Intelligence artificielle : défis et opportunités (pour l'environnement)

Marc Schoenauer
TAU team, INRIA, LISN, UP-Saclay & CNRS,

Journée 2024 du PEPR Agroécologie et Numérique
31 janvier 2024

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

Wrong

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

although ...

Observations	+ Target	+ Reward
Understand Code/compress	Predict Classification/Regression	Decide Policy/strategy
Unsupervised Learning	**Supervised** Learning	**Reinforcement** Learning

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

although …

Observations	+ Target	+ Reward
Understand Code/compress	Predict Classification/Regression	Decide Policy/strategy
Unsupervised Learning	Supervised Learning Generative AI	**Reinforcement** Learning

Agenda

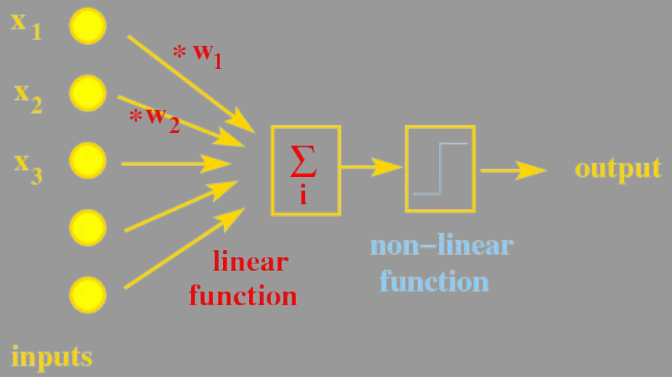
- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Agenda

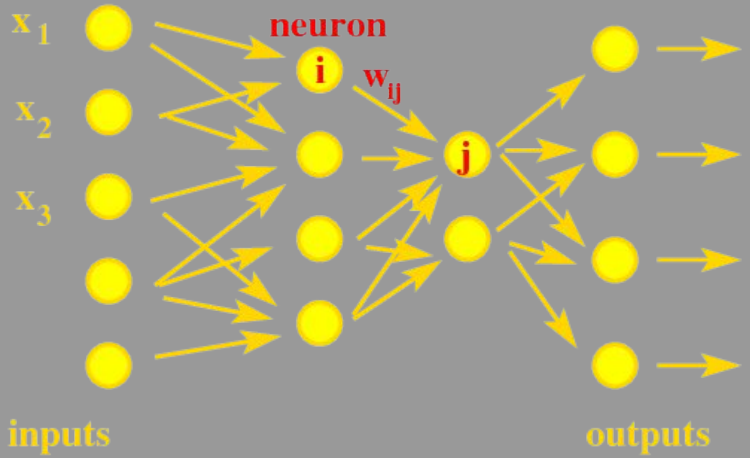
- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Supervised Learning: a Regression pb

- Given a set of labeled examples $(x^i_1, \ldots, x^i_d, y^i)_{i=1,\ldots,n}$
- Find a model **f** s.t. $f(x^i_1, \ldots, x^i_d) \approx y^i$ for all i
- A zoology of models: Polynomials, Decision trees, Random Forests, SVMs, and
- **Artificial Neural Networks**



One neuron



A network of neurons

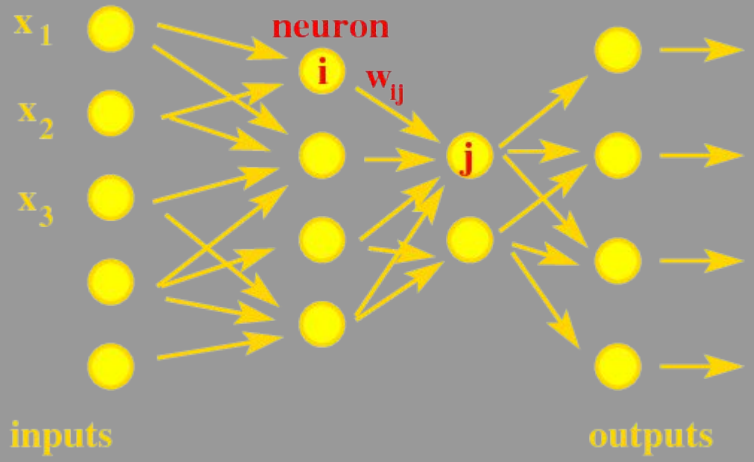
Parameters are the **weights w_{ij}**

Deep Learning in one slide

Learning Phase

Gradient **back-propagation** aka **Stochastic Gradient Descent** 60s

- Present the examples 1 by 1
  - or mini-batch by mini-batch
- **Forward** pass: Compute the **Loss**
  e.g., $L = \sum_i |y(x^i_1, \ldots, x^i_d) - NN(x^i_1, \ldots, x^i_d)|^2$
- **Backward** pass: Commute $\nabla_w L$ (chain rule)
- Modify the weights w_{ij} from $\nabla_w L$
  to decrease of the loss
- Loop



