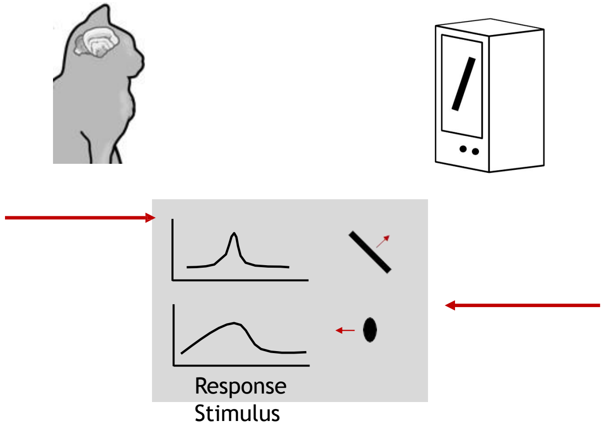
1958 Perceptron | 1969 Minsky & Papert | 1980 Neocognitron | 1985 Backprop | 1998 LeNet | 2006 Deep Learning | 2012 AlexNet | 2018 Turing Award

Perceptron



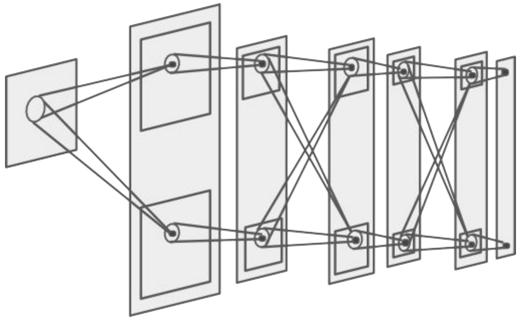
Frank Rosenblatt, ~1957

Simple and Complex cells



Hubel and Wiesel, 1959

Neocognitron



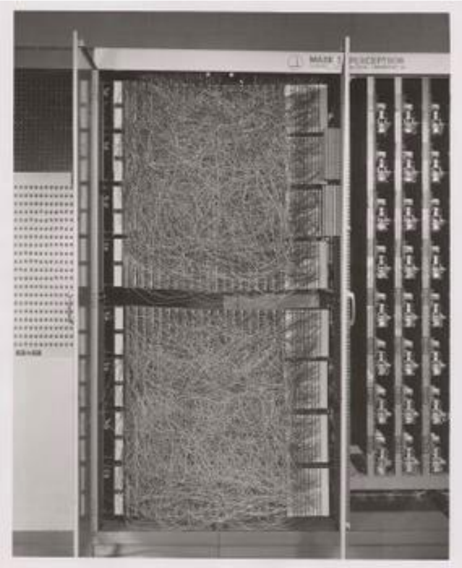
Fukushima, 1980

Deep Learning wasn't invented overnight!

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT
2001 V&J
2001 PASCAL
2009 ImageNet
2024 This class

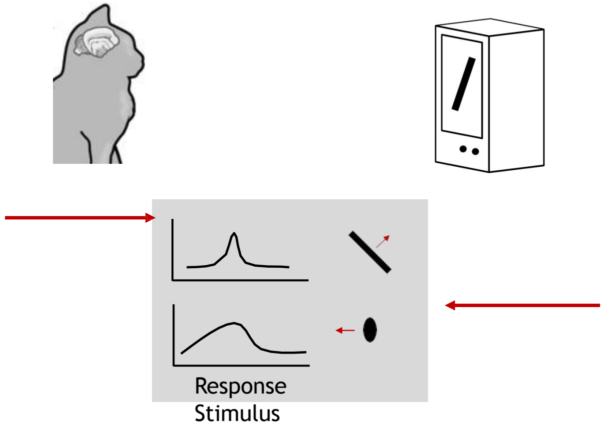
1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
1985 Backprop
1998 LeNet
2006 Deep Learning
2012 AlexNet
2018 Turing Award

Perceptron



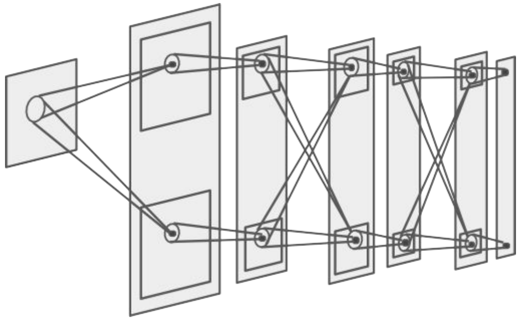
Frank Rosenblatt, ~1957

Simple and Complex cells



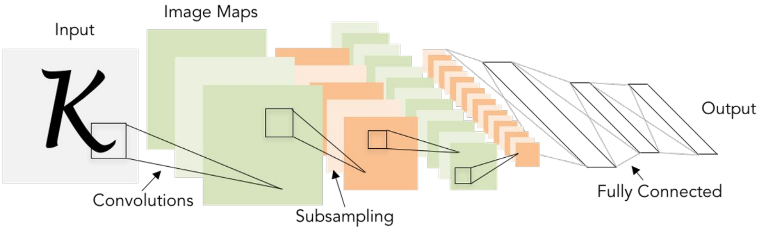
Hubel and Wiesel, 1959

Neocognitron



Fukushima, 1980

Convolutional Networks



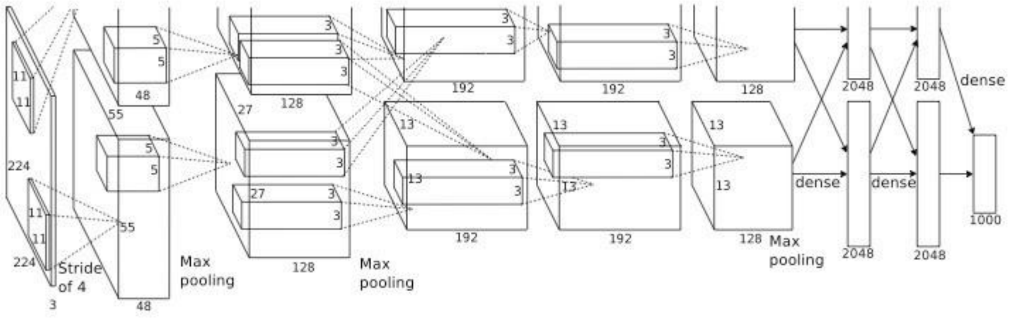
LeCun et al, 1998

Deep Learning wasn't invented overnight!

- 1959 Hubel & Wiesel
- 1963 Roberts
- 1970s David Marr
- 1979 Gen. Cylinders
- 1986 Canny
- AI Winter
- 1997 Norm. Cuts
- 1999 SIFT
- 2001 V&J
- 2001 PASCAL
- 2009 ImageNet
- 2024 This class

- 1958 Perceptron
- 1969 Minsky & Papert
- 1980 Neocognitron
- 1985 Backprop
- 1998 LeNet
- 2006 Deep Learning
- 2012 AlexNet
- 2018 Turing Award

AlexNet



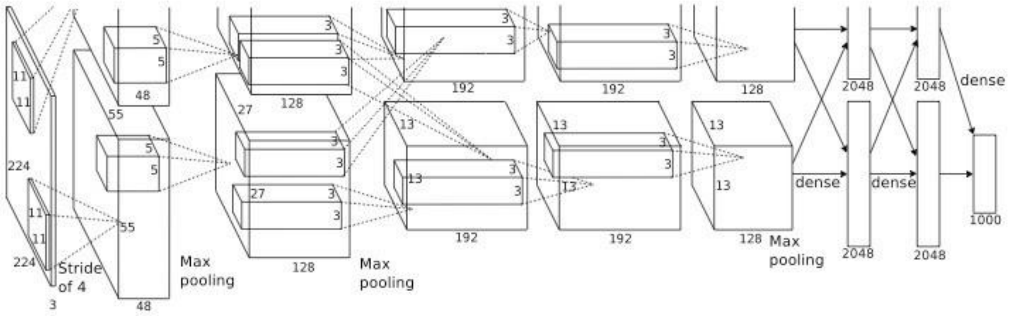
Krizhevsky, Sutskever, and Hinton, 2012

Deep Learning wasn't invented overnight!

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2001 PASCAL | 2009 ImageNet | 2024 This class

1958 Perceptron | 1969 Minsky & Papert | 1980 Neocognitron | 1985 Backprop | AI Winter | 1998 LeNet | 2006 Deep Learning | 2012 AlexNet | 2018 Turing Award

AlexNet



Krizhevsky, Sutskever, and Hinton, 2012

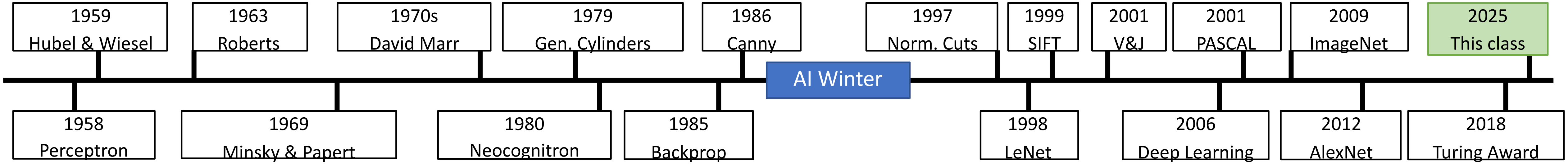
2018 Turing Award

Yoshua Bengio

Geoffrey Hinton

Yann LeCun

Deep Learning wasn't invented overnight!



Winter 2025: This class

What’s Next?

Bigger Models, More Data, More Compute

The Bitter Lesson

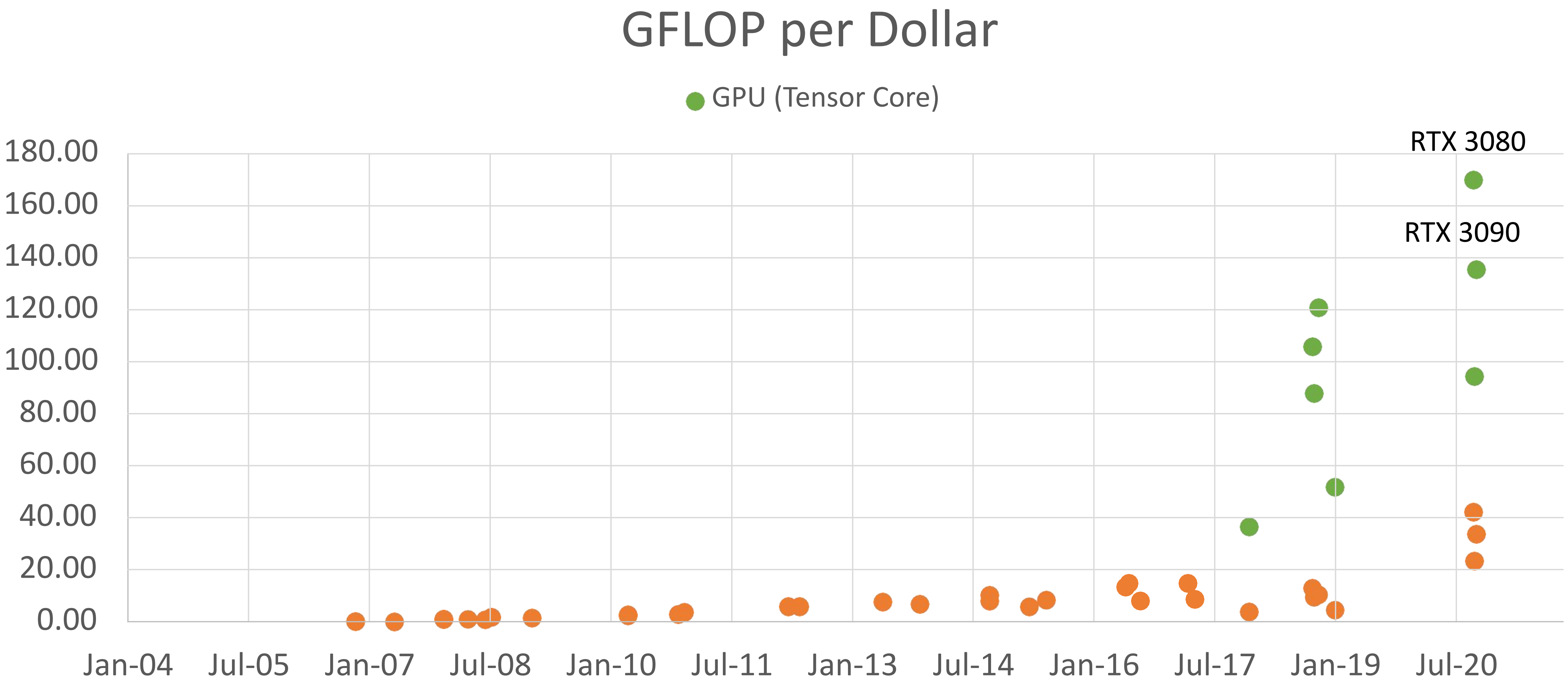
http://www.incompleteideas.net/IncIdeas/BitterLesson.html

