L'Intelligence Artificielle hier, aujourd'hui et demain

Marc Schoenauer
Equipe TAU, INRIA, LRI, UP-Sud et CNRS,
Université Paris-Saclay

29e journée CASCIMODOT
12 décembre 2018

Preamble

Artificial Intelligence is (Deep) Machine Learning

Preamble

Artificial Intelligence is (Deep) Machine Learning

FALSE

Preamble

Artificial Intelligence is (Deep) Machine Learning

although ...

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

AI: Past, Present, and Future

- Past
  - History*
  - DefinitionS
  - Recent successes*
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

(*) Many thanks to Bertrand Braunschweig

History

AI is a recent invention

History

AI is a recent invention

FALSE

History

Before 1956, some visions: Alan Turing, formal neurons, robots — AI as a mean

1956: Dartmouth workshop, first occurence of the term AI — AI as a goal

196x: *Problem solving*, games, natural langage

1968: 2001 a space odyssey, HAL

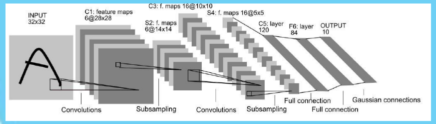
1969: *Perceptrons* (Minsky-Papert), kills research on NNs

1973: Lighthill Report, first AI Winter

198x: Prolog+FGCS; Experts Systems; Checkers (from Samuel to Chinook)

199x: Second AI Winter, but Deep Blue (chess) and first convolutional networks (CNNs)

2000: first Web applications (data)



2010: Deep learning (triumph of CNNs, AlphaGO, ...)

2018+: toward a third AI Winter?

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.

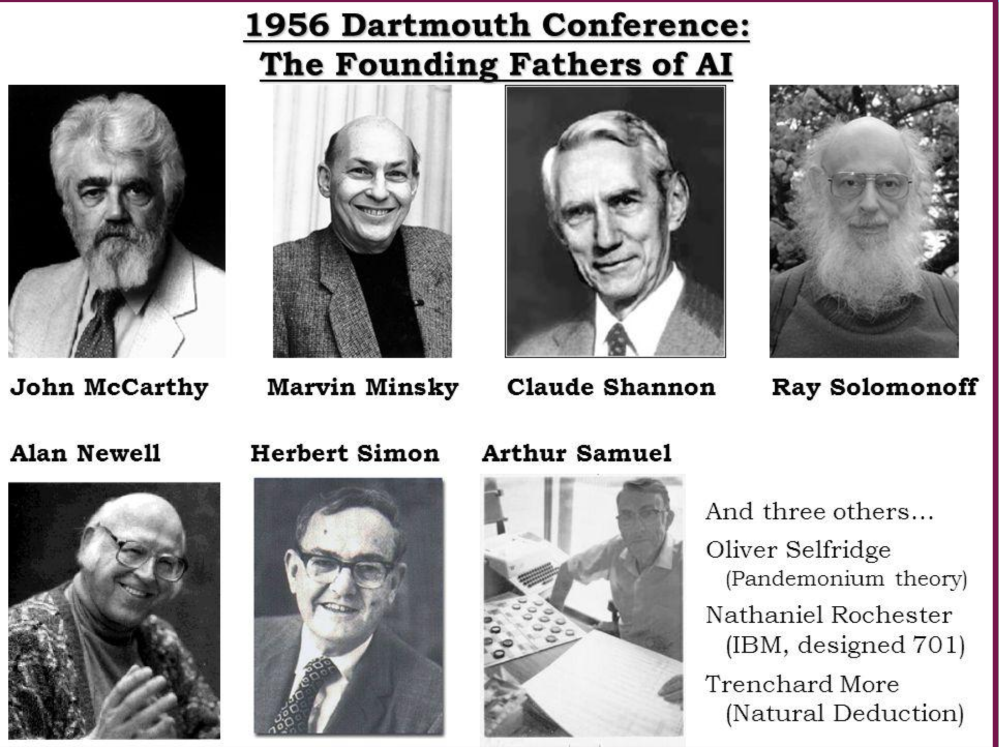
How?

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.

History

Before 1956, some visions: Alan Turing, formal neurons, robots — AI as a mean

1956: Dartmouth workshop, first occurence of the term AI — AI as a goal



We propose a study of artificial intelligence [..]. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

The vision : reasoning is a sequence of logical operations that a computer can reproduce

Goal : A *General Problem Solver*
(aka 2000+ : Artificial General Intelligence)

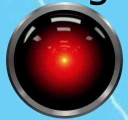
History

Before 1956, some visions: Alan Turing, formal neurons, robots — AI as a mean

1956: Dartmouth workshop, first occurence of the term AI — AI as a goal

196x: *Problem solving*, **games**, natural langage

1968: 2001 a space odyssey, HAL



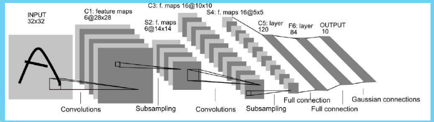
1969: *Perceptrons* (Minsky-Papert), kills research on NNs

1973: Lighthill Report, first AI Winter

198x: Prolog+FGCS; Experts Systems; **Checkers** (from Samuel to Chinook)

199x: Second AI Winter, but Deep Blue (**chess**) and first convolutional networks (CNNs)

2000: first Web applications (data)



2010: Deep learning (triumph of CNNs, Alpha**GO**, ...)

2018+: toward a third AI Winter?

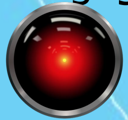
History

Before 1956, some visions: Alan Turing, formal neurons, robots — AI as a mean

1956: Dartmouth workshop, first occurence of the term AI — AI as a goal

196x: *Problem solving*, games, natural langage

1968: 2001 a space odyssey, HAL



1969: *Perceptrons* (Minsky-Papert), **kills research** on NNs

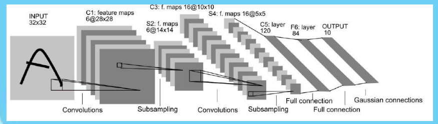
1973: Lighthill Report, **first AI Winter**

198x: Prolog+FGCS; Experts Systems; Checkers (from Samuel to Chinook)

199x: **Second AI Winter**, but Deep Blue (chess) and first convolutional networks (CNNs)

2000: first Web applications (data)

2010: Deep learning (triumph of CNNs, AlphaGO, ...)



2018+: toward a third AI Winter?

AI: Past, Present, and Future

- Past
  - History*
  - DefinitionS
  - Recent successes*
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

(*) Many thanks to Bertrand Braunschweig

Definition ?

Have machines that accomplish tasks related to (human) intelligence - possibly better than humans.

Definition ?

Have machines that accomplish tasks related to (human) intelligence - possibly better than humans.

BOF …

Definition ?

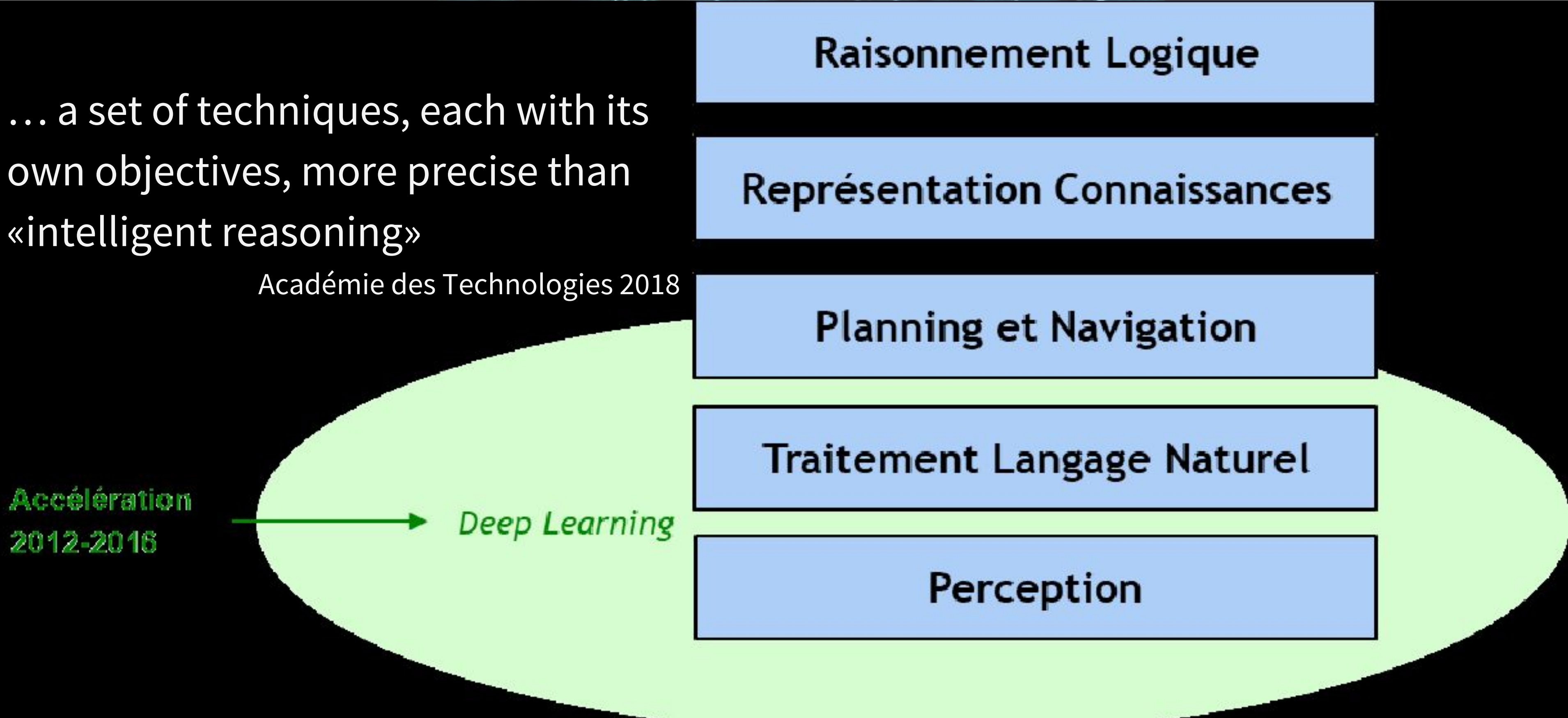
Have machines that accomplish tasks no machine ever did

- Jean-Louis Laurière, 80s
- Philippe Kahn, late 80s
- Gérard Sabah, 2017
  (rapport de l’OPECST)

Definition ?

… a set of techniques, each with its own objectives, more precise than «intelligent reasoning»

Académie des Technologies 2018



AI: Past, Present, and Future

- Past
  - History*
  - DefinitionS
  - Recent successes*
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

(*) Many thanks to Bertrand Braunschweig

Autonomy and Robotics

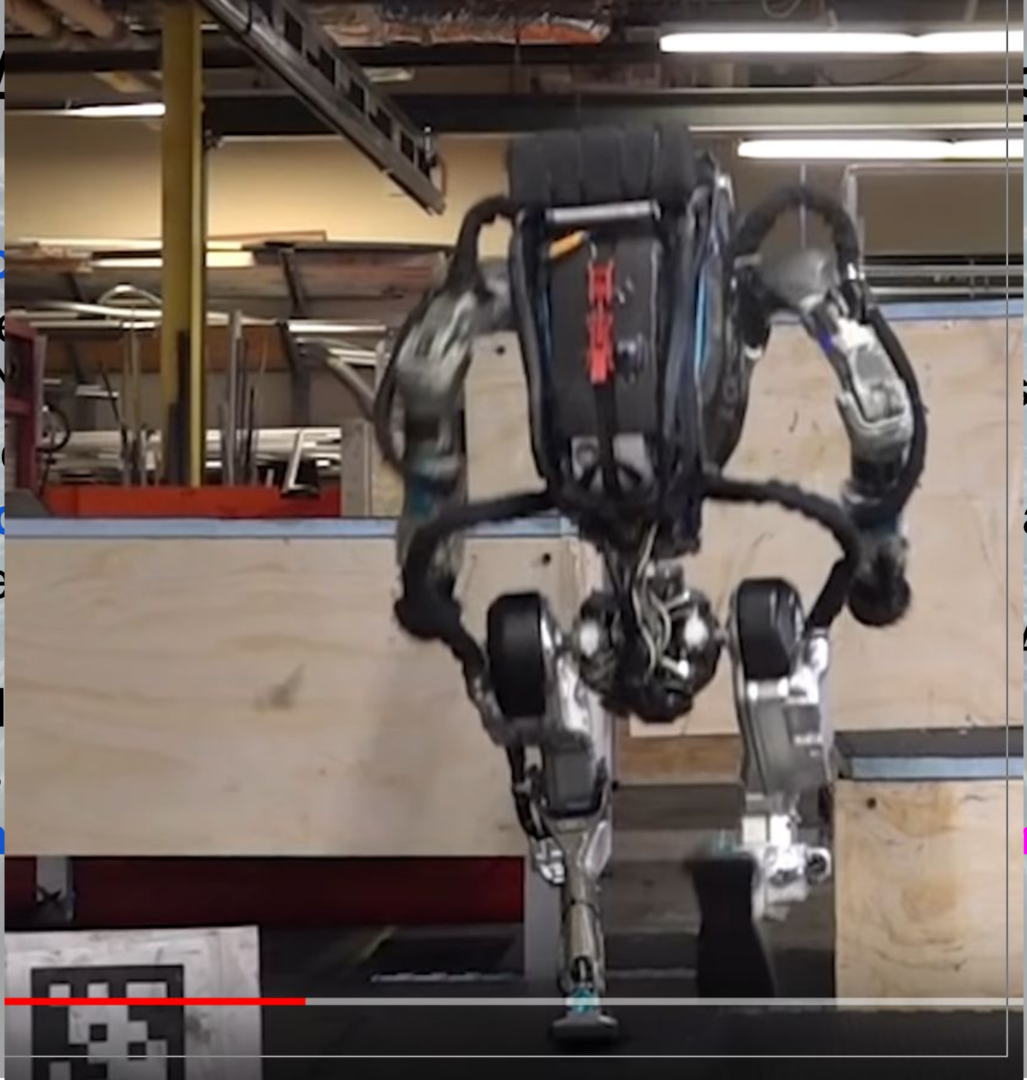
- *DARPA Autonomous Vehicle Challenge 2004-07
  - in the désert, then in urban context
- LeNet (Deep Neural Network) outperforms all challengers from Computer Vision in image recognition 2012-
- *DARPA Rescue Challenge robots who drive, walk in chaotic context, climb stairs, repair broken machines, etc 2015
- **Psibernetix shoots down (in simulation :-) the best US Air Force pilots
  - genetic algorithms and fuzzy logic ... on a Raspberry Pi! 2015
- Intel bought Israeli company MobilEye for 15 billions 2017
- **Boston Dynamics robots better and better performing 1998-

A s

- *DARPA Auto
  - in the dése
- LeNet (Deep N s from Computer Vision in image
- *DARPA Resc aotic context, climb stairs, re
- **Psibernetix Air Force pilots
  - genetic alg Pi! 2015
- Intel bought Isr
- **Boston Dyna 1998-

A s

- *DARPA Auto
  - in the dése
- LeNet (Deep N … from Computer Vision in image
- *DARPA Resc … aotic context, climb stairs, re
- **Psibernetix … Air Force pilots
  - genetic al … Pi! 2015
- Intel bought Is
- **Boston Dyn … 1998-



Games

- ****IBM Watson** beats best human players at Jeopardy 2011
  - NLP + web search + evaluation, 3 seconds on HPC
- **Deepmind** human performances on some (not all) Atari video games with Deep Reinforcement Learning 2013
  - Input: pixels; Output: joystick
- **Deepmind AlphaGo** beats World Champion of GO with a mix of Supervised and Reinforcement Learning 2016-17
- **Deepmind AlphaZéro** beats AlphaGo 100-0 using only Deep Reinforcement Learning and self-plays 2018
  - about 2 stones ahead of best human
  - AlphaZero can also be trained for other games (e.g., chess)
- ***Libratus** crushes the best Poker players of the world 2017
  - Reinforcement Learning and Bayesian techniques

NLP and and disability support

- **Microsoft Skype Translator** translates several languages in real time with Deep Learning. Similar performances for **Google Translate**, **Pilot**, ...
- **Apple Siri, Microsoft Cortana, Amazon Alexa** personal assistants use Speech Recognition and (some) Automated Reasoning
- ****Google Knowledge Graph** uses semantics to better structure the results of queries
- **Microsoft** translates from Chinese to English as good as human translators
    - with a double Deep Neural Network
- **Ava, RogerVoice** help deafs and hearing-impeached (subtitling, telephone,...)
- **Facebook** can label photos, and describe them to blind people

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

FALSE

Machine Learning

Artificial Intelligence is (Deep) Machine Learning

although ...