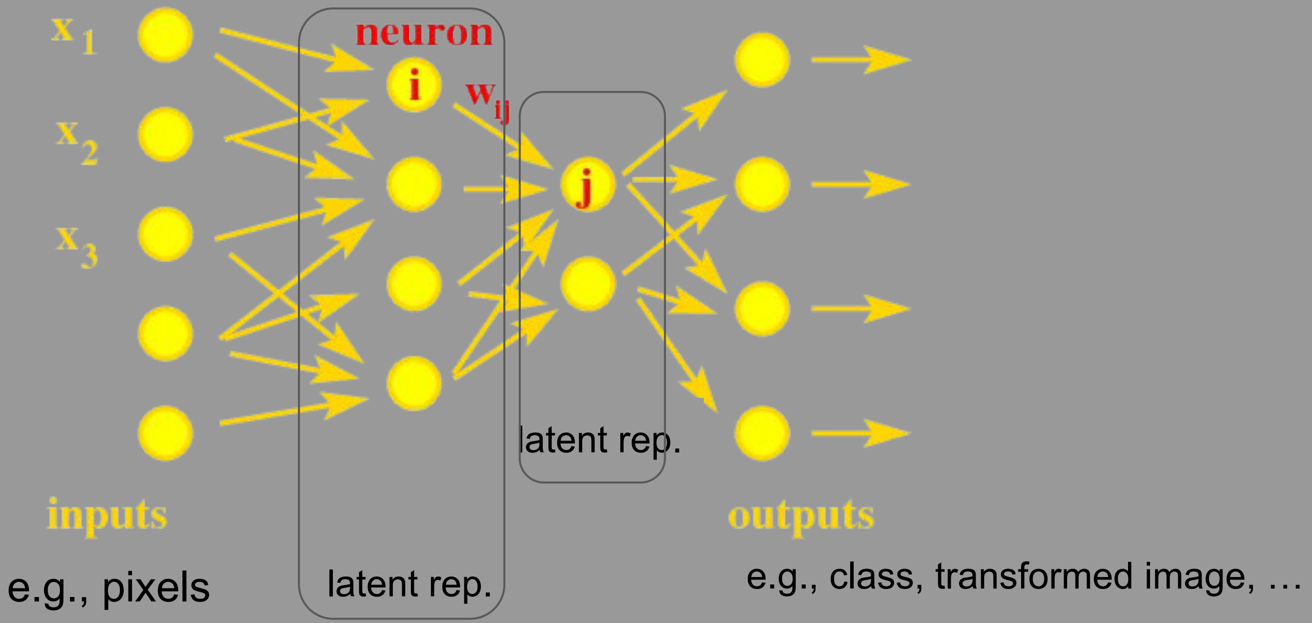
Recognition Phase aka Inference

Input an unlabelled example, the network output a predicted label

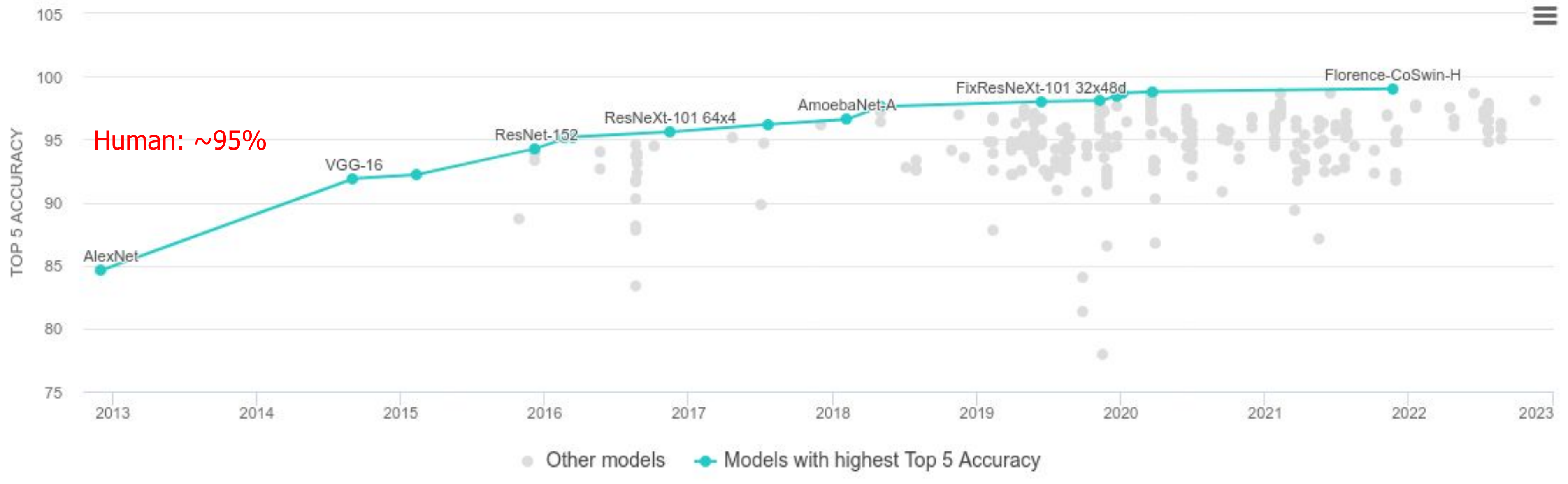
Differentiable Programming

A Deep Neural Network

- Performs **end-to-end** learning
- Learns a sequence of **representations – aka latent spaces**



Performances in Image Recognition



Leaderboard Top-5 on Imagenet (14M annotated images)

- Before 2012: non Deep Networks, ~76%
- 2012, Hinton: 84%
- Best 10 Top-1 and best 5 Top-5 are **Transformer** networks, not pure CNNs

https://paperswithcode.com/sota/image-classification-on-imagenet

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Opportunities 1/2

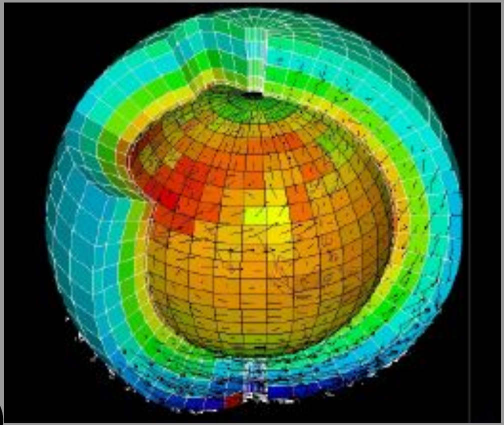
Supervised Learning for Numerical Simulations

- Numerical solutions of PDEs viewed as “images”
  - Surrogate approaches (full data-based)
  - What about error bounds?
  - Where is the physics?
- More intricate hybridizations are needed
  - but are they really?

  Will “The Data Deluge Make the Scientific Method Obsolete”?(*)

(*) C. Anderson (2008). “The End of Theory”, Wired Magazine. url: https://www.wired.com/2008/06/pb-theory/.

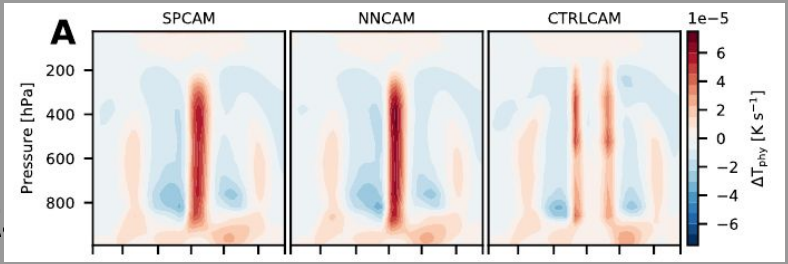
A Surrogate Approach



Learning a surrogate of sub-scale phenomena

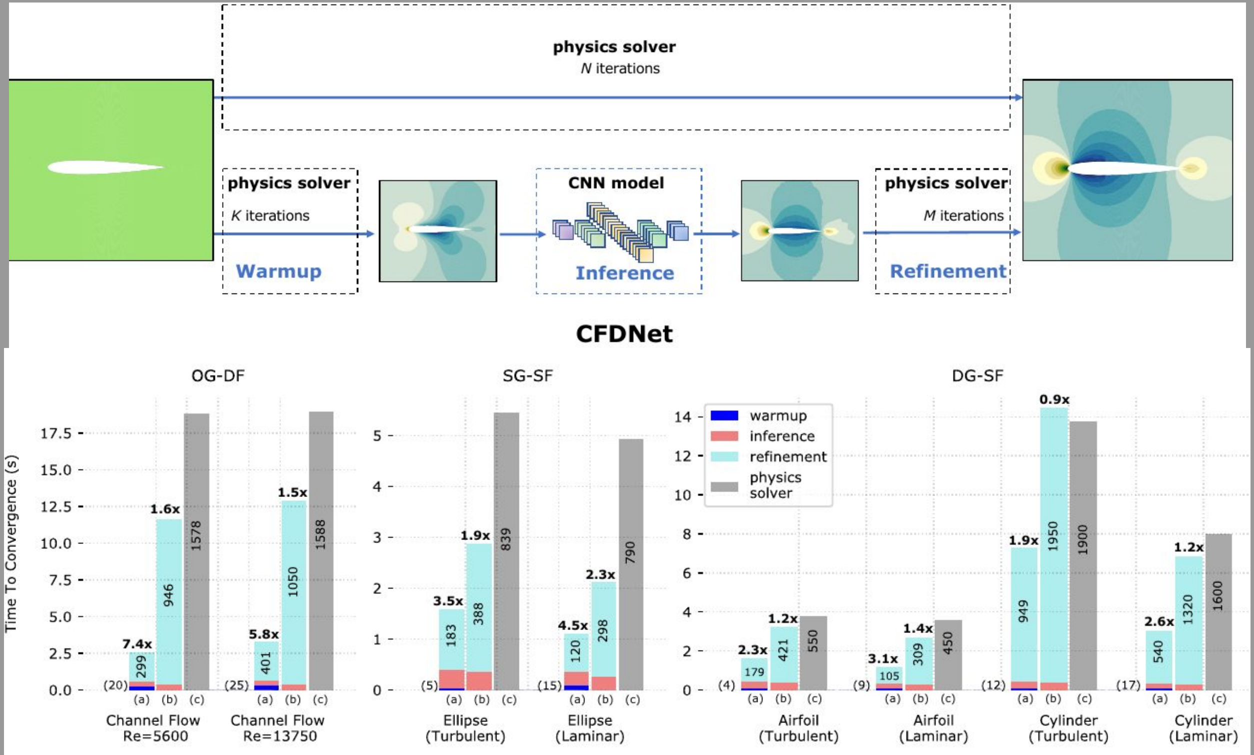
- Global climate modeling
  - 2° horizontal resolution, 30 altitude levels
  - 30mn time step
- needs to solve CRMs (Cloud Resolving Models)
  - turbulence + cloud convection + ...
  - in each column (4km-wide), at each time-step (20s)

- Train a DNN on one-year SPCAM simulations - **140M examples**
- **20x speedup**, statistics OK
- Energy conservation (post-hoc)
- Good interpolation generalization
- **Poor OoD generalization** beyond train dat
- **No error bound**



S. Rasp et al. *PNAS* 2018

Initializer Approach



Tackling the accuracy issue

Obiols-Sales, Vishnu, Malaya, et al. 2020

Physics Informed Deep Learning

Data-driven solution of PDEs

- Given a PDE: $u_t = \mathcal{N}(t, x, u, u_x, u_{xx}, \ldots)$
- Define **residual** $f := u_t - \mathcal{N}(t, x, u, u_x, u_{xx}, \ldots)$
- and **Loss** $$\frac{1}{N_u}\sum_{i=1}^{N_u} |u(t_u^i, x_u^i) - u^i|^2 + \frac{1}{N_f}\sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2$$