Rich Sutton

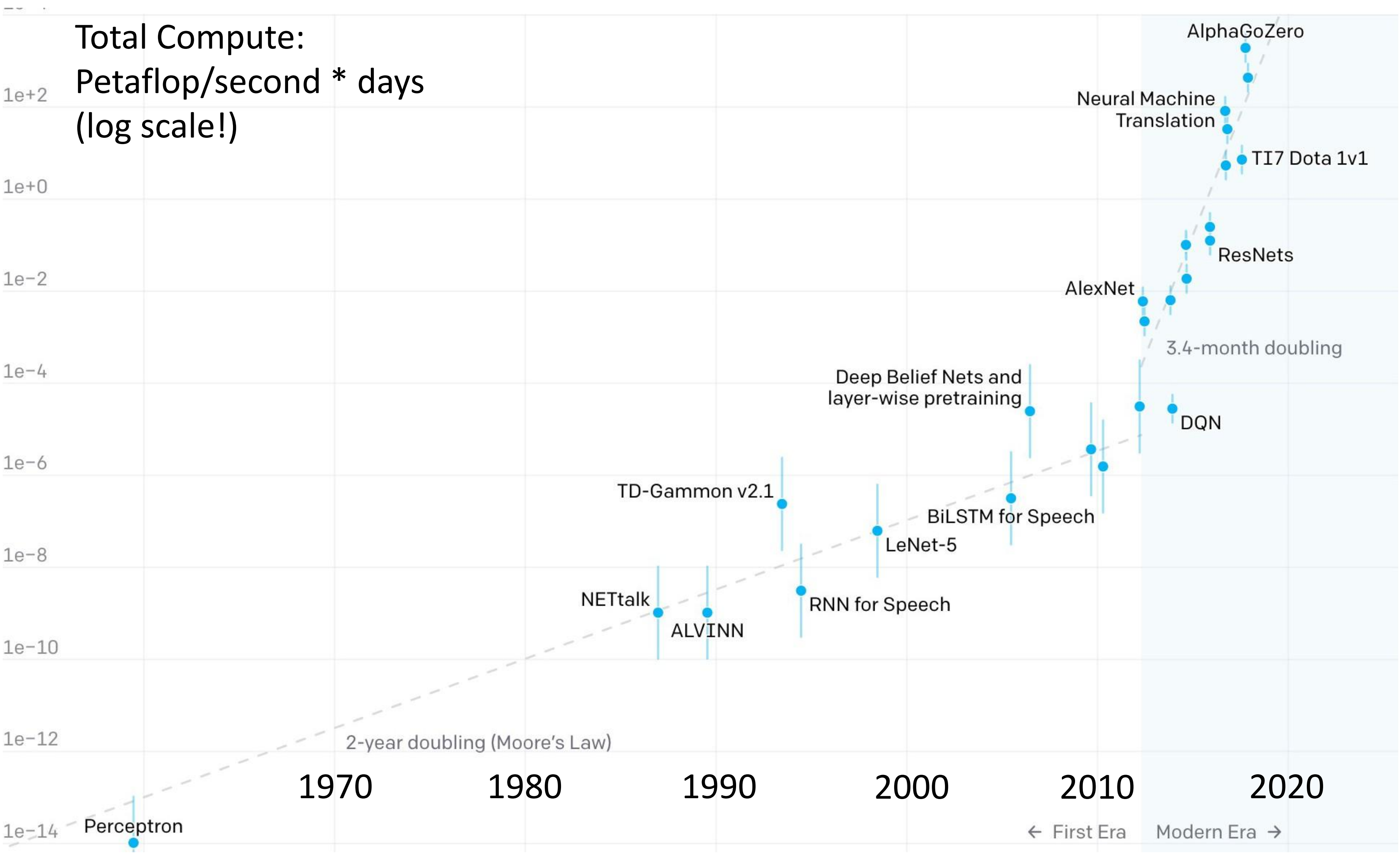
March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

Bigger Models, More Data, More Compute



Bigger Models, More Data, More Compute



Amodei et al, "AI and Compute", 2018

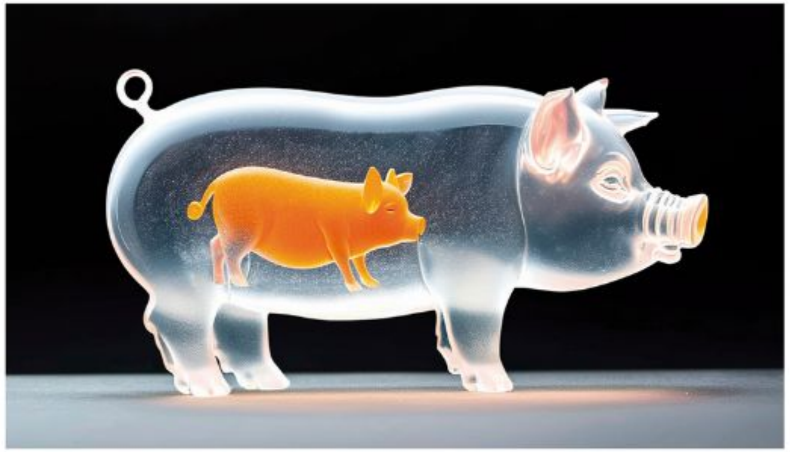
Bigger Models, More Data, More Compute

Examples that came out this year

- Stable Diffusion 3 (02/2024): text-to-image generation models from 800M to 8B.
- Sora (02/2024): text-to-image/video generation models
- LLaMA 3 (04/2024): large language model from 8B to 70B pretrained on over 15T tokens with two custom-built 24K GPU clusters
- Chameleon (05/2024): multimodal generative model, 34B parameters

Bigger Models, More Data, More Compute: Stable Diffusion 3

- Diffusion Transformer
- Rectified Flows



Prompt: Translucent pig, inside is a smaller pig.

Prompt: A massive alien space ship that is shaped like a pretzel.

Prompt: A kangaroo holding a beer, wearing ski goggles and passionately singing silly songs.

Prompt: An entire universe inside a bottle sitting on the shelf at walmart on sale.

Prompt: A cheeseburger with juicy beef patties and melted cheese sits on top of a toilet that looks like a throne and stands in the middle of the royal chamber.

Prompt: This dreamlike digital art captures a vibrant, kaleidoscopic bird in a lush rainforest

Prompt: A car made out of vegetables.

Prompt: Heat death of the universe line art

https://stability.ai/news/stable-diffusion-3-research-paper

Bigger Models, More Data, More Compute: Sora

- Diffusion Transformer

Close-up portrait shot of a woman in autumn, extreme detail, shallow depth of field

Digital art of a young tiger under an apple tree in a matte painting style with gorgeous details

https://openai.c

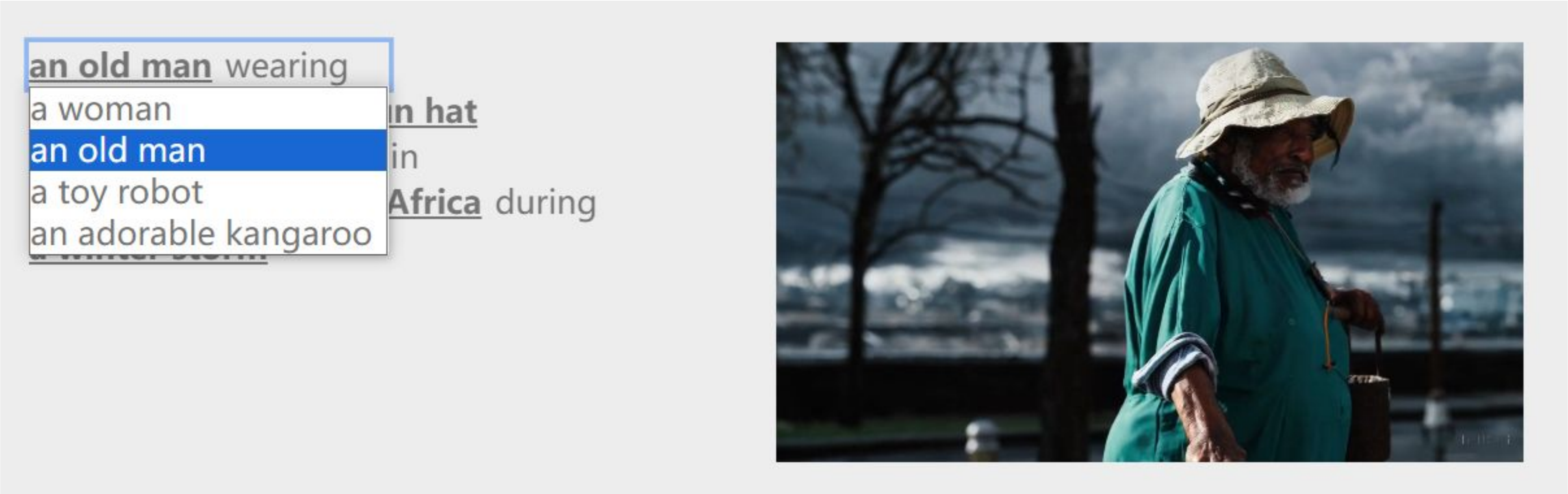
Bigger Models, More Data, More Compute: Sora

a toy robot wearing a green dress and a sun hat taking a pleasant stroll in Johannesburg, South Africa during a winter storm

https://openai.com/index/video-generation-models-as-world-simulators/

Bigger Models, More Data, More Compute: Sora

a cool tiny demo in OpenAI website



https://openai.com/index/video-generation-models-as-world-simulators/

Bigger Models, More Data, More Compute: LLaMA 3

You

What is the number that rhymes with the word we use to describe a tall plant?

Meta AI

I think I can help with that!

The word you're referring to is probably "tree", and the number that rhymes with it is "three"!

Bigger Models, More Data, More Compute: LLaMA 3

You

how to prepare Computer Vision final?

Meta AI

Preparing for a Computer Vision final exam requires a combination of understanding concepts, practicing problems, and reviewing materials. Here's a suggested plan to help you prepare:

1. Review Fundamentals
   - Linear Algebra: matrices, vectors, tensor operations
   - Calculus: derivatives, optimization
   - Probability and Statistics: probability distributions, Bayes' theorem
   - Programming skills: Python, NumPy, pandas, etc.
2. Computer Vision Concepts
   - Image processing: filtering, transforms, feature extraction
   - Object recognition: classification, detection, segmentation
   - Deep learning: CNNs, architectures (e.g., ResNet, YOLO)
   - Image generation: synthesis, GANs
3. Study Materials
   - Textbook: "Computer Vision: Algorithms and Applications" by Richard Szeliski
   - Online courses: Stanford CS231n, MIT OpenCourseWare
   - Research papers: read recent papers on arXiv, CVPR, ICCV
4. Practice Problems
   - Implement algorithms: image filtering, object detection
   - Work on projects: image classification, segmentation, generation
   - Use libraries: OpenCV, PyTorch, TensorFlow
5. Review and Practice Regularly
   - Set aside dedicated time to review and practice each week
   - Make flashcards for key terms and concepts
   - Join online communities (e.g., Kaggle, Reddit) for discussion and learning
6. Take Practice Exams
   - Find practice exams or quizzes online
   - Time yourself to simulate the actual exam experience
   - Review your mistakes and focus on weak areas

By following this plan, you'll be well-prepared for your Computer Vision final exam. Good luck!

Bigger Models, More Data, More Compute: Chameleon

Prompt: I'm bored. Could you show me some cool, quirky-looking birds? Also tell me about them in 2–3 sentences.

Response: Absolutely! Here are three birds with some unique features that you might find interesting:

Keel-Billed Toucan The Keel-Billed Toucan is a brightly colored bird native to South American rainforests and woodlands. Its most distinctive feature is its colorful beak, which can grow up to seven inches long and has a distinctive, curved shape. The beak is used for plucking fruit from trees and for social displays. Despite its large size, the Keel-Billed Toucan is a skilled flyer and can often be seen flitting through the treetops. Here is an image of this vibrant bird: <img>

Puffin Puffins are small, aquatic birds that are found in the northern hemisphere. They have distinctive black and white feathers, as well as a brightly colored beak that is used for attracting mates and defending their territory. Puffins are known for their distinctive "puffin" call, which sounds like a loud, guttural chuckle. They feed on fish, which they catch by diving into the water from the air. Puffins are colonial birds, and they breed in large colonies on rocky cliffs and islands. Here is an image of this adorable bird: <img>

Golden Pheasant The Golden Pheasant is a brightly colored bird native to the forests of China. It is a member of the pheasant family and is known for its striking plumage, which features a golden-yellow body, red face, and green tail feathers. The Golden Pheasant is a ground-dwelling bird, and it feeds on a diet of insects, seeds, and fruits. It is a popular game bird, and is often hunted for its meat and feathers. In captivity, the Golden Pheasant is sometimes kept as an ornamental bird, due to its striking appearance. Here is an image of this beautiful bird: <img>

I hope these images and descriptions have helped to brighten your day! If you have any other questions or requests, please don't hesitate to ask.

Bigger Models, More Data, More Compute: Chameleon

Prompt: <img> While chameleons are masters of camouflage in the natural world, let's see how they might fare in a man-made setting.
1. Analyze the colors, patterns, and textures in this image and identify the most challenging spot for a chameleon to blend in.
2. Then generate an image of a chameleon with the optimal pattern to camouflage itself.
3. Then, rate and explain how successful each attempt is on a scale from 1-10.

Response: The image is of a stop sign in a forested area, with some sunlight filtering through the trees. The sign has a red octagonal shape with a white border and the word "STOP" written in white letters. In the background, there are green leaves and trees.

Challenging Spot for a Chameleon: The stop sign would be the most challenging spot for a chameleon to blend in due to its bright red color and bold white letters, which would stand out in the forested environment.

Optimal Pattern for Camouflage: To camouflage itself on the stop sign, a chameleon would need to have a pattern that would break up its outline and make it less conspicuous. A pattern with shades of brown and red would help the chameleon blend in more effectively. Additionally, a mottled or dappled pattern would help to scatter light and further reduce the chameleon's visibility.

<img>
Rating and Explanation: A rating of 2/10 would be given for the chameleon's ability to blend in on the stop sign. While the chameleon's green and brown pattern helps to break up its outline, the bright red color of the stop sign will still make it stand out in the forested environment.

Bigger Models, More Data, More Compute, More problems

Hong Kong (CNN) — A British multinational design and engineering company behind world-famous buildings such as the Sydney Opera House has confirmed that it was the target of a deepfake scam that led to one of its Hong Kong employees paying out $25 million to fraudsters.

https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html

Bigger Models, More Data, More Compute, More problems

- ML Systems can encode bias
- Large models can lack common sense
- Who should control models and data?

Bigger Models, More Data, More Compute, More problems

- ML Systems can encode bias
- Large models can lack common sense
- Who should control models and data?

Stepping Back: Why Build ML Systems?

Automate decision making, so machines can make decision instead of people.