What has changed :

- Data Deluge — avent of WWW
- Moore law — or continuation
- New algorithms — or better understanding of old ones

Machine Learning

Learning from examples — recognition tasks

- Supervised — all examples are labelled
- Semi-supervised — some examples are labelled
- Unsupervised — no example is labelled

Reinforcement Learning — sequential decision making

Machine Learning

Learning from examples — recognition tasks

- Supervised — all examples are labelled
- Semi-supervised — some examples are labelled
- Unsupervised — no example is labelled

Reinforcement Learning — sequential decision making

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present
  - Supervised Learning
  - Deep Neural Networks
  - Generative Adversarial Networks
- Future: Toward Trustable Good AI?
- Conclusions

and other goodies

Supervised Learning

A toy case-study

- One example = (x_1, x_2, y), where **y** is the **label** (red or blue here)
- **Goal**: find a **model** $f(x_1, x_2)$ that separates the labels

Supervised Learning

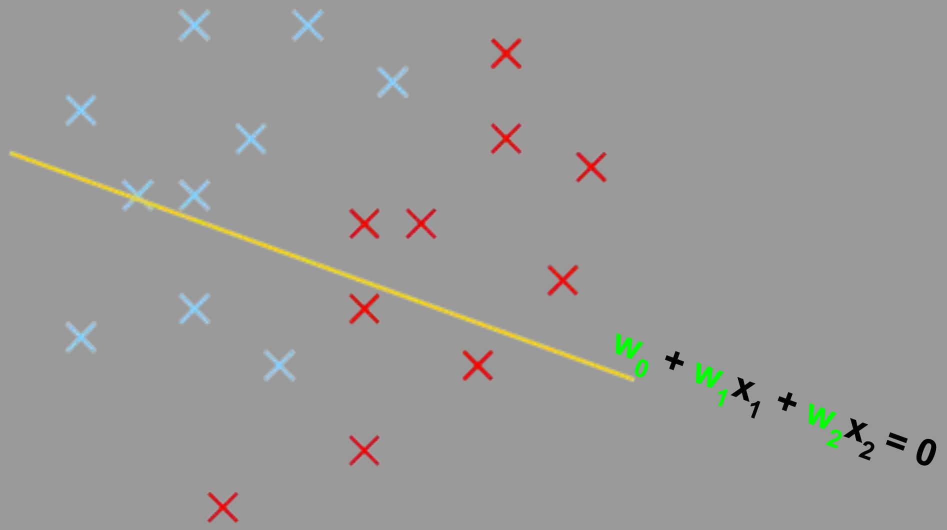
A toy case-study

- One example = (x_1,x_2, y),   where **y** is the **label** (red or blue here)
- **Goal**: find a **model** $f(x_1,x_2)$ that separates the labels
- allowing to correctly label future unlabelled example from (x_1,x_2)

Supervised Learning

A toy case-study

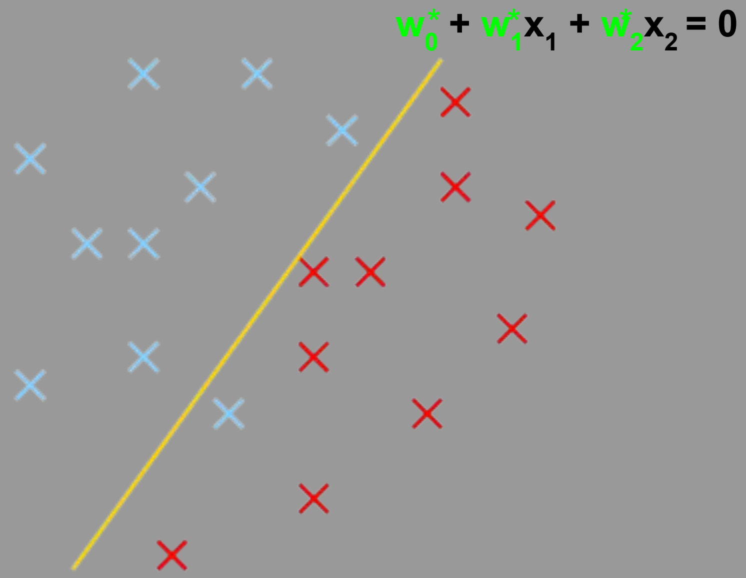
- For instance, linear models are defined by 3 parameters



Supervised Learning

A toy case-study

- For instance, linear models are defined by 3 parameters
- And we look for the parameters that best separate the data
- This is the **learning phase**



Supervised Learning

A toy case-study

- More complex models have more parameters

Supervised Learning

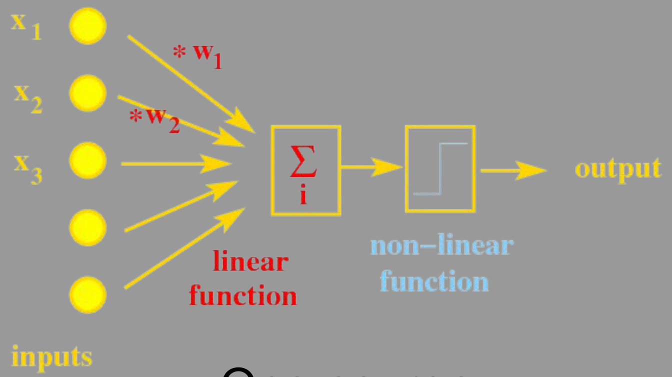
A toy case-study

- More complex models have more parameters
- with the danger of **overfitting**

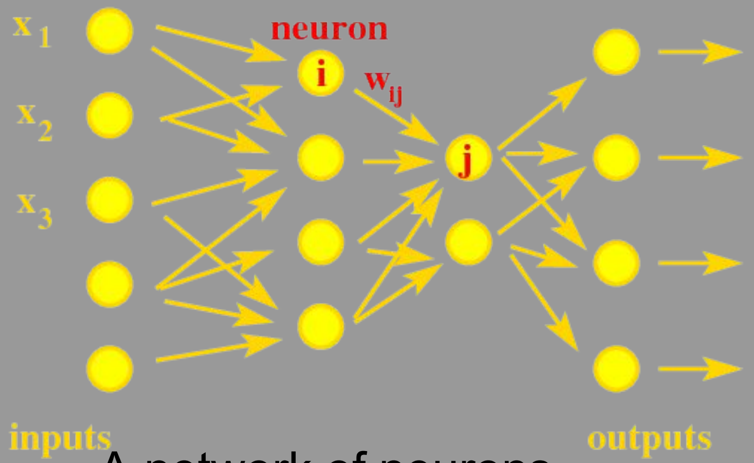
Supervised Learning

A zoology of models

- Polynoms
- Bayésiens Networks
- Decision trees and Random Forrests
- Support Vector Machine (kernel machines)
- **Artificial Neural Networks**



One neuron



A network of neurons

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

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present
  - Supervised Learning
  - Deep Neural Networks
  - Generative Adversarial Networks   and other goodies
- Future: Toward Trustable Good AI?
- Conclusions

Deep Neural Networks

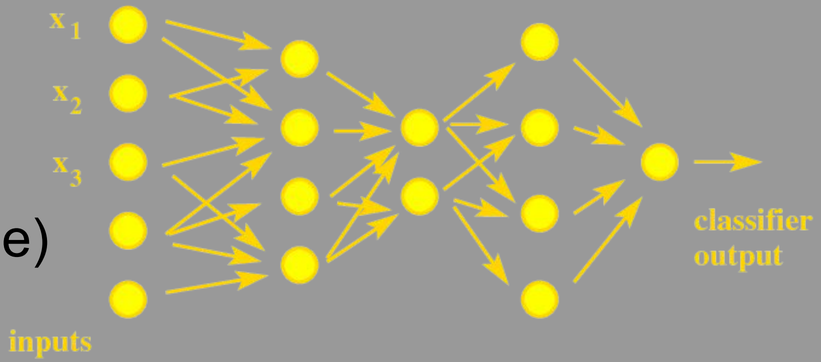
Learning Phase

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., $L = \sum |y(x_1,x_2) - NN(x_1,x_2)|^2$
- **Backward** pass: Compute $\nabla_w L$ (chain rule)
- Modify the weights w_{ij} from $\nabla_w L$

