Evolution of software for data processing and analysis

Andrey Ustyuzhanin
NRU Higher School of Economics
Yandex School of Data Analysis
ICL

Evolution of Science (abridged)

1000+ years - Empirical (Aristotle, Democritus,)
100+ years – theoretical (Newton, Kepler,)
50+ years – computational (John von Neumann)
10+ years – data-driven science (the “Fourth paraditm”, Jim Gray)

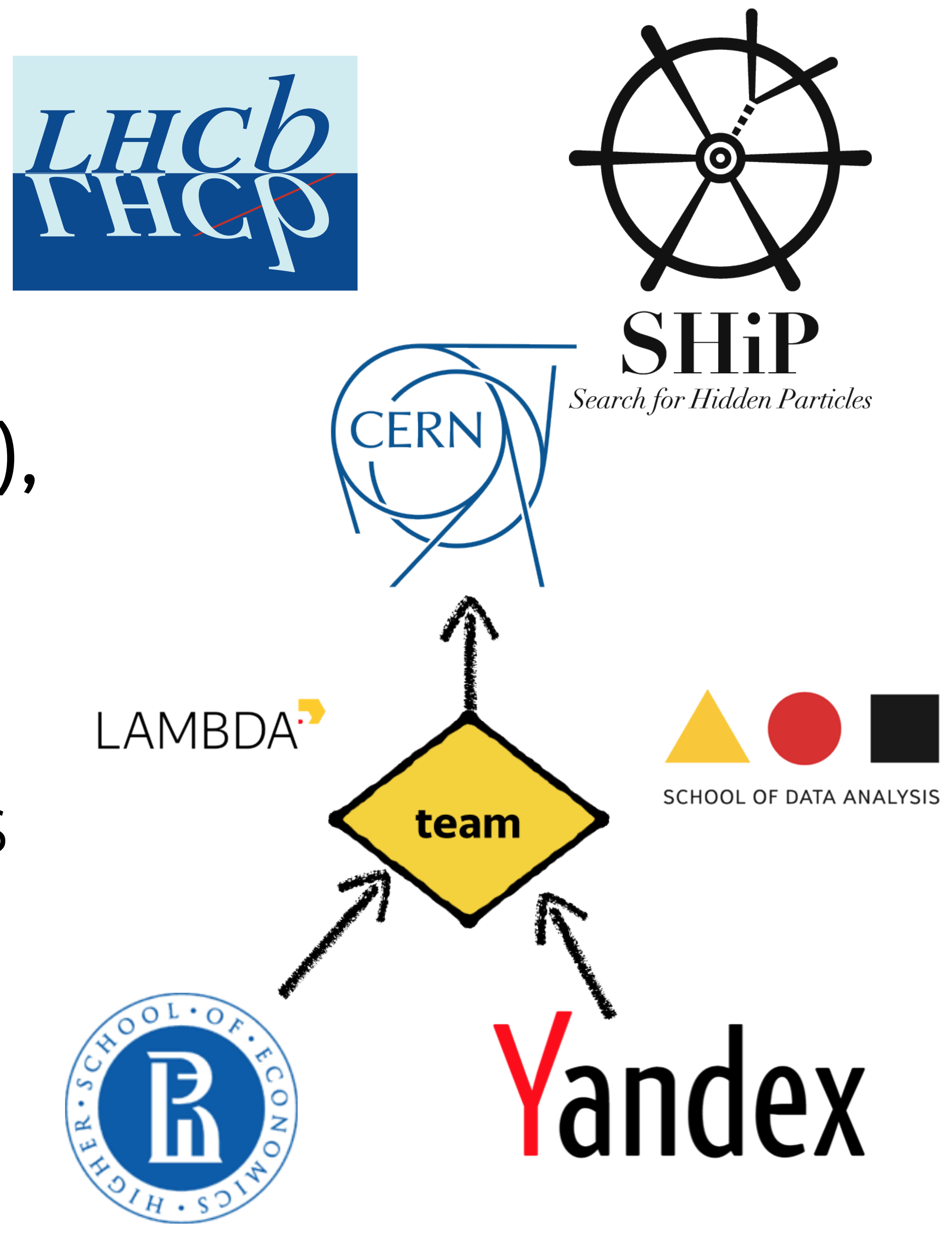
- Unify theory, experiment and simulation
- Data is captured or simulated
- Data is processed and analysed by software
- Information/knowledge is stored in computer
- Scientists analyzes database/files using data management and statistics

Less than 10 years – community-driven trend (open science, crowd-science)

Particle Physics is heavily data-driven

- Data processing
  - High-level triggers
  - Tracking
  - Particle Identification
  - Event Reconstruction
  - Signal/background separation
- Simulation
  - Pythia tuning
  - GEANT speed-up
- Detector optimization (“bang per bucks”)
  - Sensitivity / resolution
  - Cost

Quick self-intro



Head of LHCb team in Yandex School of Data Analysis (YSDA)
Head of Laboratory (link) at Higher School of Economics (HSE),

YSDA (since 2007):

- Joint master's degree in data science
- Solving High Energy Physics problems with ML approaches
- member of LHCb, SHiP, CRAYFIS

HSE, Laboratory (since 2015):

- focuses on applying ML to **natural science challenges**
- HSE has joined LHCb this summer!
- Collaborates with industry as well

Education activities (MLHEP, ML at ICL,ClermonFerrand, LaSAL, Coursera)

Machine Learning very quick intro (simplified)

Machine Learning core idea:

- Field of computer science that uses statistical techniques to give computers the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed;

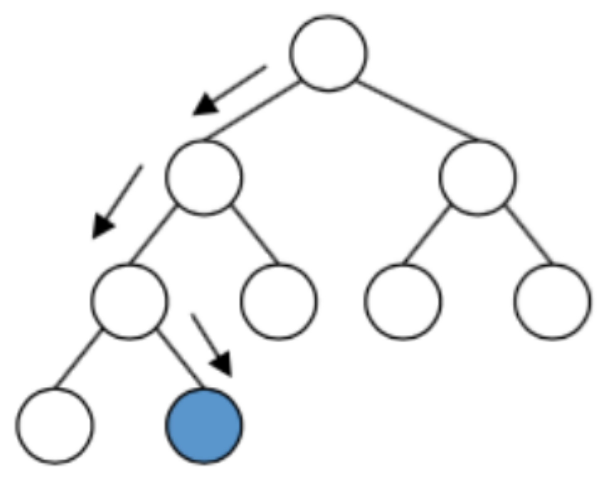
Data Science:

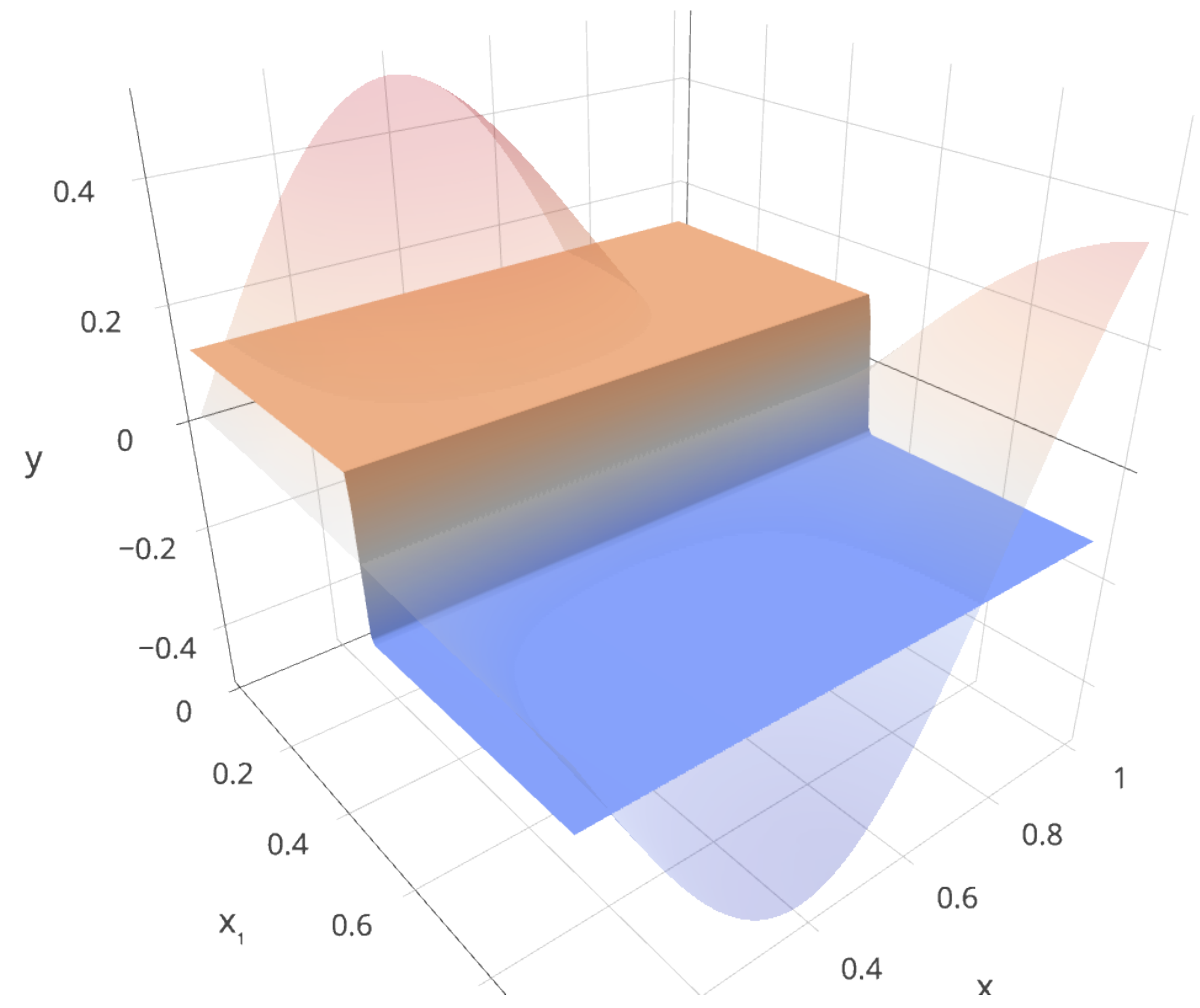
- Interdisciplinary branch of science with focus on tools and methods suitable for knowledge extraction from the raw data;

