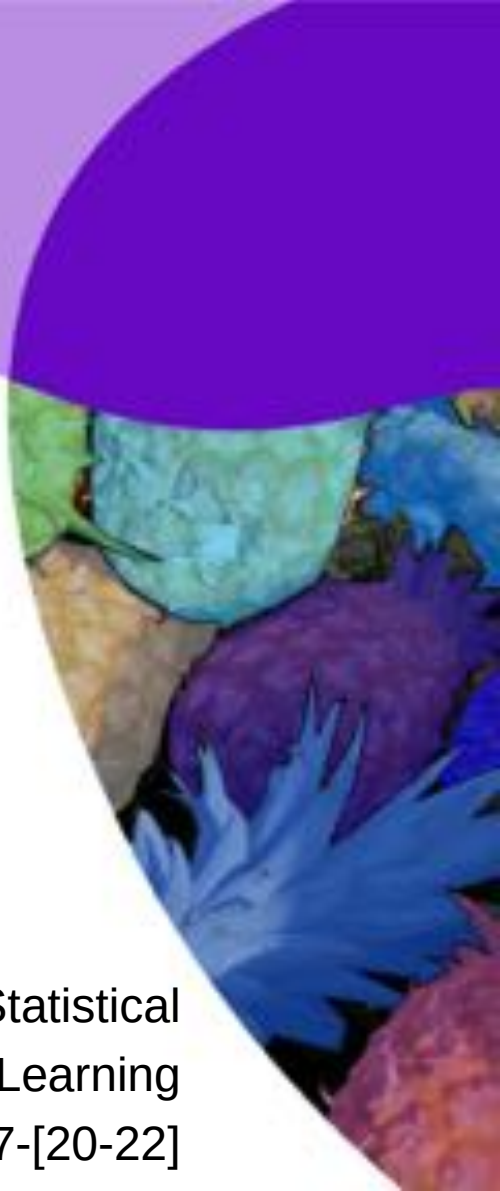
NEW YORK UNIVERSITY

From Machine Learning to Autonomous Intelligence Lecture 1

Yann LeCun
NYU - Courant Institute & Center for Data Science
Meta - Fundamental AI Research
http://yann.lecun.com



Summer School on Statistical Physics & Machine Learning
Les Houches, 2022-07-[20-22]

Plan

- **Applications of AI / ML / DL today**
  - Largely rely on supervised Deep Learning. A few on Deep RL.
  - Increasingly rely on Self-Supervised pre-training.
- **Current ML/DL sucks compare to humans and animals**
  - Humans and animals learn models of the world
- **Self-Supervised Learning**
  - Main problem: representing uncertainty, learning abstractions.
- **Energy-Based Models**
  - Sample contrastive learning methods
  - Non-contrastive learning methods

Main Messages

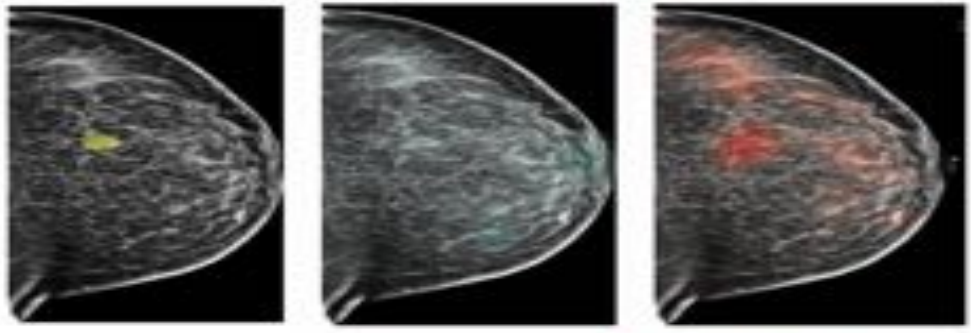
- **Deep SSL is the enabling element for the next AI revolution**
- **I'll try to convince you to:**
  - Give up on supervised and reinforcement learning
    - well, not completely, but as much as possible.
  - Give up on probabilistic modeling
    - use the energy-based framework instead
  - Give up on generative models
    - Use joint-embedding architectures instead
  - Use hierarchical latent-variable energy-based models
    - To enable machines to reason and plan.
  - See position paper: "A Path Towards Autonomous Machine Intelligence"
    - https://openreview.net/forum?id=BZ5a1r-kVsf

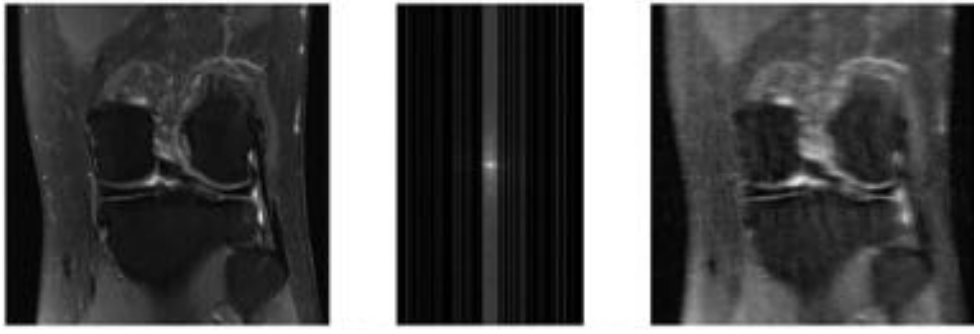
FACEBOOK AI

AI can do pretty amazing things today

Deep Learning: Protecting Lives and the Environment

- **Transportation**
  - Driving assistance / autonomous driving
- **On-line Safety / Security**
  - Filtering harmful/hateful content
  - Filtering dangerous misinformation
- **Environmental monitoring**
- **Medicine**
  - Medical imaging
  - Diagnostic aid
  - Patient care
  - Drug discovery





Deep Learning Connects People to knowledge & to each other

- **Meta (FB, Instagram), Google, YouTube, Amazon, are built around Deep learning**
  - Take Deep Learning out of them, and they crumble.
- **DL helps us deal with the information deluge**
  - Search, retrieval, ranking, question-answering
  - Requires machines to understand content
- **Translation / transcription / accessibility**
  - language ↔ language; text ↔ speech; image → text
  - People speak thousands of different languages
  - 3 billion people can’t use technology today.
  - 800 million are illiterate, 300 million are visually impaired

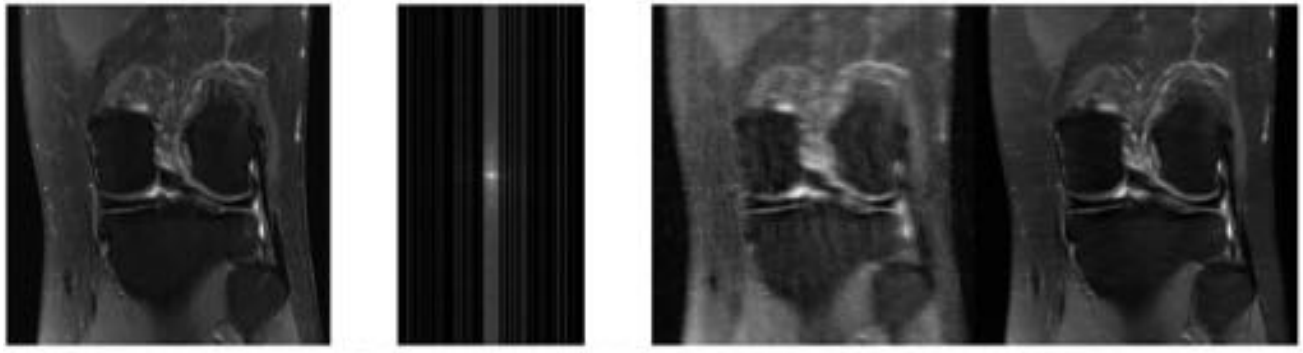
Deep Learning for On-Line Content Moderation

- **Filtering out objectionable content**
  - What constitutes acceptable or objectionable content?
  - Meta doesn’t see itself as having the legitimacy to decide
  - But in the absence of regulations, it has to do it.
- **Types of objectionable content on Facebook**
  - (with % taken down preemptively & prevalence, Q1 2022)
  - Hate Speech (95.6%, 0.02%), up from 30-40% in 2018
  - Violence incitement (98.1%, 0.03%), Violence (99.5%, 0.04%), Bullying/Harassment (67%, 0.09%), Child endangerment (96.4%), Suicide/Self-Injury (98.8%), Nudity (96.7%, 0.04%),
  - Taken down (Q1’22): Terrorism (16M), Fake accounts (1.5B), Spam (1.8B)
  - https://transparency.fb.com/data/community-standards-enforcement

Image understanding

FastMRI: 4x speed up for MRI acquisition (NYU Radiology + FAIR)

- **MRI images subsampled in k-space by 4x and 8x**
  - U-Net architecture
  - 4-fold acceleration
  - [Zbontar et al. ArXiv:1811.08839]



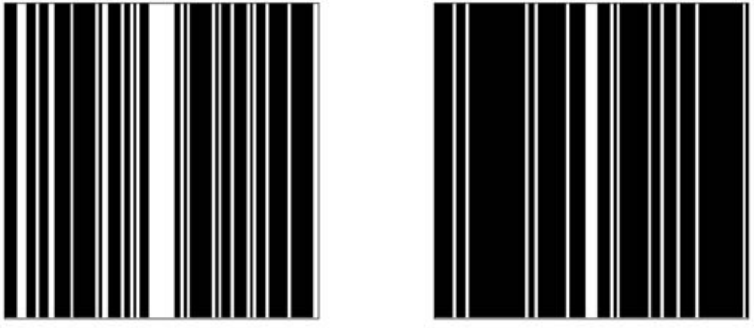
(a) Cropped and vertically flipped reconstruction from fully sampled k-space data