  to decrease of the loss
- Loop

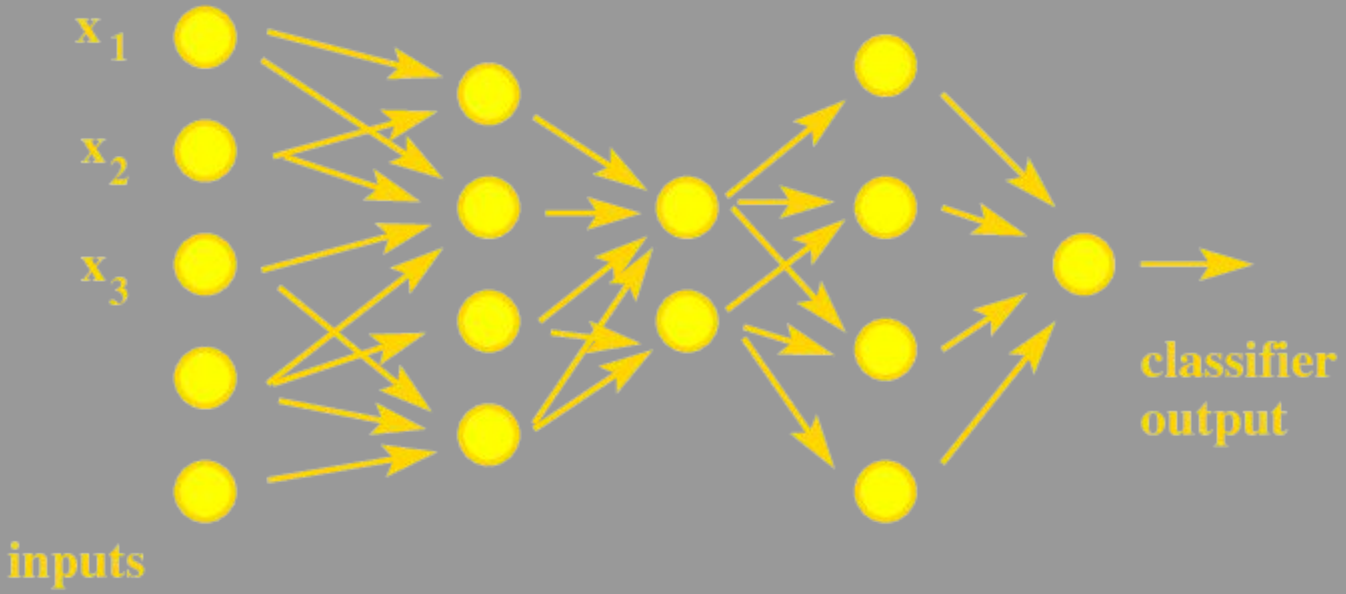


Recognition Phase aka **Inference**

Present an unlabelled example, the output of the network is the predicted label

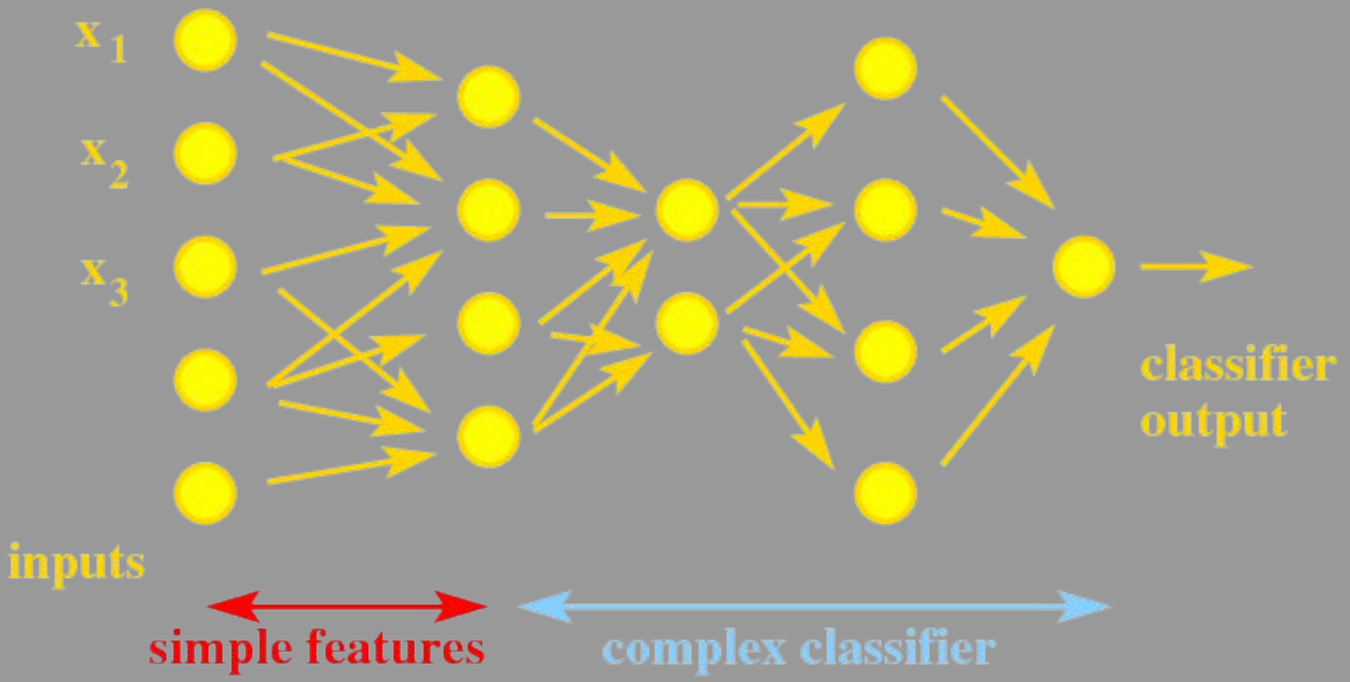
Deep Neural Networks

A Deep (layered) Neural Network is
a sequence of **representations of the data**



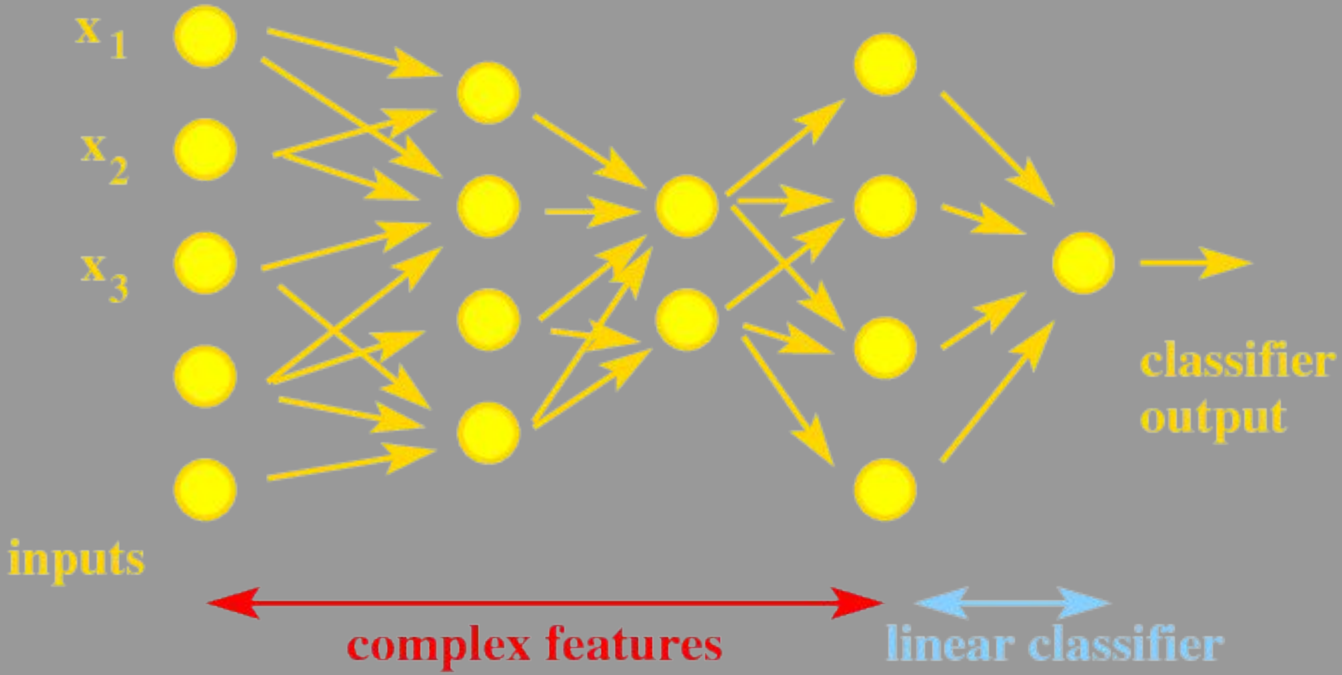
Deep Neural Networks

A Deep (layered) Neural Network is
a sequence of **representations of the data**



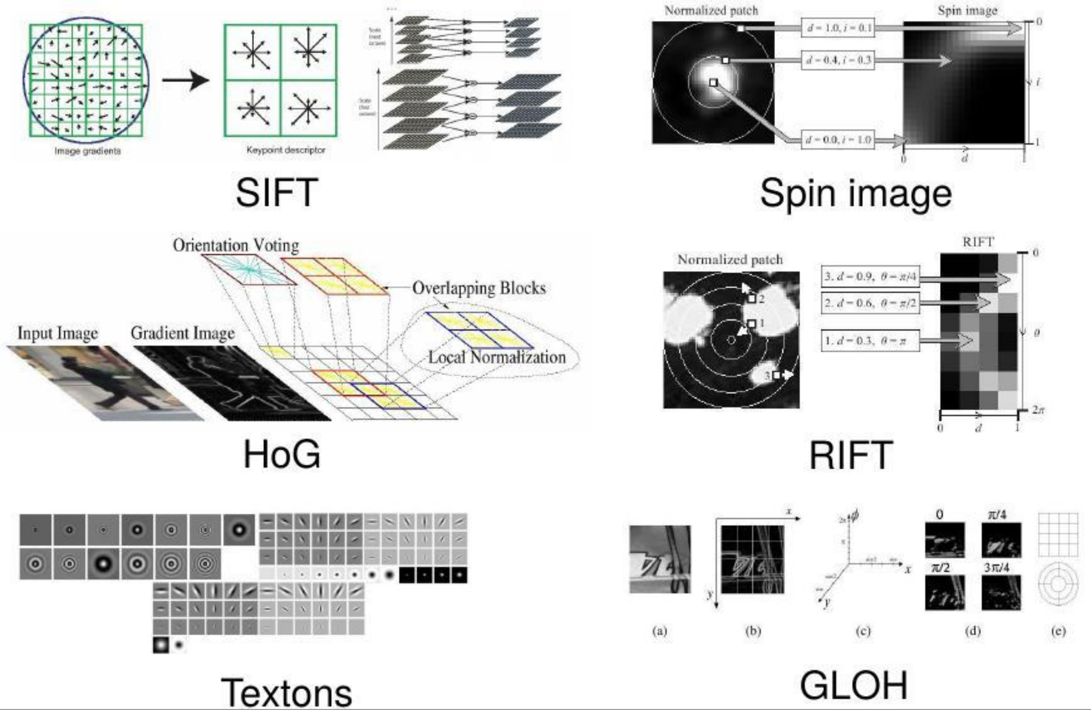
Deep Neural Networks

A Deep (layered) Neural Network is
a sequence of **representations of the data**



Good Old Computer Vision

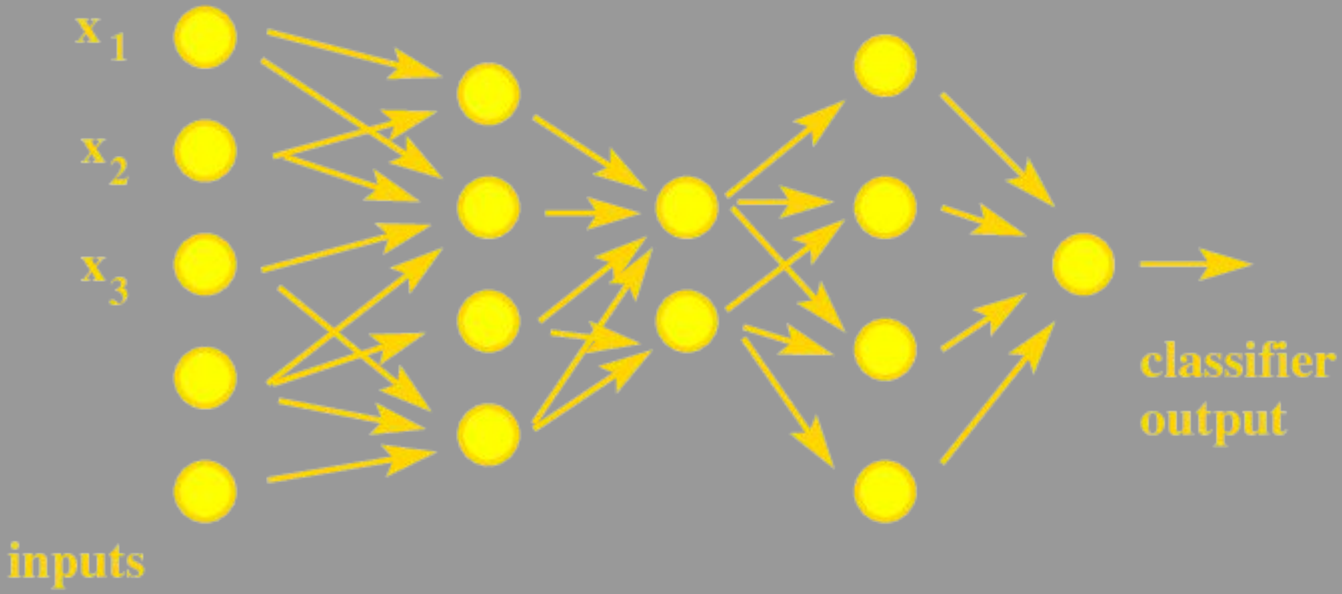
Hand-made **features** → Learned Classifier → CAT



End-to-end Learning

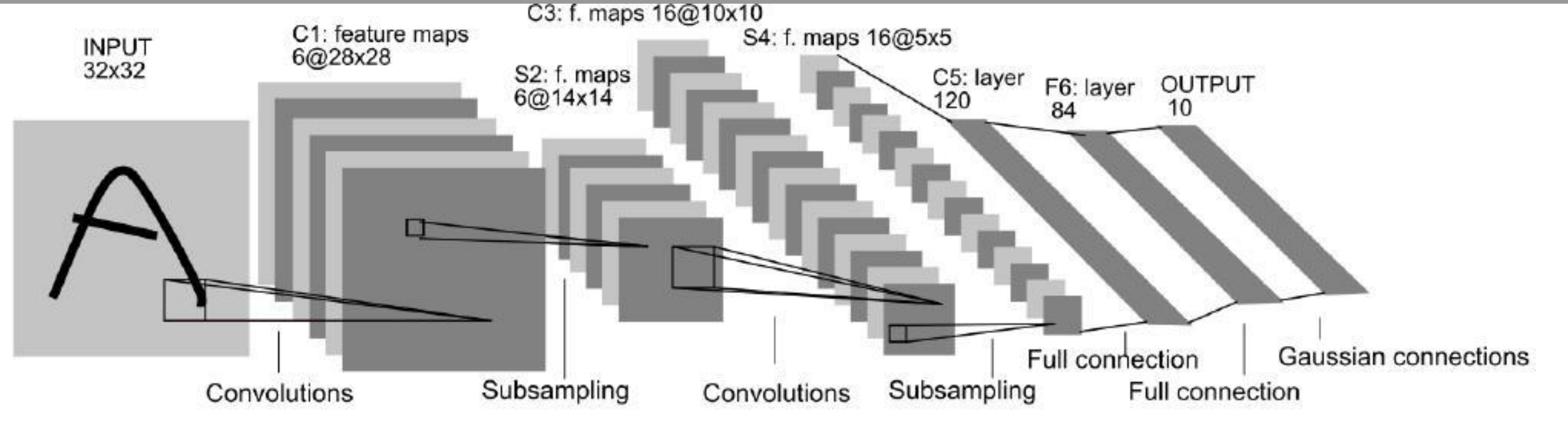
Features and Classifier are learned together