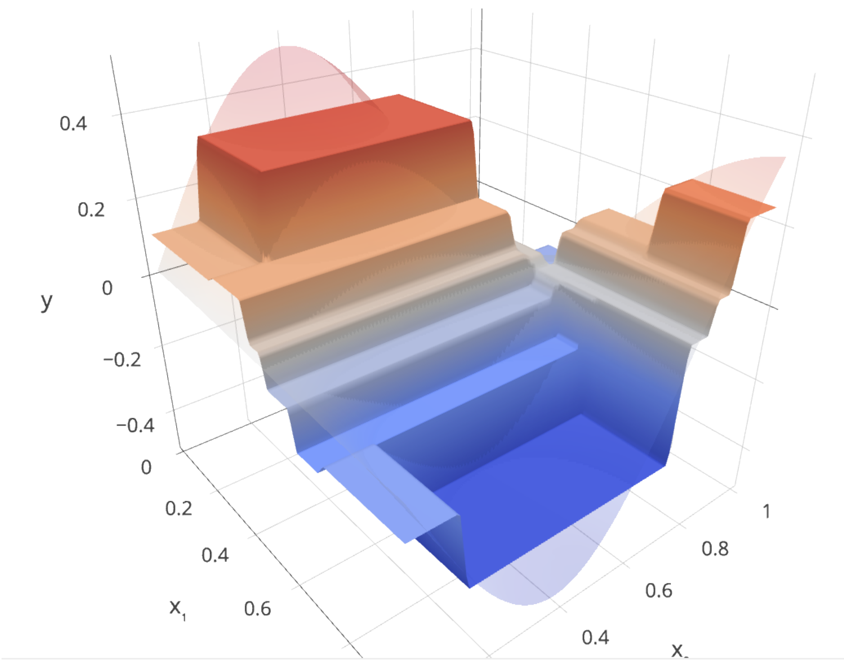
Datasets: digital (simplified) representation of our world;
Popular problem categories: supervised, unsupervised, reinforcement;
Metric – proxy of a quality we can expect from such an algorithm;
Optimization methods + learning theory: to reach prediction error minima fast;
Boosted by a) computing power, b) huge datasets (images, texts, speech, etc.)

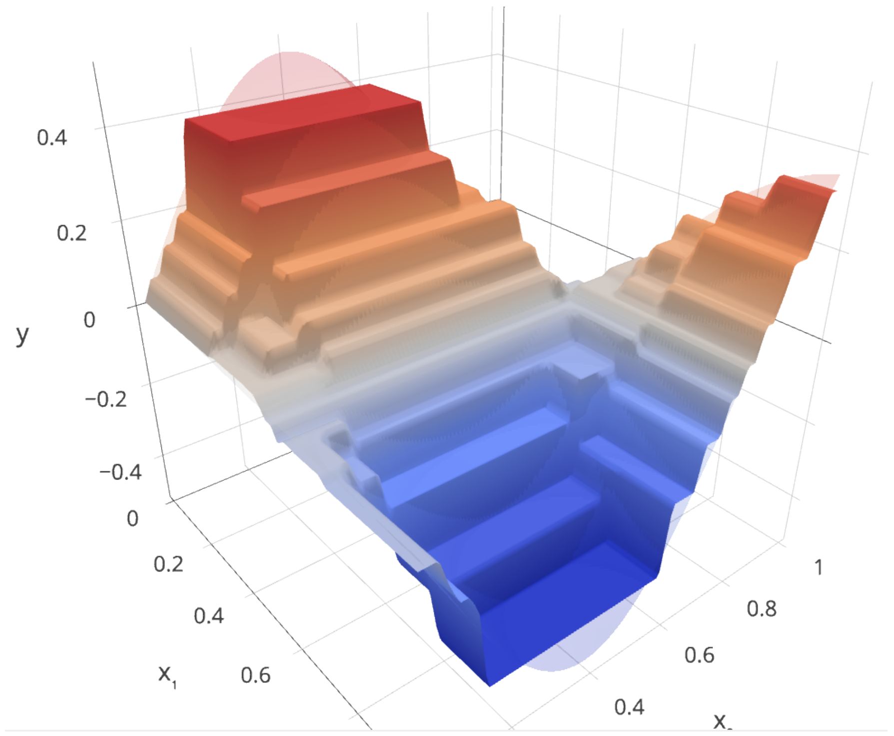
An example – boosting over decision trees

Decision tree (depth in [1, 4, 6])



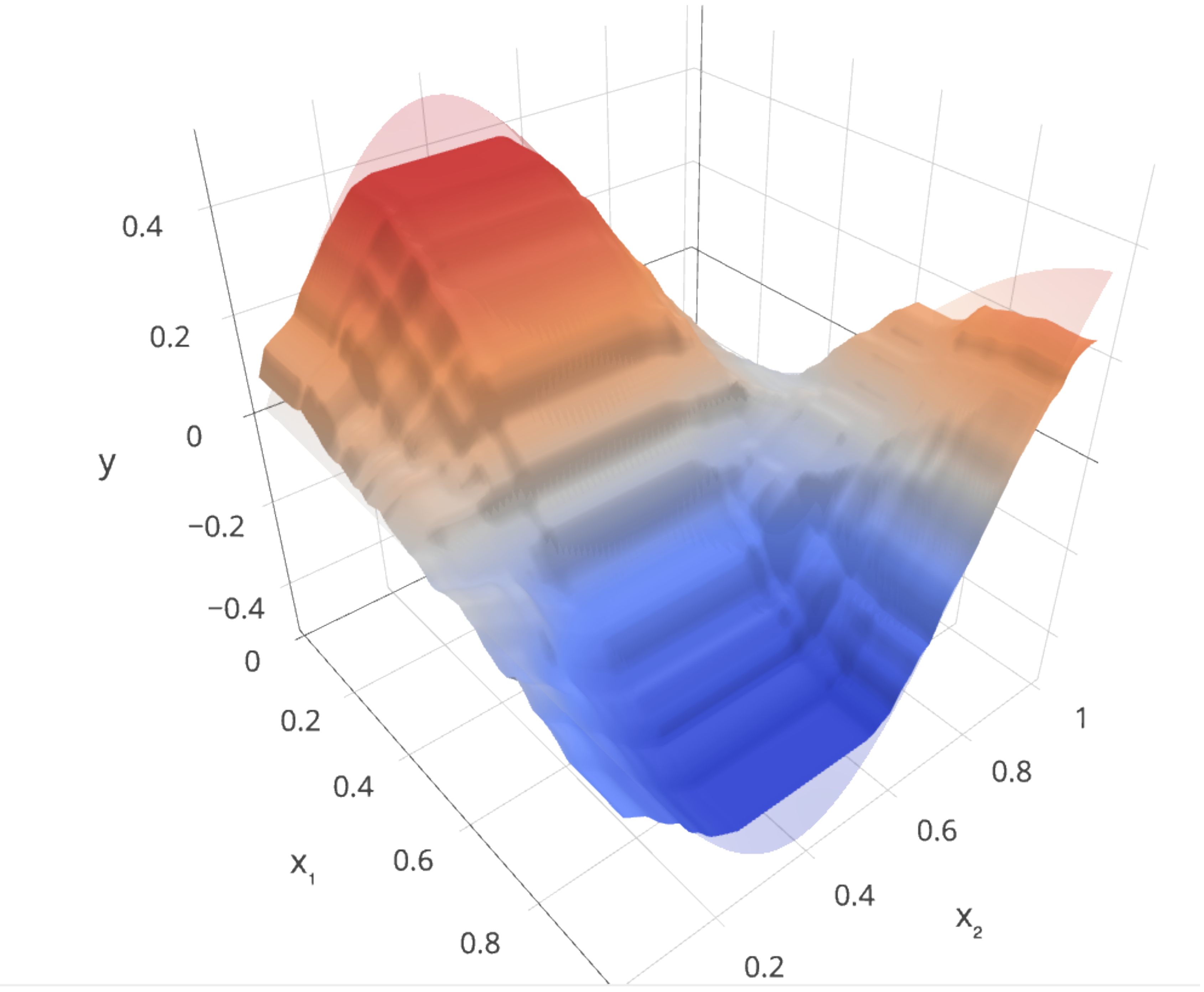






Tree ensemble (Boosted Decision Trees)