Ideal: Automated decisions can be cheaper, more accurate, more impartial, improve our lives

Reality: If we aren’t careful, automated decisions can encode bias, harm people, make lives worse

Allocative Harms

- Some systems decide how to *allocate resources*
- If the system is biased, it may allocate resources unfairly or perpetuate inequality
- Examples:
  - Insurance rates
  - College admissions
  - Job applications
  - Loan applications
  - Mortgage applications
  - Sentencing criminals

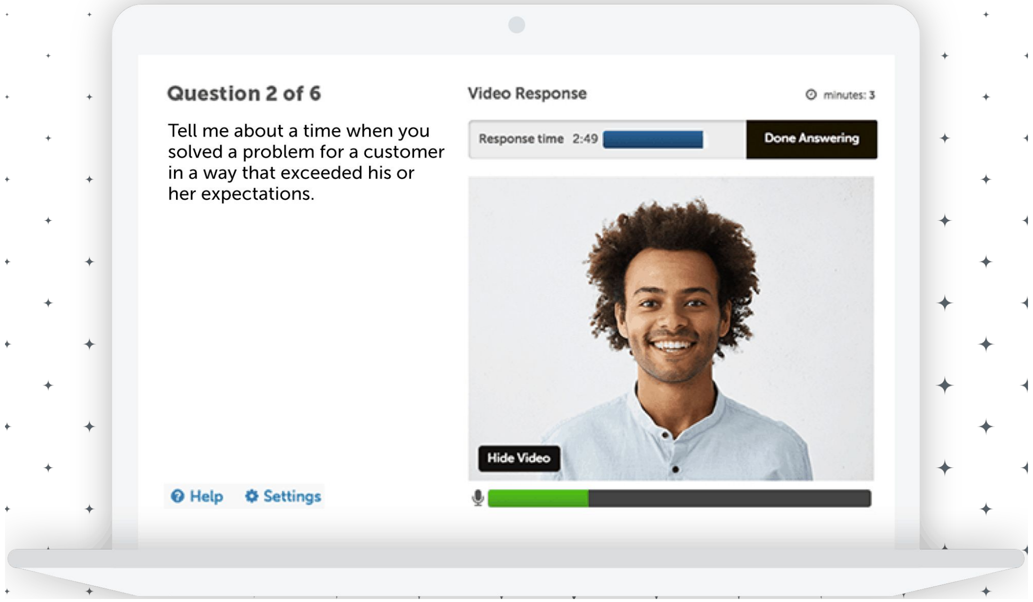
Barocas et al, "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGCIS 2017 Kate Crawford, "The Trouble with Bias", NeurIPS 2017 Keynote

Example: Video Interviewing

Technology

A face-scanning algorithm increasingly decides whether you deserve the job

HireVue claims it uses artificial intelligence to decide who's best for a job. Outside experts call it 'profoundly disturbing.'

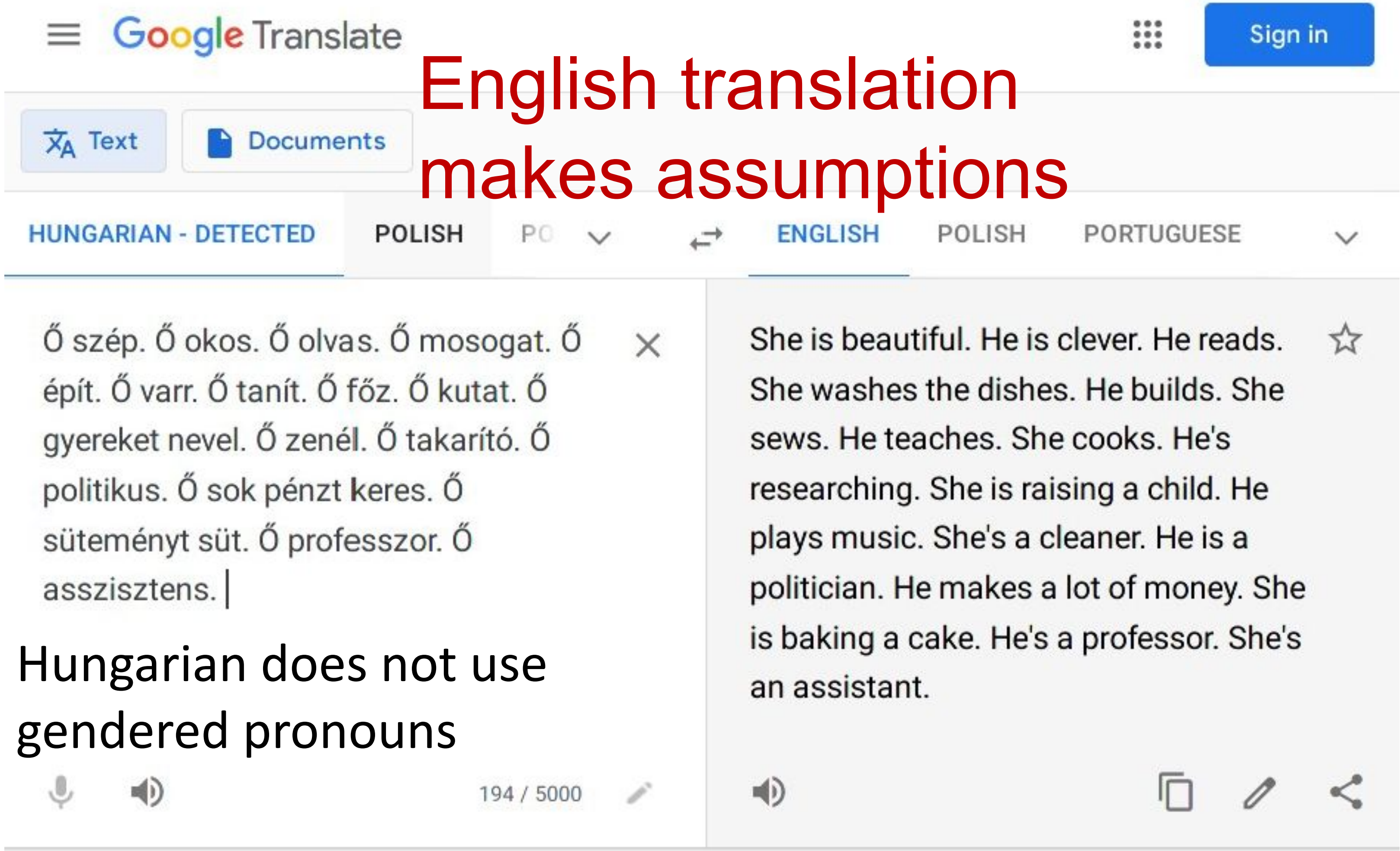


Source: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/
https://www.hirevue.com/platform/online-video-interviewing-software
Example Credit: Timnit Gebru

Representational Harms

A system reinforces harmful stereotypes

Representational Harms: Machine Translation

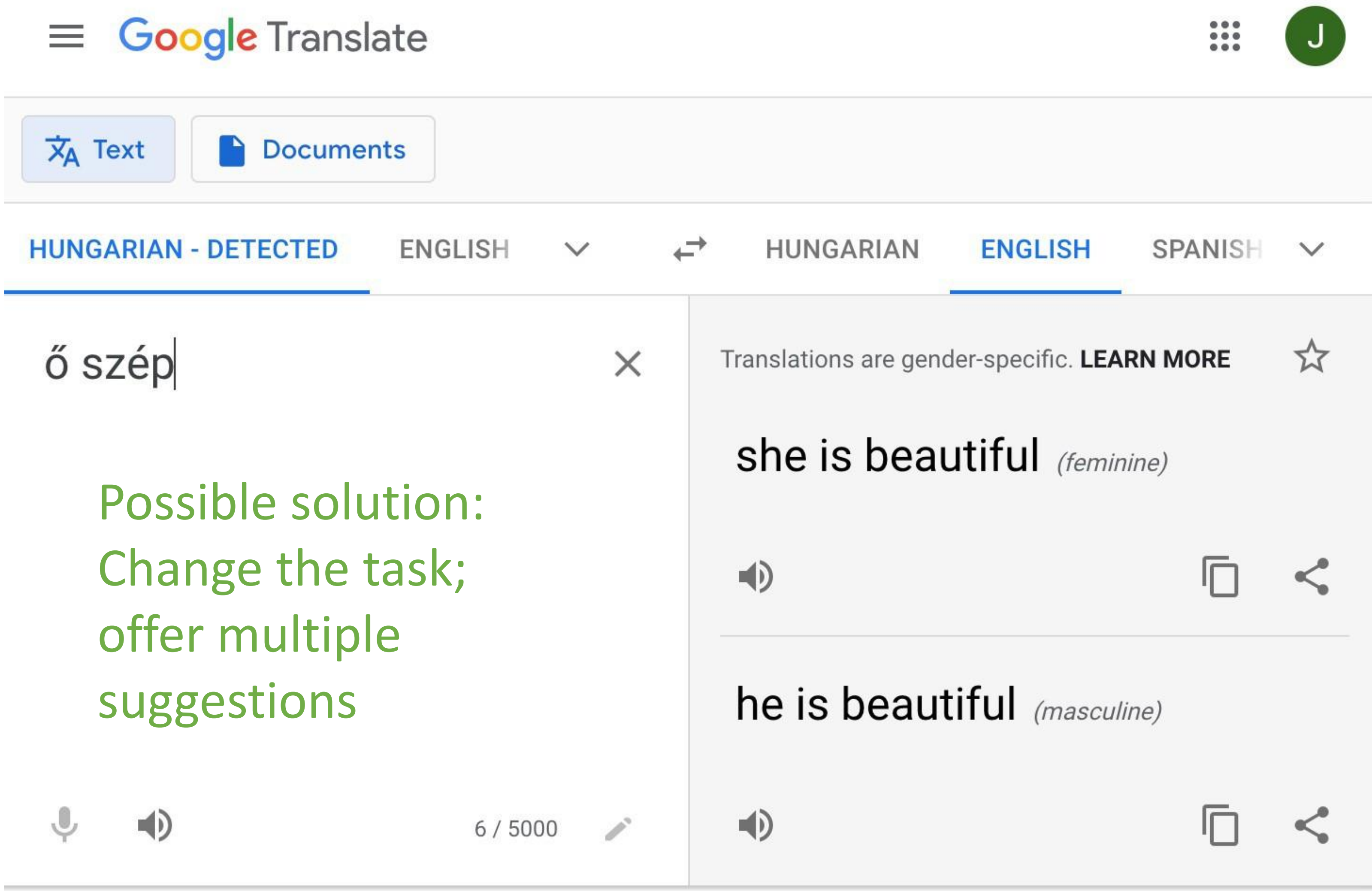


English translation makes assumptions

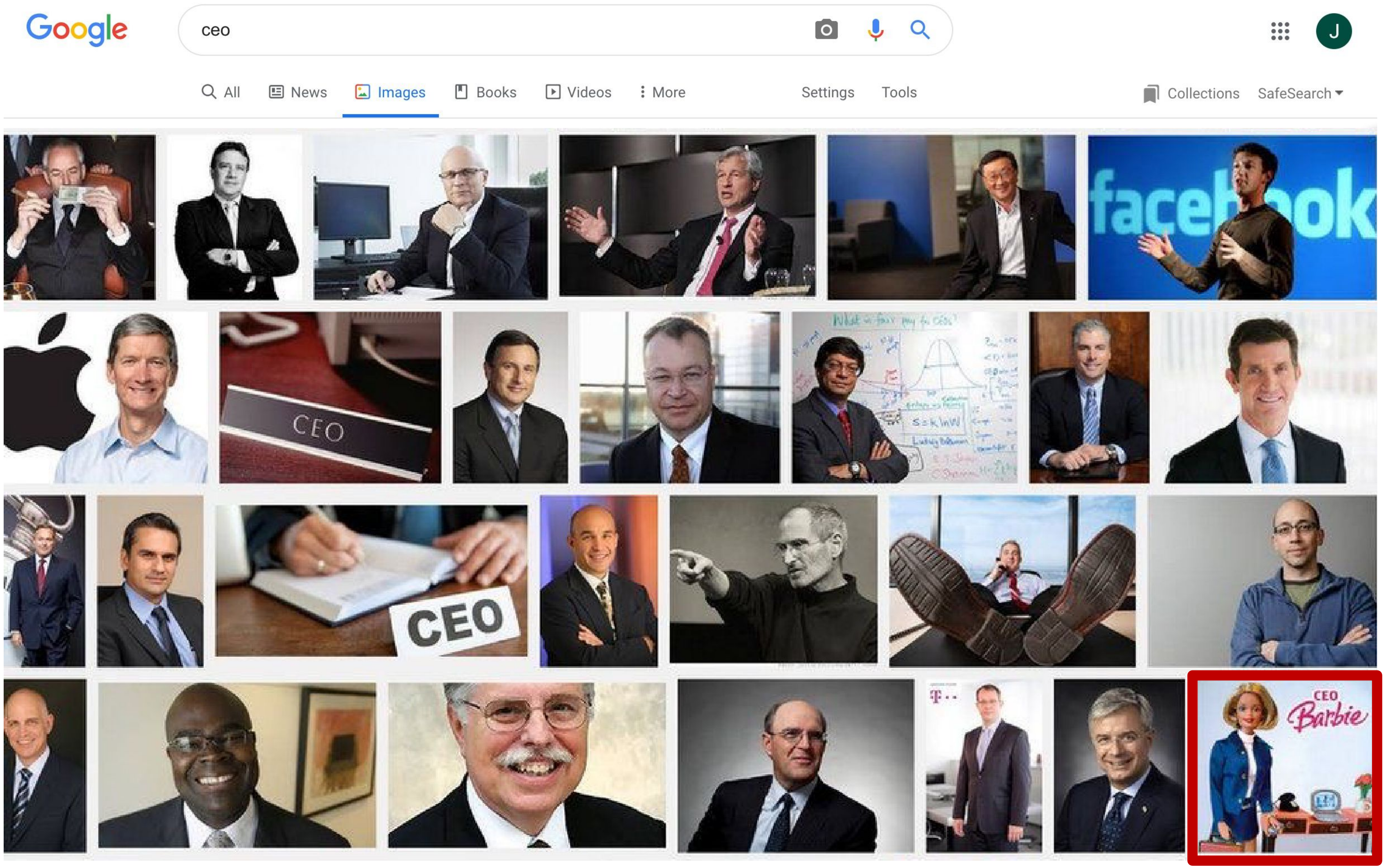
Hungarian does not use gendered pronouns