CAT



Convolutional Networks

Features and Classifier are learned together → CHAT

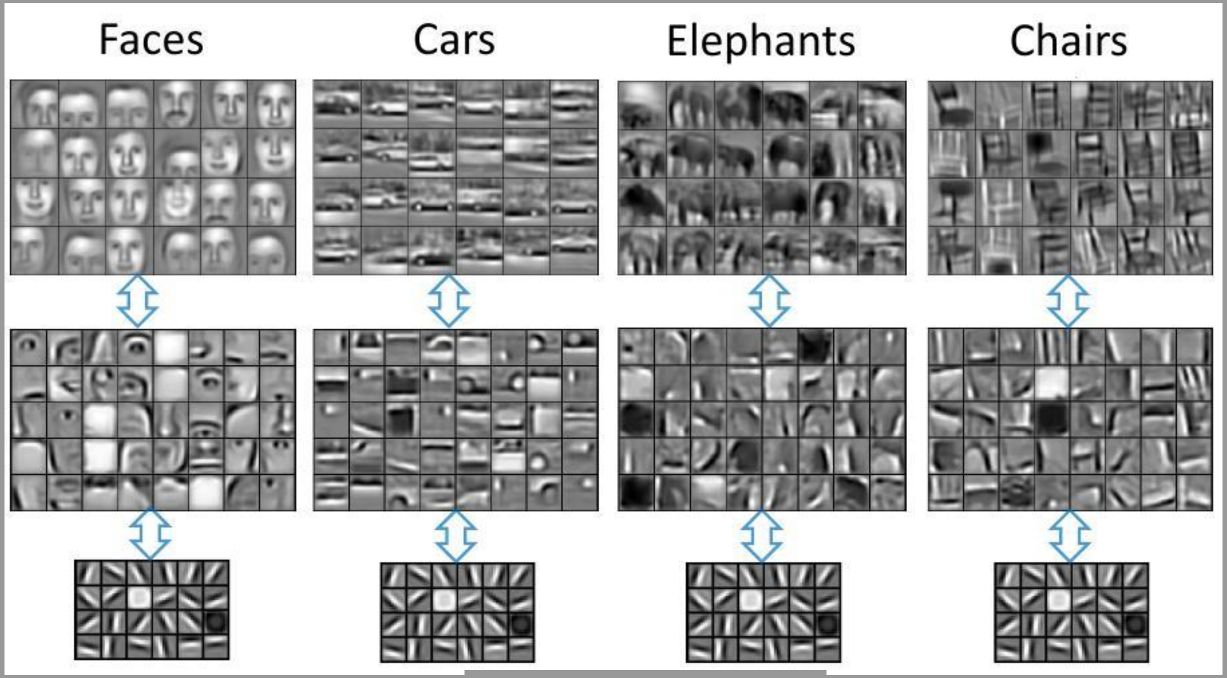


LeNet, LeCun et al., 1998

Convolutional Networks

Features and Classifier are learned together

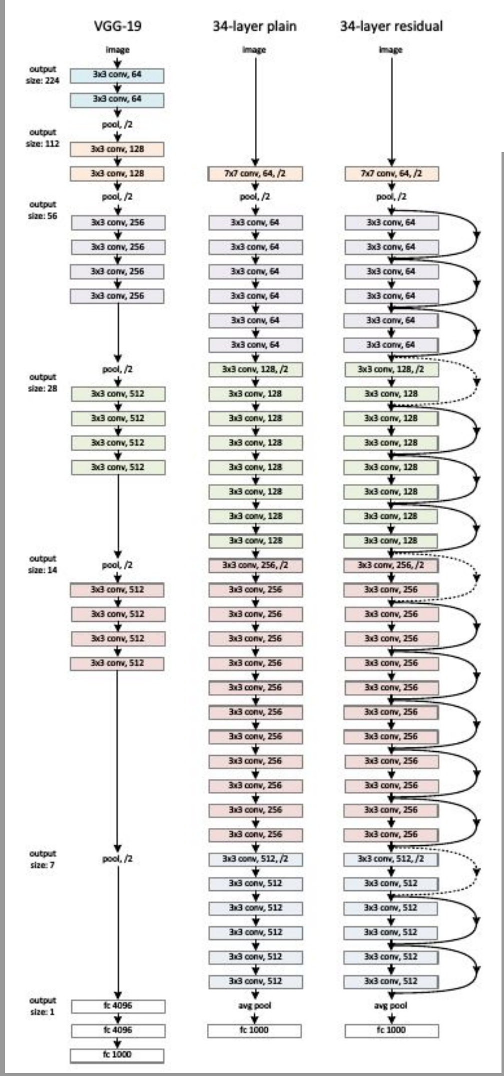
CHAT



Learned Features

State-of-the-art

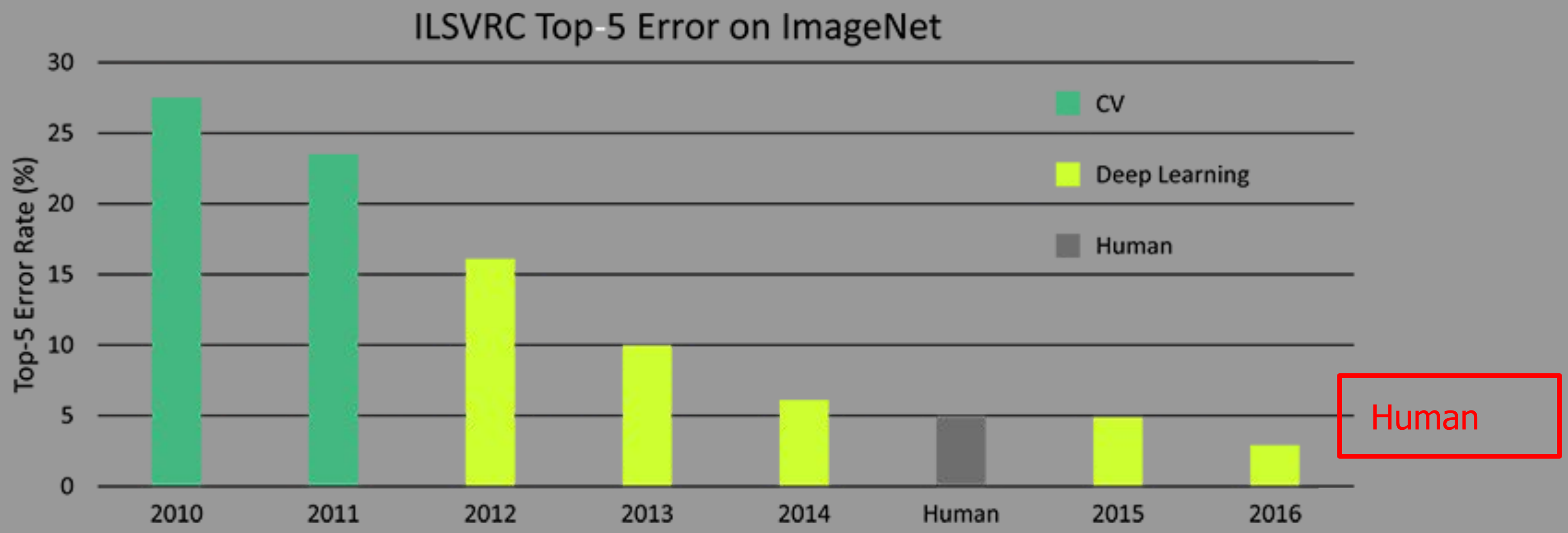
- Many datasets available
  - ImageNet : 14+ M examples, 1000 classes
- (pre-trained) networks with numerous layers
  - up to 152 !
- Millions to billions weights
  - hundreds of GPU mandatory for learning
- Several tricks of the trade
  - Dropout, residual layers, ensembles, ...



He et al., 2015

Deep Learning

Better than human learning

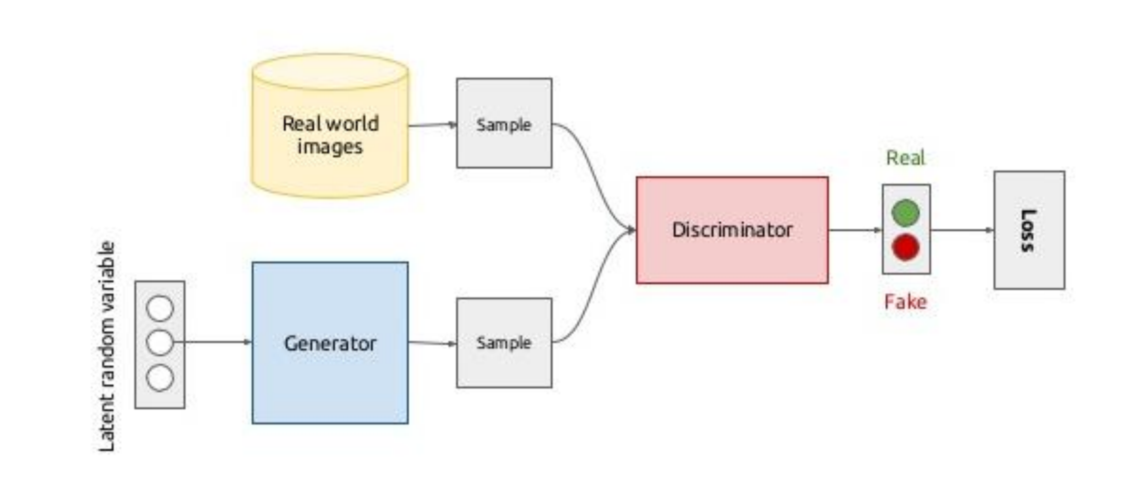


AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present
  - Supervised Learning
  - Deep Neural Networks
  - Generative Adversarial Networks   and other goodies
- Future: Toward Trustable Good AI?
- Conclusions

Generative Adversarial Networks

The most promising advance in DNN in the last 10 years (LeCun 2016)



A two-players game:

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

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Goodfellow et al., 2014

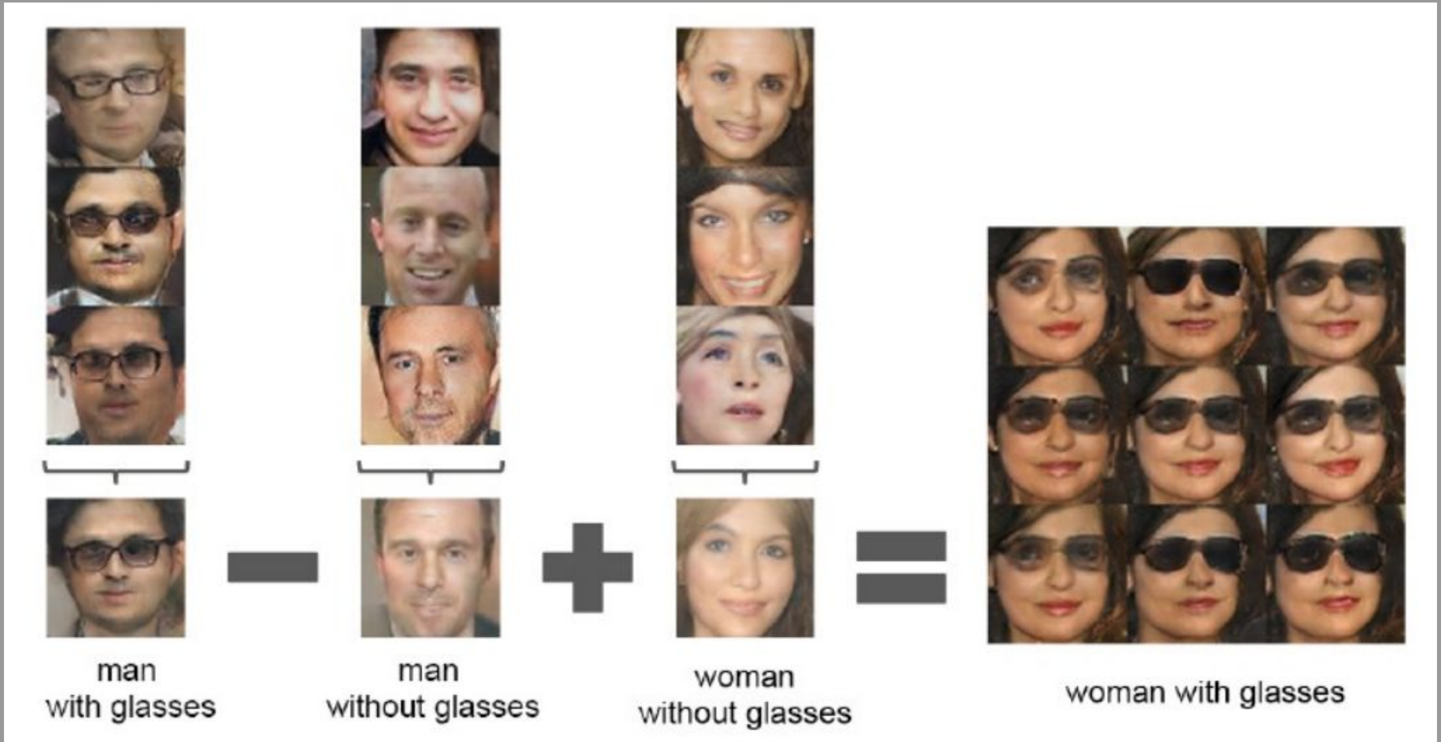
GANs

Image Generation

Goodfellows et al., 2016

GANs

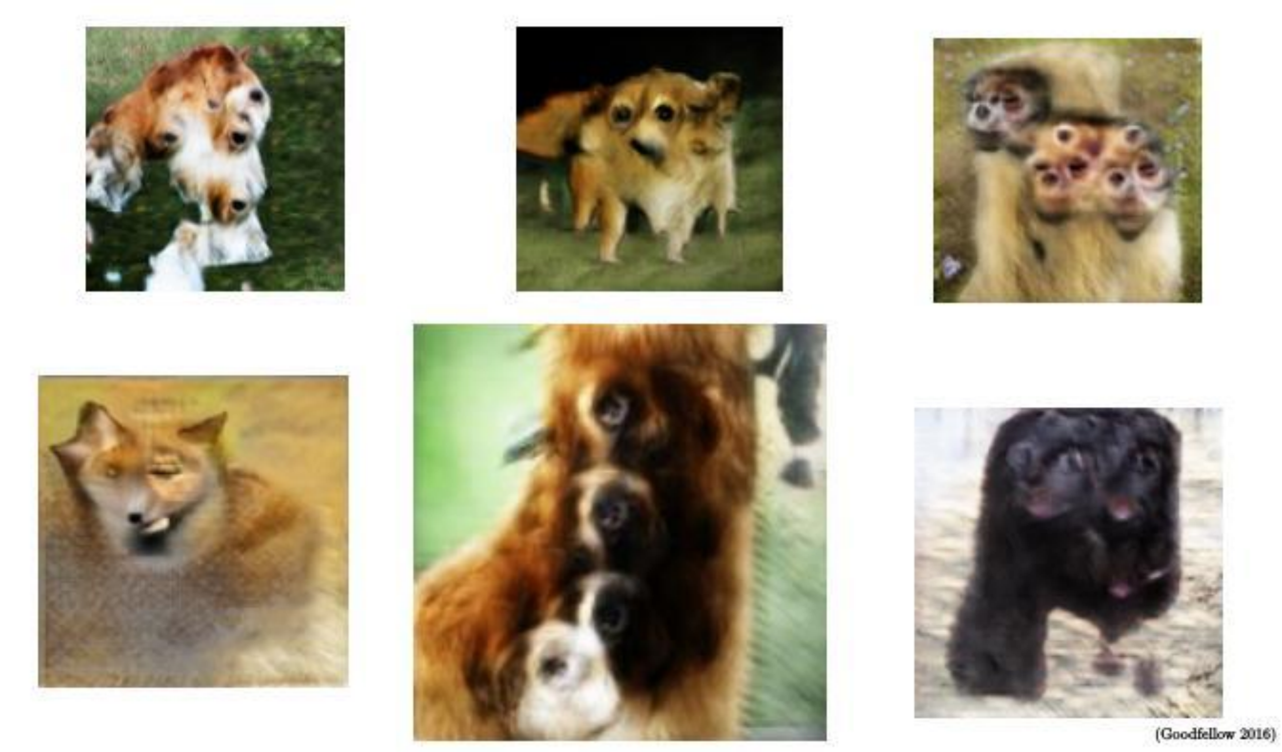
Image "Arithmetic"



Radford et al., 2015

GANs

Mode Collapse :-(

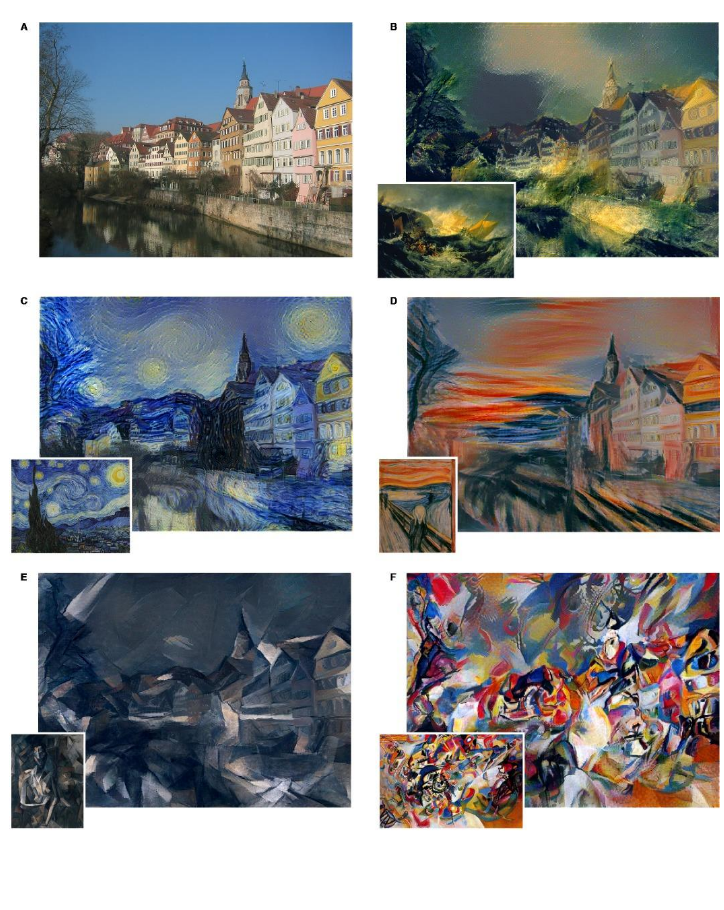


Goodfellows et al., 2016

Deep Supervised Learning

- Outstanding performances
  … in well-defined domains
  - Image recognition
  - Action identification in videos
  - Image captioning
  - Natural Language Processing
  - Automatic translation/subtitling
- Generative models (GANs)
  - lots of (fun) applications
    - e.g., style transfer →
- **Latent representations**
- **Differentiable programming**

But ...



AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

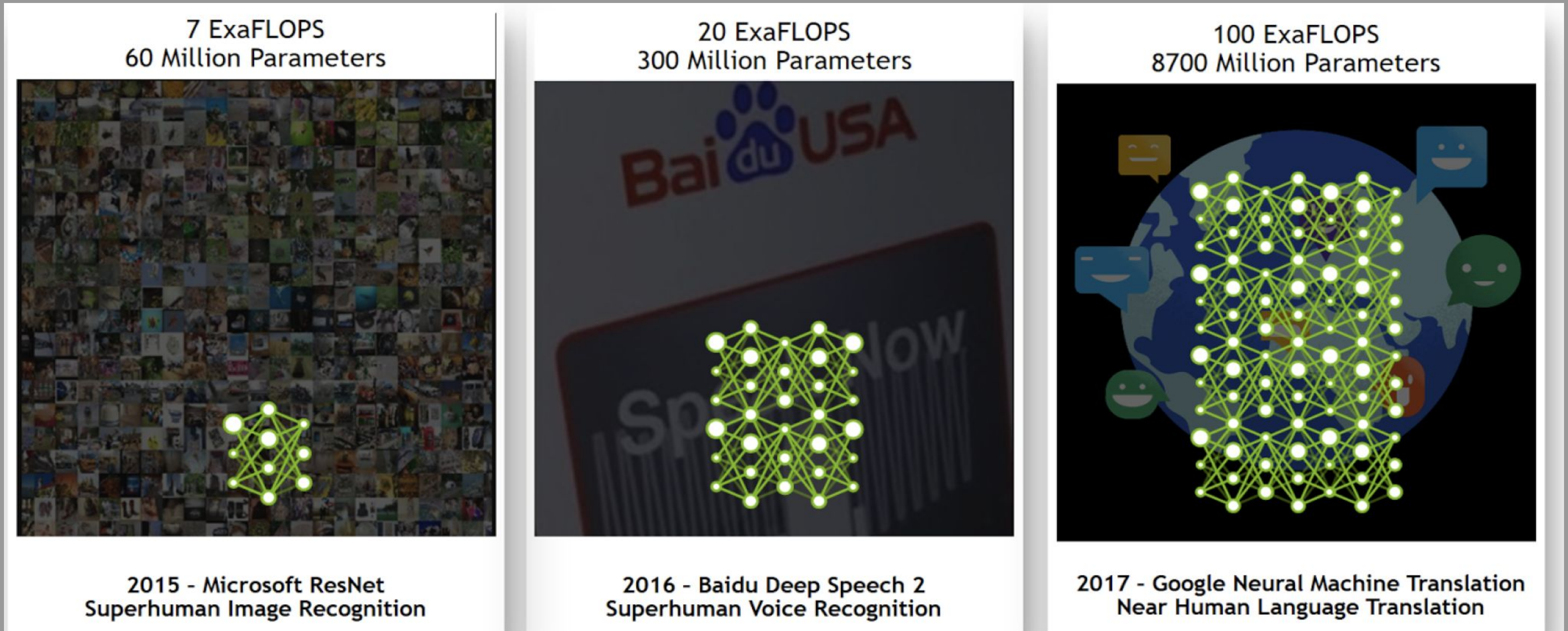
AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
  - Big Data and Irreproducible Science
  - Validation and Certification
  - Explainability and Causality
  - Fairness and Transparency
- Conclusions

Huge computational cost

Loads of data, Tons of weights

- Ecological disaster
- Irreproducible results



Meta-cost

Hyperparameters tuning

- Loss function — MSE, Cross-Entropy, ...
- Network topology — # layers, # neurons, convolutional, residual, ...
- Activation function — logistic, tanh, ReLU, ...
- Batch size
- Optimizer — SGD, w. momentum, Nesterov, Adagrad, Adam, ...
  - and its parameters (e.g., learning rate)
- Initialization
- Dropout or not dropout
- ...

No principled rules

Existing networks, empirical rules, or meta-optimization

Small Data

Deep Networks need huge datasets

- Pre-trained network

- Data augmentation

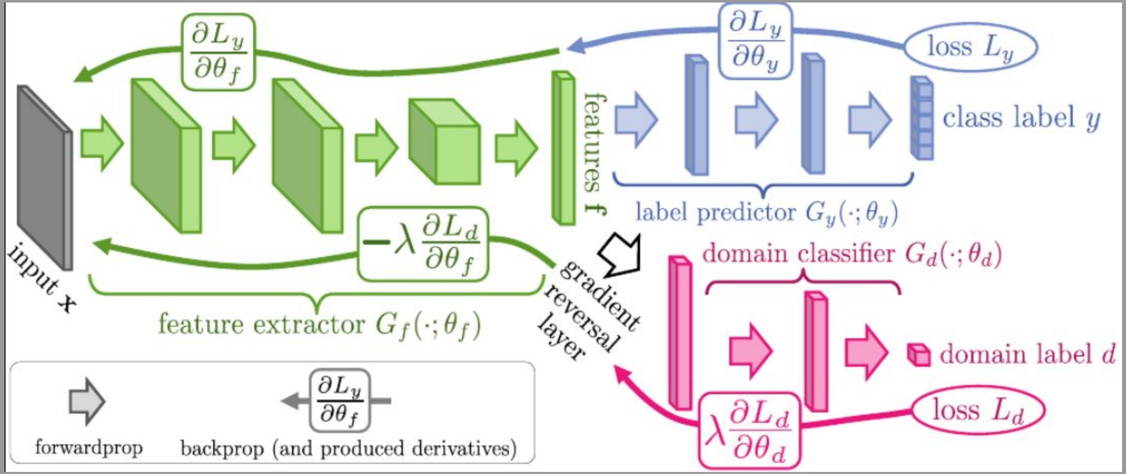
e.g., human poses: MS kinnect (not DL!), Varol et al., 2017

- One-shot learning

External memory and attention mechanism

- Transfer Learning with **Domain Adversarial NNs**

Transfer Learning DANNs



Ganin et al. 2015

- Labelled source domain
- Unlabelled target domain
- Label predictor trained to classify source
- Feature Extractor tries to fool domain classifier (reverse gradient)

More difficult to train than GANs

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
    - Big Data and Irreproducible Science
    - Validation and Certification
    - Explainability and Causality
    - Fairness and Transparency
- Conclusions

Adversarial examples

Noise at test time

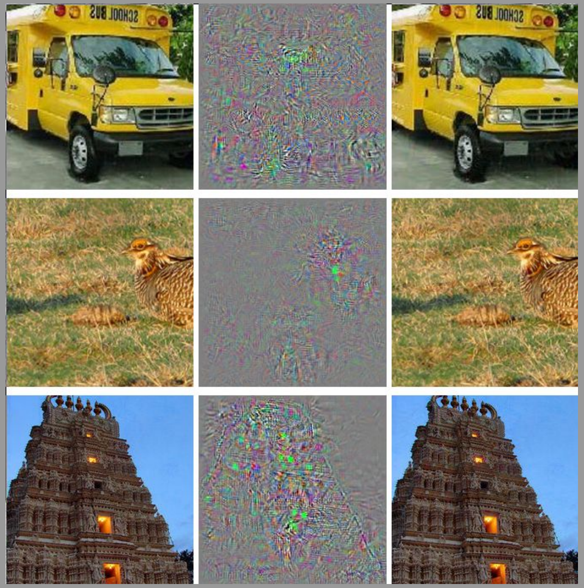
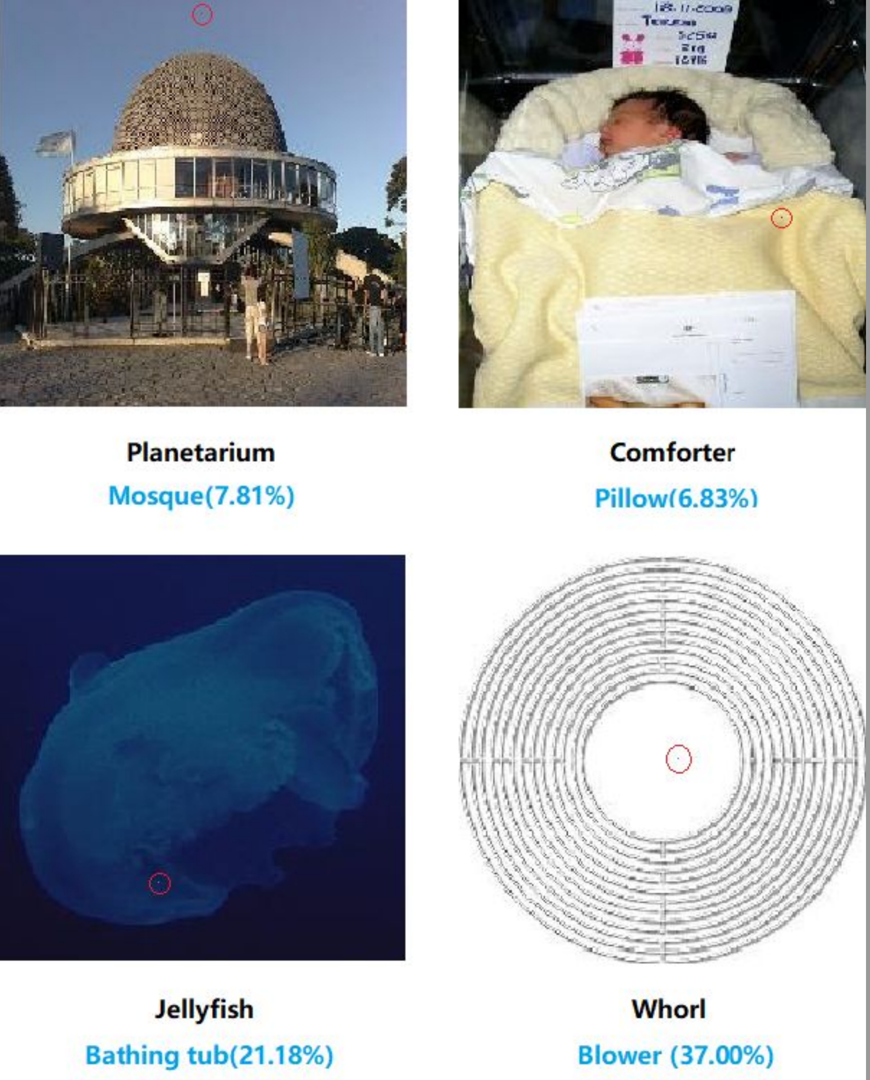


Image + chosen Noise → Ostrich

Image + random Noise → OK

Szegedy et al.,2014

Adversarial examples

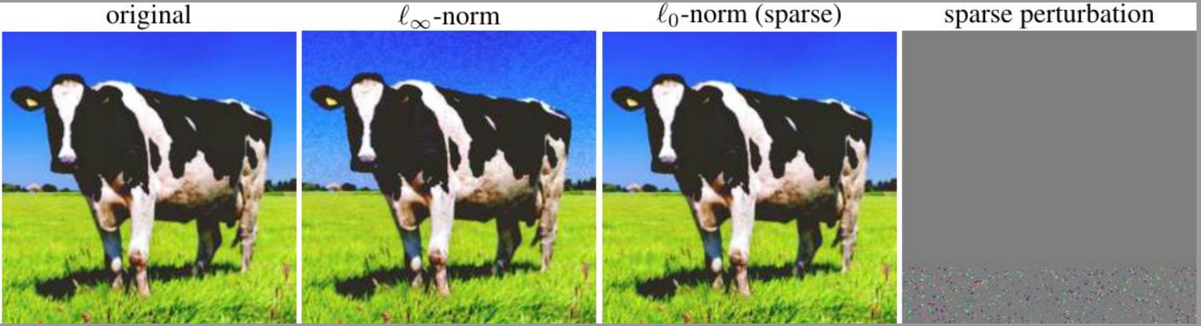


One-pixel attacks

Su et al.,2017

Adversarial examples

Sparse attack



Cow (a) classified as “Traffic Light” (b-c)