$$D^{(3)}(\mathbf{x}) = T_1(\mathbf{x}) + T_2(\mathbf{x}) + T_3(\mathbf{x})$$



$$T_4(\mathbf{x}) \leftarrow \alpha(f(\mathbf{x}) - D^3(\mathbf{x}))$$

http://bit.ly/2mtVewf

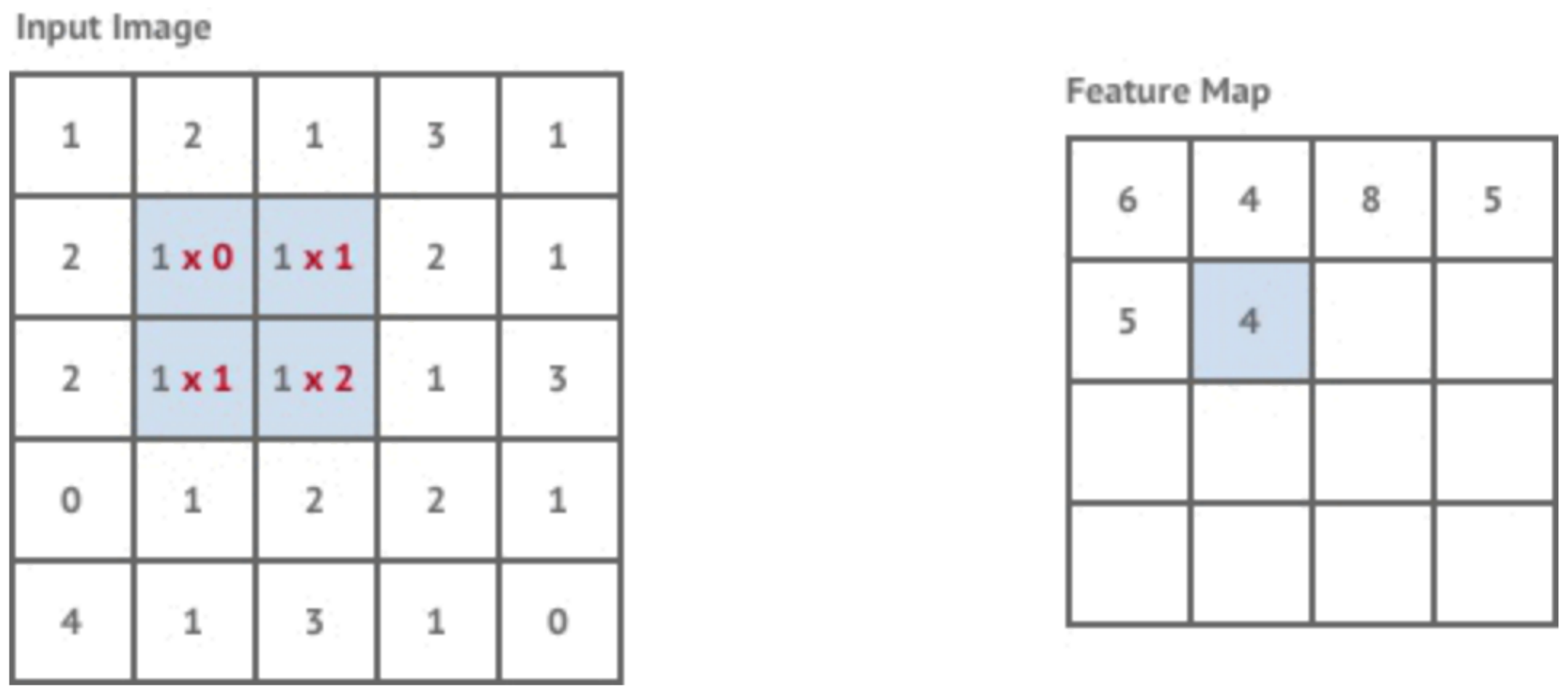
Machine Learning history (abridged)

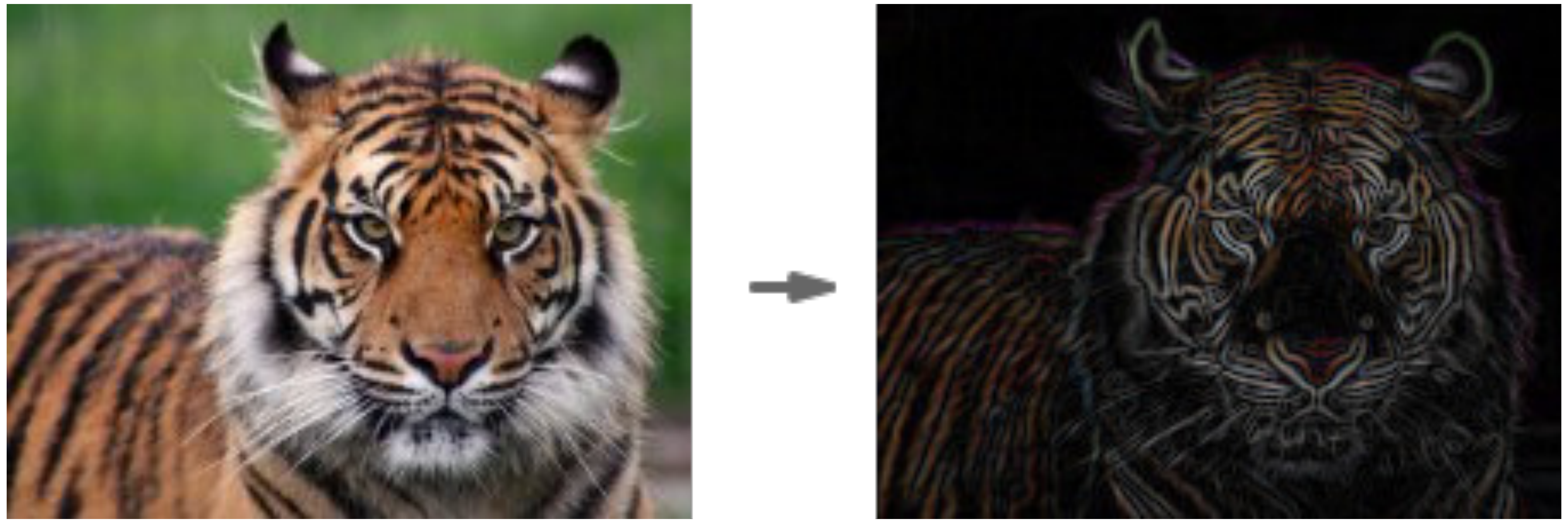
- Has passed several development stages since 1950-s
  - Expert systems
  - Knowledge representations
  - Connectioninst systems (backpropagation)
  - Data mining
- Has been adopted by several branches of domain sciences (including Physics)
- Boosted by deep learning in 2010-s field – i.e. neural-network-like processing