(b) Rectangular masked k-space

(c) Reconstruction via zero-filled IFFT

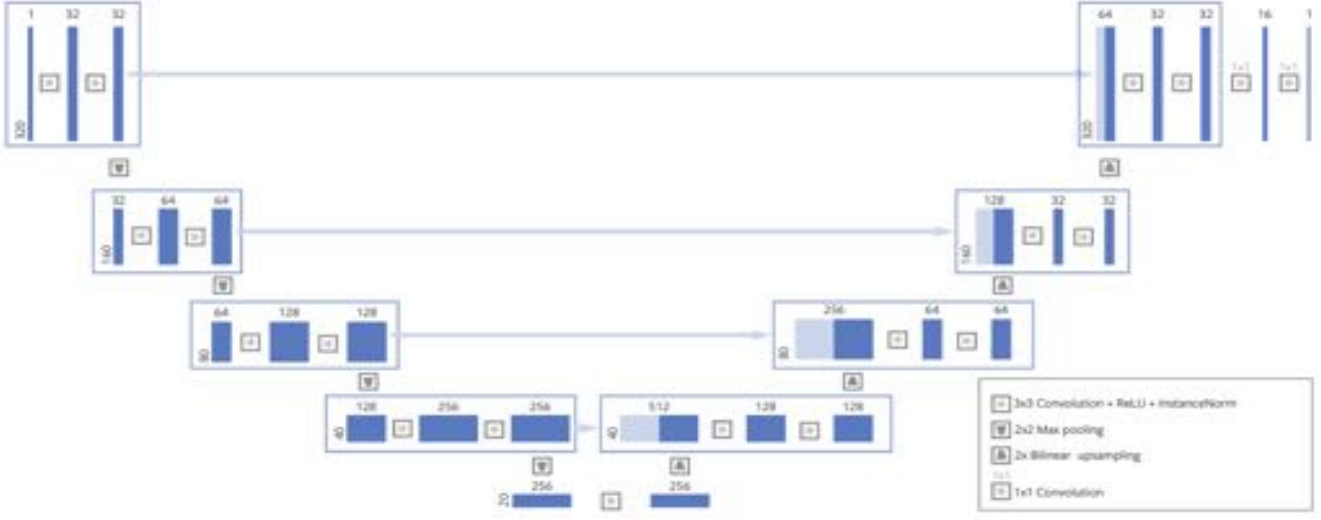
(d) Deep-learning baseline UNET reconstruction

- **K-space masks**



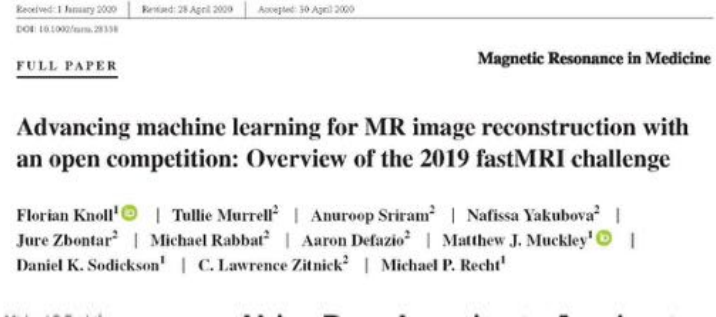
(a) 4-fold acceleration

(b) 8-fold acceleration



FastMRI (NYU Radiology+FAIR): 4x speed up for MRI acquisition

- **Radiologists could not tell the difference between clinical standard and 4x accelerated/restored images**
  - They often preferred the accelerated/restored images
  - [Recht et al., American Journal of Roentgenology 2020]
- **Similar systems are now integrated in new MRI machines.**



Received: 1 January 2020 | Revised: 28 April 2020 | Accepted: 30 April 2020

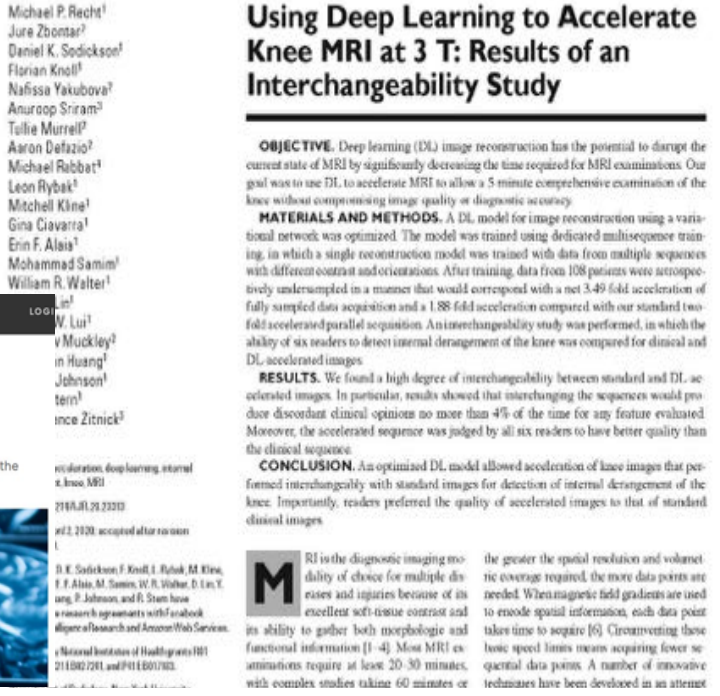
DOI: 10.1002/mrm.28338

FULL PAPER

Magnetic Resonance in Medicine

Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge

Florian Knoll[1] | Tullie Murrell[2] | Anuroop Sriram[2] | Nafissa Yakubova[2] | Jure Zbontar[2] | Michael Rabbat[2] | Aaron Defazio[2] | Matthew J. Muckley[1] | Daniel K. Sodickson[1] | C. Lawrence Zitnick[2] | Michael P. Recht[1]



Michael P. Recht[1]
Jure Zbontar[2]
Daniel K. Sodickson[1]
Florian Knoll[1]
Nafissa Yakubova[2]
Anuroop Sriram[2]
Tullie Murrell[2]
Aaron Defazio[2]
Michael Rabbat[2]
Leon Rybak[1]
Mitchell Kline[1]
Gina Ciavarra[1]
Erin F. Alaia[1]
Mohammad Samim[1]
William R. Walter[1]

Using Deep Learning to Accelerate Knee MRI at 3 T: Results of an Interchangeability Study

OBJECTIVE. Deep learning (DL) image reconstruction has the potential to disrupt the current state of MRI by significantly decreasing the time required for MRI examinations. Our goal was to use DL to accelerate MRI to allow a 5 minute comprehensive examination of the knee without compromising image quality or diagnostic accuracy.

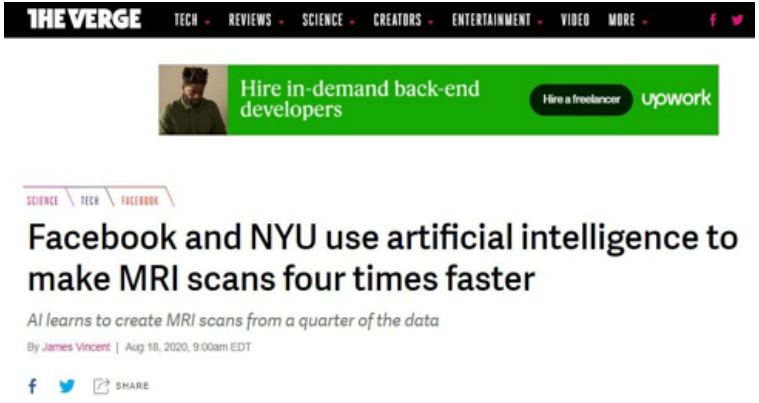
MATERIALS AND METHODS. A DL model for image reconstruction using a variational network was optimized. The model was trained using dedicated multisequence training, in which a single reconstruction model was trained with data from multiple sequences with different contrast and orientations. After training, data from 108 patients were retrospectively undersampled in a manner that would correspond with a net 3.49 fold acceleration of fully sampled data acquisition and a 1.88-fold acceleration compared with our standard two-fold accelerated parallel acquisition. An interchangeability study was performed, in which the ability of six readers to detect internal derangement of the knee was compared for clinical and DL accelerated images.

RESULTS. We found a high degree of interchangeability between standard and DL accelerated images. In particular, results showed that interchanging the sequences would produce discordant clinical opinions no more than 4% of the time for any feature evaluated. Moreover, the accelerated sequence was judged by all six readers to have better quality than the clinical sequence.

CONCLUSION. An optimized DL model allowed acceleration of knee images that performed interchangeably with standard images for detection of internal derangement of the knee. Importantly, readers preferred the quality of accelerated images to that of standard clinical images.

MRI is the diagnostic imaging modality of choice for multiple diseases and injuries because of its excellent soft-tissue contrast and its ability to gather both morphologic and functional information [1–4]. Most MRI examinations require at least 20–30 minutes, with complex studies taking 60 minutes or

the greater the spatial resolution and volumetric coverage required, the more data points are needed. When magnetic field gradients are used to encode spatial information, each data point takes time to acquire [6]. Circumventing these basic speed limits means acquiring fewer sequential data points. A number of innovative techniques have been developed in an attempt