Shafahi et al.,2018

All classified as “Speed limit 45”
under various angles/distances

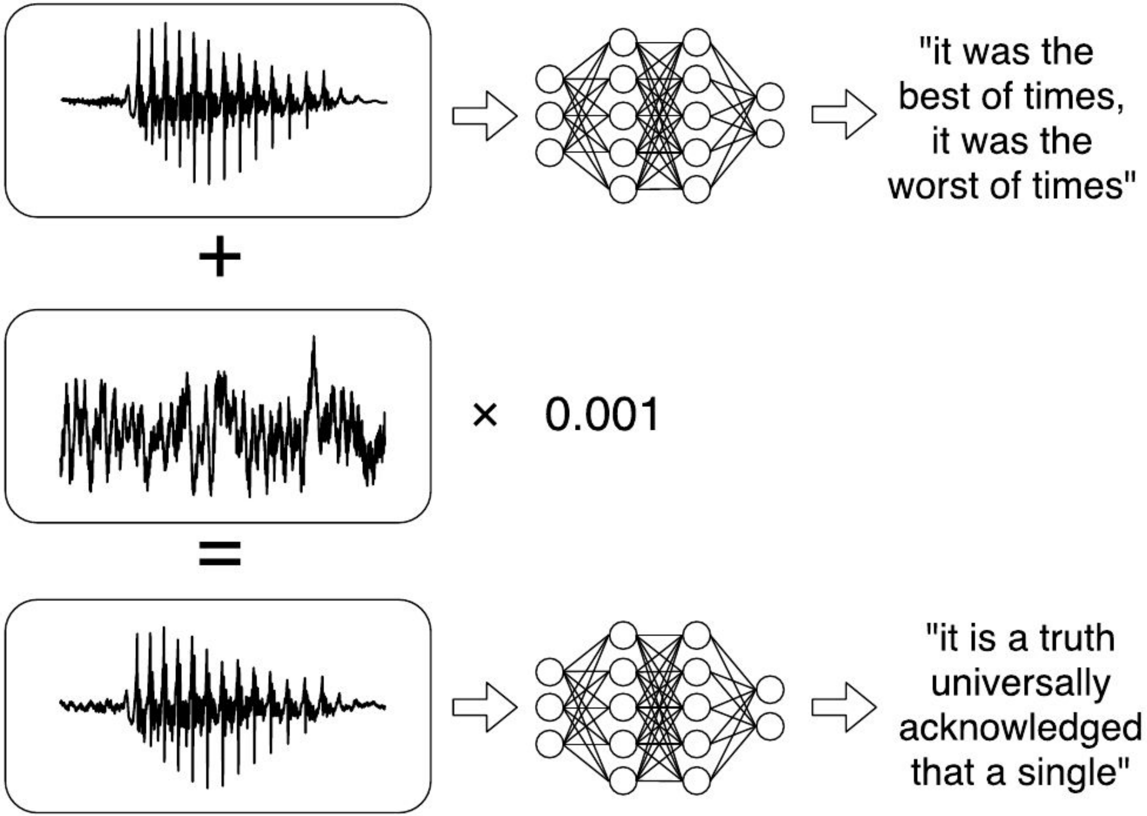
Eykholt et al.,2017

Adversarial examples

Over a series of transformations (and in 3D)

Athalye et al. 2017

Adversarial examples



In Audio domain

Carlini & Wagner 2018

Adversarial examples

A never-ending (?) arms race

- Principled attacks are proposed — see above
- Principled defenses … to these attacks — e.g., robust optim., Madry et al. 2017
- But are broken by stronger attacks … — see e.g., Athalye et al. 2018

But

- Adversarial examples are inevitable — for some problems Shafahi et al. 2018
- An intrinsic property of high-dimensional inputs — Simon-Gabriel et al. 2018

Unseen contexts

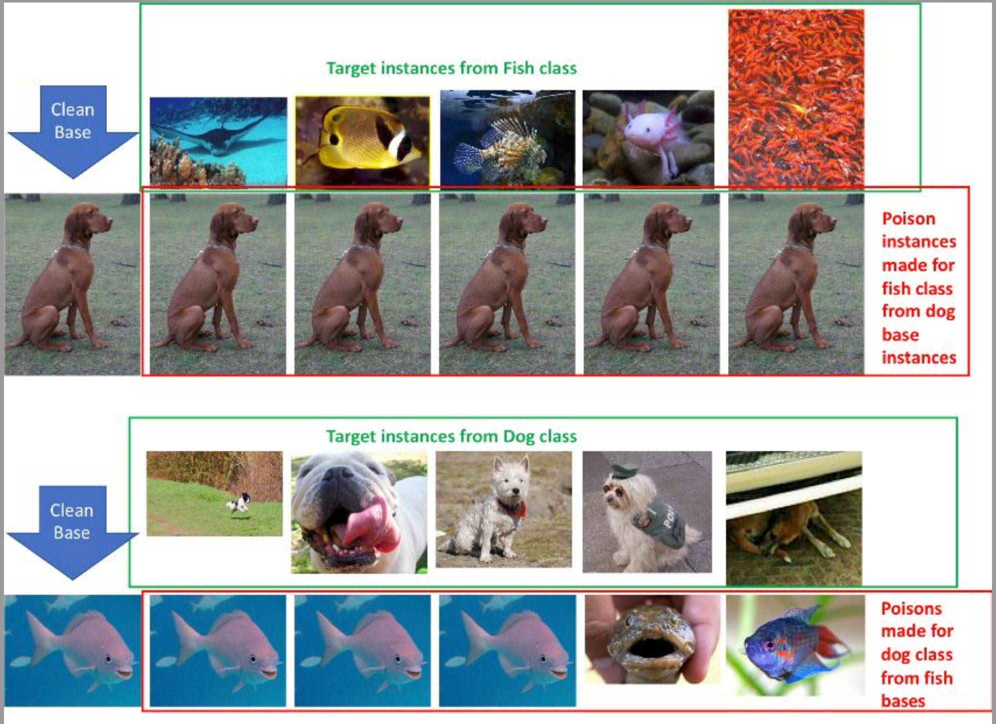
Out of sample **examples**

A cow doesn’t go to the beach

Bottou et al., 2017

Poisoned examples

Noise at learning time



Shafahi et al., 2018

Validation of DNNs

- An experimental science
- No formal validation of learned models
- Completeness issue for statistical validation

- Need to validate the training data
  - Traceability

regulations

- Guaranteed bounds
- Toward formal proofs for AI

e.g., Asimov's robotic laws

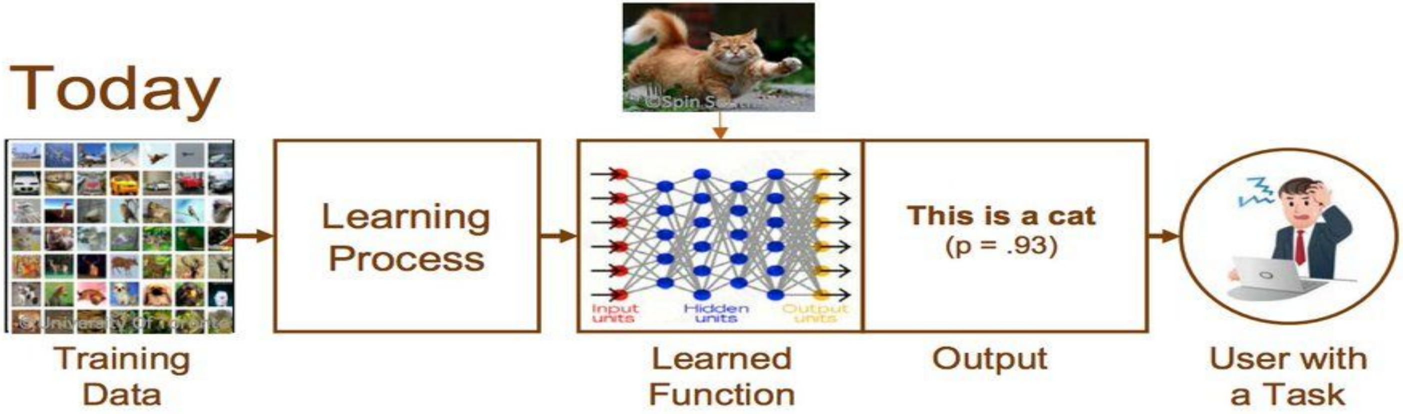
e.g., Mirman et al., 2018

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
    - Big Data and Irreproducible Science
    - Validation and Certification
    - Explainability and Causality
    - Fairness and Transparency
- Conclusions

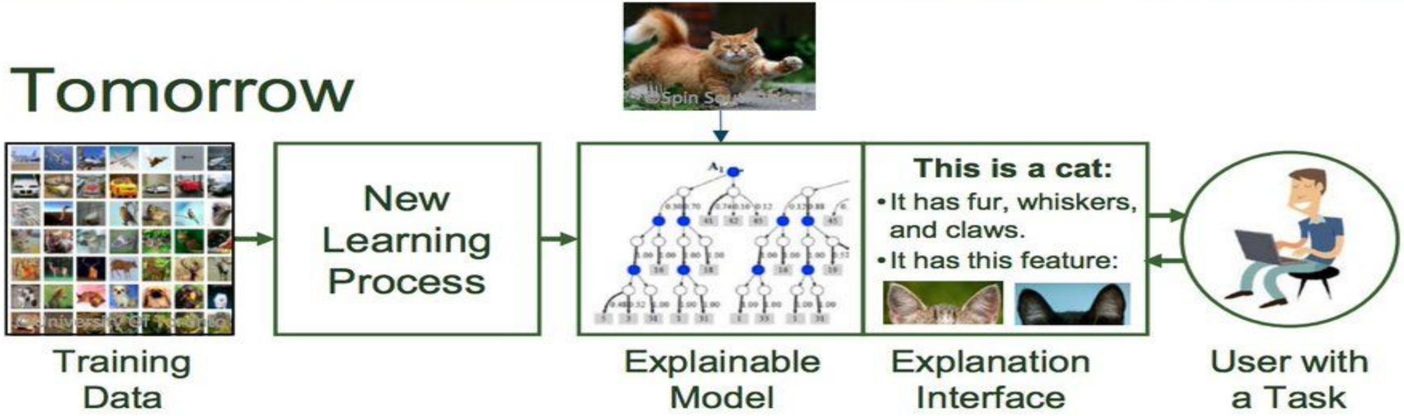
Explainable AI – What Are We Trying To Do?

Today



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Tomorrow



- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

Explainability and Interpretability

Learned models are black boxes

- Ill-defined and subjectives concepts
- Depends on the type of model
  - moderately: decision trees are ok
  - … not random forests

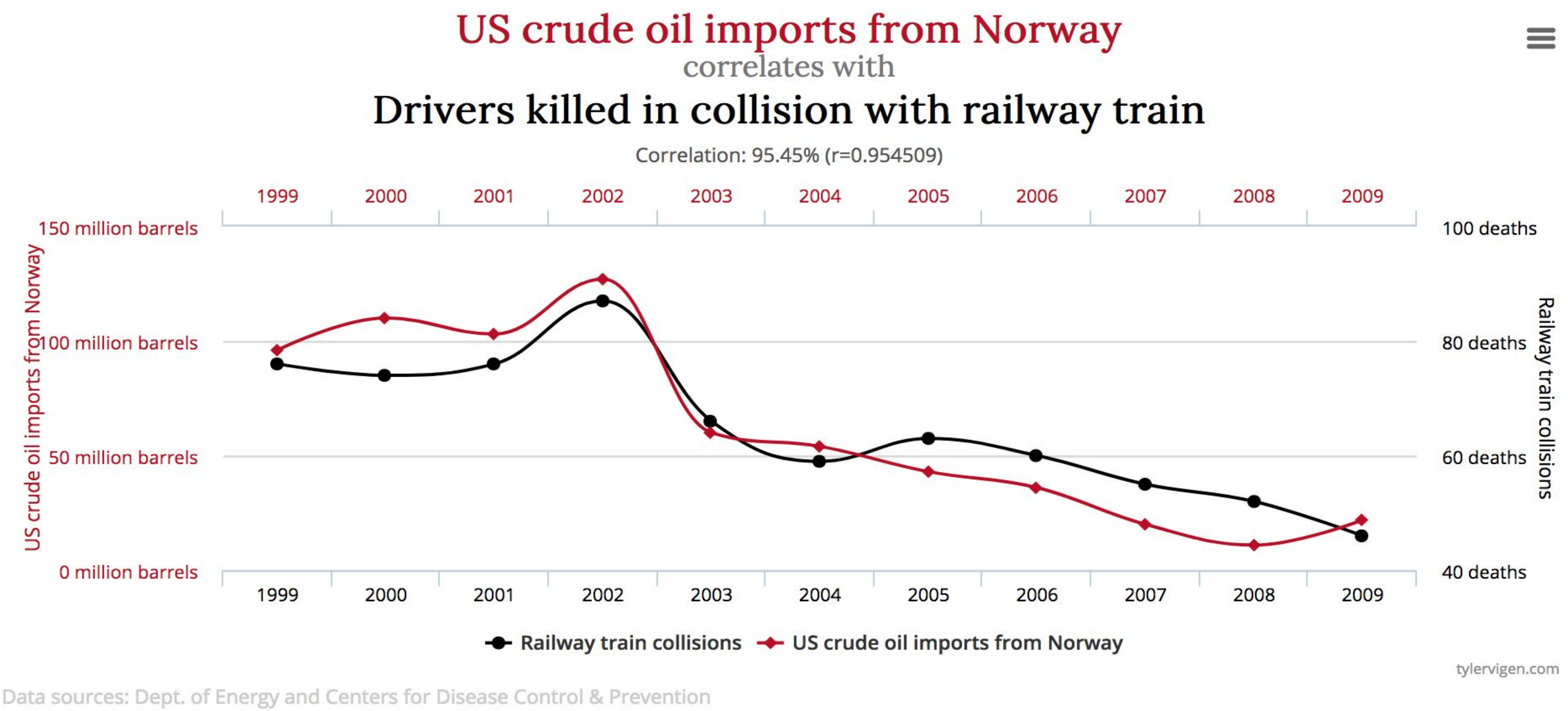
The debate

- How much are you ready to lose in accuracy?
- Cite the nearest known examples e.g., influence fns, Koh & Liang, 2017
- Well, we trust our doctor, don't we …

Symbolic AI to the rescue?

Correlation vs causality

http://www.tylervigen.com/spurious-correlations



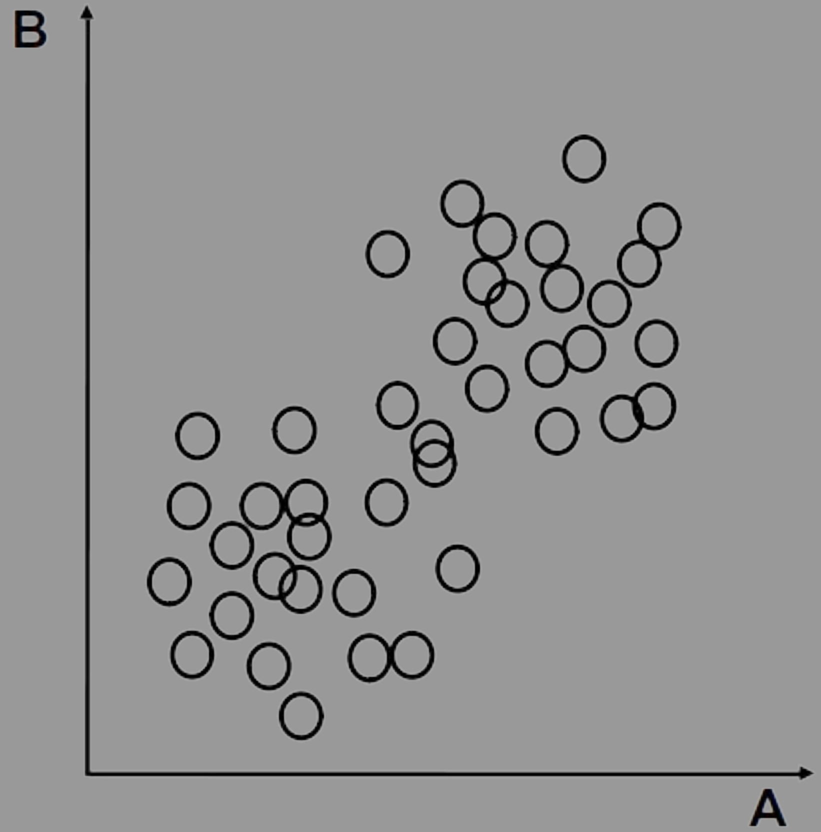
Correlation vs causality

Supervised learning doesn’t make a difference

- “What if” scenarios needed for decision making
- Causality usually from common sense
- Learn from data?