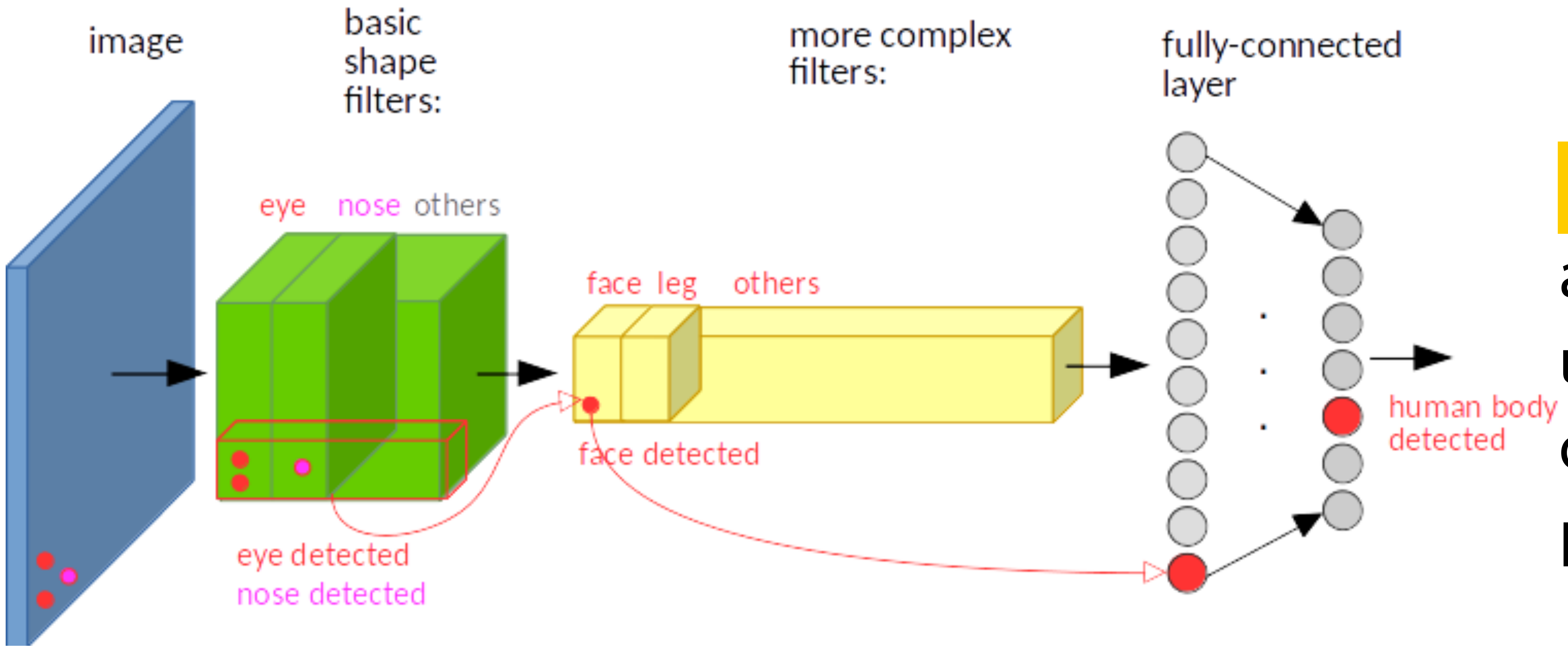
Relation to statistics: Leo Breiman distinguished two statistical modelling paradigms: data model and algorithmic model, wherein "algorithmic model" means more or less the machine learning algorithms.

https://wiki2.org/en/Machine_learning

Convolutional neural network (NN) for vision



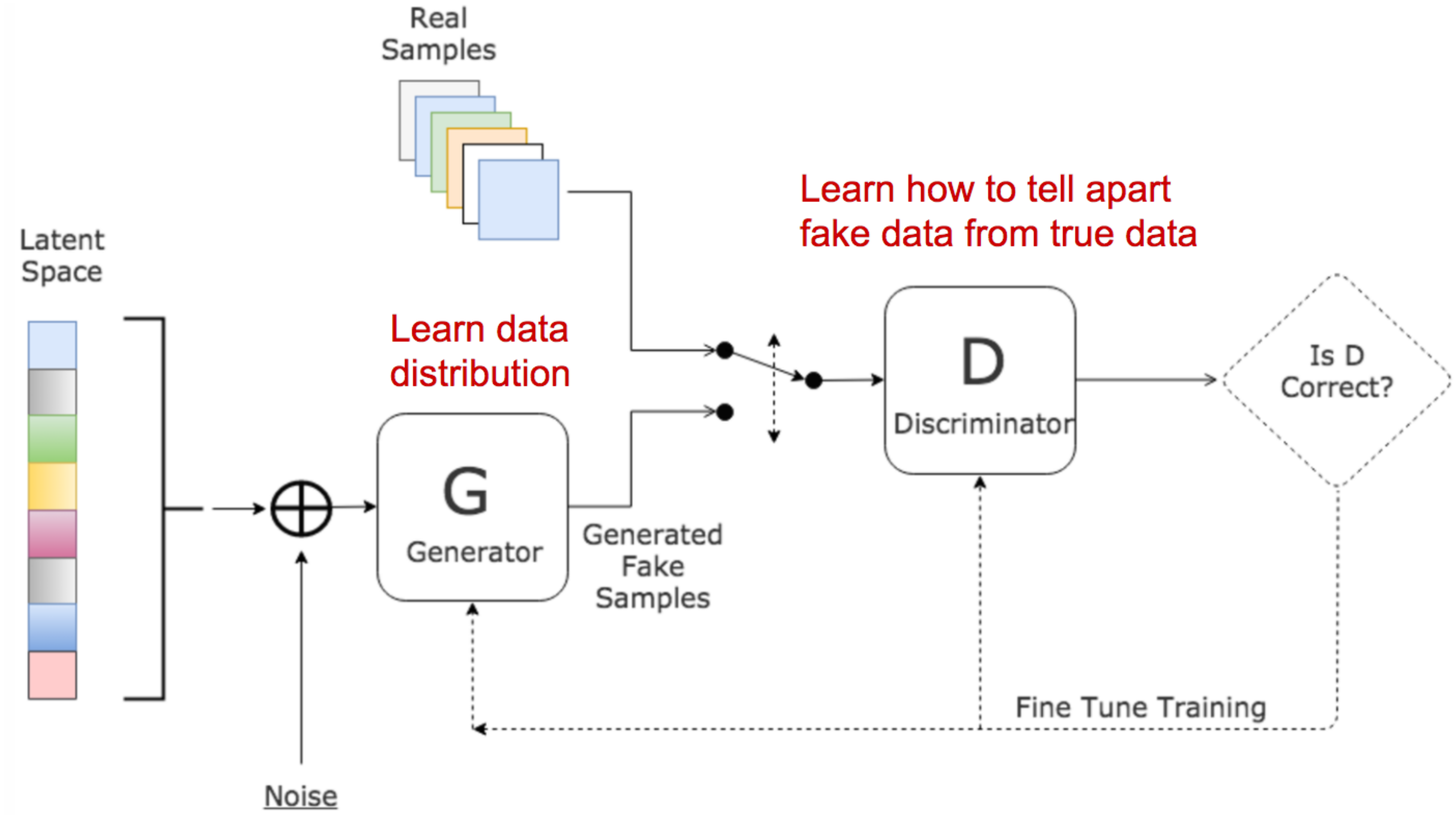




All data transformations are differentiable, so we can use gradient to adjust NN coefficients to mimize the prediction error

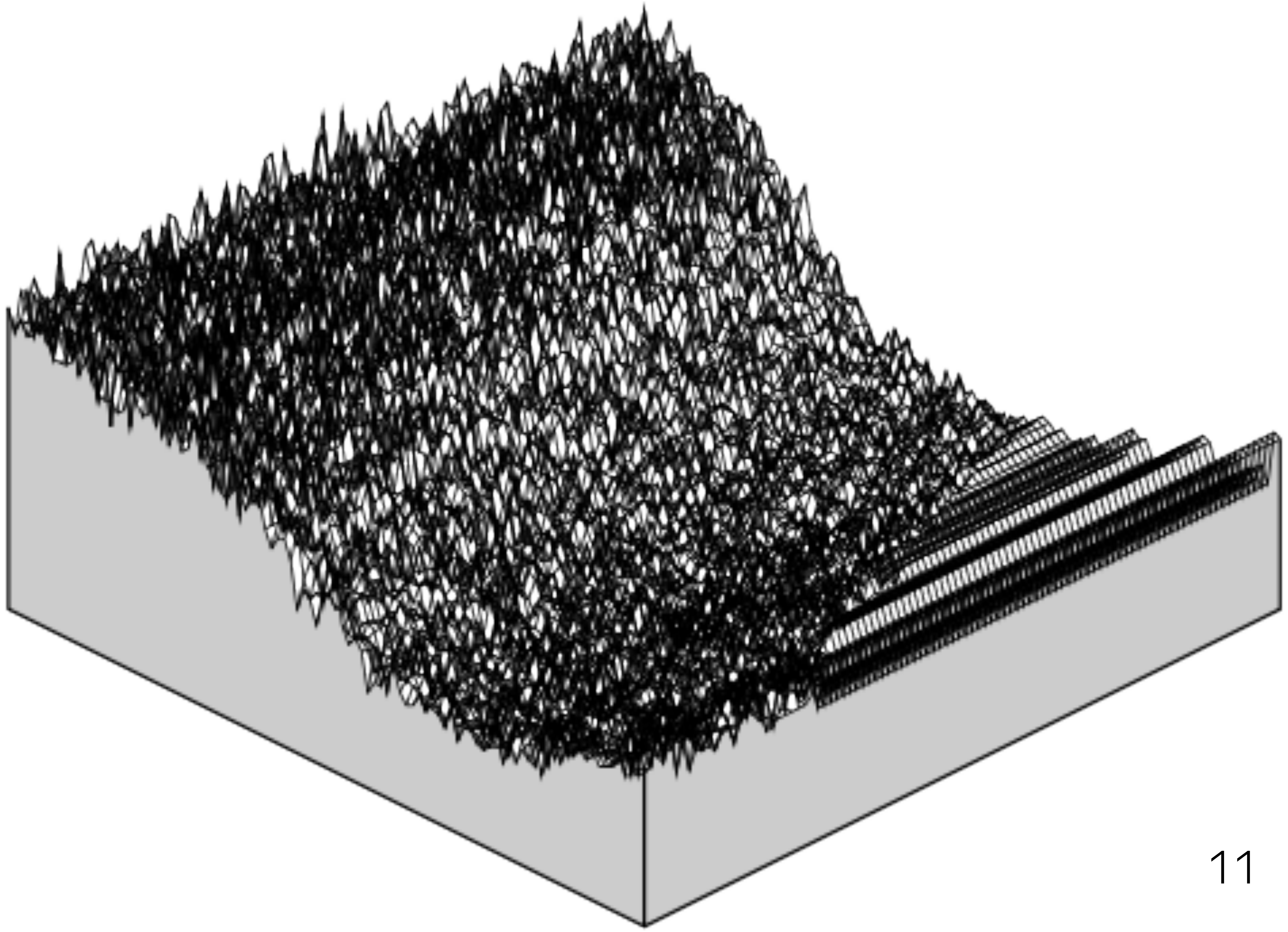
http://bit.ly/2muXG5Z http://bit.ly/2uMSSwl