THE VERGE TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO MORE

Hire in-demand back-end developers
Hire a freelancer upwork

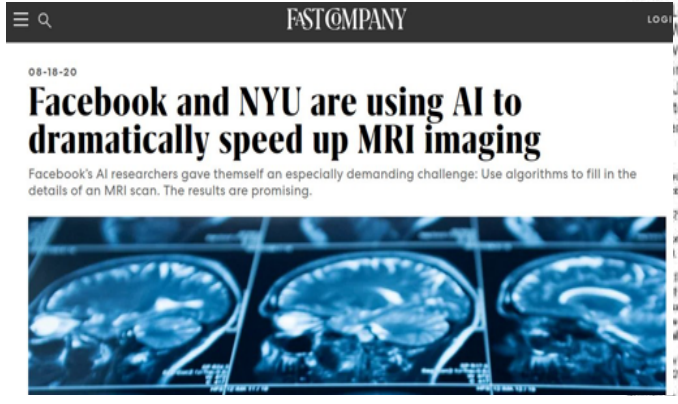
SCIENCE \ TECH \ FACEBOOK

Facebook and NYU use artificial intelligence to make MRI scans four times faster

AI learns to create MRI scans from a quarter of the data

By James Vincent | Aug 18, 2020, 9:00am EDT

SHARE



FAST COMPANY

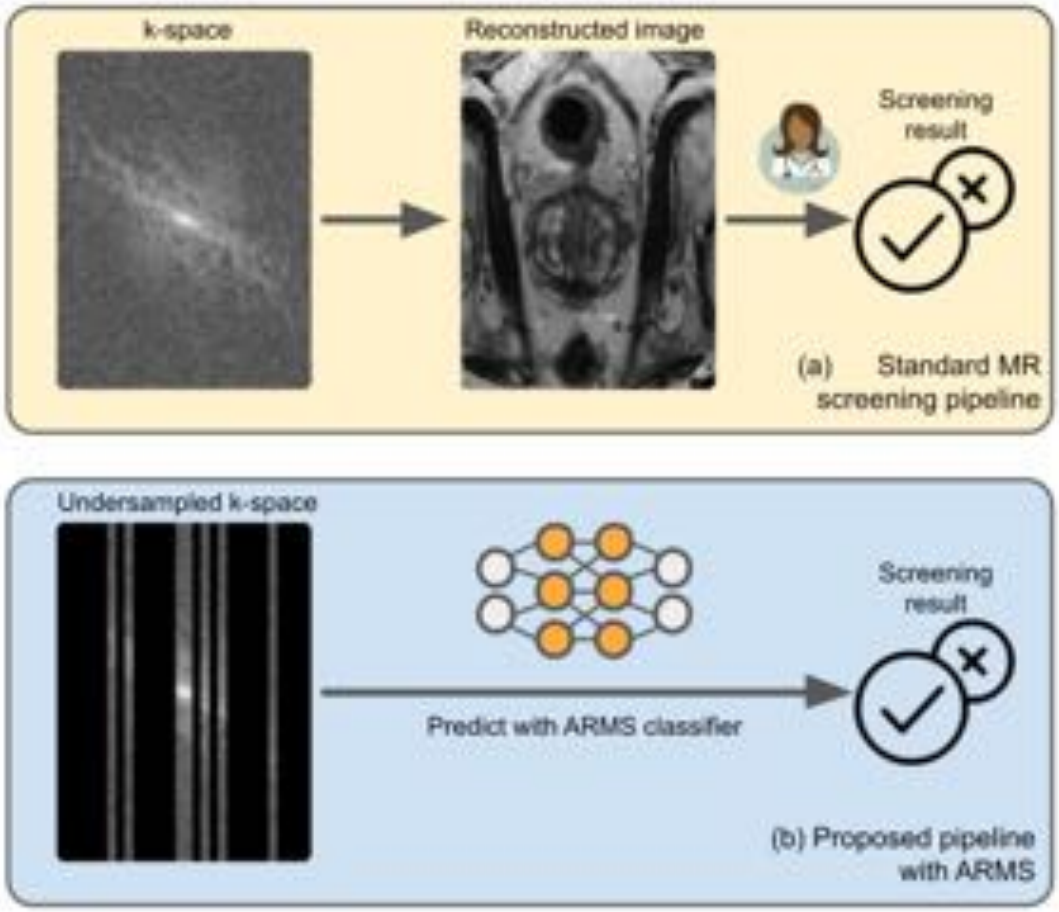
08-18-20

Facebook and NYU are using AI to dramatically speed up MRI imaging

Facebook's AI researchers gave themself an especially demanding challenge: Use algorithms to fill in the details of an MRI scan. The results are promising.

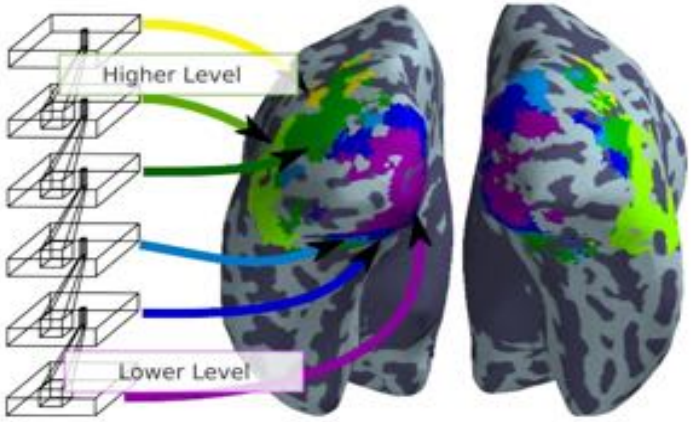
Why produce an image at all? [S. Chopra's group, NYU]

- **Why not directly from raw data to diagnosis / screening?**
  - Humans need 2D image sliced displayed on a monitor
  - DL systems can accept grossly undersampled (10-20x) or low-field raw data representing the entire volume.
  - They can be trained to directly produce a screening result



AI accelerates progress of biomedical sciences

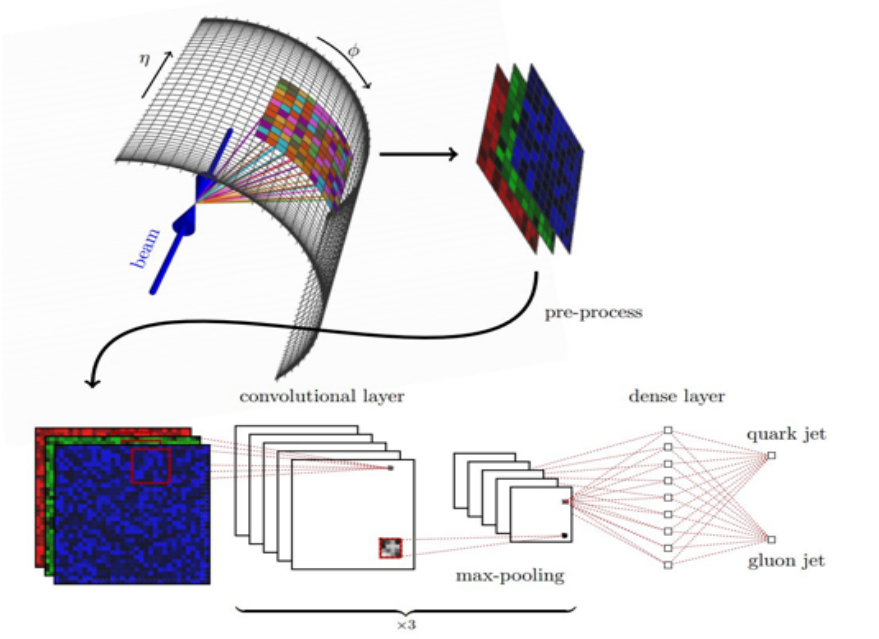
- **Neuroscience**
  - Neural nets as models of the brain
  - Models of vision, audition, & speech understanding
- **Genomics**
  - Identifying gene regulation networks
  - Curing genetic diseases?
- **Biology / biochemistry**
  - Predicting protein structure and function
  - Designing proteins
  - Drug discovery

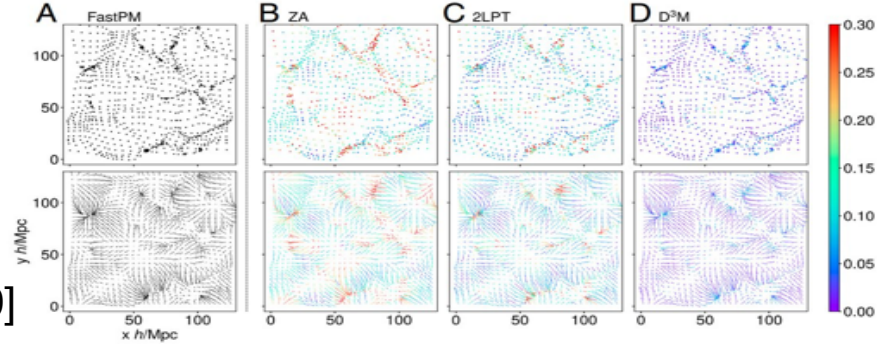


[DeepMind, AlphaFold]

AI accelerates the progress of physical sciences

- **Physics**
  - Analyzing particle physics experiments
  - Accelerating complex simulations: fluids, aerodynamics, atmosphere, oceans,….
  - Astrophysics: enabling universe-wide simulations, classifying galaxies, discovering exoplanets….
- **Chemistry**
  - Finding new compounds
- **Material science**
  - Predicting new material properties
  - Designing new meta-materials