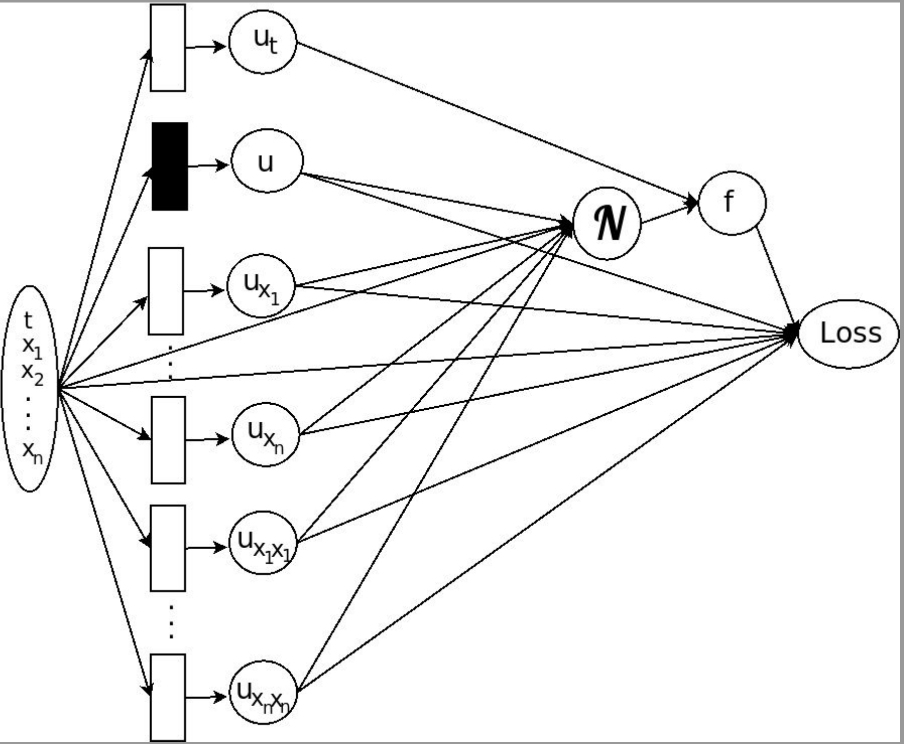
with

- $\{t_u^i, x_u^i, u^i\}_{i=1}^{N_u}$ initial and boundary training data
- $\{t_f^i, x_f^i\}_{i=1}^{N_f}$ collocation training points

Raissi 2017

Physics Informed Deep Learning

Thanks to differentiable programming

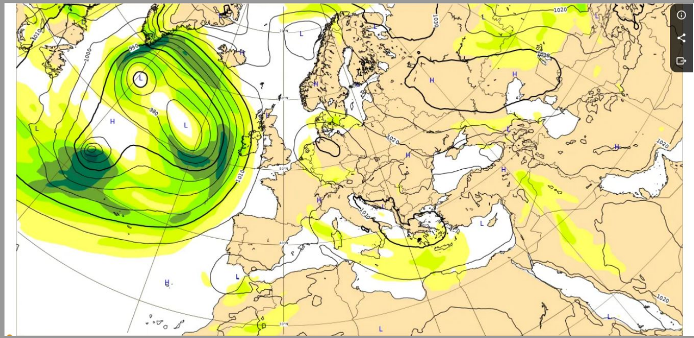


Issues

- Requires new learning for new boundary conditions or source term
- Data-hungry
- Physics still approximate
    - and not constrained

Raissi et al., 2017

Surrogates strike back: GraphCast



Google DeepMind data-based weather model

- **Input**: Two states (6 hours ago and now)
- **Output**: next state (6 hours ahead)
  - and iterate (autoregressive)
- **State**: 235 M-variables (900 Mb)
  - 28×28 km grid (721 × 1440) × 37 vertical levels
- 36.7 M weights - encode/process/decode
- Trained on 39 years of ECMWF's ERA5 data
- **Loss**: MSE for N auto-regressive steps
- 3 weeks on 32 TPU v4 devices

Outperforms ECMWF's IFS on 10-days forecast

Lam et al. (DeepMind), Science, 2022

Surrogates strike back: recent models

	FourCastNet	Pangu-Weather	GraphCast
AI technique	AFNO (transformer)	3DEST (transformer)	Graph neural network
Hardware – train (inference)	64 A100 (1 A100)	192 V100 (1 V100)	32 TPU v4 (1 TPU v4)
Speed – train (inference[1])	16 hours (2.8 s)	16 days (14 s)	3 weeks (60 s)
Forecast scores[2]	Comparable to IFS	Better than IFS	Better than IFS
# of variables	20	69	227
Open-source	Yes	Yes	No

Th. Rieutord (Met Éireann), A review of the recent papers on fully data-driven NWP with AI, March 2023

Hybridization

Still some issues

- Out of Distribution generalization (e.g., due to climate change)
- Certification / error bounds
- Robustness and replicability
- Explainability (what science is about)
- Huge training set required
- Sustainable (frugal) learning?

but things are moving faster than ever

Opportunities 2/2

Correlation vs causality

- Supervised learning learns **deductive** models
  - If umbrellas are open, it is raining
- and not **prescriptive** models
  - if people open their umbrellas, is it going to rain?
- Causality is often implicit, or common sense
  - But what to do when it is unknown?
- and can come from hidden variables
  - Correlation between wealth of company and well-being of employees

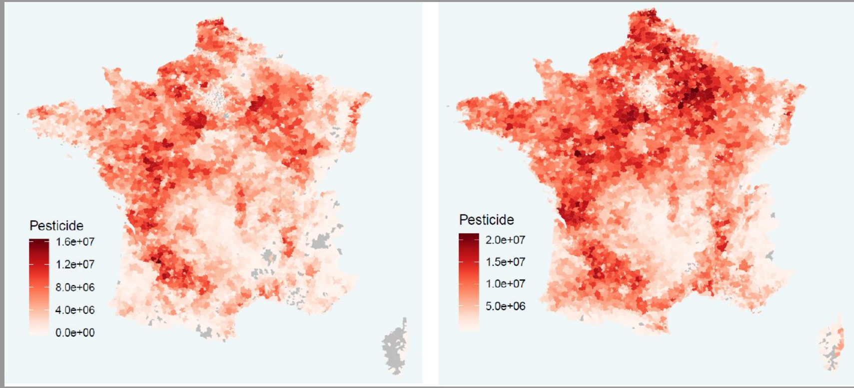
It can (often) be learned from data

Causality: Use cases

Causal modeling (not deep learning :-)

Horapest

- Links between pesticides and neonatal disorders
- Coll. Inria – CHU Toulouse
- Data: SNDS + ventes (BNV-D)



Nutriperso

- Long-term goal: personalized nutritional recommendations
- Coll. INRAE, CEA, AgroParistech, … Inria
- Kantar panel dataset: 170 000 food items, 20 000 households + BMI
- Difficulty: preserve the details of the food items

e.g., 387 types of pizza

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Risks of Supervised (Deep) Learning

Even before the blooming of Generative AI,

- Transparency: Explainability and Interpretability
- Robustness
- Verification and validation

and of course dramatic lack of sustainability, worsened with generative AI

in the following, AI Index = AI Index Report 2023, Stanford Human-centered AI (2022 and before)

Robustness w.r.t. context/noise

Funny

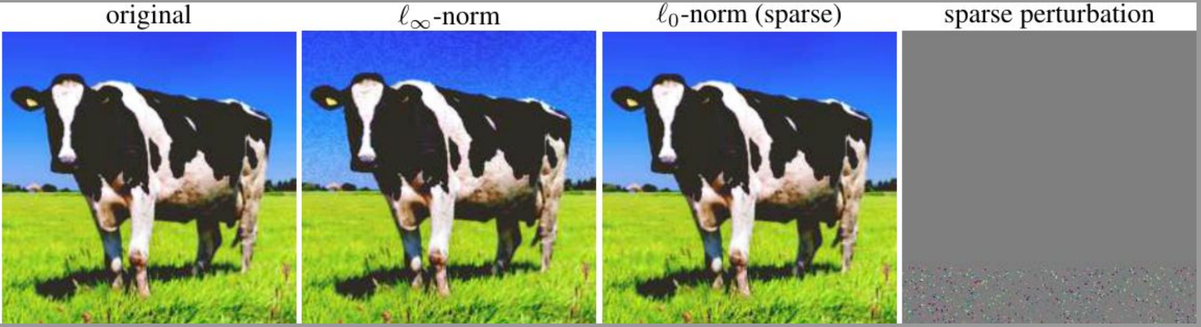
The Verge, Nov. 2020

or not

Issue: completeness of the training set

Robustness w.r.t. attacks

Well chosen noise → wrong label



Cow (a) classified “Traffic light” (b-c)

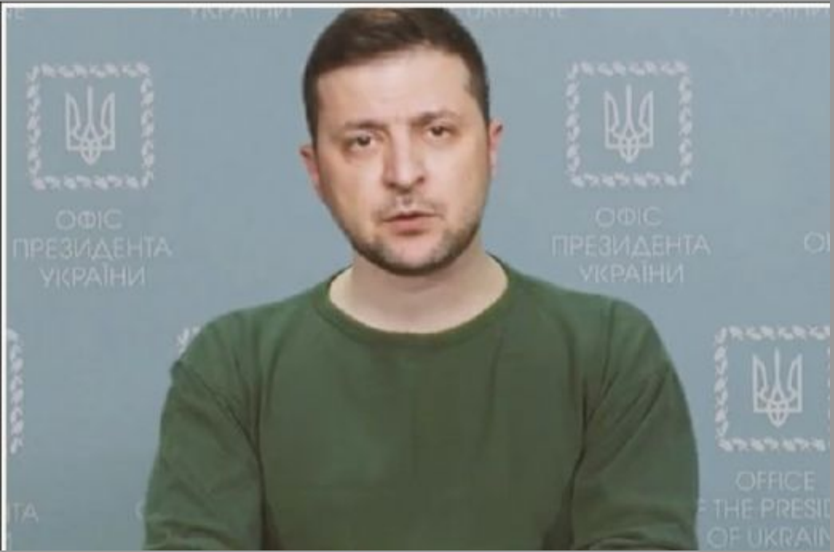
Shafahi et al.,2018

All are recognized “Speed limit 45” from different distances and angles.