Generative Adversarial Networks (GAN)



Optimize sophisticated functions

- We'd better prepare for the following challenges:
  - Non-smooth or even discontinuous objective
  - Multi-modality (i.e. oscillating wrt features)
  - Noise-resistance
  - High dimensionality, e.g., $d \gg 1000$
  - Constraints (possibly non-smooth, ...)
- Approaches that work:
  - Kriging
  - Baeysian optimization
  - Natural evolution strategy
- HEP applications:
  - Design of detectors
  - MC generator parameter tuning
  - Planning of experiment



Deep Learning summary

- Versatile way (language) for expressing solution templates for data-driven problems, e.g.:
  - Shift-invariance – convolutional layer
  - Sequence memory for – recurrent layer
- Sophisticated optimization methods
- Now driven mainly by industry (Apple, Goole, Facebook, Amazon, Microsoft)
- Is not going away anytime soon
- Prominent challenges:
  - Model Inrepretation
  - Few-shot learning

ML examples in HEP

(abridged and **heavily** biased)

LHCb Trigger Optimization for Run-II

40 MHz bunch crossing rate

↓

L0 Hardware Trigger:
1 MHz readout, high E_T/P_T signatures

- 450 kHz $h^{\pm}$
- 400 kHz $\mu/\mu\mu$
- 150 kHz e/γ

↓

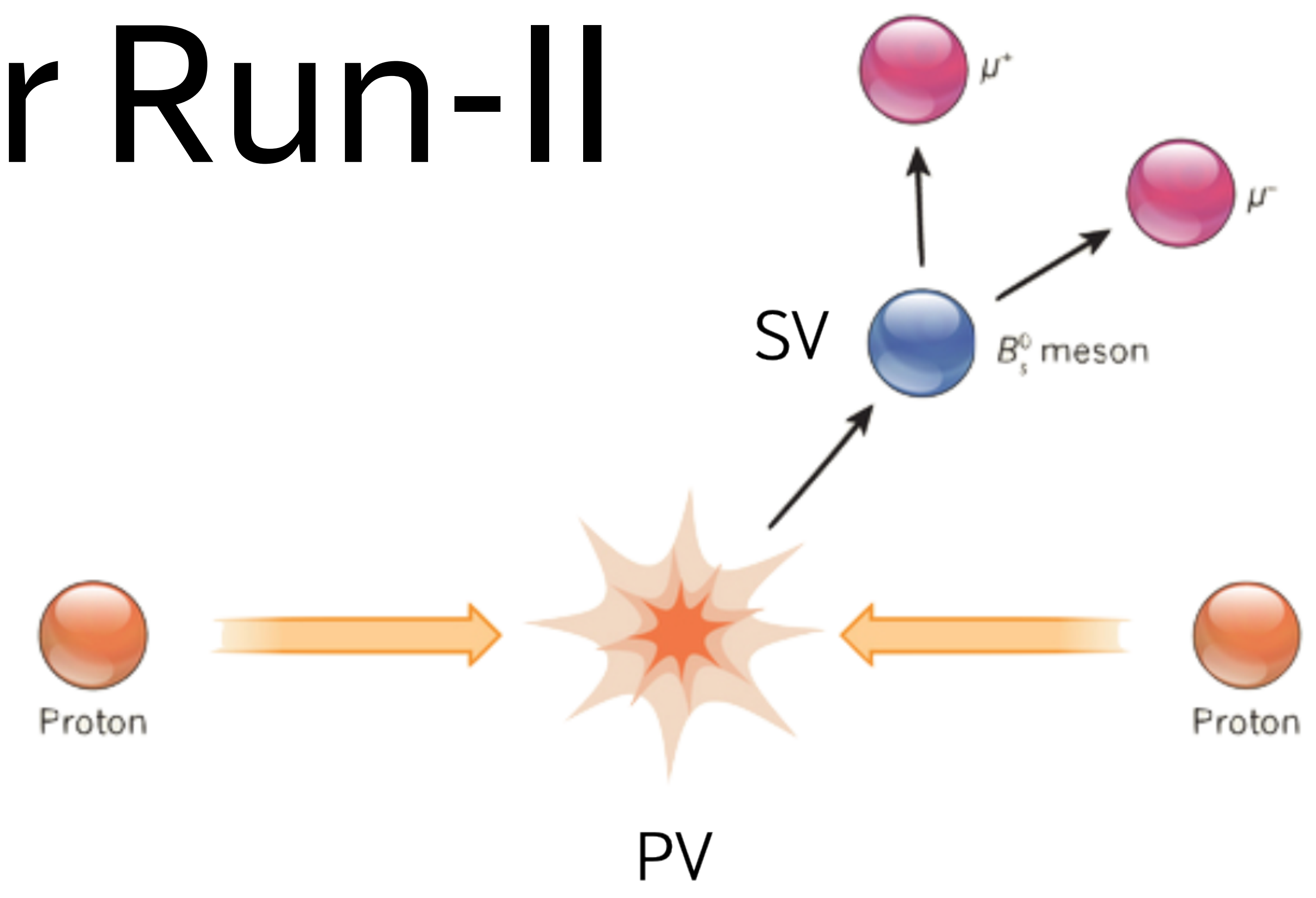
Software High Level Trigger
- 29000 Logical CPU cores
- Offline reconstruction tuned to trigger time constraints
- Mixture of exclusive and inclusive selection algorithms

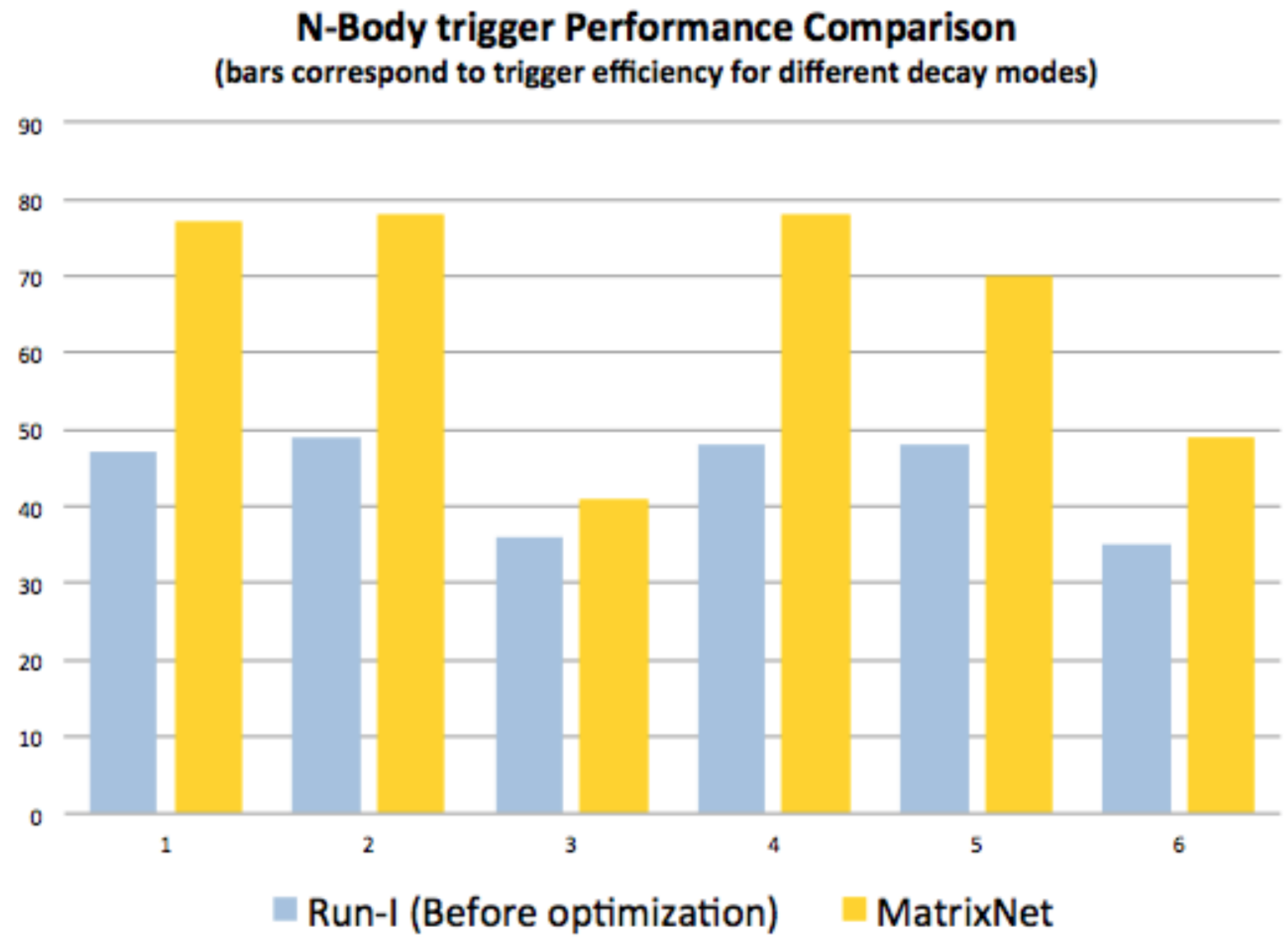
↓

5 kHz Rate to storage

- 2 kHz Inclusive Topological
- 2 kHz Inclusive/ Exclusive Charm
- 1 kHz Muon and DiMuon