Source: https://www.reddit.com/r/europe/comments/m9uphb/hungarian_has_no_gendered_pronouns_so_google

Representational Harms: Machine Translation

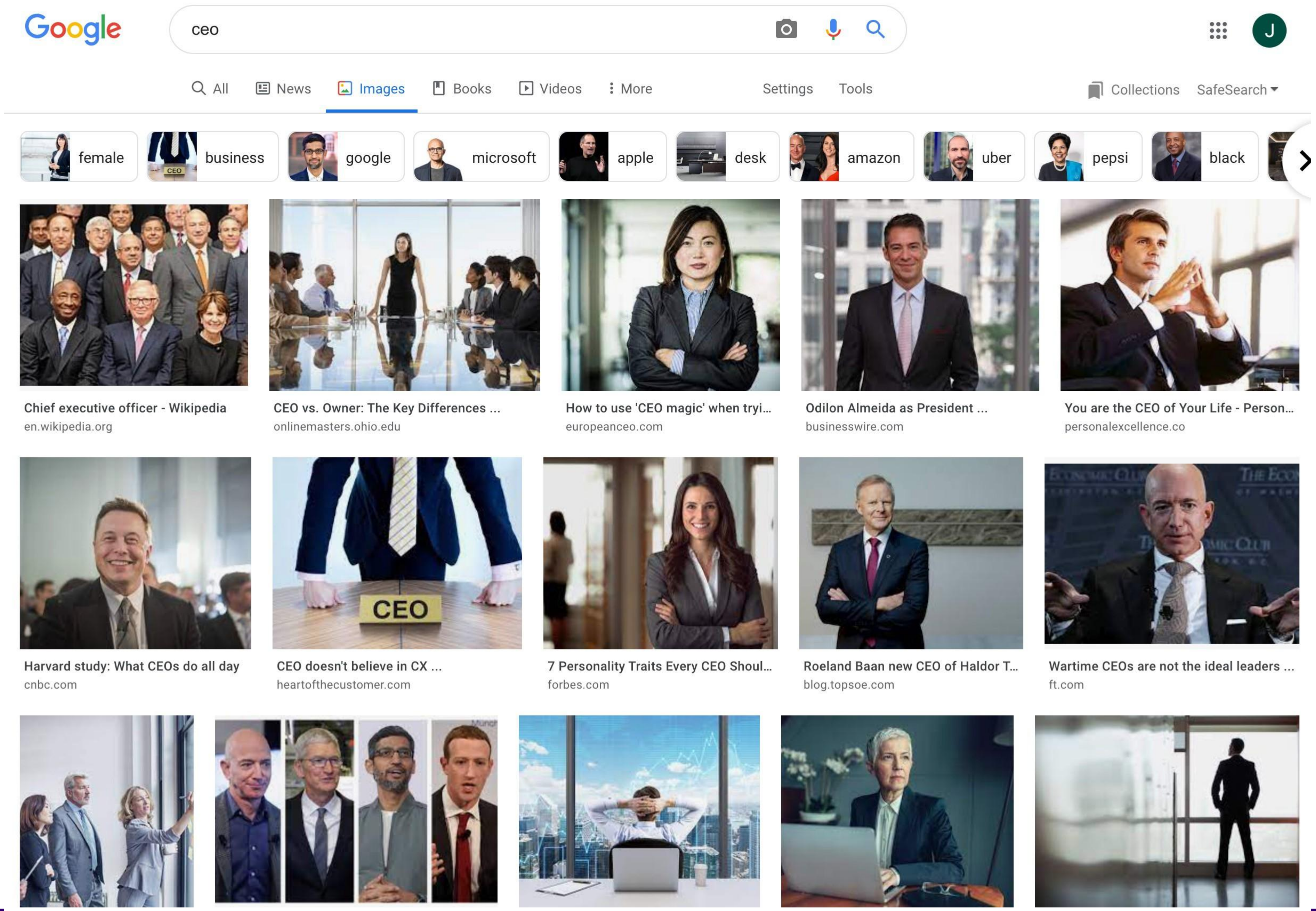


Possible solution:
Change the task;
offer multiple
suggestions



First woman: CEO Barbie =(

Source: https://www.bbc.com/news/newsbeat-32332603



Recent results more diverse

Representational Harm in Super-Resolution

Input: Low-Resolution Face

Output: High-Resolution Face

Menon et al, “PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models”, CVPR 2020
Example source: https://twitter.com/Chicken3gg/status/1274314622447820801

Representational Harm in DALL-E 2

Text Prompt: “lawyer”

Ramesh et al, “Hierarchical Text-Conditional Image Generation with CLIP Latents”, arXiv 2022
https://github.com/openai/dalle-2-preview/blob/main/system-card.md

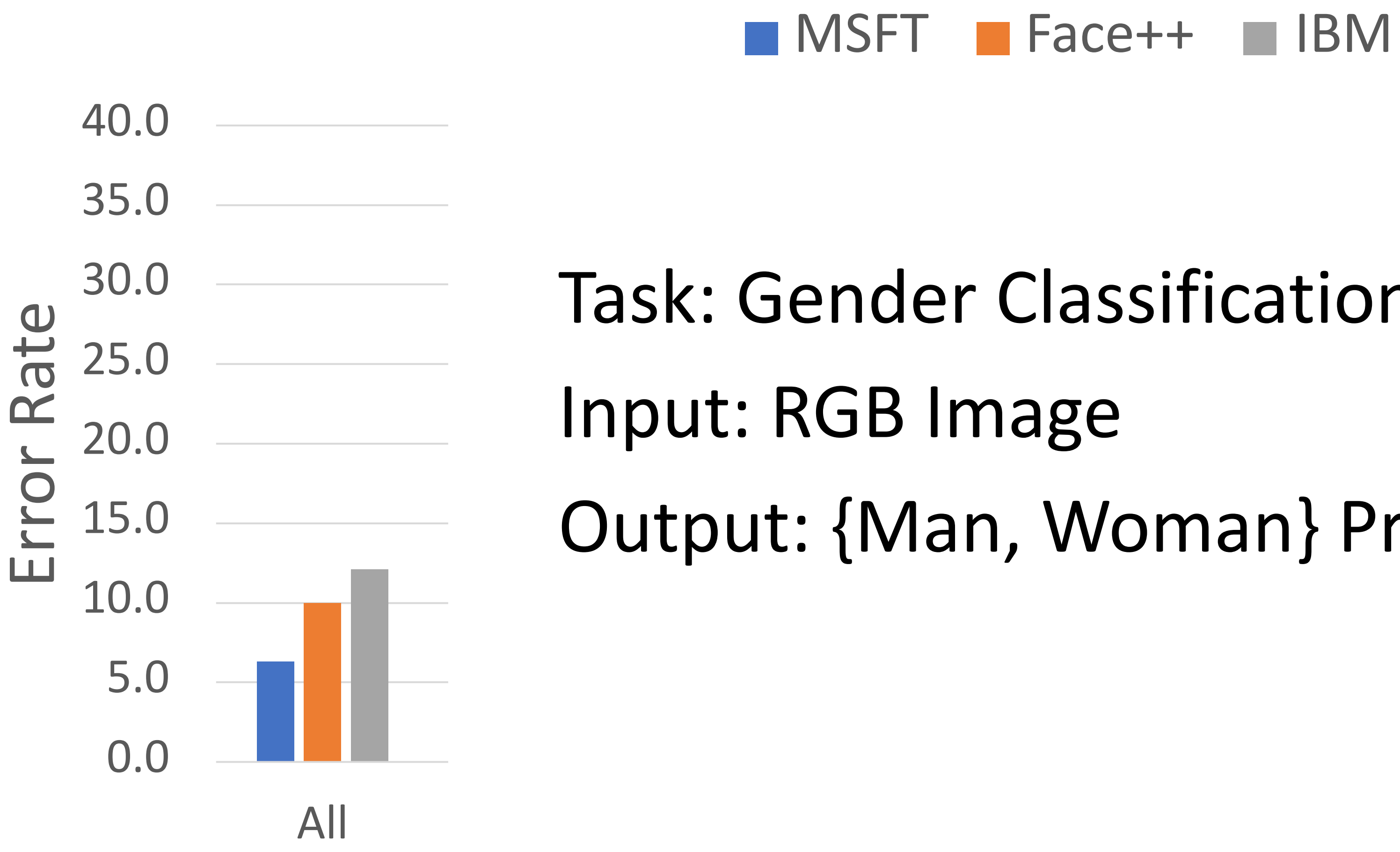
Representational Harm in DALL-E 2

Text Prompt: “nurse”

Ramesh et al, “Hierarchical Text-Conditional Image Generation with CLIP Latents”, arXiv 2022