Eykholt et al.,2017

latent rep.

Deep Fakes



March 2022: Ukraine surrenders!
AI Index p134

PRIVACY AND SECURITY

Scammer Successfully Deepfaked CEO's Voice To Fool Underling Into Transferring $243,000

Jennings Brown
9/03/19 11:20am • Filed to: AUDIO DEEPFAKES

71.3K 45 7

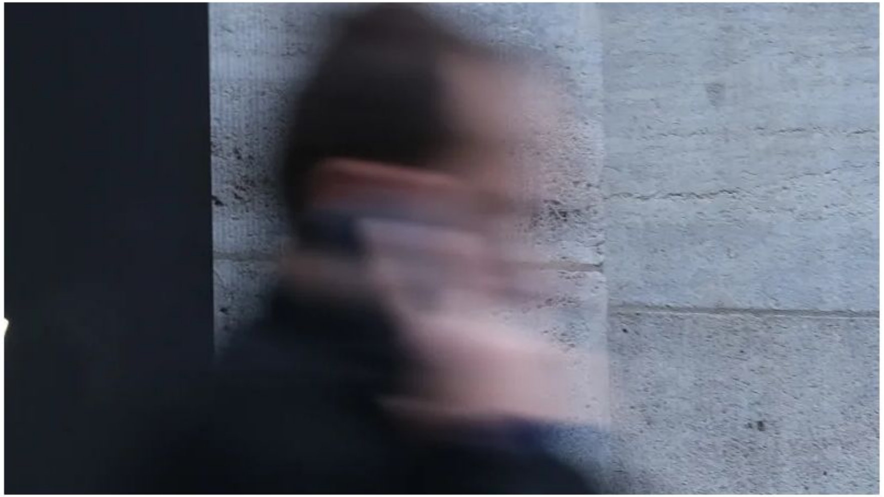


Photo: Sean Gallup (Getty)

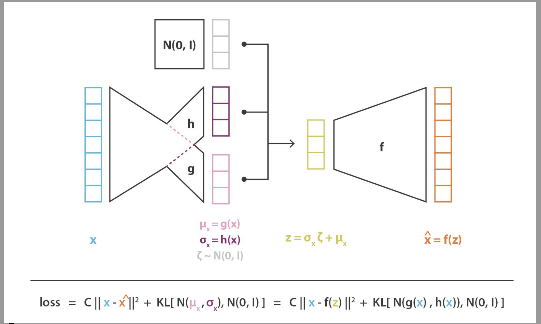
Gizmodo ← Wall Street Journal 30/08/2019

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Early approaches

Kingma and Welling, 2014



Variational Auto-Encoders

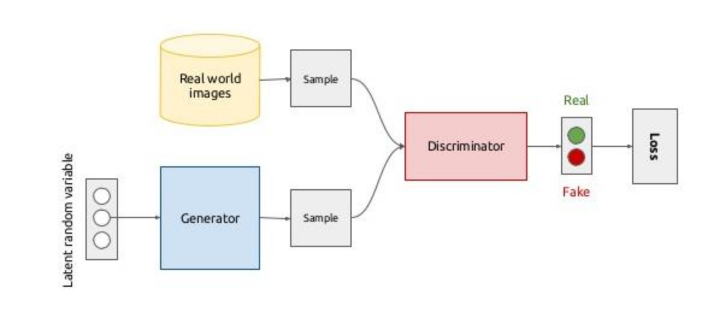
- Non-linear dimension reduction
- Regularization of latent space → Gaussian

but

- Sample latent space for new images, but no easy control
- Works fine … for low-res images only

Generative Adversarial Networks: a 2-player game

- Standard Backprop for Discriminator
- Inverted Backprop for Generator
- Difficult balance in practice

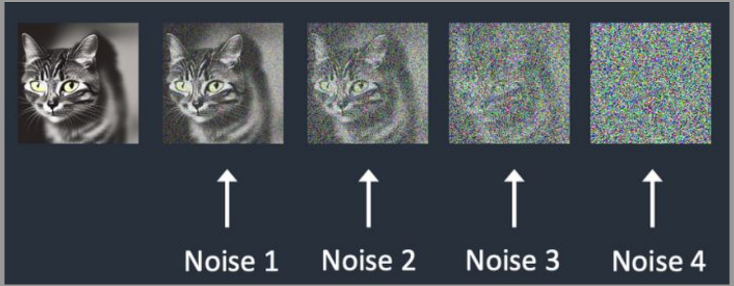


Goodfellow et al., 2014

Diffusion Models

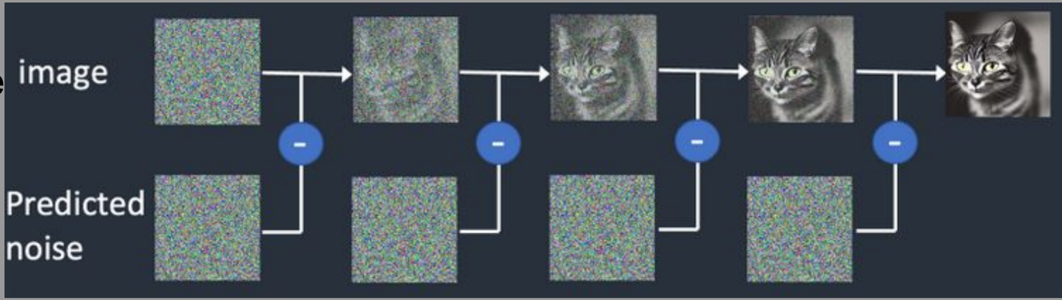
Forward diffusion

- Add gradual levels of noise to the images in the dataset
- Train a noise predictor to predict the noise added (supervised)
  - examples = all steps of diffusion
  - U-net architecture



Reverse diffusion

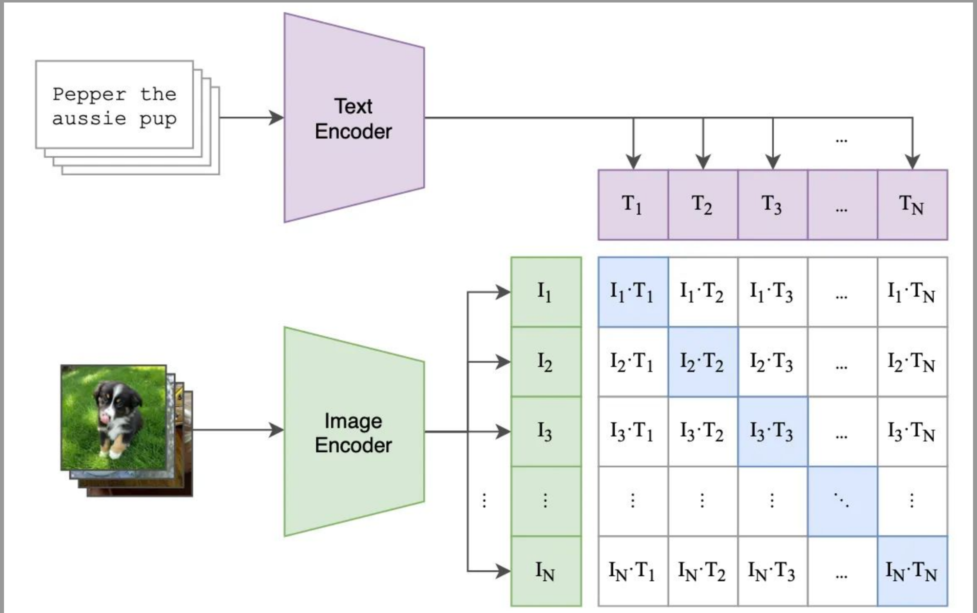
- Start from random image
- Iterate
  - Use noise predictor to predict noise
  - Remove from image
- Until convergence
- Result is similar (but different) from images in the original dataset
  - but undirected
- **Control**: through conditioning (e.g., with text/caption)