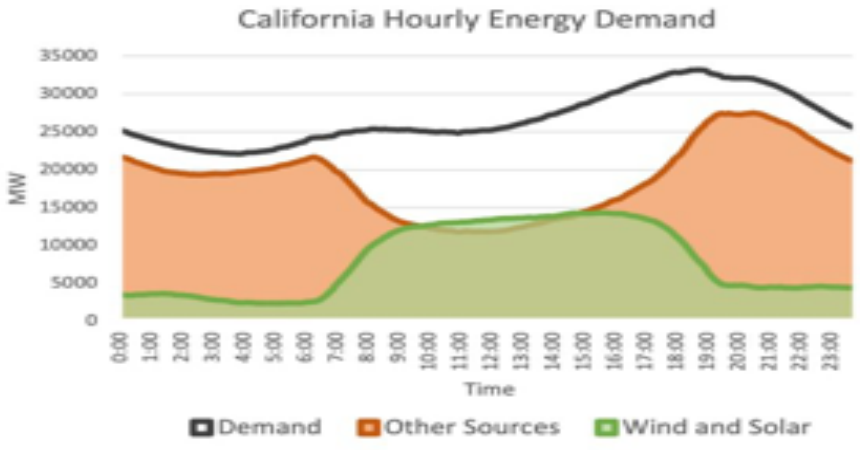


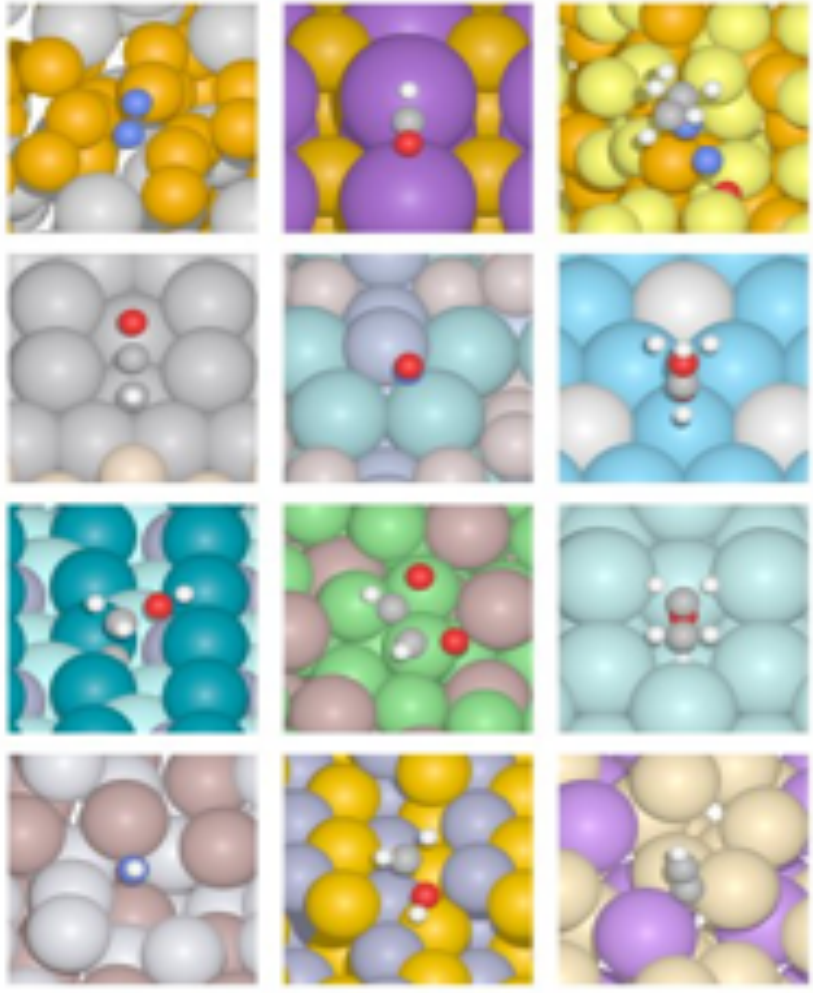


[He 2019]

Open Catalyst Project: open competition

- **Want to solve climate change?**
- **Discovering new materials to enable large-scale energy storage**
  - Efficient & scalable extraction of hydrogen from water through electrolysis
- **Sponsored by FAIR & CMU**
  - [Zitnick https://arxiv.org/abs/2010.09435]

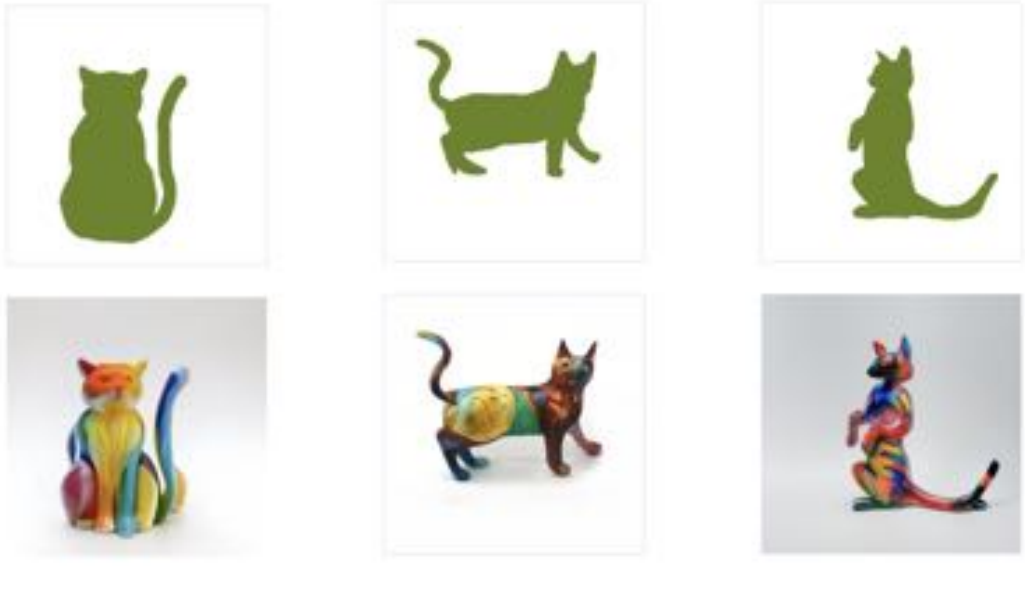


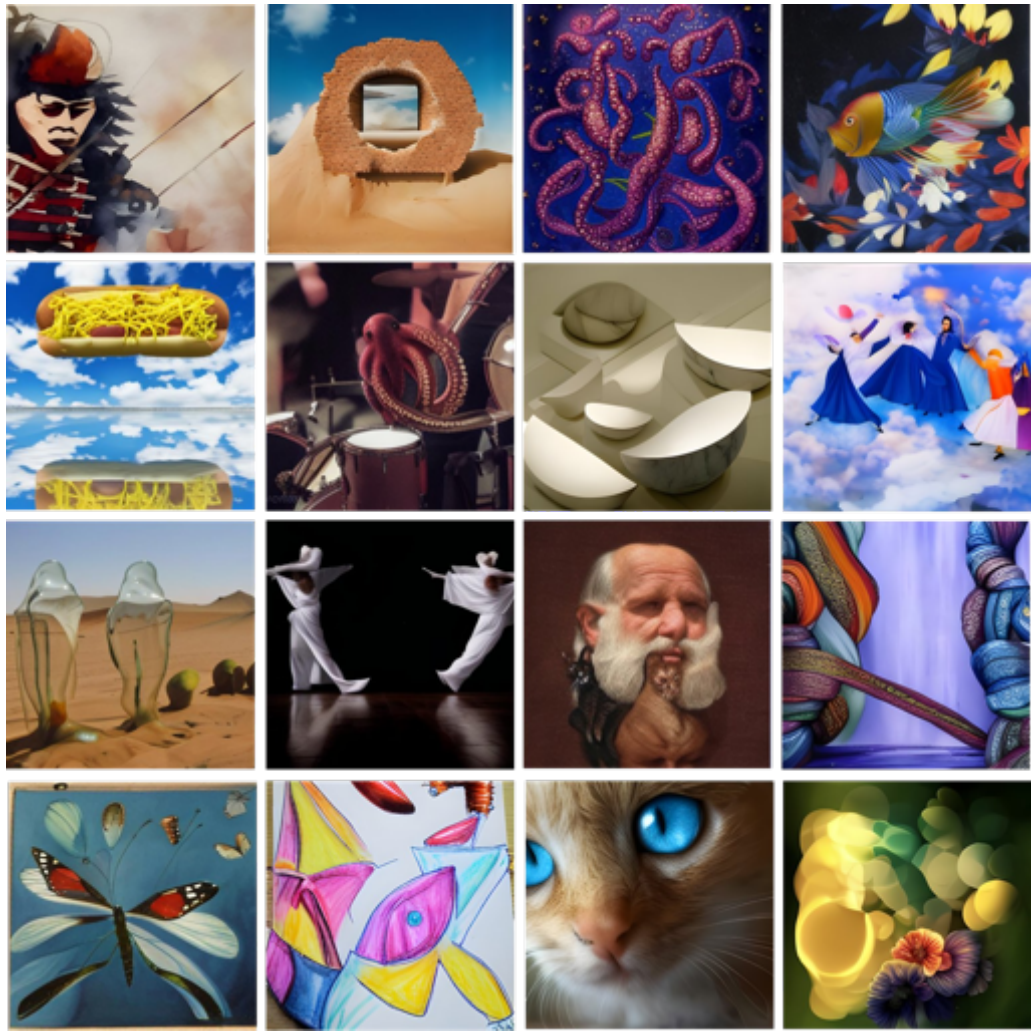


Make-A-Scene: making art with the help of AI

- **1. Type a text description,**
- **2. Draw a sketch**

“A colorful sculpture of a cat”

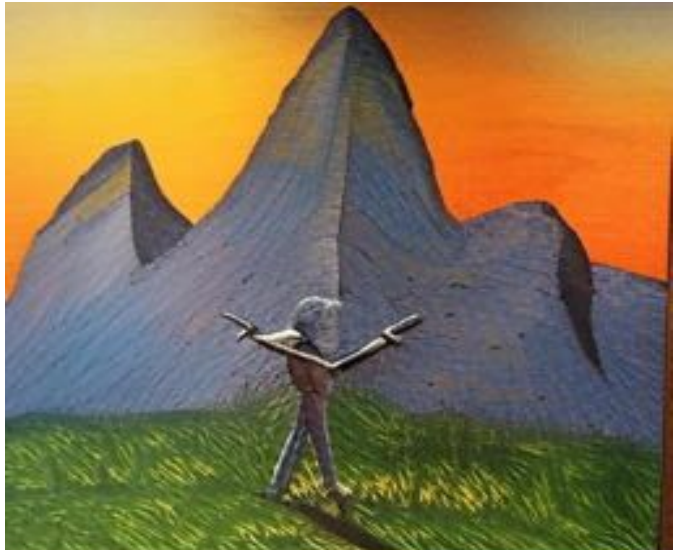




Playing with Make-A-Scene

Drinking a glass of Burgundy by the sea

painting of a physicist on a mountain path watching the sunset, in the style of Van Gogh



FACEBOOK AI

Current ML Sucks!

Where are my self-driving car,
virtual assistant, domestic robot?