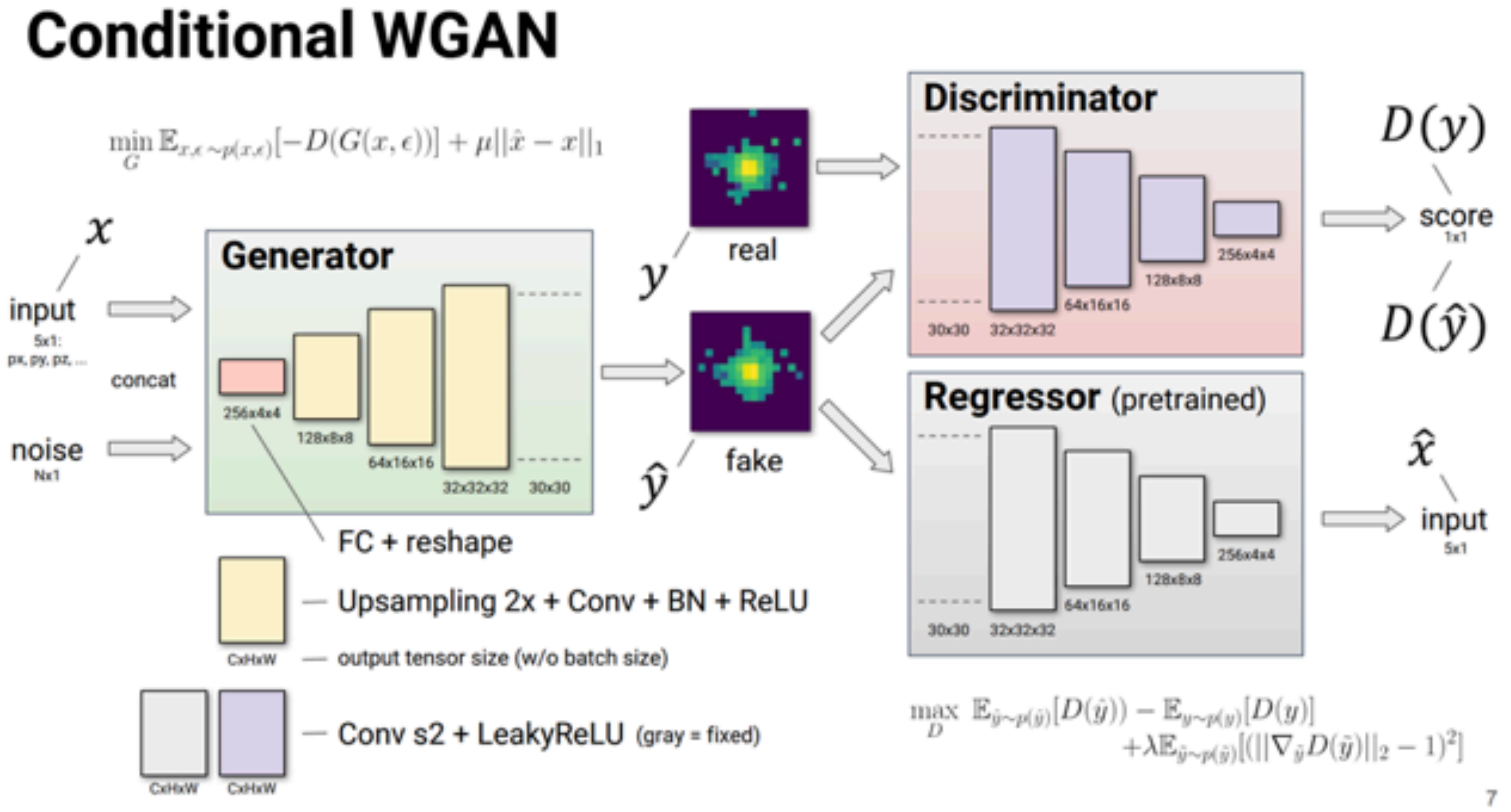
http://stacks.iop.org/1742-6596/664/i=8/a=082025





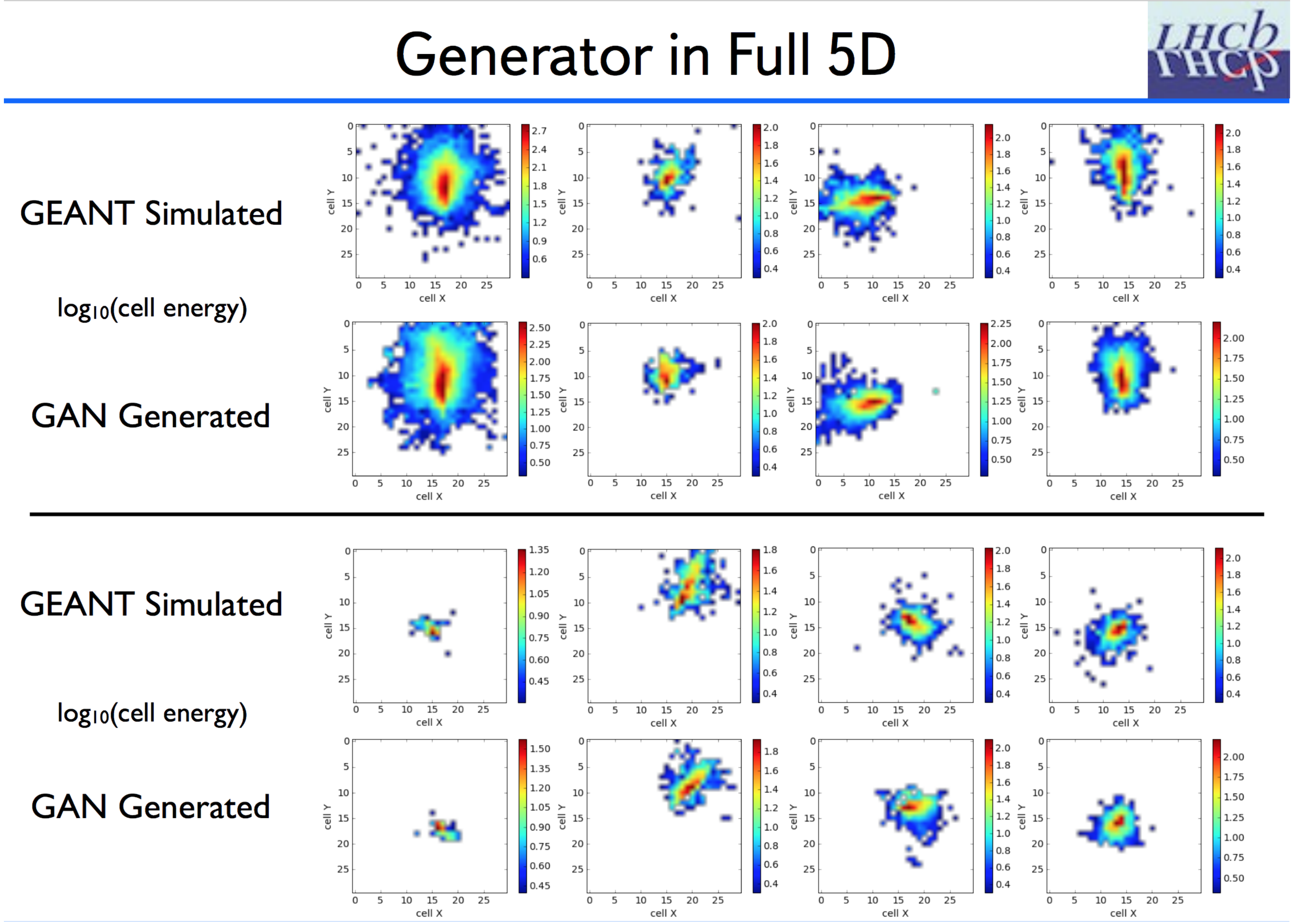
LHCb Fast Calorimeter Response Generation