Sohl-Dickstein et al., 2015

Text-to-image: CLIP

Contrastive Language-Image Pre-Training

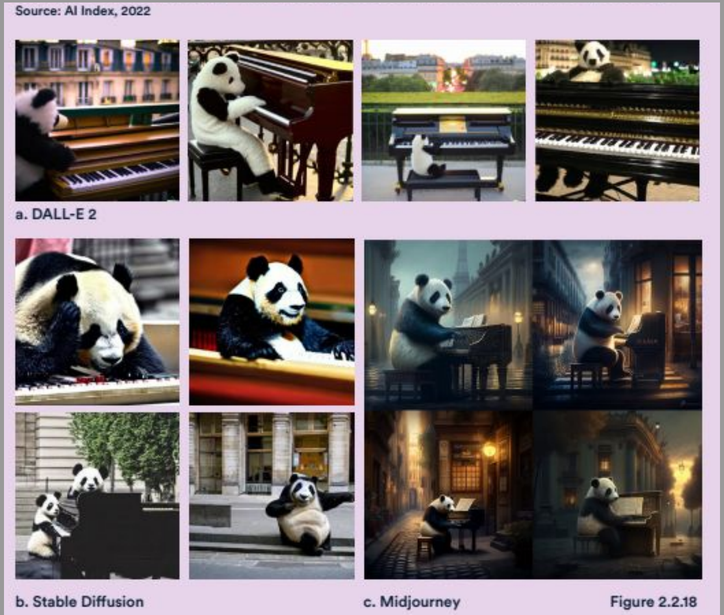


- Text and images encoded in the same latent space using a **contrastive loss**:
  - maximize cosine similarity for matching pairs, minimize it for others
- Trained on 400M (image, caption) pairs (30 days on 592 V100, ~1M$ AWS)

Sohn, 2016; OpenAI 2021

Text-to-image

- Open AI's *Dalle-e 2*
- Stability AI's *Stable Diffusion*
- Midjourney's *Midjourney*
- Meta's *Make-a-scene*
- Google's *Imagen*



A panda playing piano on a warm evening in Paris

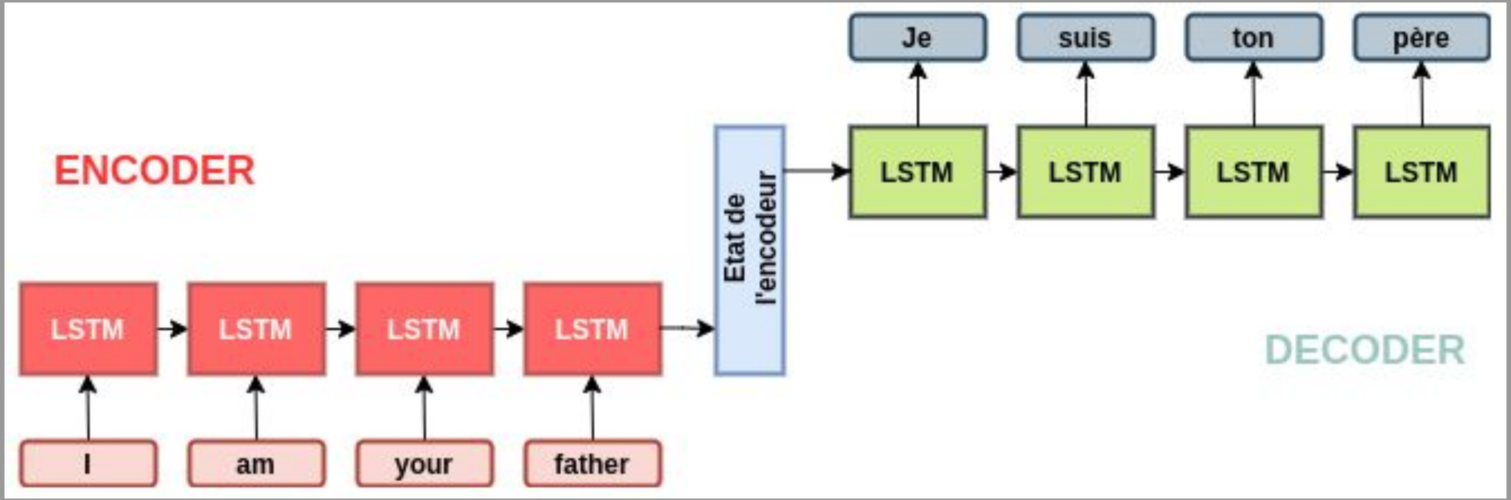
AI Index 2023 p90

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP, and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Early models: Sequence to Sequence

- **Word embeddings** (CBOW, GloVe, …)
- **Recurrent** neural networks (e.g., LSTM)
- Only the encoder’s internal state goes to the decoder



- A revolution in Machine Translation

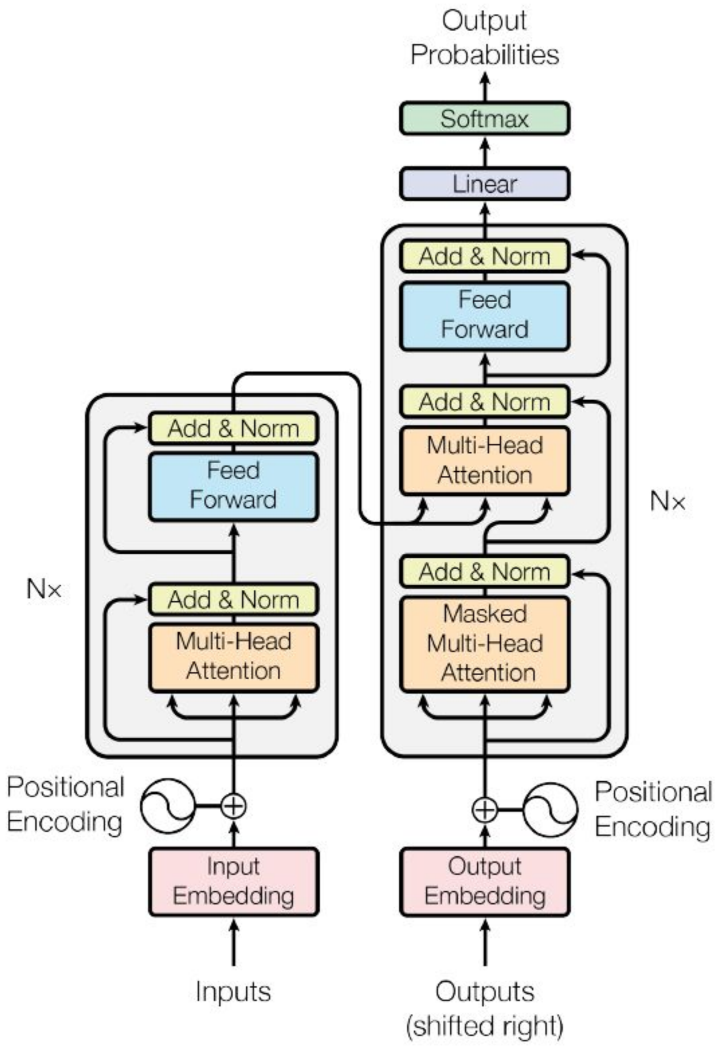
Sutskever et al., 2014

Attention is all you need

- **No recurrence**
  - Handles (chunk of) sequences
    - e.g., NLP data
  - based on (word) embeddings
  - with positional information
- **Attention** mechanisms handles links between any 2 positions of the sequence