Requirements for Future ML/AI Systems

- **Understand the world, understand humans, have common sense**
- **Level-5 autonomous cars**
  - That learn to drive like humans, in about 20h of practice
- **Virtual assistants that can help us in our daily lives**
  - Manage the information deluge (content filtering/selection)
  - Understands our intents, takes care of simple things
  - Real-time speech understanding & translation
  - Overlays information in our AR glasses.
- **Domestic Robots**
  - Takes care of all the chores
- **For this, we need machines near-human-level AI**
  - Machines that understand how the world works

"Her" (2013)

Machine Learning sucks! (compared to humans and animals)

- **Supervised learning (SL) requires large numbers of labeled samples.**
- **Reinforcement learning (RL) requires insane amounts of trials.**
- **SL/RL-trained ML systems:**
  - are specialized and brittle
  - make “stupid” mistakes
- **Machines don’t have common sense**

- **Animals and humans:**
  - Can learn new tasks **very** quickly.
  - Understand how the world works
- **Humans and animals have common sense**

Machine Learning sucks! (plain ML/DL, at least)

- **Machine Learning systems (most of them anyway)**
  - Have a constant number of computational steps between input and output.
  - Do not reason.
  - Cannot plan.

- **Humans and some animals**
  - Understand how the world works.
  - Can predict the consequences of their actions.
  - Can perform chains of reasoning with an unlimited number of steps.
  - Can plan complex tasks by decomposing it into sequences of subtasks

Three challenges for AI & Machine Learning

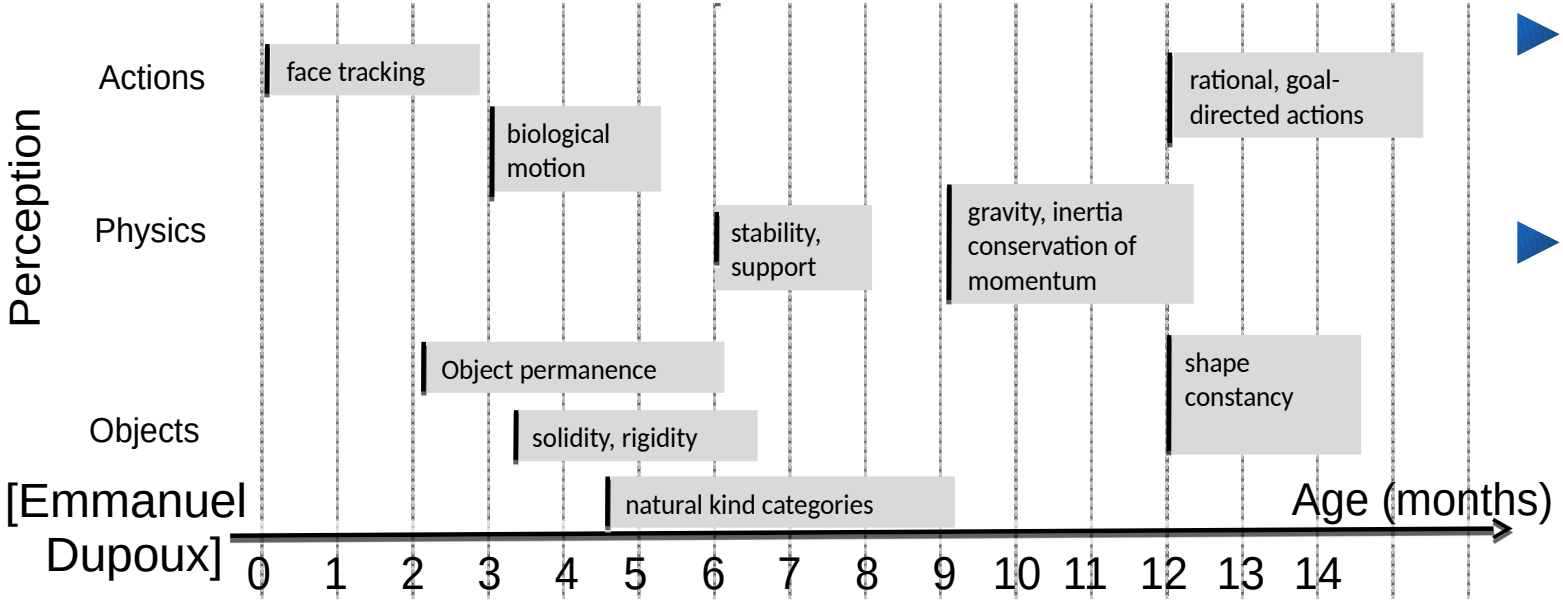
- **1. Learning representations and predictive models of the world**
  - Supervised and reinforcement learning require too many samples/trials
  - **Self-supervised learning** / learning dependencies / to fill in the blanks
    - learning to represent the world in a non task-specific way
    - Learning predictive models for planning and control
- **2. Learning to reason,** like Daniel Kahneman's "System 2"
  - Beyond feed-forward, System 1 subconscious computation.
  - Making reasoning compatible with learning.
    - Reasoning and planning as energy minimization.
- **3. Learning to plan complex action sequences**
  - Learning hierarchical representations of action plans

FACEBOOK AI

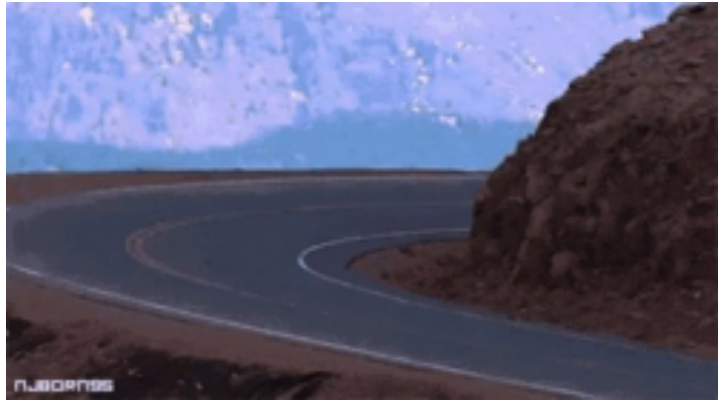
How do humans and animals learn so quickly?

Not supervised.
Not Reinforced.

How could machines learn like animals and humans?



- **How do babies learn how the world works?**
- **How can teenagers learn to drive with 20h of practice?**



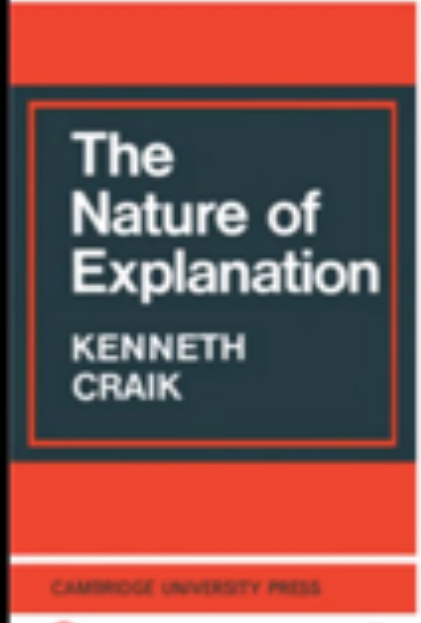
How do Human and Animal Babies Learn?

- **How do they learn how the world works?**
- **Largely by observation, with remarkably little interaction (initially).**
- **They accumulate enormous amounts of background knowledge**
  - About the structure of the world, like intuitive physics.
- **Perhaps common sense emerges from this knowledge?**

Photos courtesy of Emmanuel Dupoux

Common sense is a collection of models of the world

AI systems need to build "mental models"



If the organism carries a `small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it (Craik, 1943,Ch. 5, p.61)