Pairwise causality

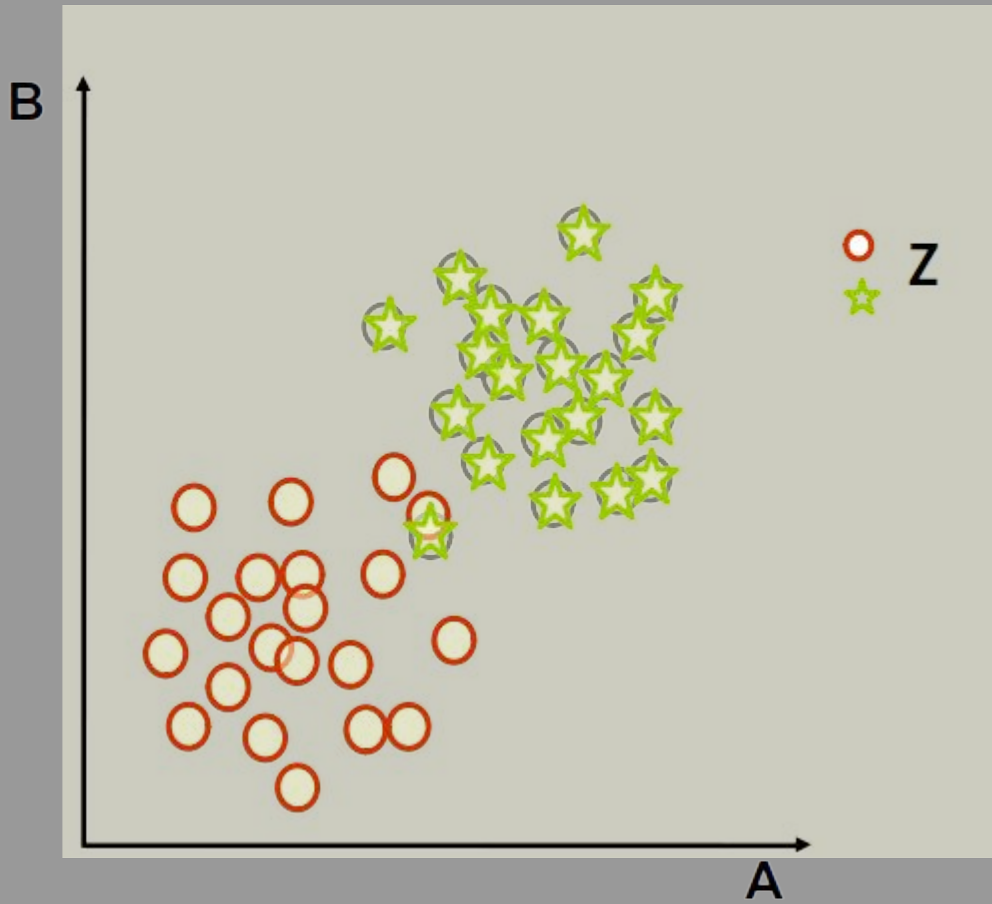
Are A and B related by
some causality dependance?



Pairwise causality

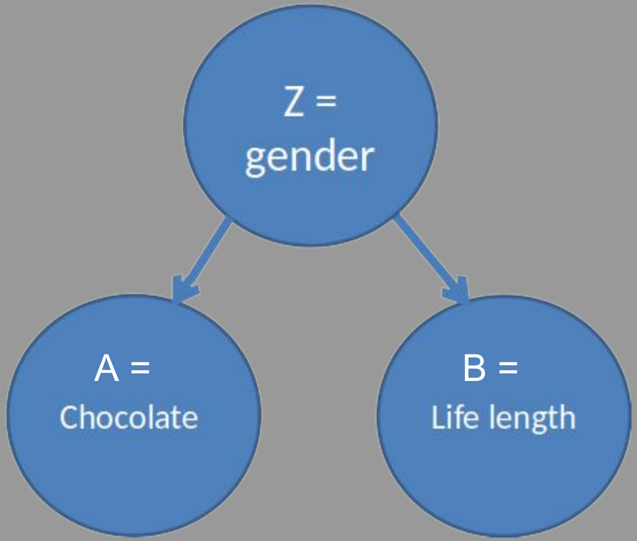
When conditioned on Z,
they are independant

$$A \perp B \mid Z$$

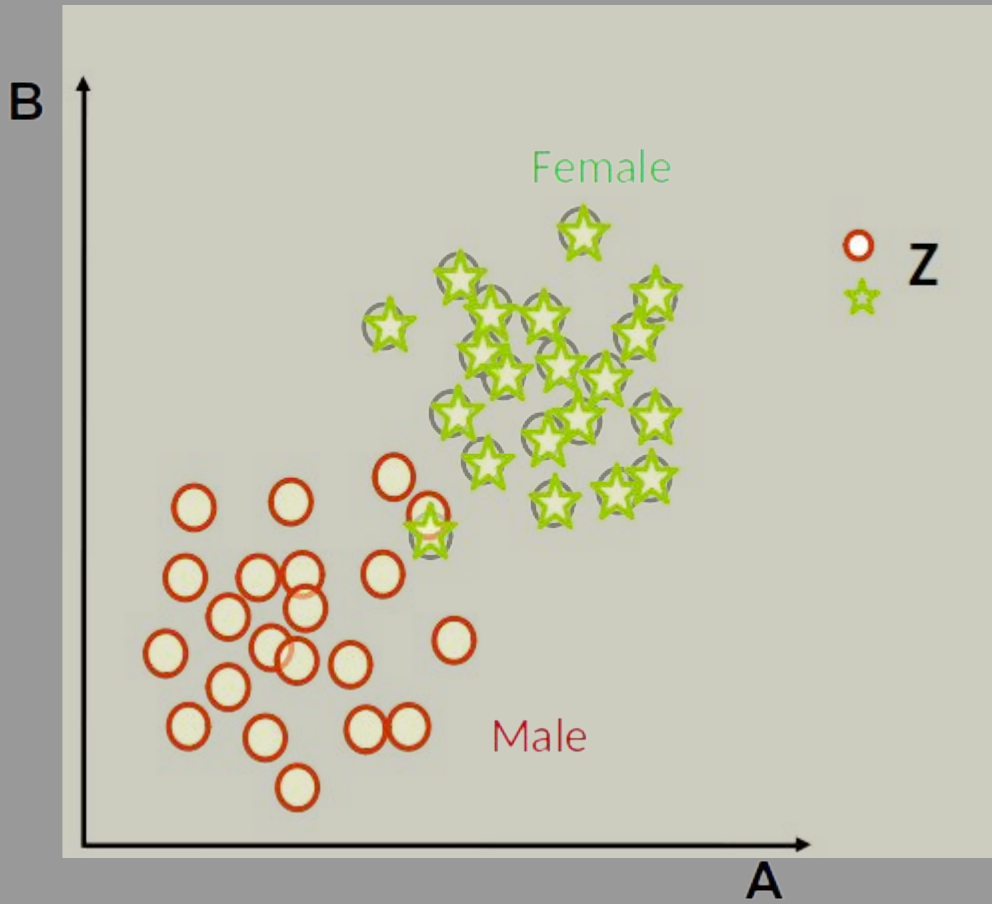


Pairwise causality

Maybe **NO**

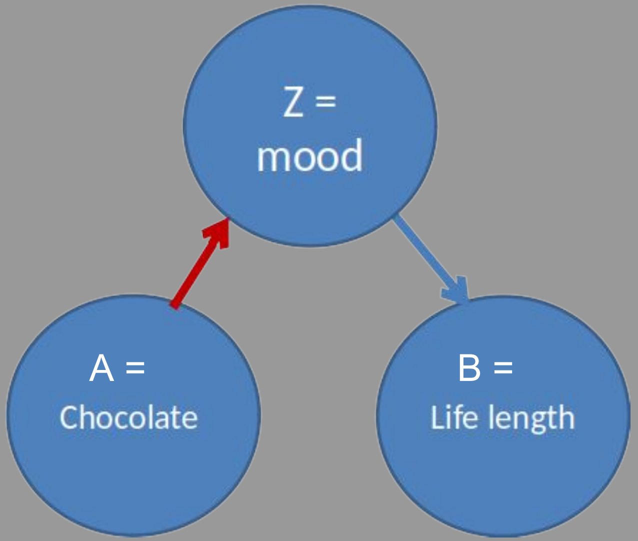


$A \perp B \mid Z$

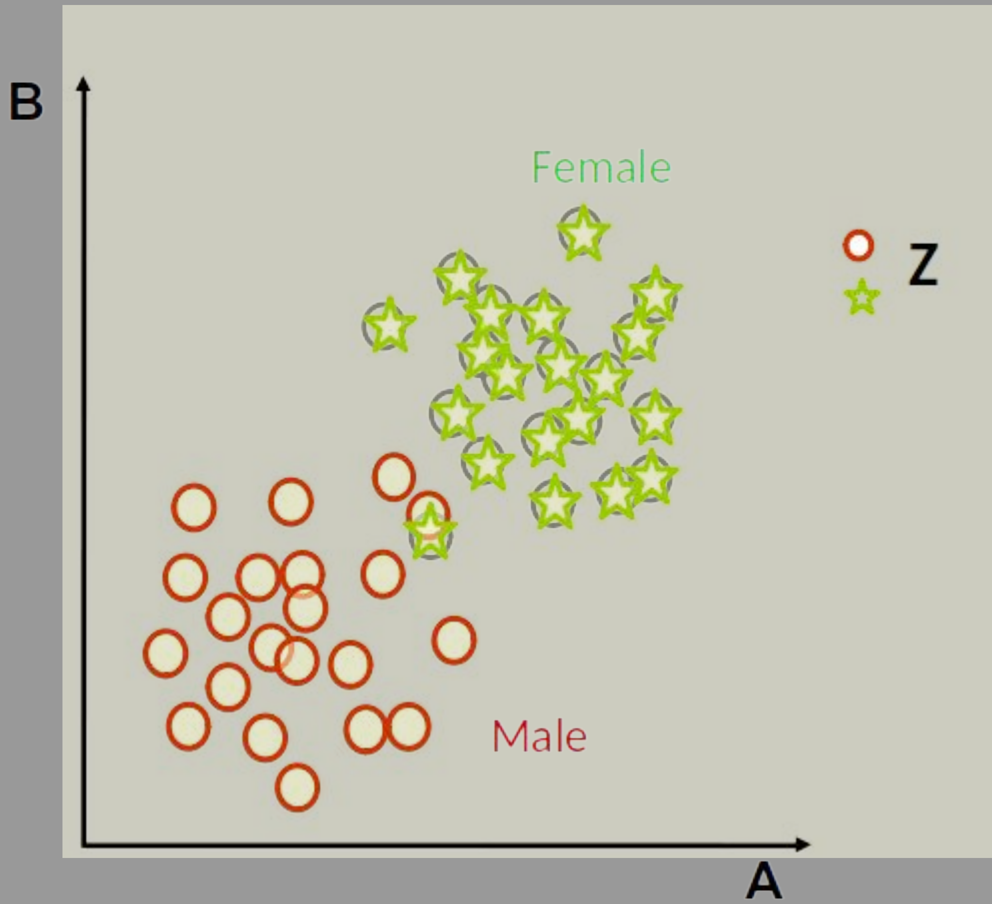


Pairwise causality

Maybe **YES**

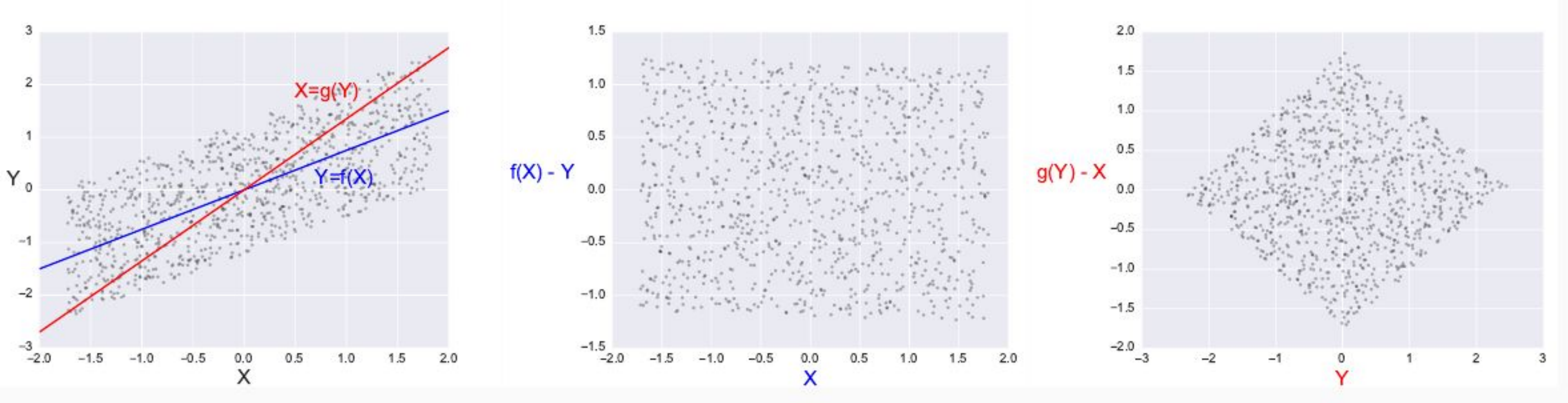


$$A \perp B \mid Z$$



Parwise causality

Identifiable case thanks to asymmetry

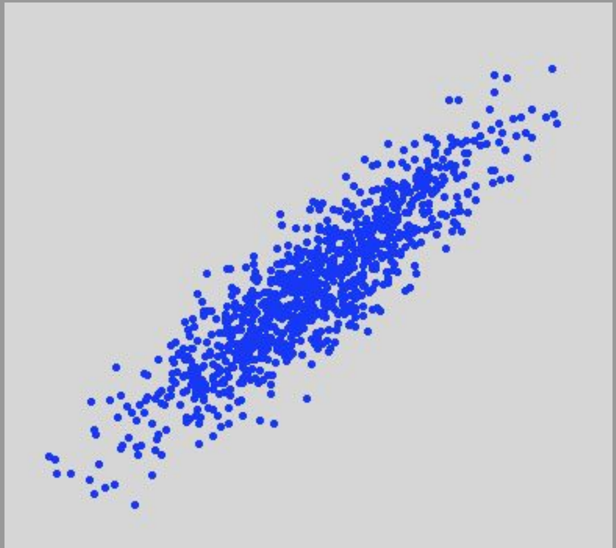


- Causal additive noise: Y = f(X) + E with X ⊥ E
- Perform a regression, check independence of the residual and the cause