https://github.com/openai/dalle-2-preview/blob/main/system-card.md

Gender Shades: Intersectionality



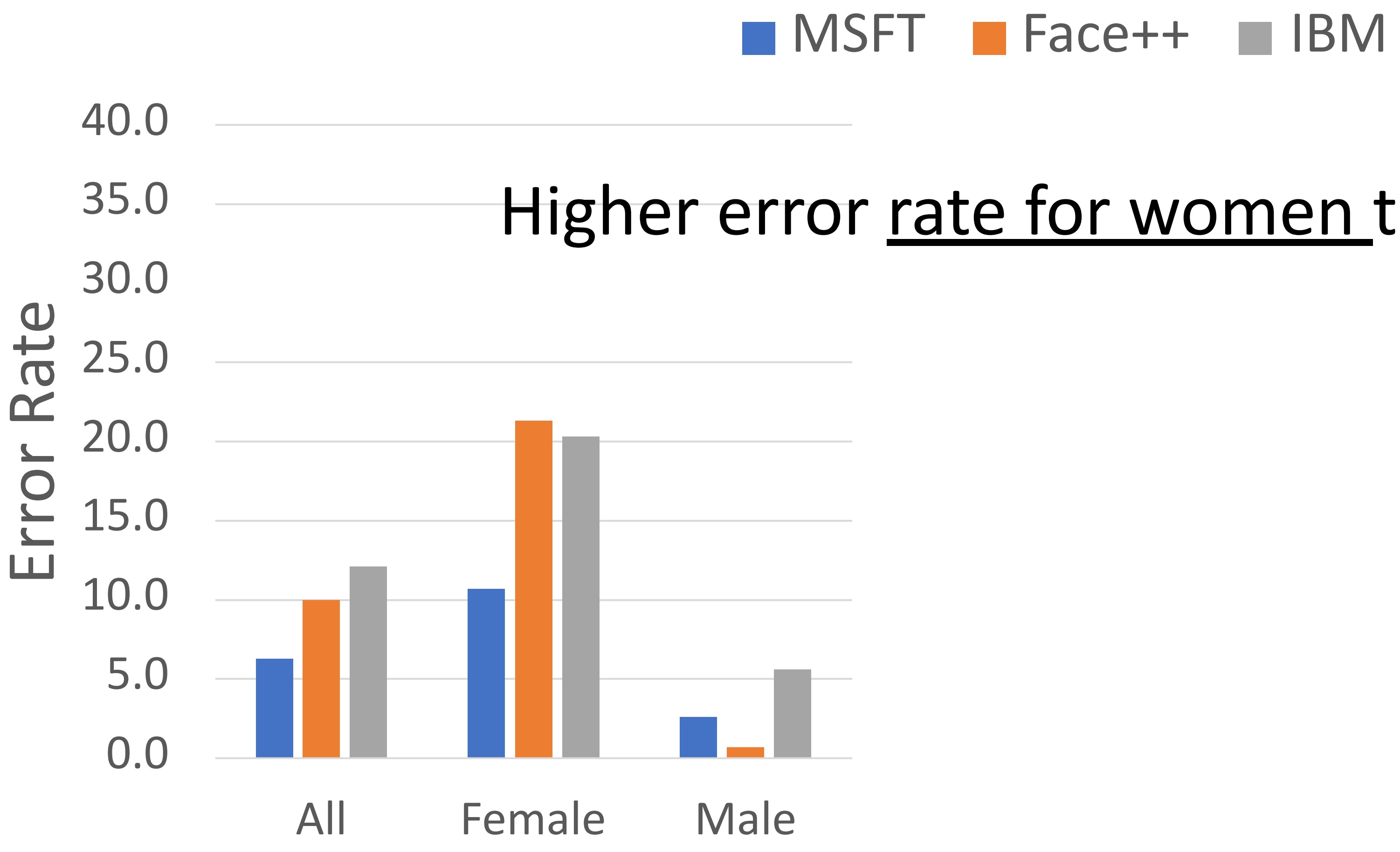
Task: Gender Classification

Input: RGB Image

Output: {Man, Woman} Prediction

Buolamwini and Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, FAT* 2018

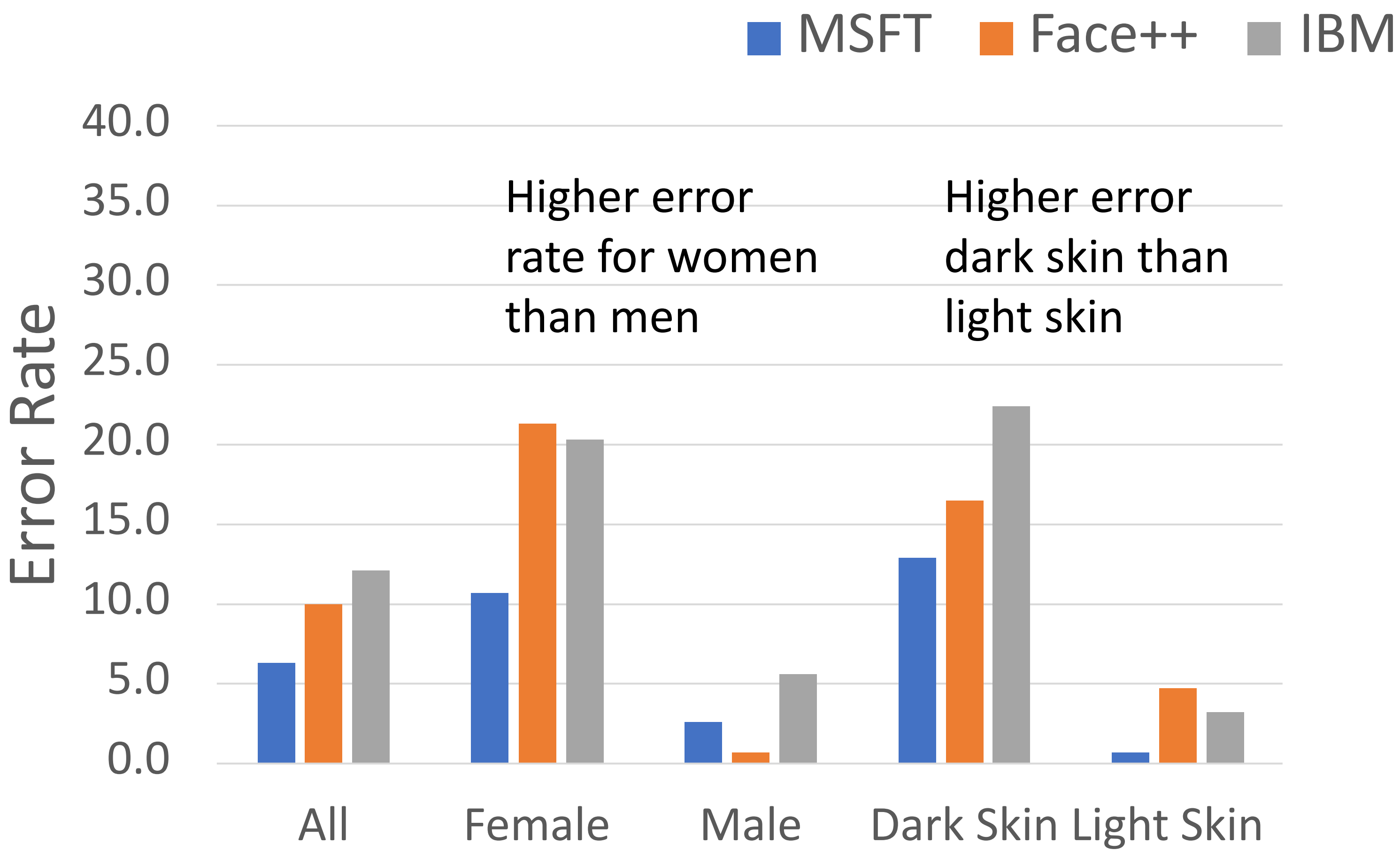
Gender Shades: Intersectionality



Higher error rate for women than men

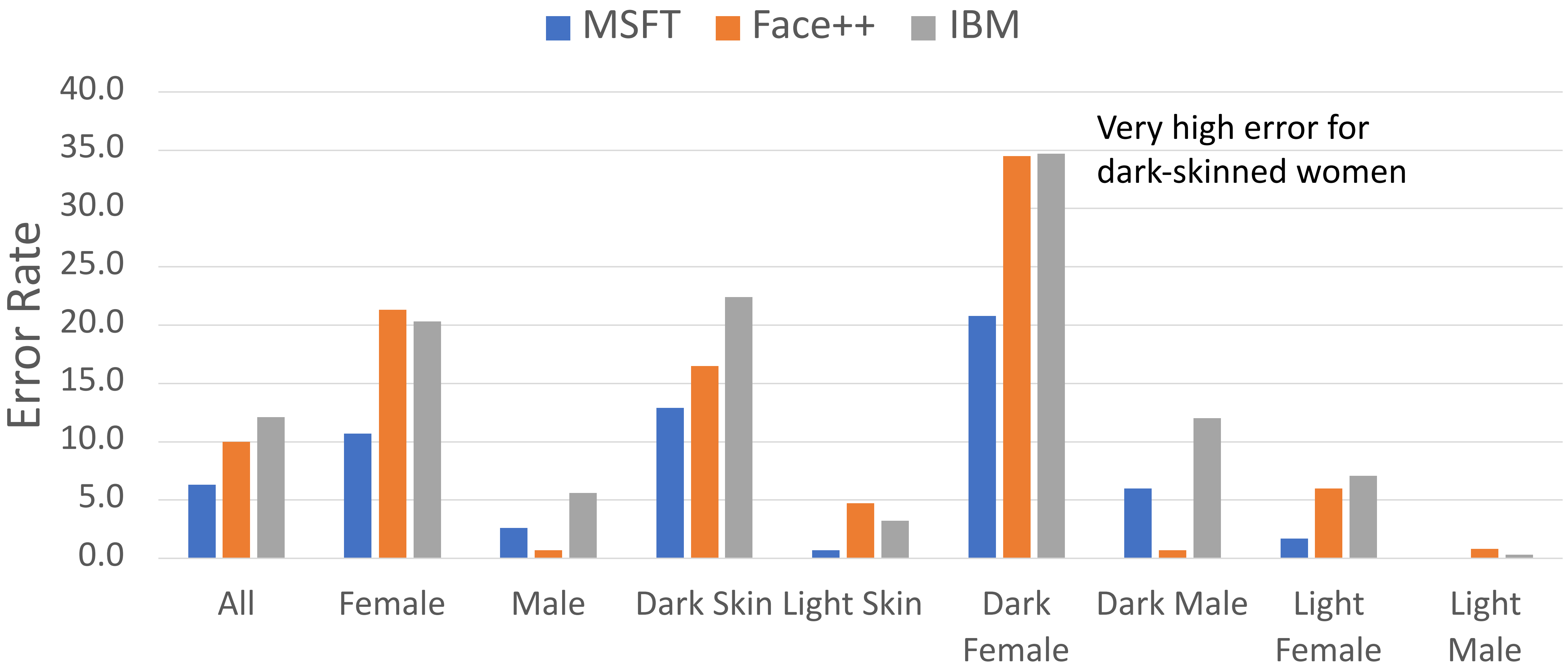
Buolamwini and Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", FAT* 2018

Gender Shades: Intersectionality



Buolamwini and Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, FAT* 2018

Gender Shades: Intersectionality



Buolamwini and Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, FAT* 2018

Same solutions

Datasheets for Datasets

Idea: A standard list of questions to answer when releasing a dataset. Who created it? Why? What is in it? How was it labeled?

A Database for Studying Face Recognition in Unconstrained Environments — **Labeled Faces in the Wild**

Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

Labeled Faces in the Wild was created to provide images that can be used to study face recognition in the unconstrained setting where image characteristics (such as pose, illumination, resolution, focus), subject demographic makeup (such as age, gender, race) or appearance (such as hairstyle, makeup, clothing) cannot be controlled. The dataset was created for the specific task of pair matching: given a pair of images each containing a face, determine whether or not the images are of the same person.[1]

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The initial version of the dataset was created by Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, most of whom were researchers at the University of Massachusetts Amherst at the time of the dataset's release in 2007.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

The construction of the LFW database was supported by a United States National Science Foundation CAREER Award.

The dataset does not contain all possible instances. There are no known relationships between instances except for the fact that they are all individuals who appeared in news sources on line, and some individuals appear in multiple pairs.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)or features? In either case, please provide a description.

Each instance contains a pair of images that are 250 by 250 pixels in JPEG 2.0 format.

Is there a label or target associated with each instance? If so, please provide a description.

Each image is accompanied by a label indicating the name of the person in the image.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included in the dataset.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

There are no known relationships between instances except for

Gebru et al, "Datasheets for Datasets", FAccT 2018

Model Cards

Idea: A standard list of questions to answer when releasing a trained model. Who created it? What data was it trained on? What should it be used for? What should it **not** be used for?

Model Card

- **Model Details**. Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use**. Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors**. Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics**. Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data**. Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data**. May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

Mitchell et al, “Model Cards for Model Reporting”, FAccT 2019

Model Cards

Some models are just for research and not to be deployed. **Make it clear!**

Out-of-Scope Use Cases

Any deployed use case of the model - whether commercial or not - is currently out of scope. Non-deployed use cases such as image search in a constrained environment, are also not recommended unless there is thorough in-domain testing of the model with a specific, fixed class taxonomy. This is because our safety assessment demonstrated a high need for task specific testing especially given the variability of CLIP's performance with different class taxonomies. This makes untested and unconstrained deployment of the model in any use case currently potentially harmful.

Certain use cases which would fall under the domain of surveillance and facial recognition are always out-of-scope regardless of performance of the model. This is because the use of artificial intelligence for tasks such as these can be premature currently given the lack of testing norms and checks to ensure its fair use.

https://github.com/openai/CLIP/blob/main/model-card.md

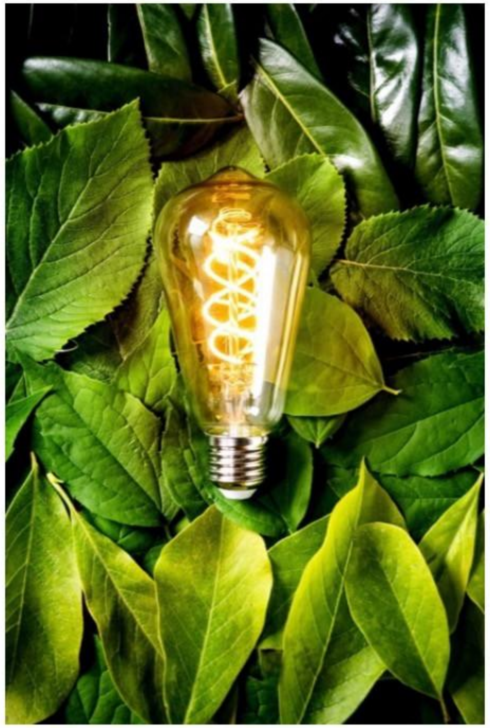
Bigger Models, More Data, More Compute, More problems

- ML Systems can encode bias
- Large models can lack common sense
- Who should control models and data?

Large Models Lack Common Sense

Some plants surrounding a lightbulb

A lightbulb surrounding some plants



Large vision + language models cannot correctly pair images with captions

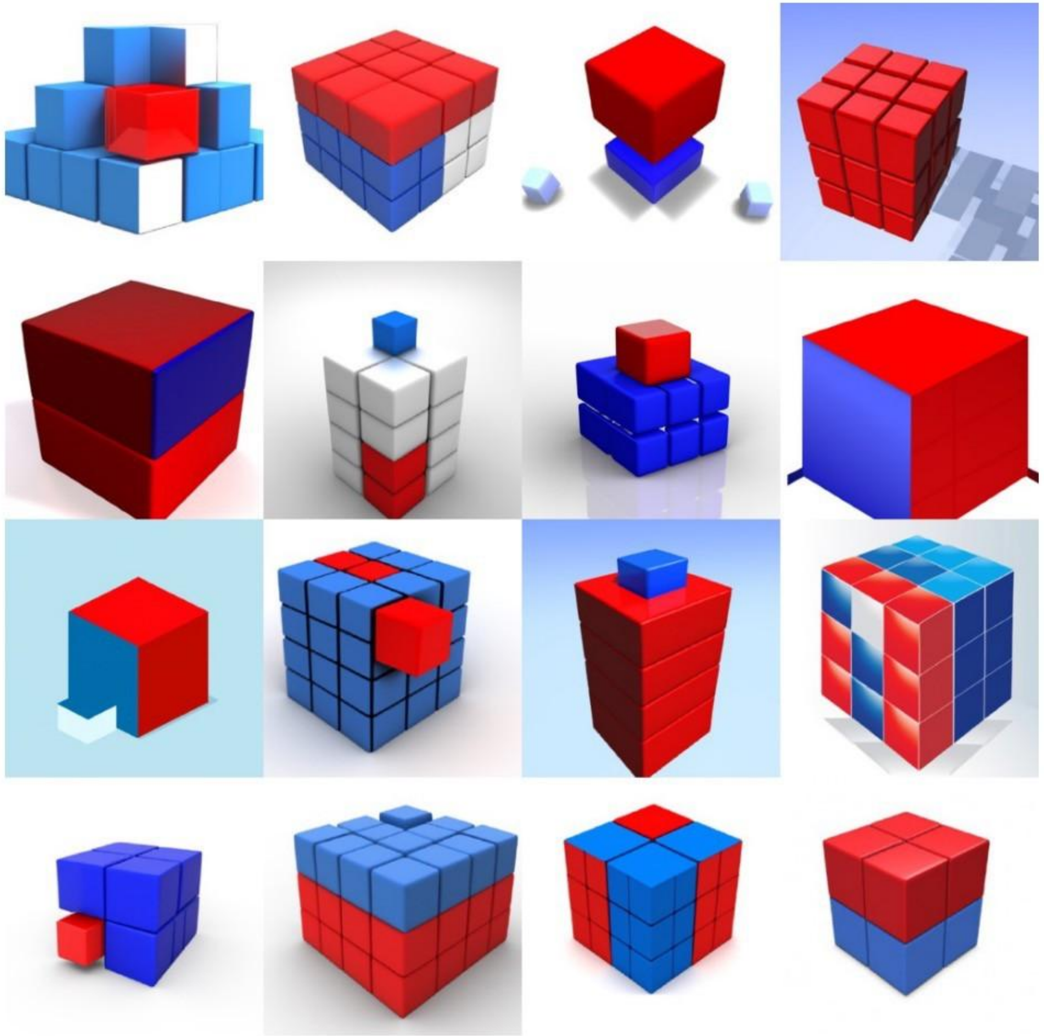
Thrush et al, “Winoground: Probing Vision and Language Models for Visio-Linguistic Compositionality”, CVPR 2022

Large Models Lack Common Sense

Samples from DALL-E 2 for the prompt:

"a red cube on top of a blue cube"

Simple compositions of objects, attributes, relationships often not respected



Ramesh et al, "Hierarchical Text-Conditional Image Generation with CLIP Latents", arXiv 2022

Large Models Lack Common Sense: GPT-2

Bold = prompt written by human
Italic = completion written by GPT-2

I was born in 1950. In the year 2025 my age will be *35. That was only a few years ago. Most things in life just continue to improve.*

I see a black dog and a brown horse. The bigger animal's color is *black, and the smaller is brown.*

Examples generated using https://talktotransformer.com/

Large Models Lack Common Sense: GPT-3

Bold = prompt written by human
Italic = completion written by GPT-3

At the party, I poured myself a glass of lemonade, but it turned out to be too sour, so I added a little sugar. I didn't see a spoon handy, so I stirred it with a paper napkin. But that turned out to be a bad idea because *the napkin disintegrated in the glass. After I finished the drink, I threw the napkin away in a wastebasket—but when I picked up the glass to wash it out, there was a big black spot on my hand.*

Source:
https://cs.nyu.edu/~davise/papers/GPT3CompleteTests.html

Large Models Lack Common Sense

Open question: Can large models learn common sense about the world from lots of (internet) data? Or are there fundamental limitations?