- A revolution in Machine Translation

→ **Transformers**

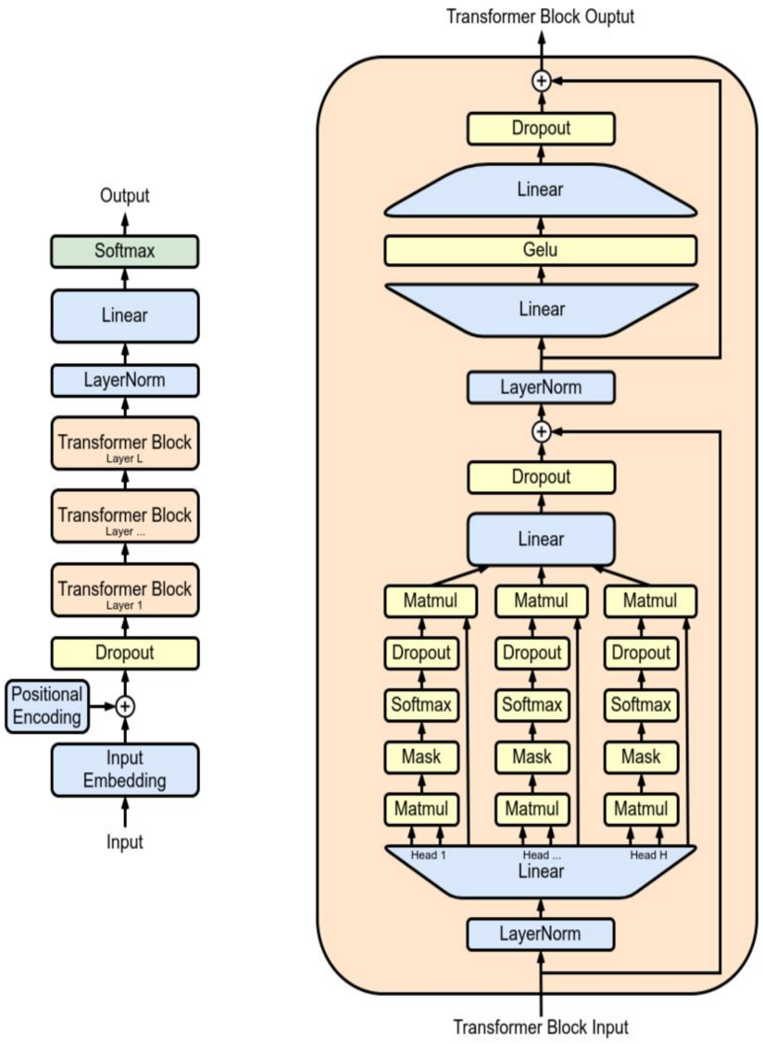


Vaswani et Google team, NeurIPS 2017 (>100k cites)

LLMs

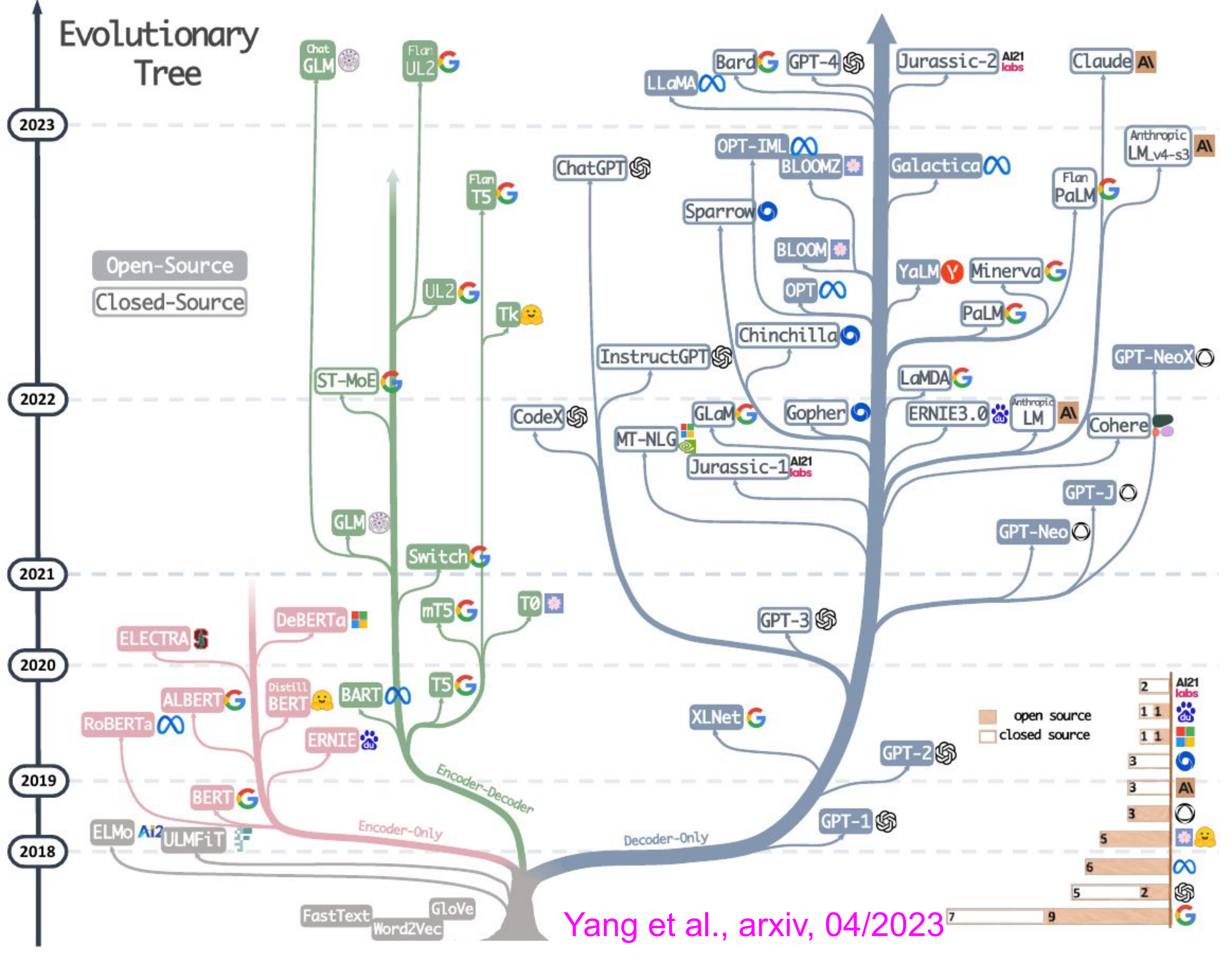
- Pre-trained transformers
- Trained **to predict next word**
  - among ~50000 tokens
  - with window size ~4000 (GPT4: ~32000)
  - on huge corpuses (400B tokens)
- At immense computational cost
- Need to be fine-tuned to specific tasks
- Huge models
  - **GPT3**: 175B weights (**GPT4** undisclosed)
  - **BloomZ**: same, but open
  - **PalM**: 540B weights
  - **Llama 2**: 7,13 or 70B, open

Also available for **image**, **music**, …

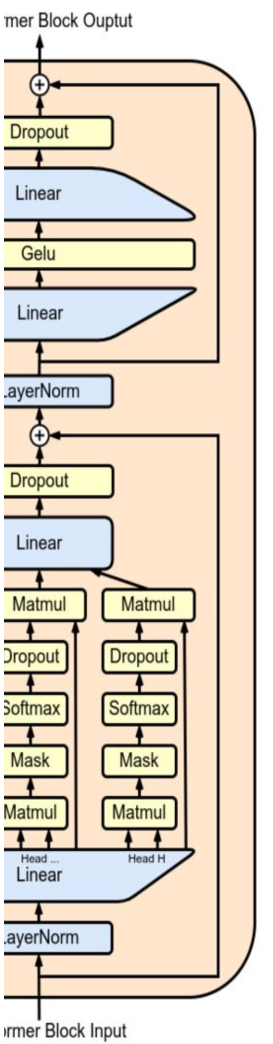


- Pre-tra
- Trainec
  - am
  - wit
  - on
- At imm
- Need t
- Huge r
  - GF
  - Bl
  - Pa
  - Lla

Also availa



Yang et al., arxiv, 04/2023



Chatbots

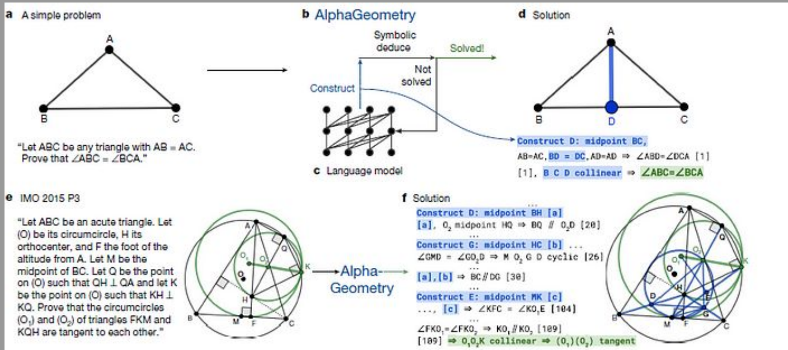
- **ChatGPT3** = GPT3.5 +
  - Fine-tuning with supervised learning (→ Instruct GPT)
  - Robustified with Reinforcement Learning from Human Feedback
    - Safety filters (against racist, sexist, negationist temptations)
  - Smart (hidden) pre-prompt
- Open interface: To date, **100M+ users**
  - with ranking possibilities → samples to improve robustness
- **ChatGPT4**, based on GPT4, is a closed system (by Open AI :-)
- **Llama 2-chat**, same, but open, with focus on safety

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Opportunities

- Easy & cheap **fine-tuning** (BloomZ: few hours of one A100 40Gb; Oracle, 03/2023)
  - Public services
  - Private confidential company knowledge
  - … RightWingGPT (5000 well-chosen new examples, < $300 on AWS)
- A productivity gain in many domains
  - All NLP systems (translation, summary, routine documents, …)
  - Code synthesis (e.g., MS co-pilot, Meta CodeLlama, …)
  - Robotics (perception, interaction, planning, control)
- Fabulous **exploration tools**
  e.g., **AlphaGeometry** (DeepMind):
  - an LLM generate hypotheses, that
  - a formal deduction software (in)validates