Parwise causality

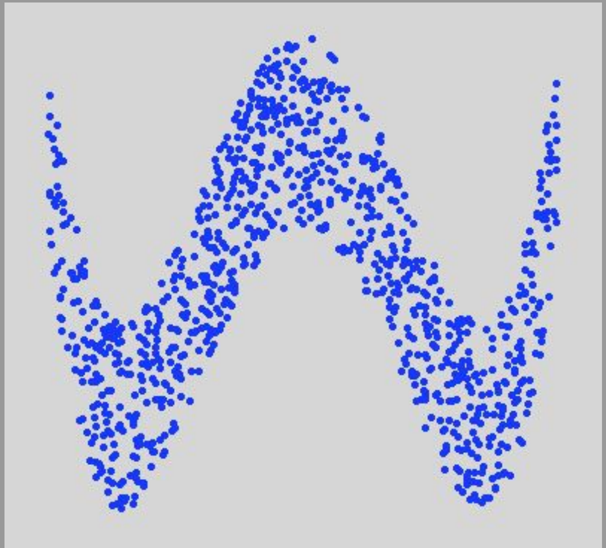
A non-identifiable case

- Linear
- Gaussian input
- Gaussian noise

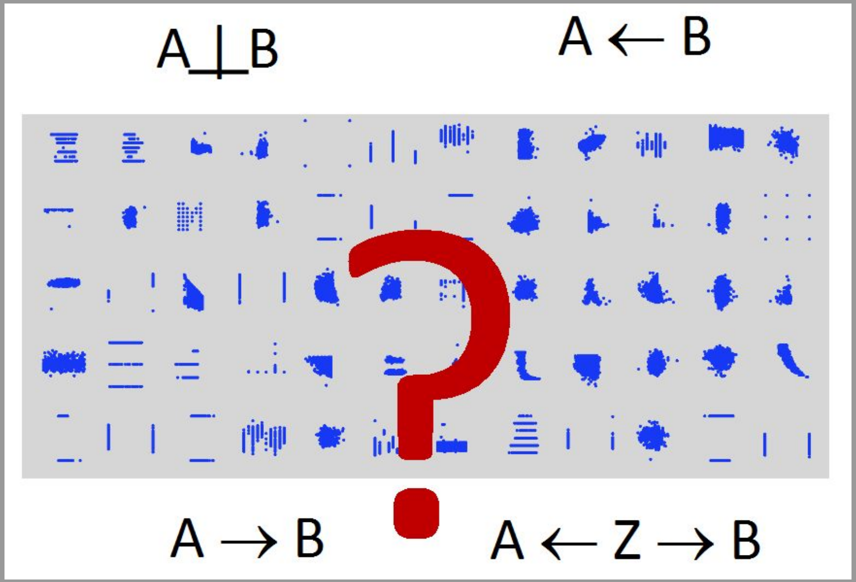


An identifiable case

- Non-invertible OR
- Non-Gaussian input OR
- Non-Gaussian noise

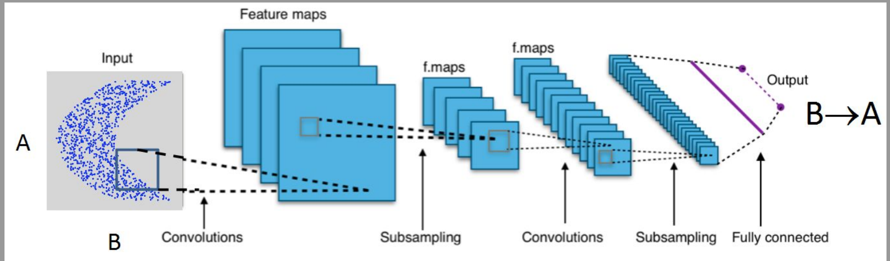


Cause-effect challenge



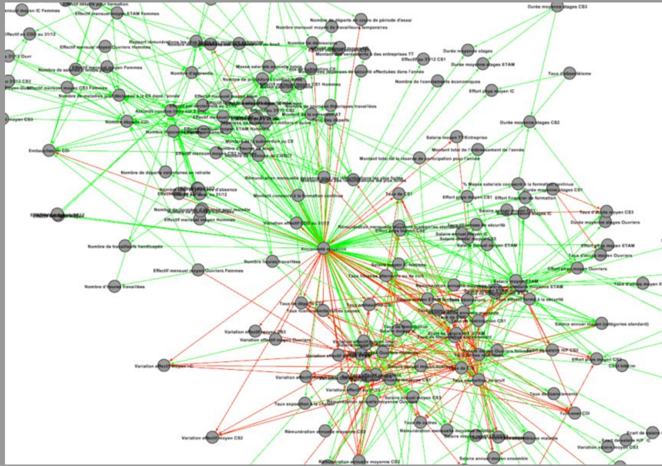
I. Guyon et al, 2013-2015

- A classification problem
- Data ~ images
- ‘easy’ if enough sample datasets (with known causality relation)

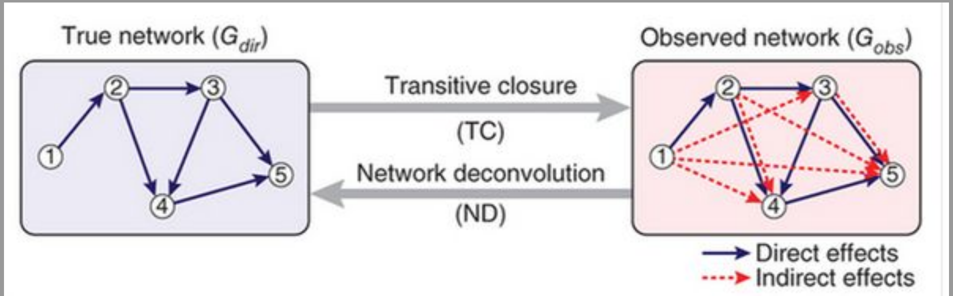


using DNNs: Lopez-Paz et al, 2016

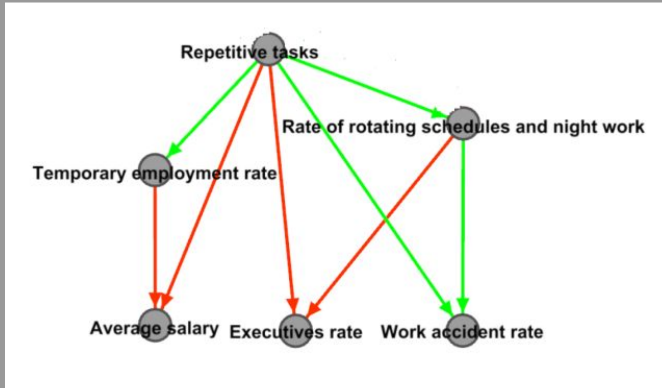
General case



- Generate first network sketch using pairwise dependencies
  - doesn’t scale up
- Remove spurious (indirect) connections



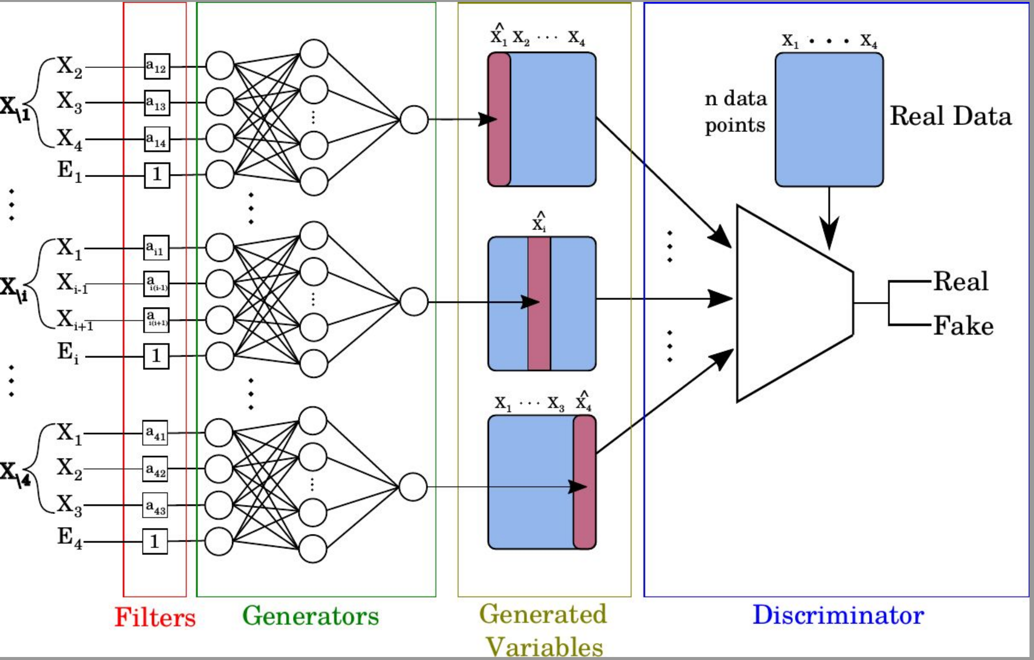
- Orient and weight edges



A global approach

SAM: Structure Agnostic Model

Kalainathan et al., 2018



Ingredients (generated variables)

- Causal filter
  $a_{i,j}$ = impact of X_i on X_j
- Dense layer
- Non-linear activation
- Linear readout

$$\hat{X}_i = m_i^\top \tanh\left(\bar{W}_i^\top (a_i \odot X) + n_i E_i + b_i\right) + \beta_i$$

SAM

Loss function *d* sparse causal mechanisms vs shared discriminator

- Adversarial learning

$$L_i = \mathbb{E}_{x_i, x_{\setminus i}}[\log D(x_i, x_{\setminus i})] + \mathbb{E}_{e_i, x_{\setminus i}}[\log(1 - D(\hat{f}_i(e_i, x_{\setminus i}), x_{\setminus i}))]$$

- + L1 regularization on filters (→ sparsity)

$$L_\lambda = \sum_{i=1}^{d} L_i + \lambda \sum_{i=1}^{d} \|a_i\|_1 \quad , \lambda \geq 0$$

Discussion

- No combinatorial search
- Possible cycles: genuine, or non-identifiability

Still far from a unified accepted framework

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
    - Big Data and Irreproducible Science
    - Validation and Certification
    - Explainability and Causality
    - Fairness and Transparency
- Conclusions

Fairness

- Biased data lead to biased models and decisions
  - Predictive Justice, Bank loans, ...
- No formal definition of fairness
  - In fact, too many! Barocas and Hardt, NIPS17 tutorial
- Hiding discriminative variables insufficient Proxy variables

Beyond scientific research

- Recognized labels e.g., FDU, Maathics (Toulouse)
- Discrimination Impact Assessment?

Transparency

Mandatory for societal acceptance

- Open Source not enough
    - Open Data
    - controlled experiments

auditability by law

- Explainability
- See the **TransAlgo** platform

Trustable Good AI

Beyond research issues

- Scientific and legal advances
  - Human in control
  - Accountability
- Ethical rules

enforce European humanist values

  - By design
    - Public debate, CCNE-bis, ...

CERNA, COERLE, ...

  - Control
    - Citizen crowd control, independent institution, ...

Without trust, societal AI winter ahead

AI: Past, Present, and Future

- Past: History, definitionS, and recent successes
- Present: Around Deep Learning
- Future: Toward Trustable Good AI?
- Conclusions

Deep Learning

Missing Scientific Issues

- Unsupervised/predictive learning — Predict the world from observations
- Reinforcement learning — The cherry on the cake, LeCun 2017

The main advances

- Discover relevant latent representations
- Differentiable programming — but is the world differentiable?

The debate

J. Schmidhuber

- Are handcrafted (explainable) features useless?
- What does the learnt representations mean?
- Any invariance/physical properties in these representations?

Other approaches

Trendy Scientific Issues

- Probabilistic Programming
- Artificial General AI — and should we fear it?

The debate about Machine Learning

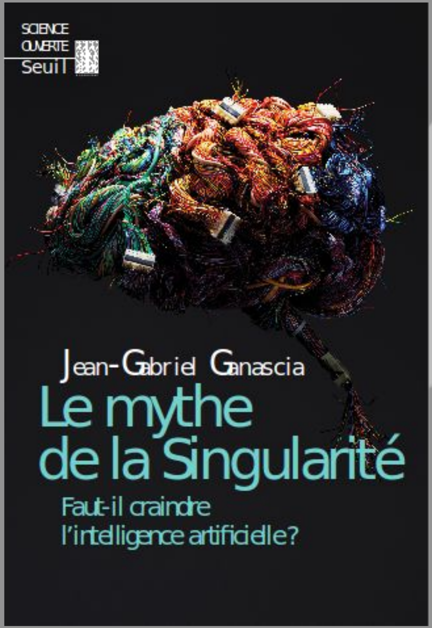
- Machine Learning **vs**
- Symbolic AI — Knowledge representations, Semantic Web, AI Planning, …
- In France and in Europe — not in USA and China

Beyond the scientific debate, hidden political agendas

The singularity

Jean-Gabriel Ganascia

When intelligent machines take over the world



Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all.
Stephen Hawking 1er mai 2014

but also Elon Musk, Bill Gates, ...

The singularity

When intelligent machines take over the world

Fantasy

- weak AI: can do a single task
- strong AI (Artificial General Intelligence)

AI = mean

AI = goal

- Today, only exist weak AIs
  - even AlphaZéro
  - even automatic translators
  - even personal assistants
  - ...

Toward an AGI ?

What seems to be crucially missing

- Intrinsic motivation — On-going (see e.g., Pierre-Yves Oudeyer)
- Embodiment — On-going (see e.g., Rolf Pfeifer)
- Common sense
- Self-consciousness

Other approaches

Trendy Scientific Issues

- Probabilistic Programming
- Artificial General AI — and should we fear it?

The debate about Machine Learning

- Machine Learning **vs**
- Symbolic AI — Knowledge representations, Semantic Web, AI Planning, …
- In France and in Europe — not in USA and China

Beyond the scientific debate, hidden political agendas

Collaboration, not Competition

Questions ?