Bigger Models, More Data, More Compute, More problems

- ML Systems can encode bias
- Large models can lack common sense
- Who should control models and data?

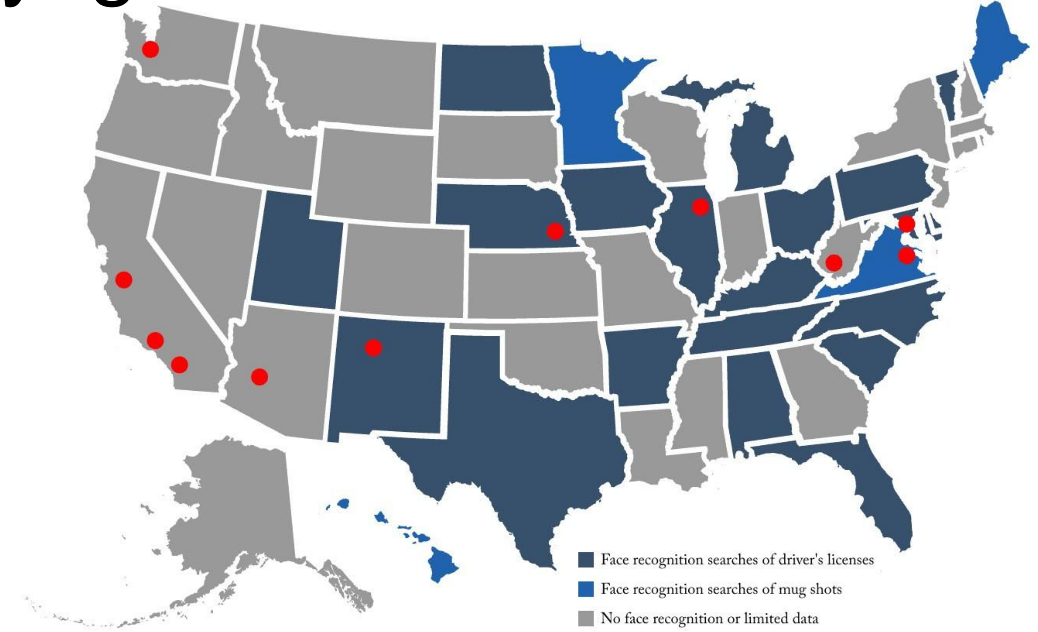
Who should control data?

Image copyright != Consent to use in a dataset

Birhane and Prabhu, "Large Image Datasets: A Pyrrhic Win for Computer Vision?", WACV 2021

Who should control data?

Image copyright != Consent to use in a dataset



"One in two American adults is in a law enforcement face recognition network."

Birhane and Prabhu, "Large Image Datasets: A Pyrrhic Win for Computer Vision?", WACV 2021
Garvie, Bedoya, and Frankle: "The Perpetual Line-Up", 2016, https://www.perpetuallineup.org/

Who should control models?

The largest models can only be trained by large non-academic institutions. Is this a problem?

Who should control models?

The largest models can only be trained by large non-academic institutions. Is this a problem?

Should governments regulate the use of ML-based solutions?

Bigger Models, More Data, More Compute

What's the most critical bottleneck in a long run? model? data? compute?

Bigger Models, More Data, More Compute

I vote for data

Why data?

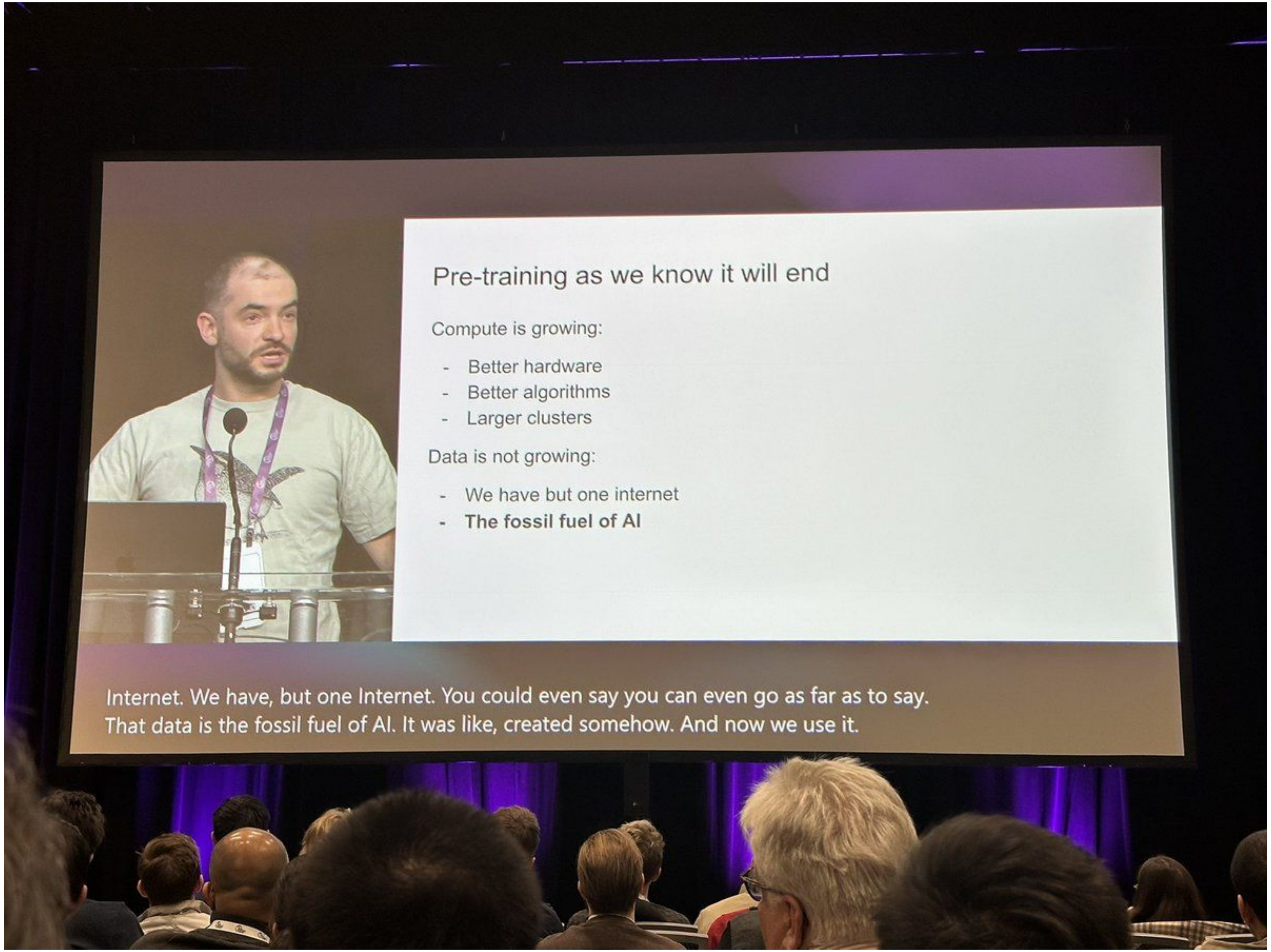
- The data we used to train Large Language Model
  - web crawl data
  - Github
  - Wikipedia
  - Books
  - Arxiv
  - StackExchange
  - ……

Why data?

- The data we used to train Large Language Model
  - web crawl data
  - Github
  - Wikipedia
  - Books
  - Arxiv
  - StackExchange
  - ……

The internet isn't big enough to provide all the data we need

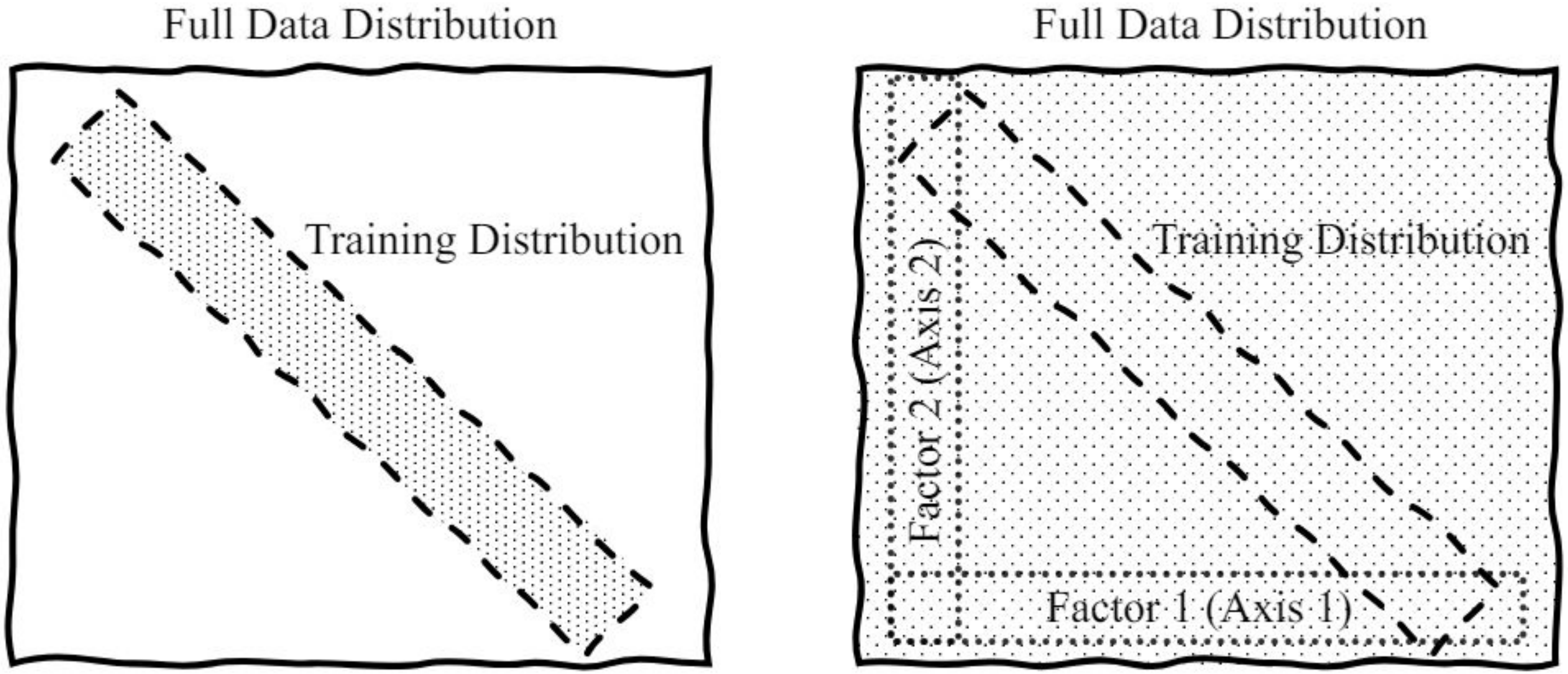
Why data?


Pre-training as we know it will end
Compute is growing:
- Better hardware
- Better algorithms
- Larger clusters
Data is not growing:
- We have but one internet
- The fossil fuel of AI
Internet. We have, but one Internet. You could even say you can even go as far as to say.
That data is the fossil fuel of AI. It was like, created somehow. And now we use it.

Solutions?

- **Start over**, find something else as data-efficient as human
  - human don’t need the internet-scale data to be intelligent
  - but we haven’t figured out exactly why
- **Synthetic data**
  - the power of compositionality
  - the power of programing
  - the power of trained models

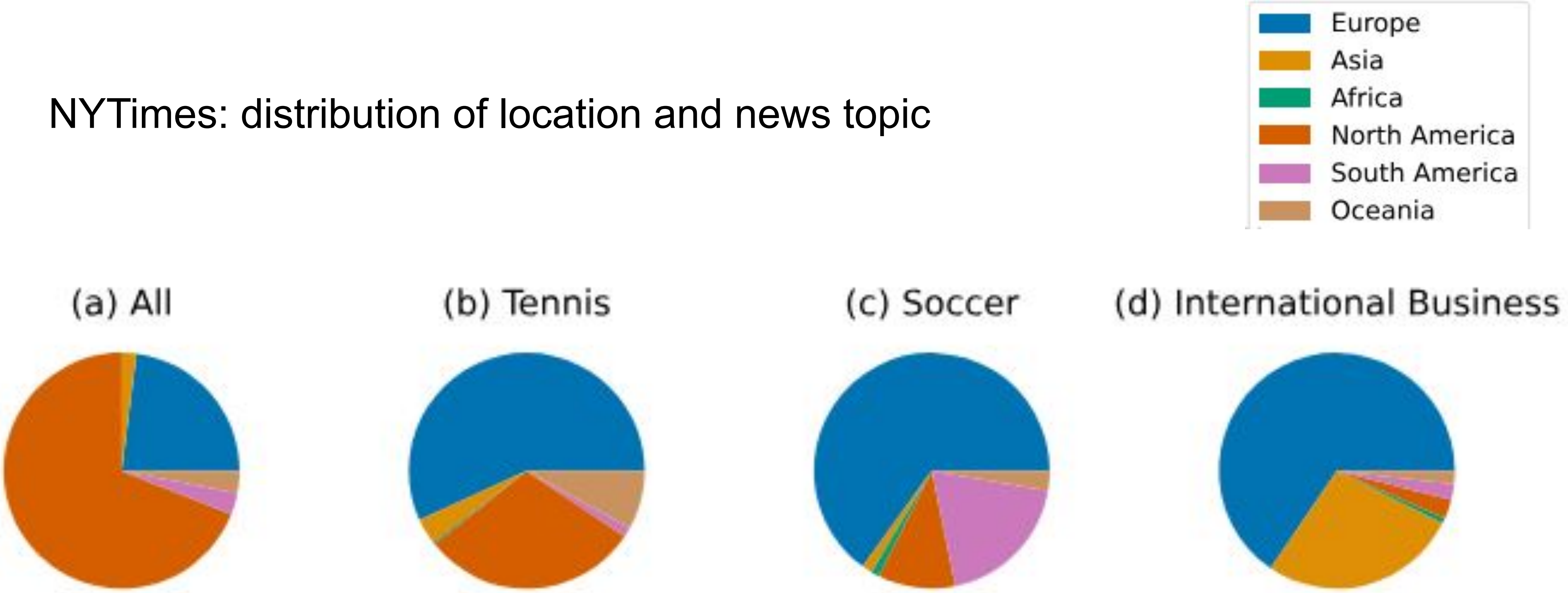
Synthetic data: the power of compositionality



https://arxiv.org/pdf/2402.01103

Synthetic data: the power of compositionality

NYTimes: distribution of location and news topic



Yu, Yue*, Yuchen Zhuang*, **Jieyu Zhang***, Yu Meng, Alexander J. Ratner, **Ranjay Krishna**, Jiaming Shen and Chao Zhang. “Large Language Model as Attributed Training Data Generator: A Tale of Diversity and Bias. NeurIPS 2023.