- Goal: approximate Resource Consuming simulation by NN trained by WGAN
- Demand for simulation will increase 100x



http://bit.ly/2JDM9tV

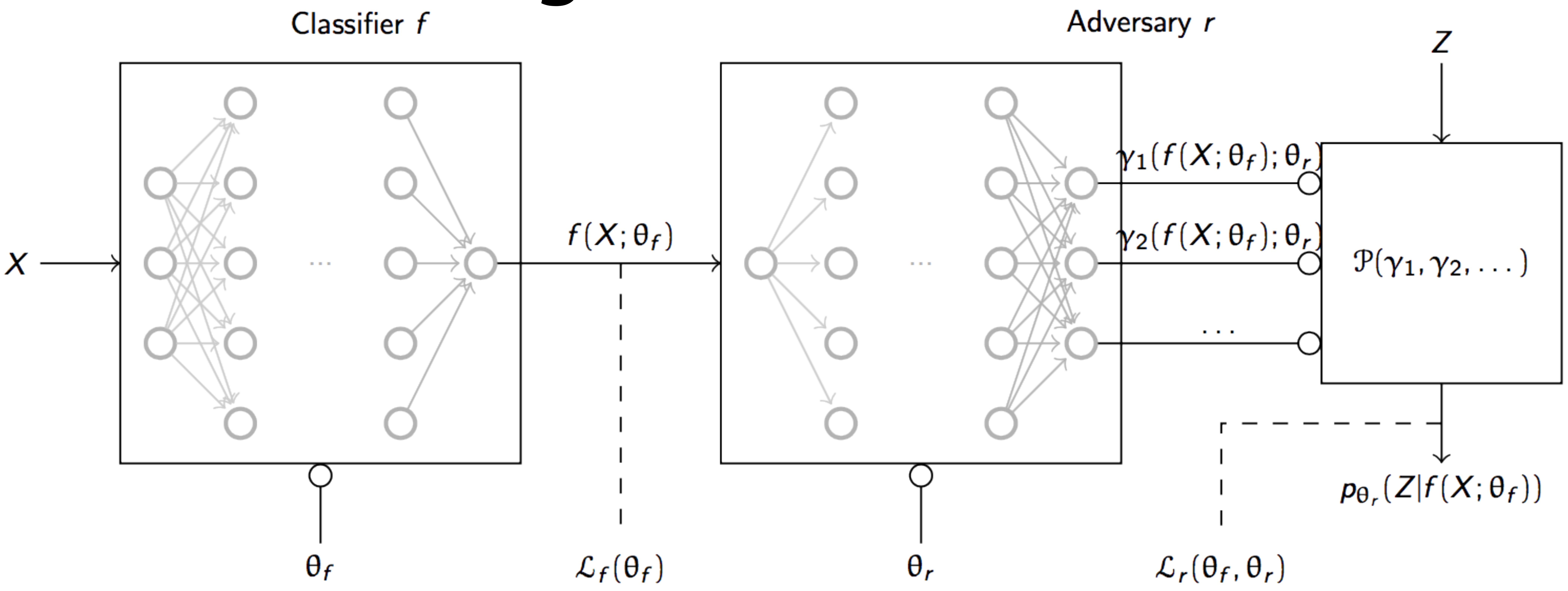
LHCb Fast Calorimeter Response Generation



- Goal: approximate Resource Consuming simulation by NN trained by WGAN
- Demand for simulation will increase 100x

http://bit.ly/2JDM9tV

GANs for training uncorrelated models



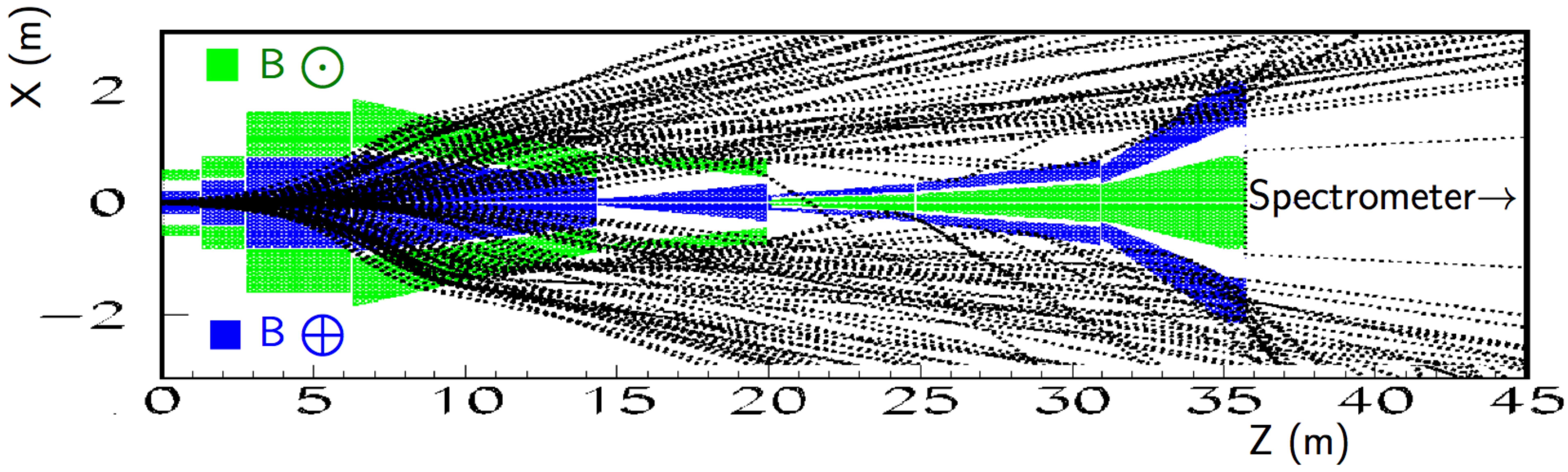
$$\hat{\theta}_f, \hat{\theta}_r = \arg\min_{\theta_f} \max_{\theta_r} \mathcal{L}_f(\theta_f) - \mathcal{L}_r(\theta_f, \theta_r)$$

Adversary part identifies PDF parameters that can be used to infer Z (decor. feature) from *f*.

Intuitively, *r* penalizes *f* so it is impossible to reconstruct Z.