Commonsense is not just facts, it is a collection of models

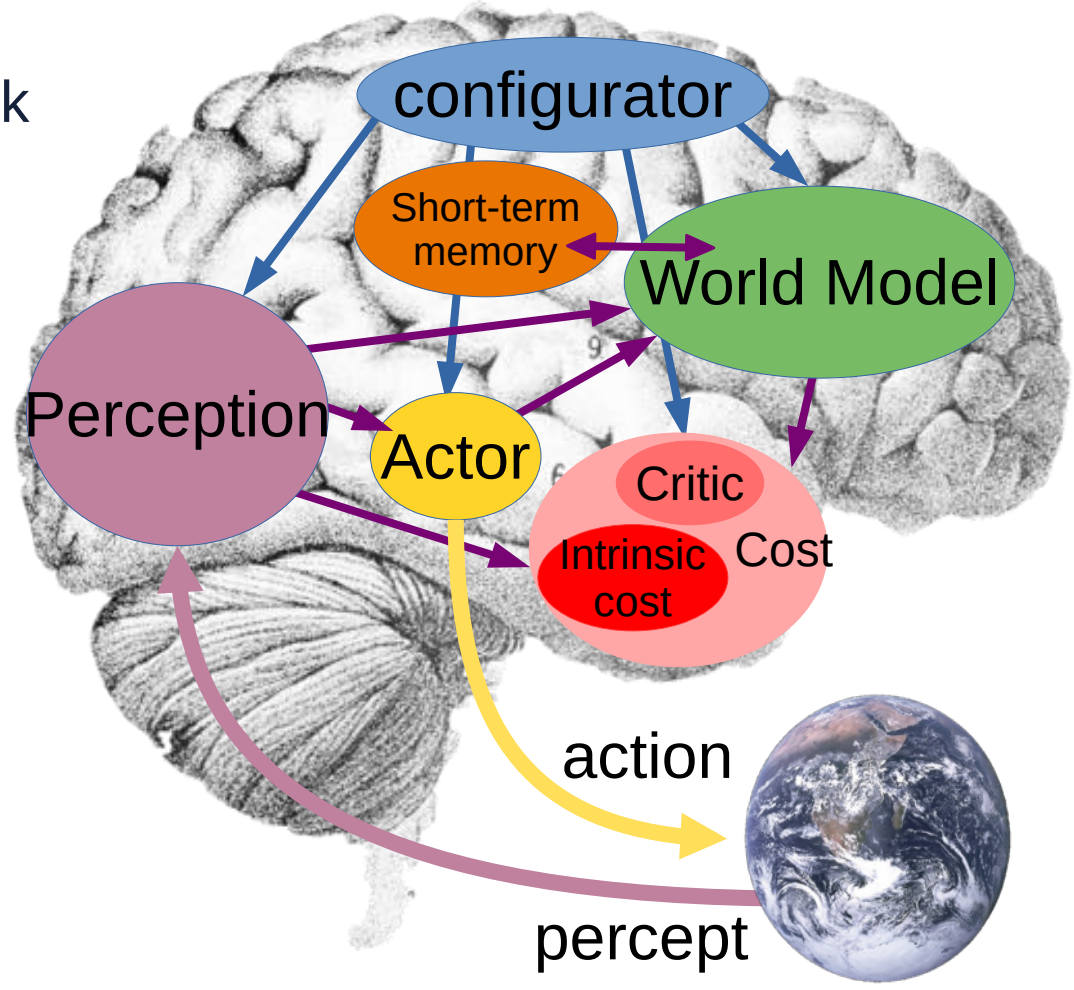
Jitendra Malik

FACEBOOK AI

Architecture of Autonomous AI

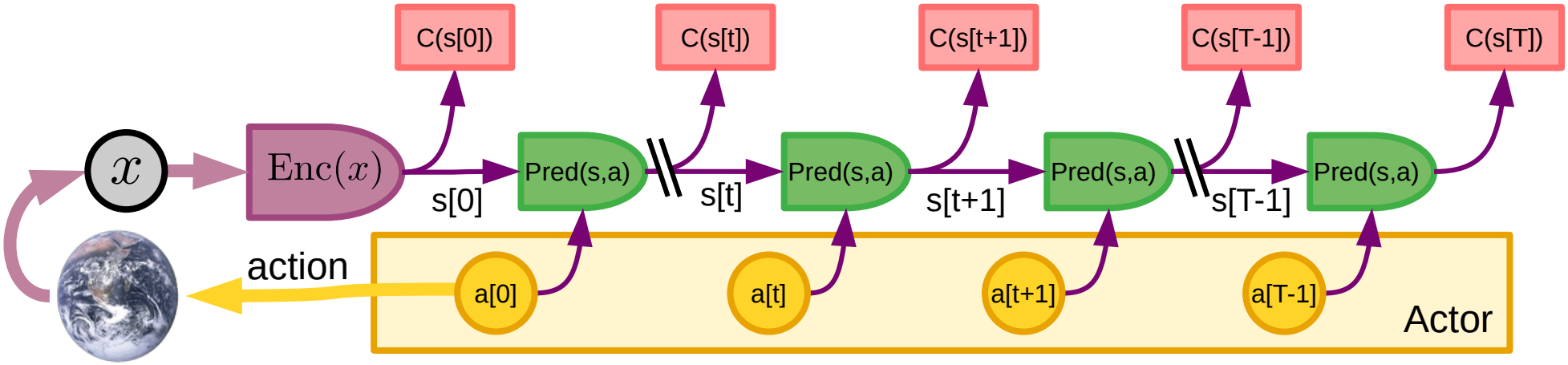
Modular Architecture for Autonomous AI

- **Configurator**
  - Configures other modules for task
- **Perception**
  - Estimates state of the world
- **World Model**
  - Predicts future world states
- **Cost**
  - Compute “discomfort”
- **Actor**
  - Find optimal action sequences
- **Short-Term Memory**
  - Stores state-cost episodes



Mode-2 Perception-Planning-Action Cycle

- **Akin to Model-Predictive Control (MPC) in optimal control.**
- **Actor proposes an action sequence**
- **World Model imagines predicted outcomes**
- **Actor optimizes action sequence to minimize cost**
  - e.g. using gradient descent, dynamic programming, MC tree search…
- **Actor sends first action(s) to effectors**



FACEBOOK AI

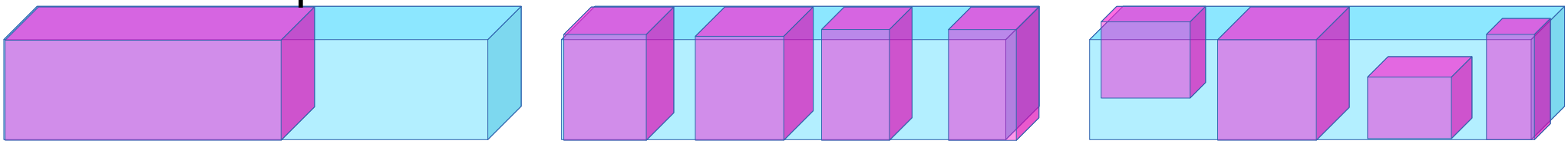
Training the World Model with Self-Supervised Learning

Capturing dependencies between inputs.
Representing uncertainty.

Self-Supervised Learning = Learning to Fill in the Blanks

- **Reconstruct the input or Predict missing parts of the input.**

time or space →

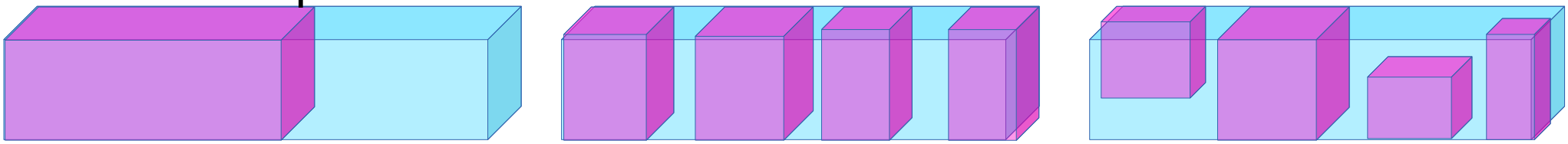


This is a [...] of text extracted [...] a large set of [...] articles

Self-Supervised Learning = Learning to Fill in the Blanks

- **Reconstruct the input or Predict missing parts of the input.**

time or space →



This is a piece of text extracted from a large set of news articles

Two Uses for Self-Supervised Learning

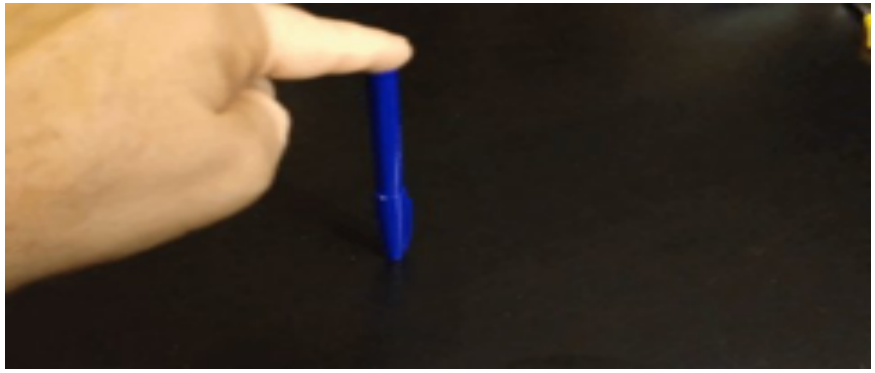
- **1. Learning hierarchical representations of the world**
  - SSL pre-training precedes a supervised or RL phase
- **2. Learning predictive (forward) models of the world**
  - Learning models for Model-Predictive Control, policy learning for control, or model-based RL.
- **Question: how to represent uncertainty & multi-modality in the prediction?**

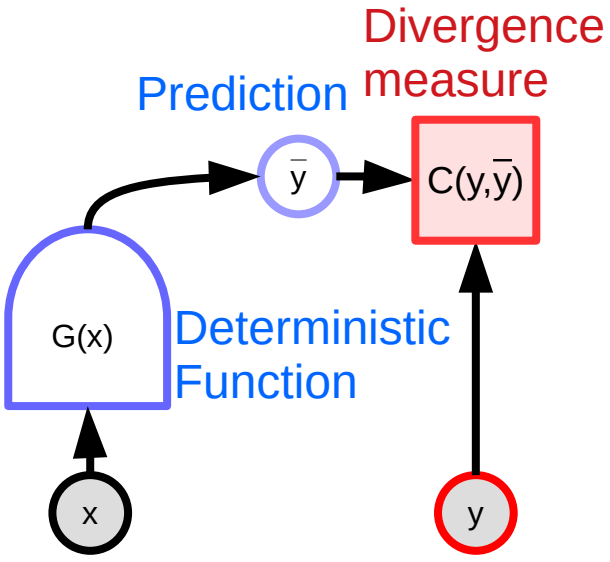
Learning Paradigms: information content per sample

- **"Pure" Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

The world is stochastic

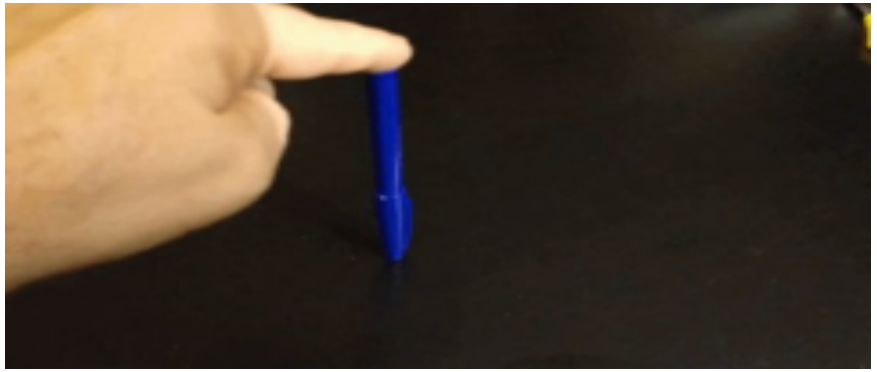
- Training a system to make a single prediction makes it predict the average of all plausible predictions
- **Blurry predictions!**