Synthetic data

- generate synthetic data that ensures balanced distribution

Synthetic data: the power of programing

Question : How many white step stool are there in the image?

Choices : 2, 5, 7, 6

Answer : 2

Question : How many brown clock are there in the image?

Choices : 3, 7, 4, 1

Answer : 3

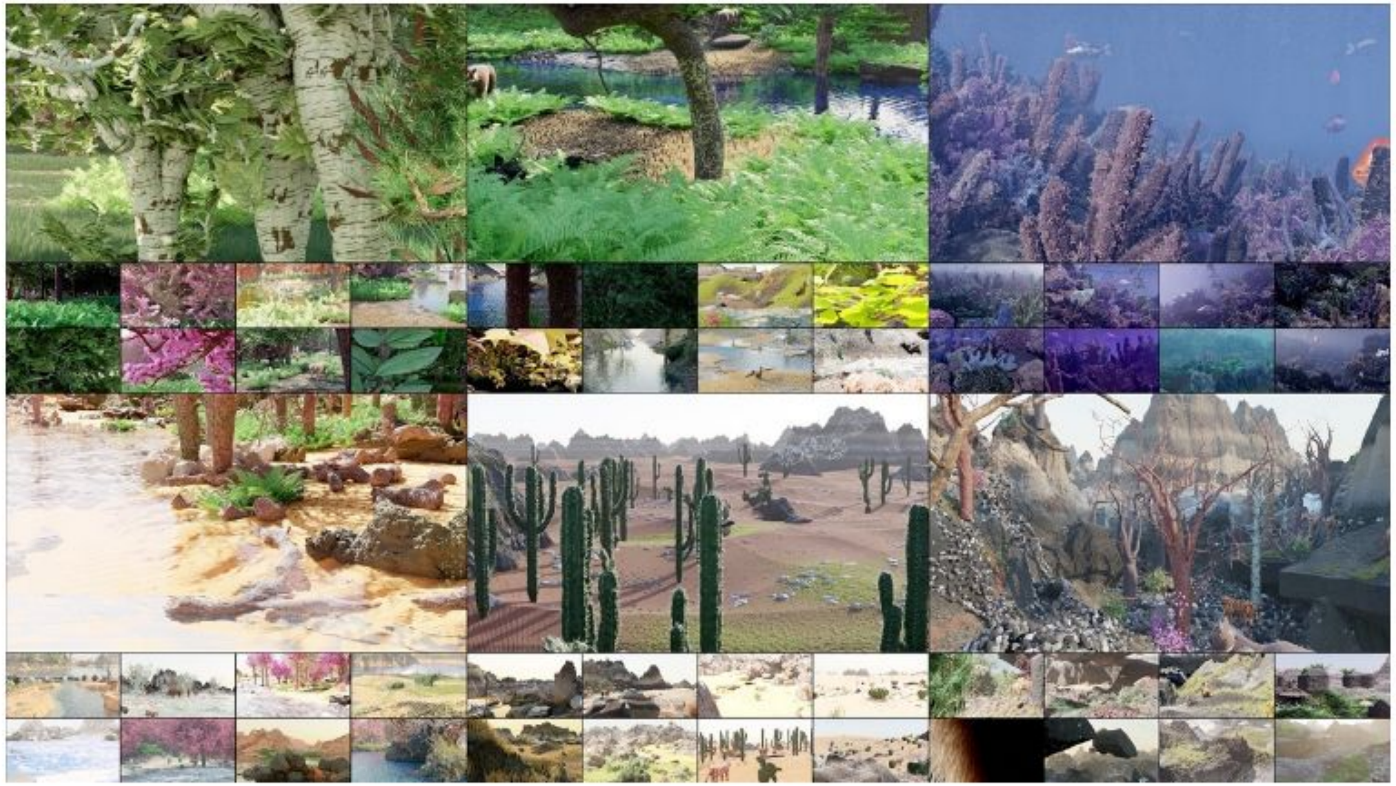
Question : How many wood highchair are there in the image?

Chocies : 3, 1, 2, 4

Answer : 3

ongoing work from our group

Synthetic data: the power of programing



demo: https://youtu.be/6tgspeI-GHY

About Infinigen

Infinigen is a procedural generator of 3D scenes, developed by Princeton Vision & Learning Lab. Infinigen is optimized for computer vision research and generates diverse high-quality 3D training data. Infinigen is based on Blender and is free and open-source (BSD 3-Clause License). Infinigen is being actively developed to expand its capabilities and coverage. Everyone is welcome to contribute.

Posts from @PrincetonVL Follow

Princeton Vision & Lea @Prir · Dec 15, 2023

Infinigen v1.1.1 is out: updated to Blender3.6 and revamped installation to be more robust. See github.com/princeton-vl/i... for more, including our "minimal" install variant which lets you

https://infinigen.org/

Synthetic data

- generate synthetic data that ensures balanced distribution
- generate synthetic data programmatically as we like and scale up

Synthetic data: the power of trained models

Meta uses LLama 2 to help data filtering for LLama 3

To ensure Llama 3 is trained on data of the highest quality, we developed a series of data-filtering pipelines. These pipelines include using heuristic filters, NSFW filters, semantic deduplication approaches, and text classifiers to predict data quality. We found that previous generations of Llama are surprisingly good at identifying high-quality data, hence we used Llama 2 to generate the training data for the text-quality classifiers that are powering Llama 3.

https://ai.meta.com/blog/meta-llama-3/

Synthetic data: the power of trained models

OpenAI trained Sora with model-generated video captions

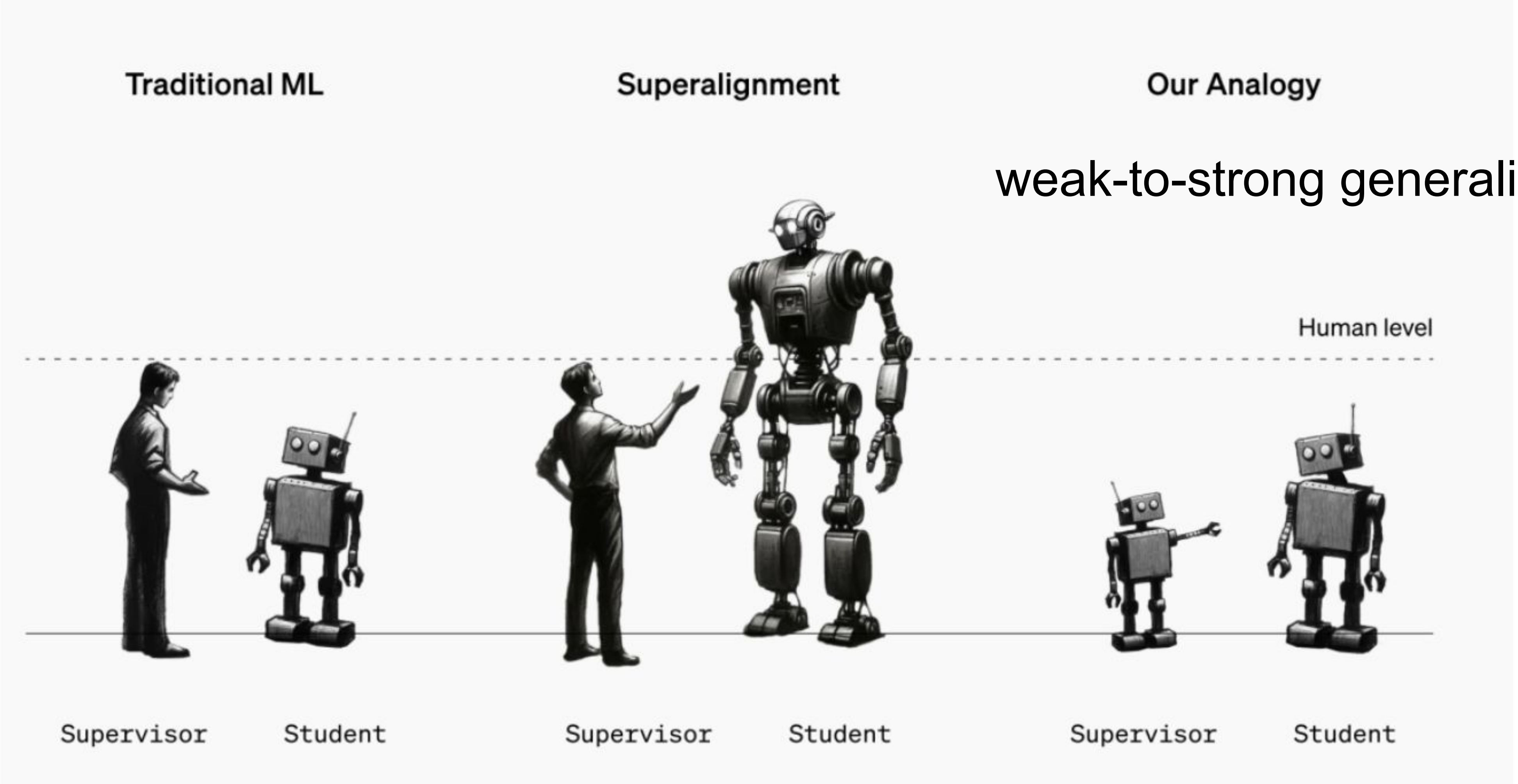
Language understanding

Training text-to-video generation systems requires a large amount of videos with corresponding text captions. We apply the re-captioning technique introduced in DALL·E 3[30] to videos. We first train a highly descriptive captioner model and then use it to produce text captions for all videos in our training set. We find that training on highly descriptive video captions improves text fidelity as well as the overall quality of videos.

https://openai.com/index/video-generation-models-as-world-simulators/

Synthetic data: the power of trained models

weak-to-strong generalization



OpenAI. https://openai.com/index/weak-to-strong-generalization/

Synthetic data: the power of trained models

We can significantly improve generalization in many settings. We use a simple method that encourages the strong model to be more confident—including confidently disagreeing with the weak supervisor if necessary. **When we supervise GPT-4 with a GPT-2-level model using this method on NLP tasks, the resulting model typically performs somewhere between GPT-3 and GPT-3.5.** We are able to recover much of GPT-4's capabilities with only much weaker supervision.

OpenAI. https://openai.com/index/weak-to-strong-generalization/

Synthetic data

- generate synthetic data that ensures balanced distribution
- generate synthetic data programmatically as we like and scale up
- let weaker model supervise strong model