Trinh et al., Nature, 01/2024

Risks

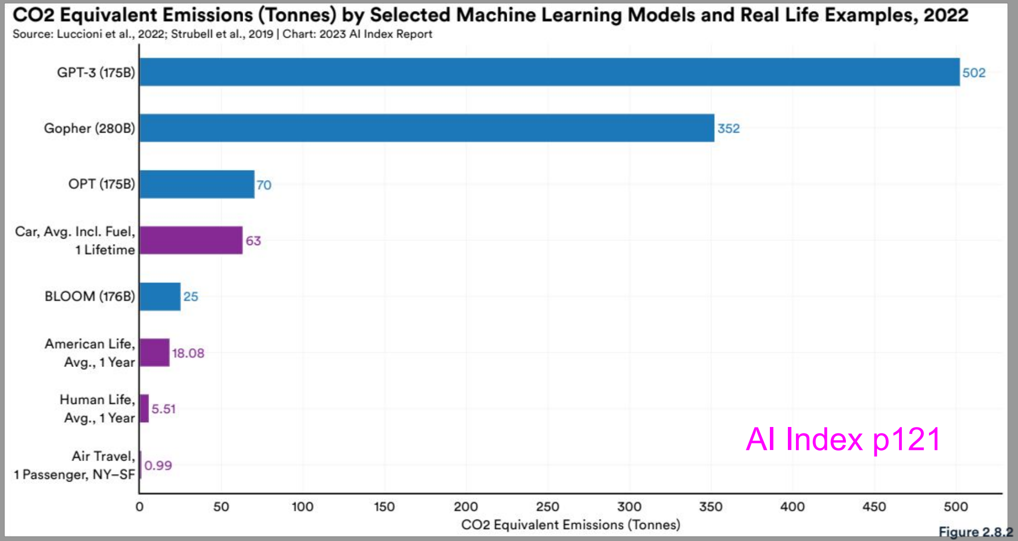
- Size matters
  - **Environmental** cost
  - Accessibility (even for inference)
  - Loss of sovereignty for Europe
    - In spite of, e.g. **BloomZ, OpenGPT-X, …**
- **Fake**/undesired information
  - Hallucinations, adversarial prompting
- Biases (gender, race, …) →
- **Stereotyped** outputs
- Societal risks on jobs (music composers, teachers, …)
- No identified training sources
  - No transparency (where does this information come from?)
  - Not GDPR compliant (copyrighted information was indeed used)

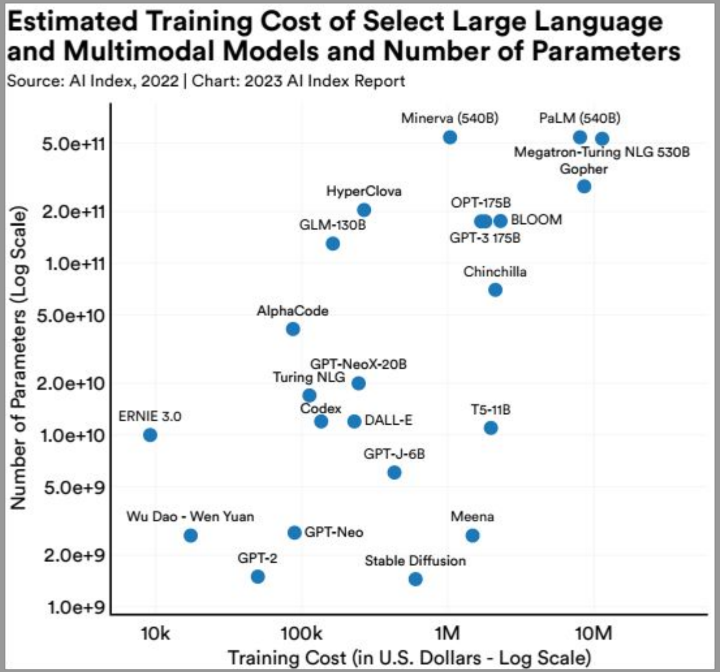
Assertive CEOs by SD

Need to **regulate** and **educate**

The Sustainability Paradox

- Not sustainable
- Irreproducible science



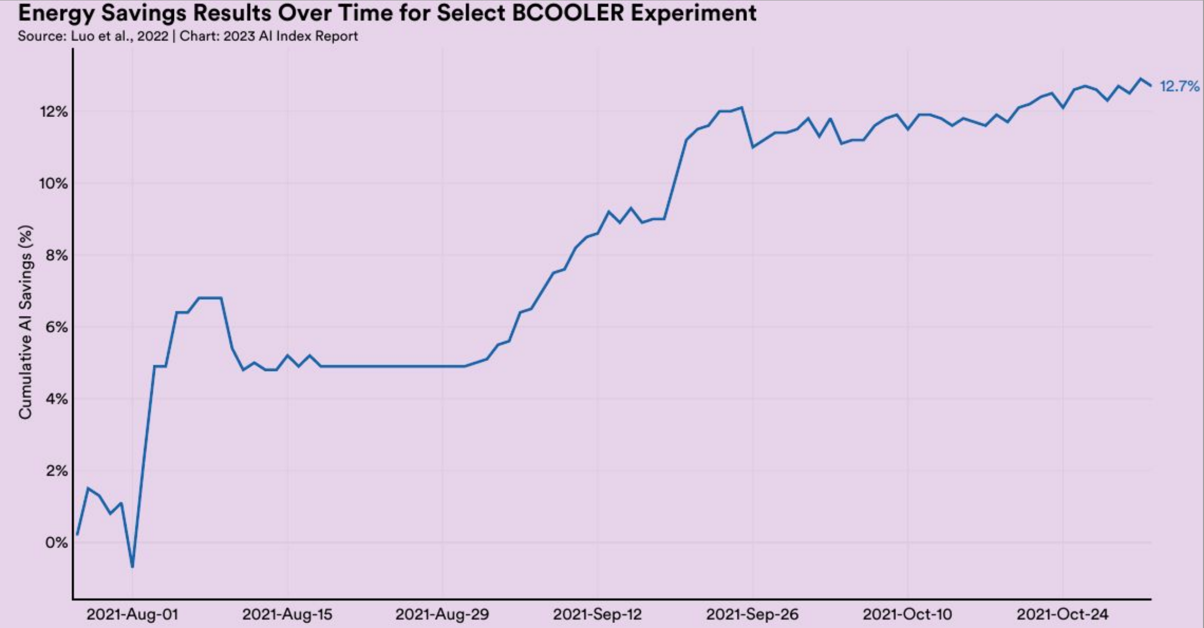


- Recent LLMs less greedy
- **Frugal learning** gaining momentum

e.g. Falcon**40b** (TII) Open LLM Leaderboard

The Sustainability Paradox (2)

DeepMind's BCOOLER: RL to optimize cooling of Google's data centers



AI Index p122

- Is there a sustainable trade-off?

Agenda

- Background and History
- Supervised Learning
  - Opportunities and Risks
- Generative AI for images, for NLP and beyond
  - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Back to History

Before 1956, some visions: Alan Turing, formal neurons, robots

AI as a mean

Can Machines Think?

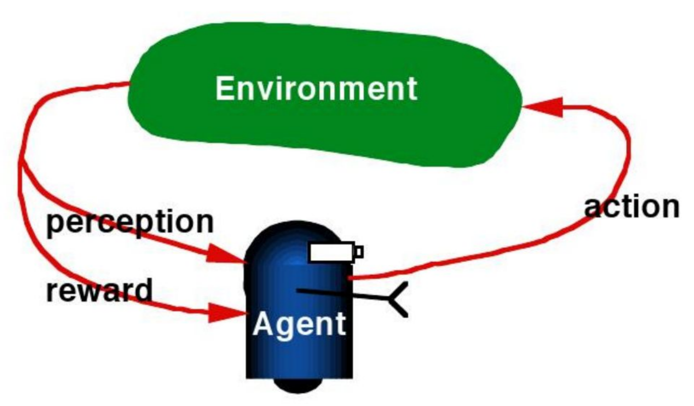
The problem is mainly one of programming. [...] brain estimates: 10^{10} to 10^{15} bits. [...] I can produce about a thousand digits of programme lines a day, so that about sixty workers, working steadily through the fifty years, might accomplish the job, if nothing went into the wastepaper basket. Some more expeditious method seems desirable.

How?

by (...) mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands.

One could carry through the organization of an intelligent machine with only two interfering inputs, one for pleasure or reward, and the other for pain or punishment.