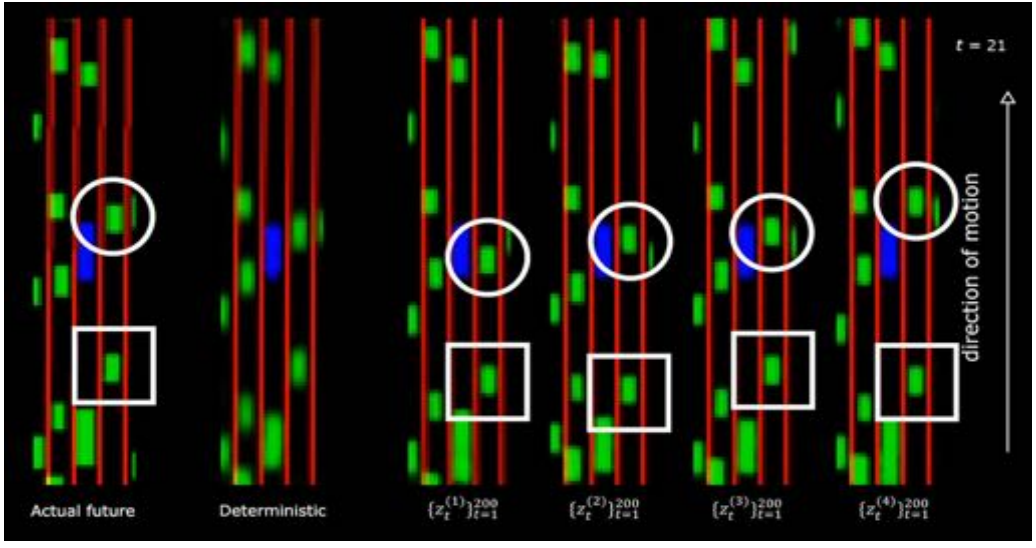
The world is unpredictable. Output must be multimodal.

- Training a system to make a single prediction makes it predict the average of all plausible predictions
- **Blurry predictions!**



How do we represent uncertainty in the predictions?

- **The world is only partially predictable**
- **How can a predictive model represent multiple predictions?**
- **Probabilistic models are intractable in continuous domains.**
- **Generative Models must predict every detail of the world**
- **My solution: Joint-Embedding Predictive Architecture**



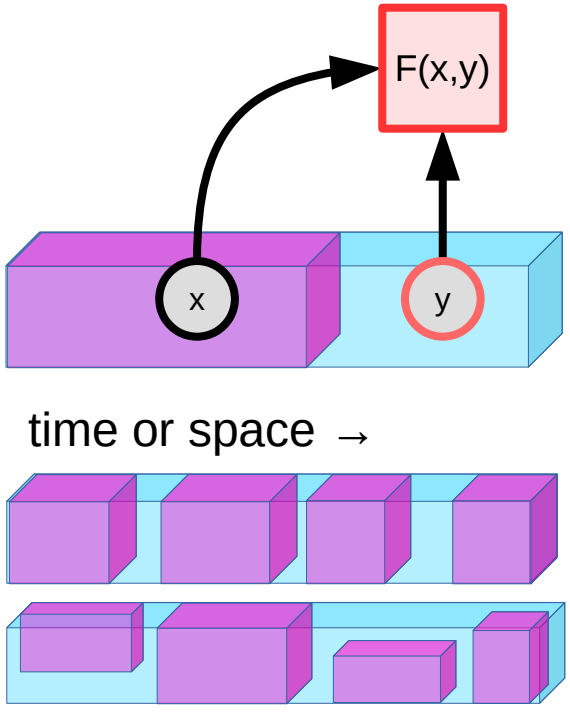
FACEBOOK AI

Energy-Based Models

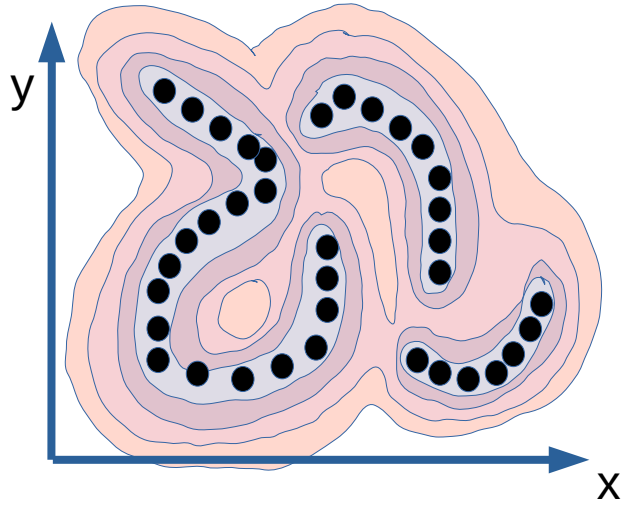
Capture dependencies through an energy function.
See “A tutorial on Energy-Based Learning” [LeCun et al. 2006]

Energy-Based Models: Implicit function

- **Gives low energy for compatible pairs of x and y**
- **Gives higher energy for incompatible pairs**

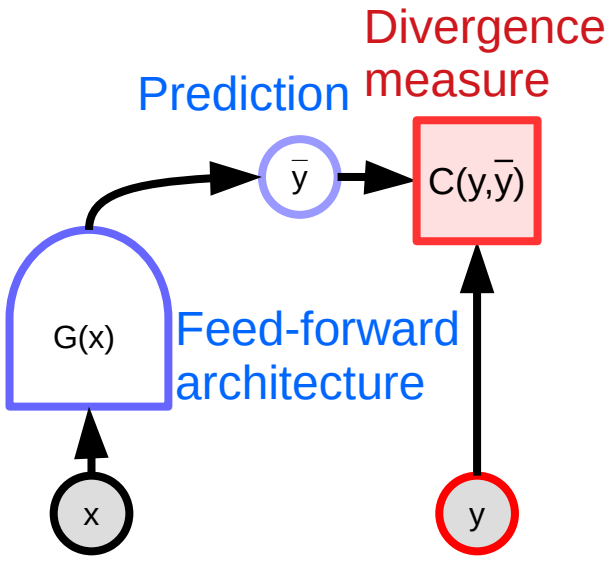


Energy
Function

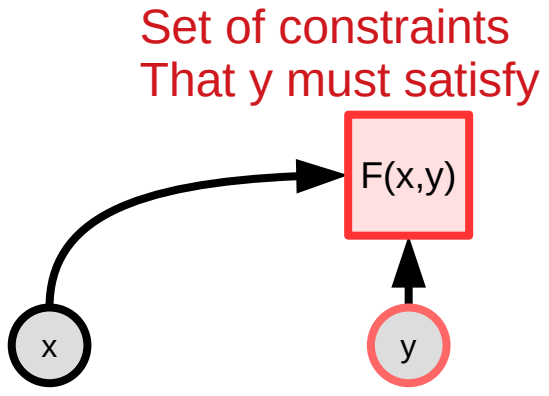


Energy-Based Models

- **Feed-forward nets use a finite number of steps to produce a single output.**
- **What if…**
  - The problem requires a complex computation to produce its output? (complex inference)
  - There are multiple possible outputs for a single input? (e.g. predicting future video frames)



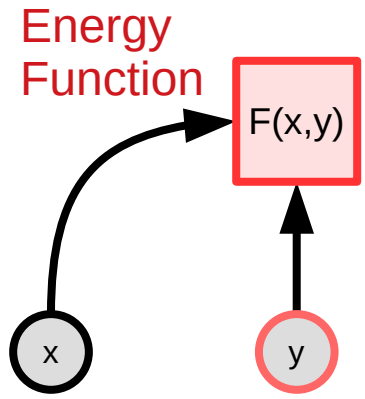
- **Inference through constraint satisfaction**
  - Finding an output that satisfies constraints: e.g a linguistically correct translation or speech transcription.
  - Maximum likelihood inference in graphical models



Energy-Based Models (EBM)

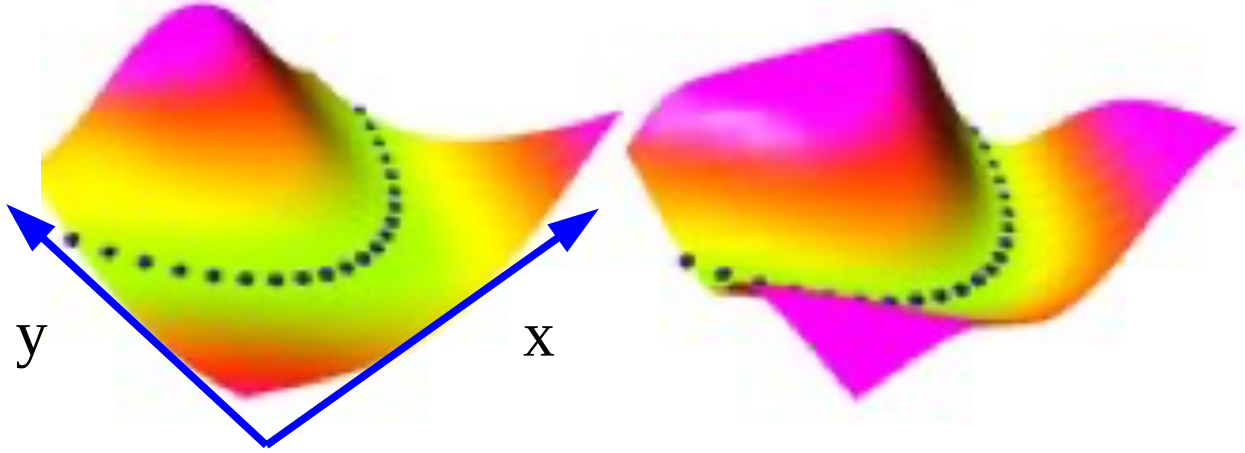
- **Energy function F(x,y) scalar-valued.**
  - Takes low values when y is compatible with x and higher values when y is less compatible with x
- **Inference: find values of y that make F(x,y) small.**
  - There may be multiple solutions

$$\check{y} = \text{argmin}_y F(x, y)$$



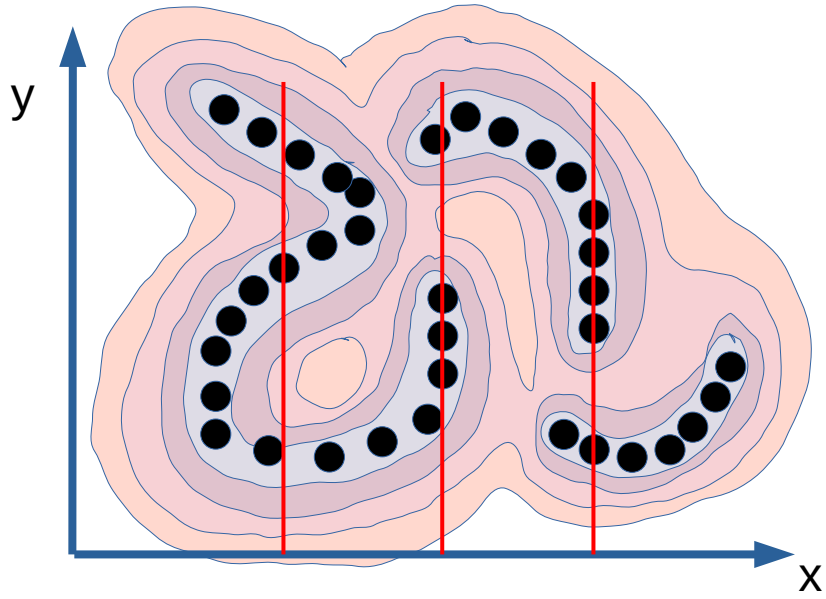
- **Note: the energy is used for inference, not for learning**