Data-centric AI

Introduction to Data-Centric AI

MIT

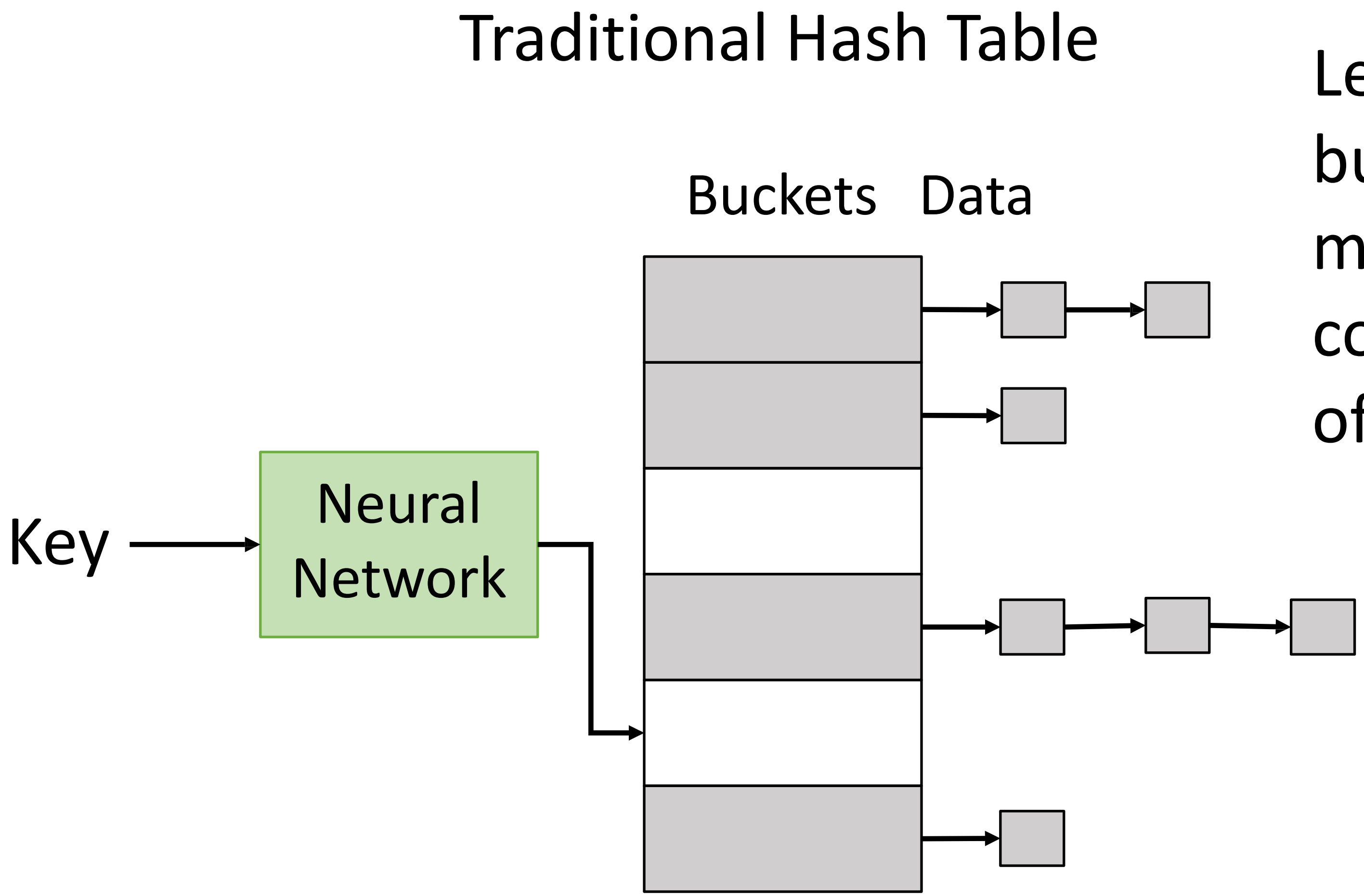
https://dcai.csail.mit.edu/

Deep Learning is Here to Stay

Deep Learning is Here to Stay

and will impact more than computer vision

Deep Learning for Computer Science

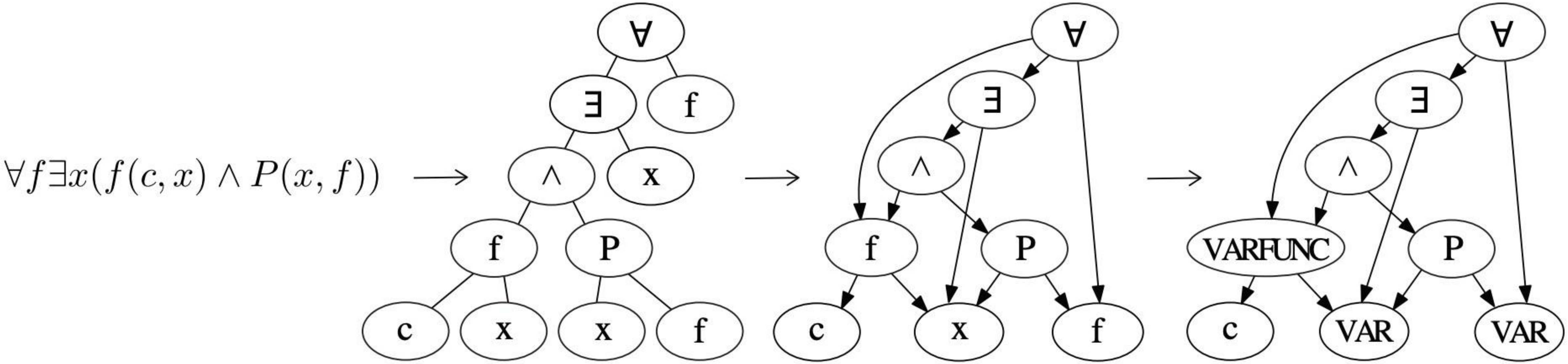


Learn to assign keys to buckets in a way that minimizes hash collisions for the types of data you encounter

Kraska et al, "The Case for Learned Index Structures", SIGMOD 2018

Deep Learning for Mathematics

Convert mathematical expressions into graphs, process then with graph neural networks!



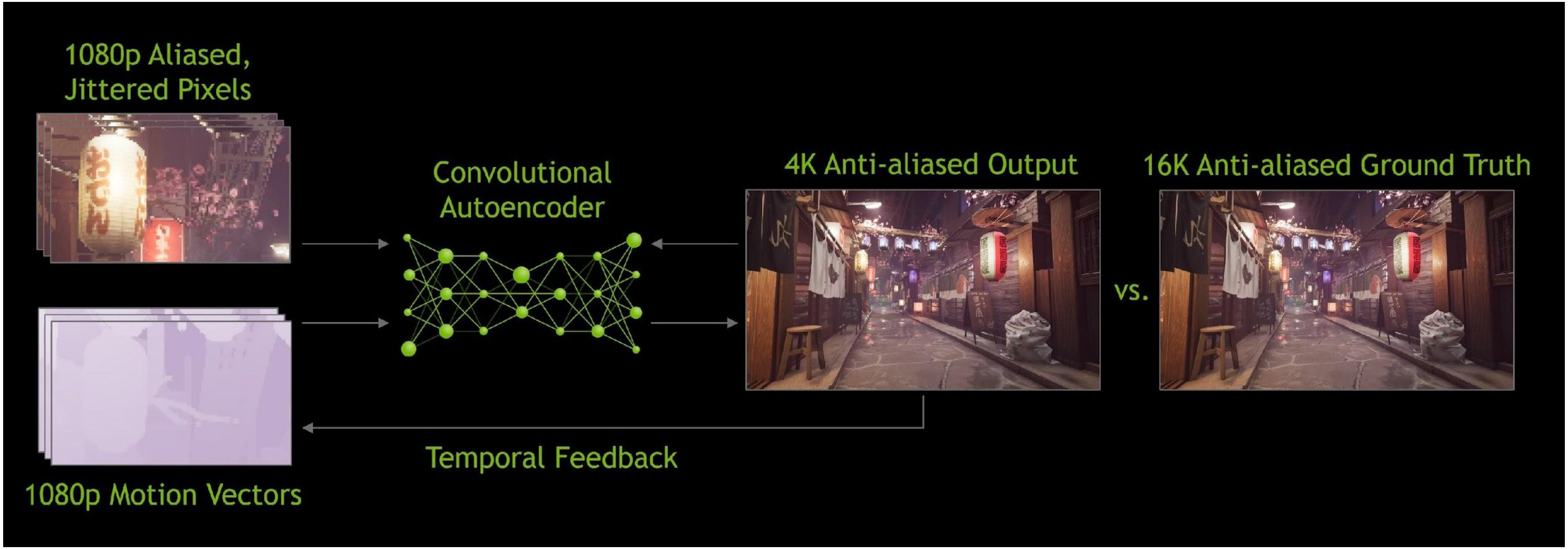
Applications: Theorem proving, symbolic integration

Wang et al, “Premise Selection for Theorem Proving by Deep Graph Embedding”, NeurIPS 2017

Kaliszyk et al, “Reinforcement Learning of Theorem Proving”, NeurIPS 2018

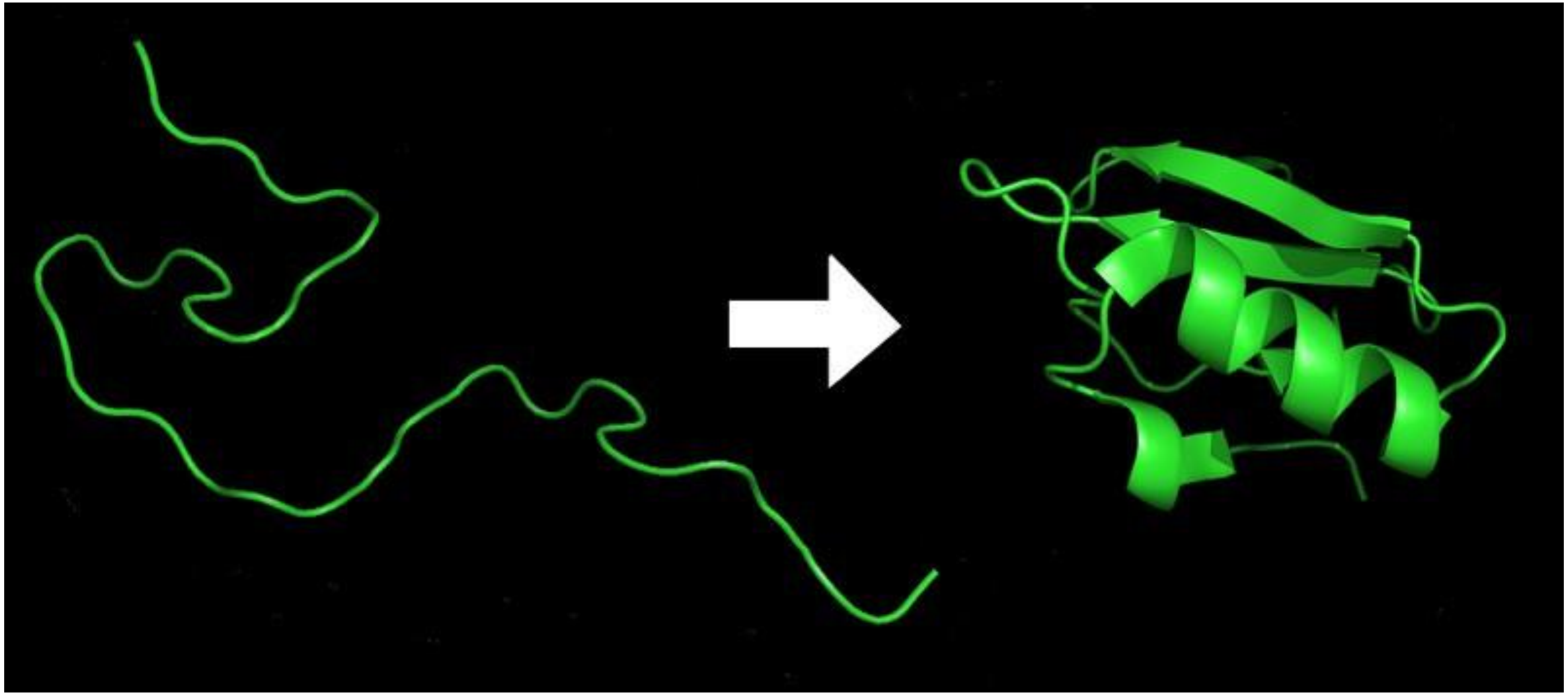
Lample and Charton, “Deep Learning for Symbolic Mathematics”, arXiv 2019

Deep Learning for Graphics: NVIDIA DLSS



https://www.nvidia.com/en-us/geforce/news/nvidia-dlss-2-0-a-big-leap-in-ai-rendering/

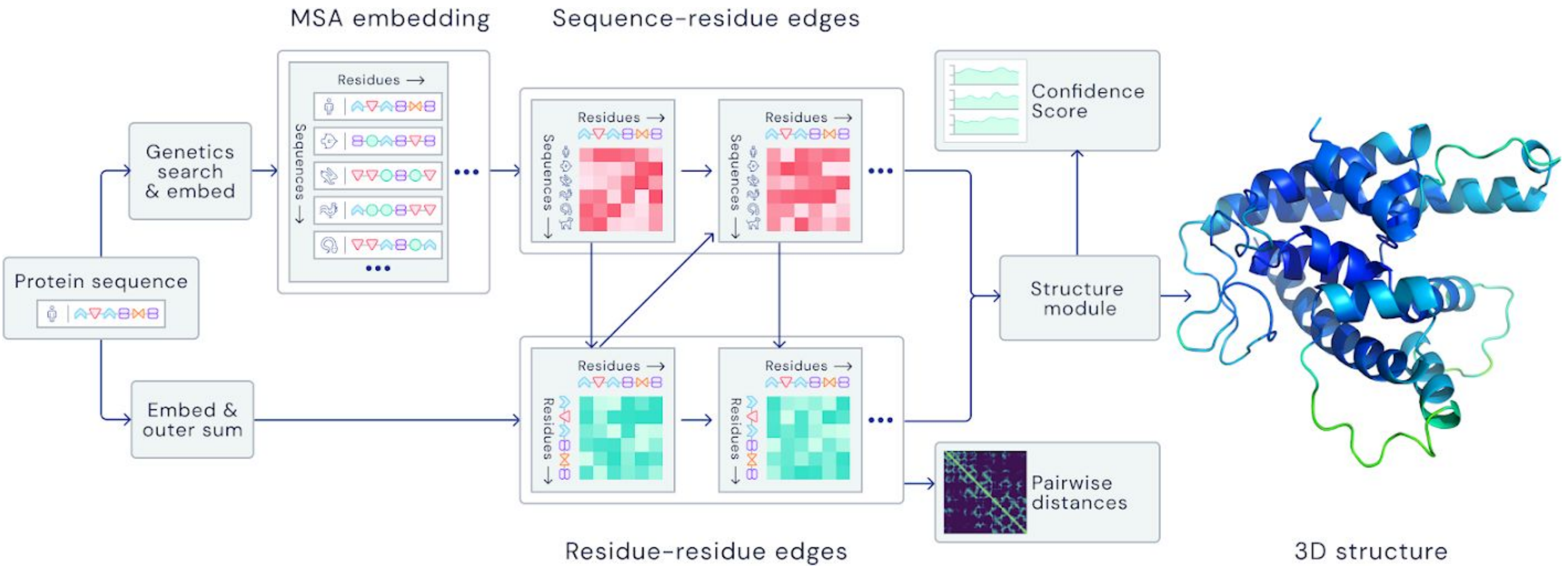
Deep Learning for Science: Protein Folding



Input: 1D sequence of amino acids

Output: 3D protein structure

Deep Learning for Science: AlphaFold 2



https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

Computer Vision Technology Can Better Our Lives

Computer Vision Technology Can Better Our Lives

Now is a great time to be working in computer vision and deep learning!

The End!