Reinforcement Learning



- Agent maintains a state
- In a given state, it performs an action
- Action modifies the environment
- Agent receives a reward, perceives the new environment
- and updates its internal state accordingly

Goal: learn a policy [state → action] to maximize cumulated reward

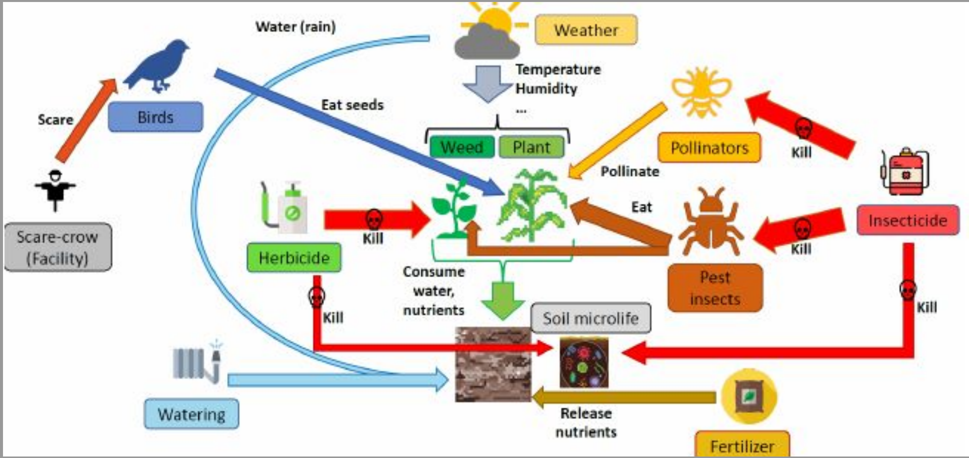
Popular approaches:

- (Deep) Q-learning
- Policy gradient methods, in particular Proximal Policy Optimization (**PPO**)
- Multi-armed bandits

RL for agriculture

Farm-Gym

- An RL platform for solving gamified agronomy challenges
- **Entities**: Plant, Weeds, Soil, Fertilizer, Cides, …
- highly **coupled** in **stochastic** interaction
- Ecosystem is a POMDP
- **Goal**: learn interventions to maximize the yield
- Hand-made expert agent outperforms PPO significantly



O. Maillard et al. *Artificial Intelligence for Agriculture and Food Systems*, 2023

RL for agriculture

- **Gym-DSSAT**, a gamified version of DSSAT, dynamic crop growth simulation models for over 42 crops
  - **Use case**: rainfed maize production in southern Mali (*)
  - Choice among preselected **nitrogen management practices**
  - Improves over multi-location multi-year field trials (pre-defined practices are tested in an equiproportional way during a fixed number of years)

(*) Gautron et al., *Field Crops Research*, 2024

- **WeG@rden**

Inria AEx by O-A Maillard

  - a collaborative platform for (market) gardening data collection
  - a will-be recommendation tool for gardners using RL

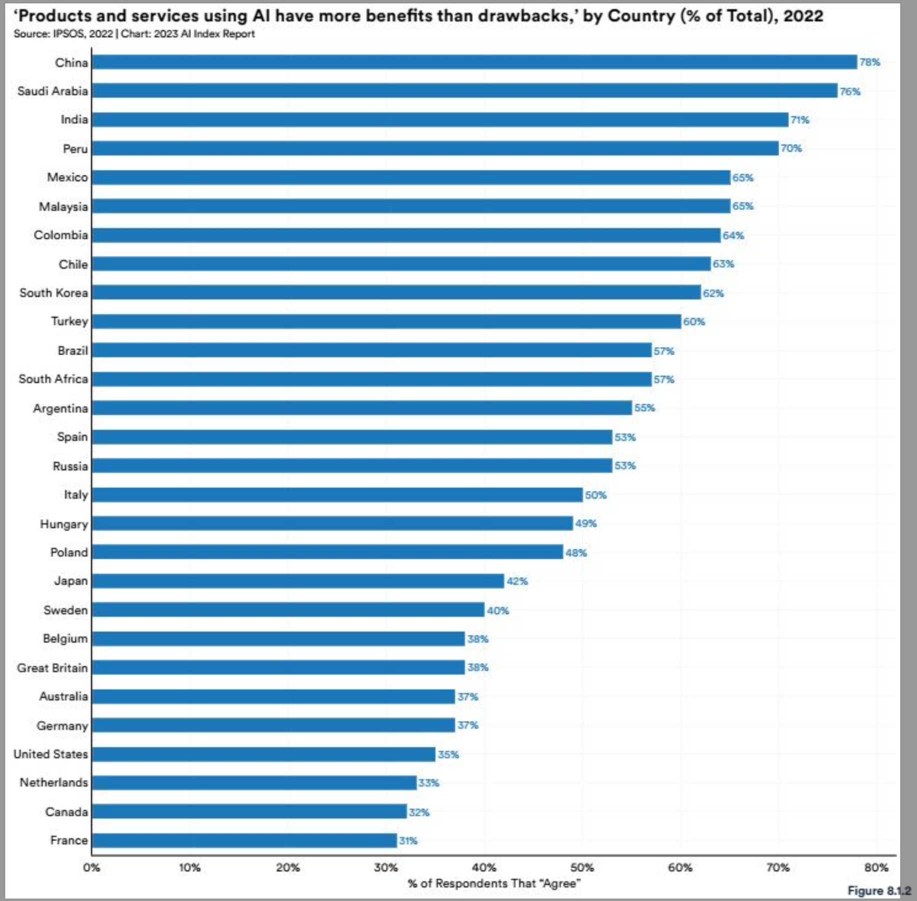
Agenda

- Background and History
- Supervised Learning
    - Opportunities and Risks
- Generative AI for images, for NLP and beyond
    - Opportunities and Risks
- Reinforcement Learning
- Societal Risks as Conclusion

Societal Acceptance

Public acceptance rates

- China 78%
- South Korea 62%
- Japan 42%
- France 31%



AI Index p324

Regulations toward Trustworthy AI

RGPD

- Consent on data
- Human decisions only - but …

Wall Street J. 15 June 23

- Data traceability still missing

AI Act

- Based on risk evaluation
- LLMs don’t exactly fit in

Ethics

- Public debate, CCNE-bis, ...

CERNA, COERLE, …

- Trust Labels
- A posteriori control
  - Citizens, independent institution (e.g., Inria Regalia)

Toward an AGI ?

AI as a goal

What is critically missing

- Intrinsic motivation — On-going (e.g., Pierre-Yves Oudeyer)
- Embodiment — On-going (e.g., Rolf Pfeifer)
- Common sense — can it be statistically acquired?
- Self-consciousness

Who profits from the crime?

J.-G. Ganascia

A **scarecrow**,
to hide the real dangers?

- Increased control of our activities
- Loss of sovereignty of individuals by lack of digital skills
- Loss of sovereignty of states w.r.t. tech giant companies

The new opium of the masses

We need a strong regulation, not a moratorium

A first use case for regulation?

Ban Autonomous Lethal Weapons

Un drone tueur russe aurait été aperçu en Ukraine

Par Vonintsoa
Mis à jour: Mars 2022
18 mars 2022, 10h20

L'apparition du drone tueur russe en Ukraine soulève les inquiétudes quant à l'implication de l'IA dans la guerre.

Les USA vont fournir aux Ukrainiens des « drones tueurs » Switchblade de nouvelle génération

Par Matthias Bertrand
Publié le jeudi 17 mars 2022 à 11:09 • *Il y a 11j*
4 min de lecture

"Not the right time", USA says at UN meeting (03/2023)

Wrap-up

Take home

Scientific Highlights

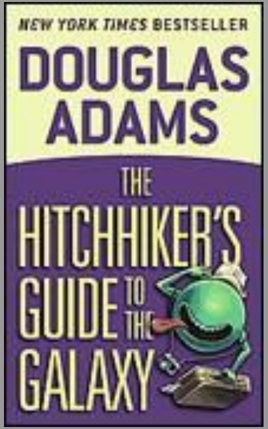
- Differentiable programming
- Latent representations
- Creative losses
- More and more complex hierarchical architectures
- An ever-growing zoology of (pre-trained) tools
- Next step: Smart Hybridization → use state-of-the-art ideas/tools!

Societal Issues

- Sustainability, safety, trustworthiness: Research and early education
- Regulations lagging behind technology
- Sovereignty vs private interests and foreign countries

Take home

42



Scientific Highlights

- Differentiable programming
- Latent representations
- Creative losses
- More and more complex hierarchical architectures
- An ever-growing zoology of (pre-trained) tools
- Next step: Smart Hybridization → use state-of-the-art ideas/tools!

Thank you

Societal Issues

- Sustainability, safety, trustworthiness: Research and early education
- Regulations lagging behind technology
- Sovereignty vs private interests and foreign countries

Adams 99