- **Example**
  - Blue dots are data points

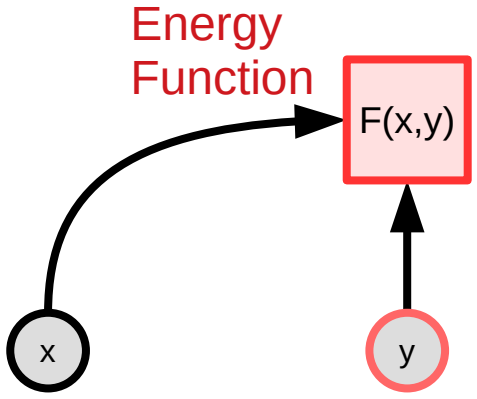


Energy-Based Model: implicit function

- **Energy function that captures the x,y dependencies:**
  - Low energy near the data points. Higher energy everywhere else.
  - If y is continuous, F should be smooth and differentiable, so we can use gradient-based inference algorithms.

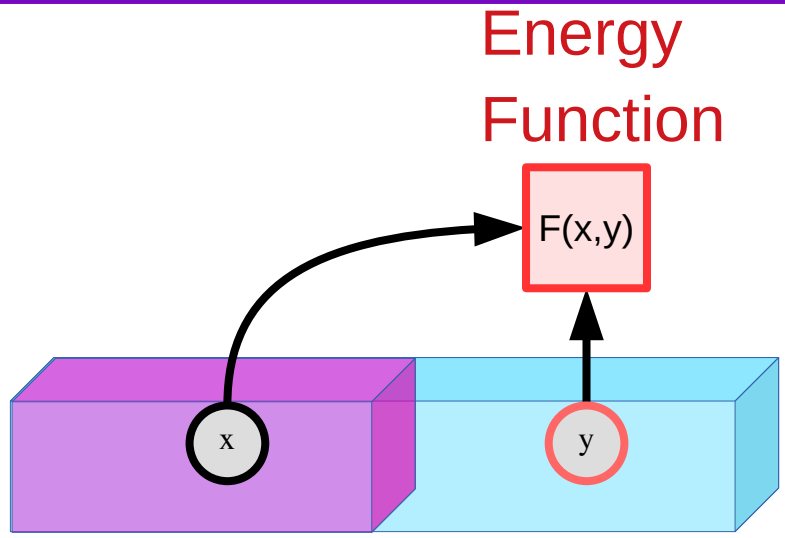


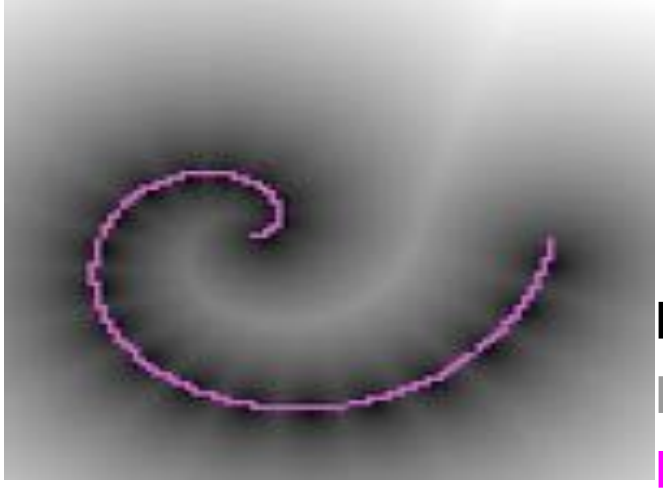
$$\check{y} = \text{argmin}_y F(x, y)$$



Energy-Based Model: unconditional version

- **Conditional EBM: F(x,y)**
- **Unconditional EBM: F(y)**
  - measures the compatibility between the components of y
  - If we don't know in advance which part of y is known and which part is unknown

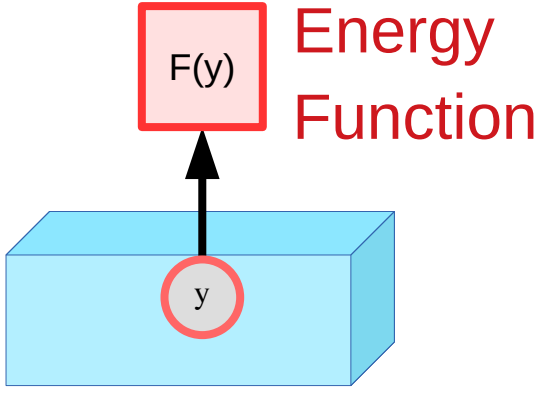




Dark = low energy (good)

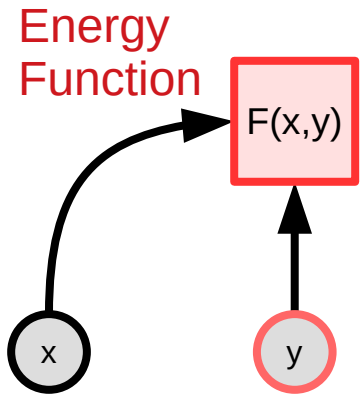
Bright = high energy (bad)

Purple = data manifold



Energy-Based Models vs Probabilistic Models

- **Probabilistic models are a special case of EBM**
  - Energies are like un-normalized negative log probabilities
- **Why use EBM instead of probabilistic models?**
  - EBM gives more flexibility in the choice of the scoring function.
  - More flexibility in the choice of objective function for learning
- **From energy to probability: Gibbs-Boltzmann distribution**
  - Beta is a positive constant

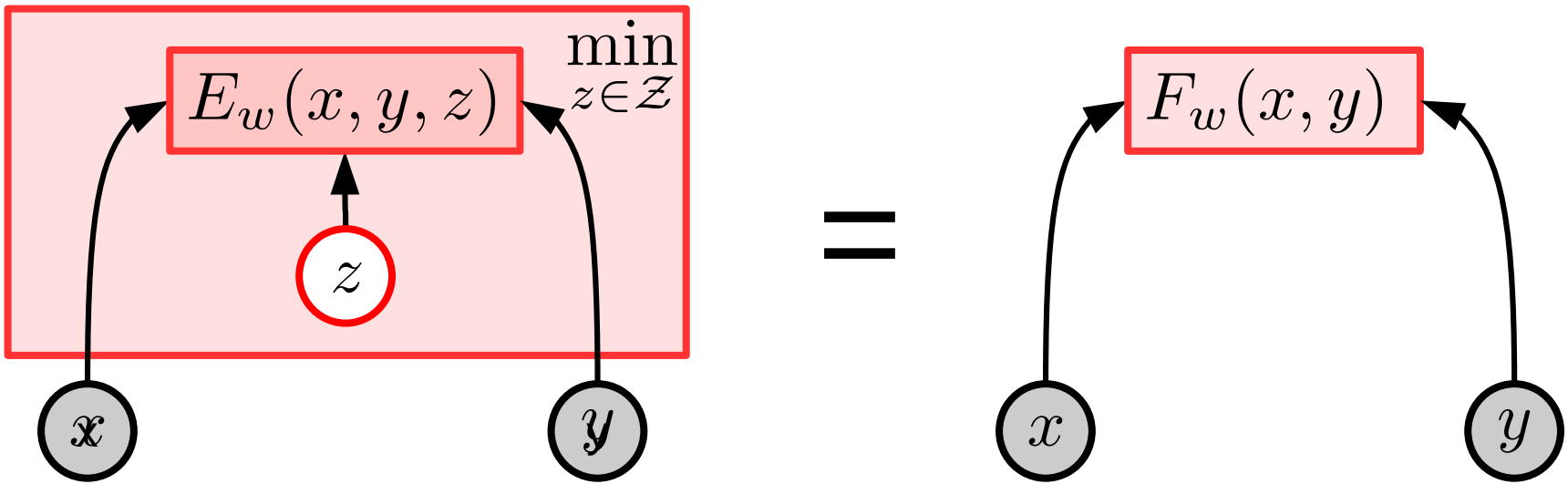


$$P(y|x) = \frac{e^{-\beta F(x,y)}}{\int_{y'} e^{-\beta F(x,y')}}$$

Latent-Variable EBM

- **Latent variable z:**
  - Captures the information in y that is not available in x
  - Computed by minimization

$$\check{z} = \text{argmin}_{z \in \mathcal{Z}} E_w(x, y, z) \qquad F_w(x, y) = E_w(x, y, \check{z})$$



Latent-Variable Generative EBM Architecture

- **Latent variables:**
  - parameterize the set of predictions
- **Ideally, the latent variable represents independent explanatory factors of variation of the prediction.**
- **The information capacity of the latent variable must be minimized.**
  - Otherwise all the information for the prediction will go into it.