G. Louppe et al, http://bit.ly/2GejPgY

SHiP Active Muon Shield Optimization

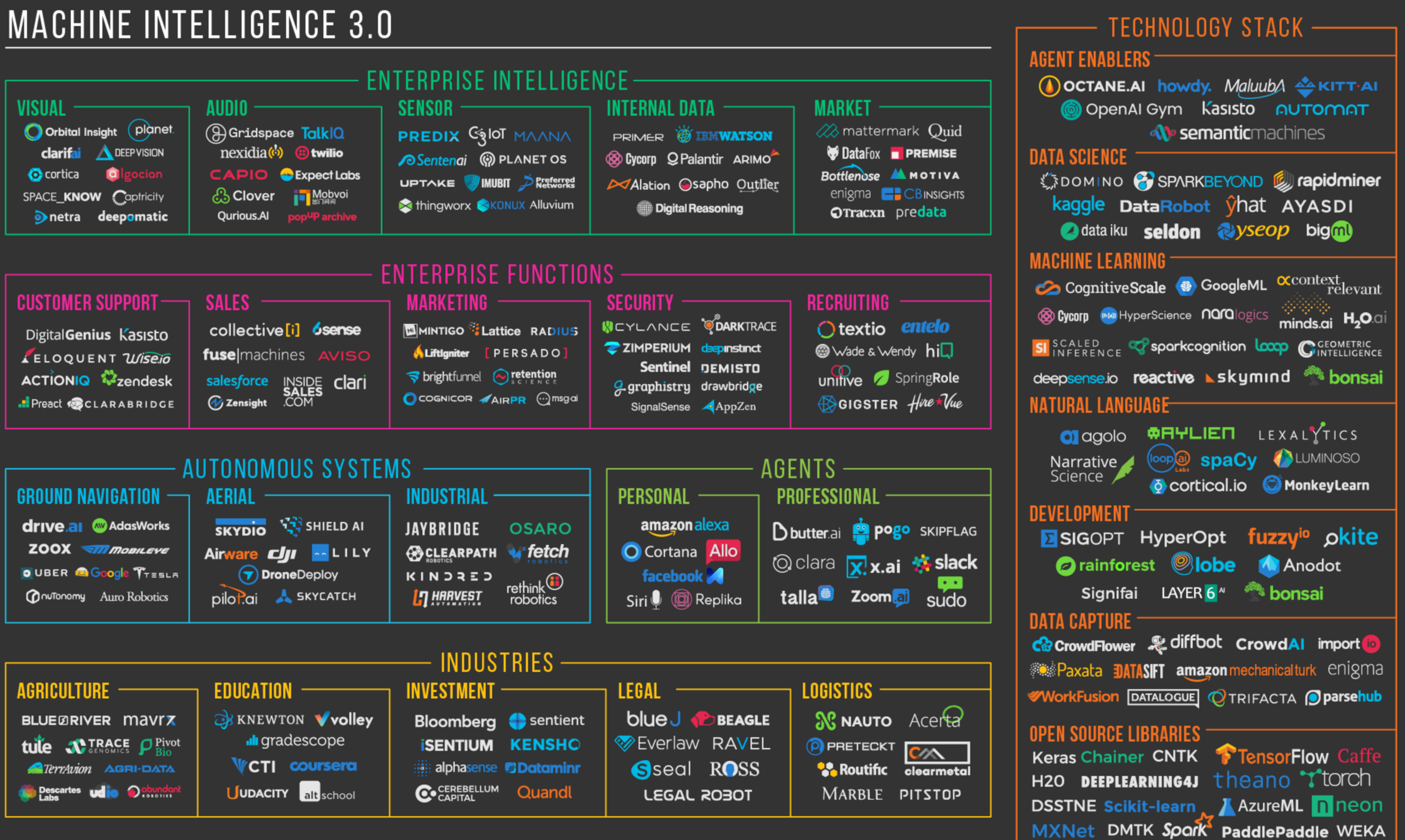


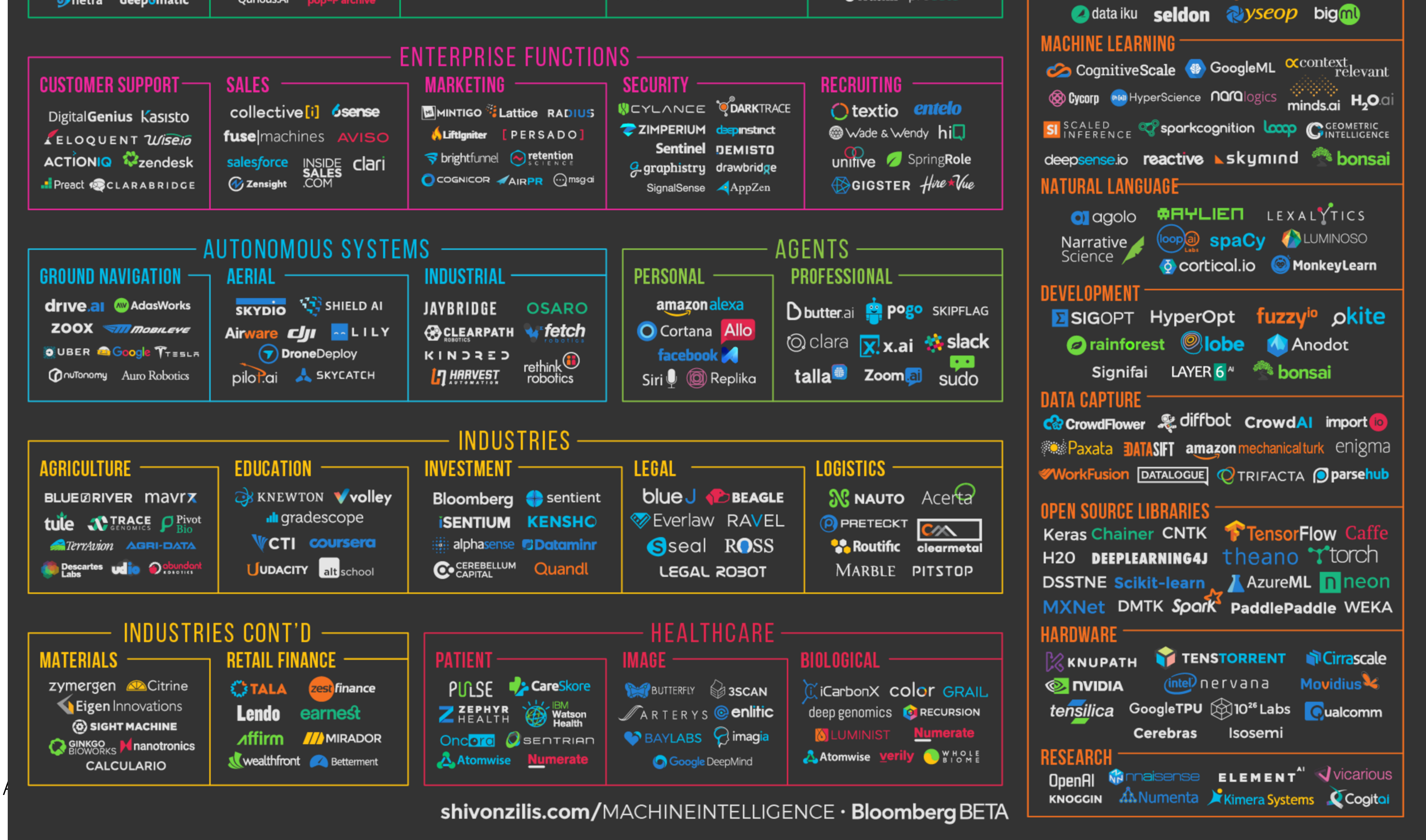
50—dimensional optimization problem

Bayesian optimization: “designed” magnets that have same efficiency by are 25% lighter – converts to 1M CHF

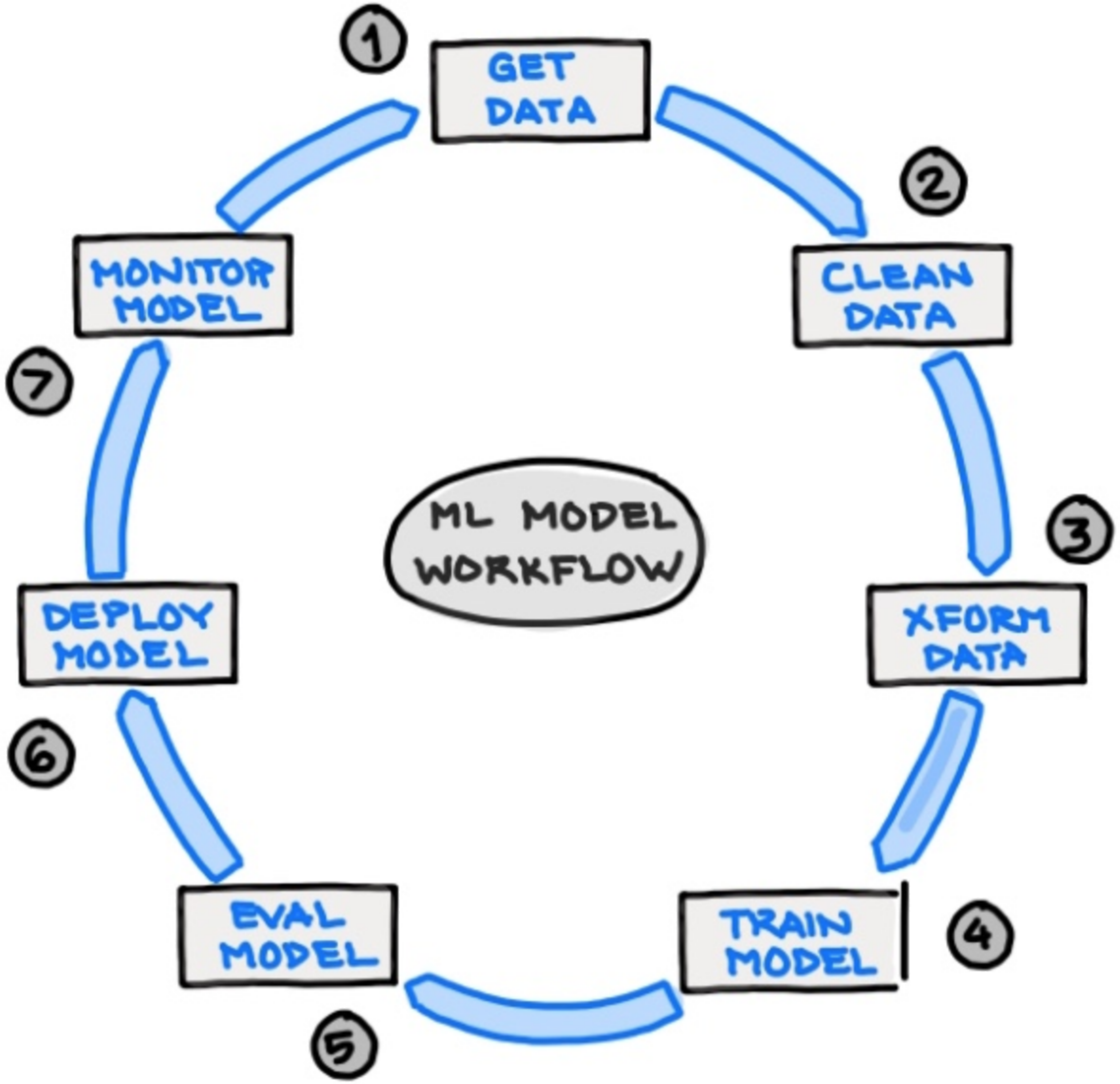
http://iopscience.iop.org/article/10.1088/1742-6596/934/1/012050/meta

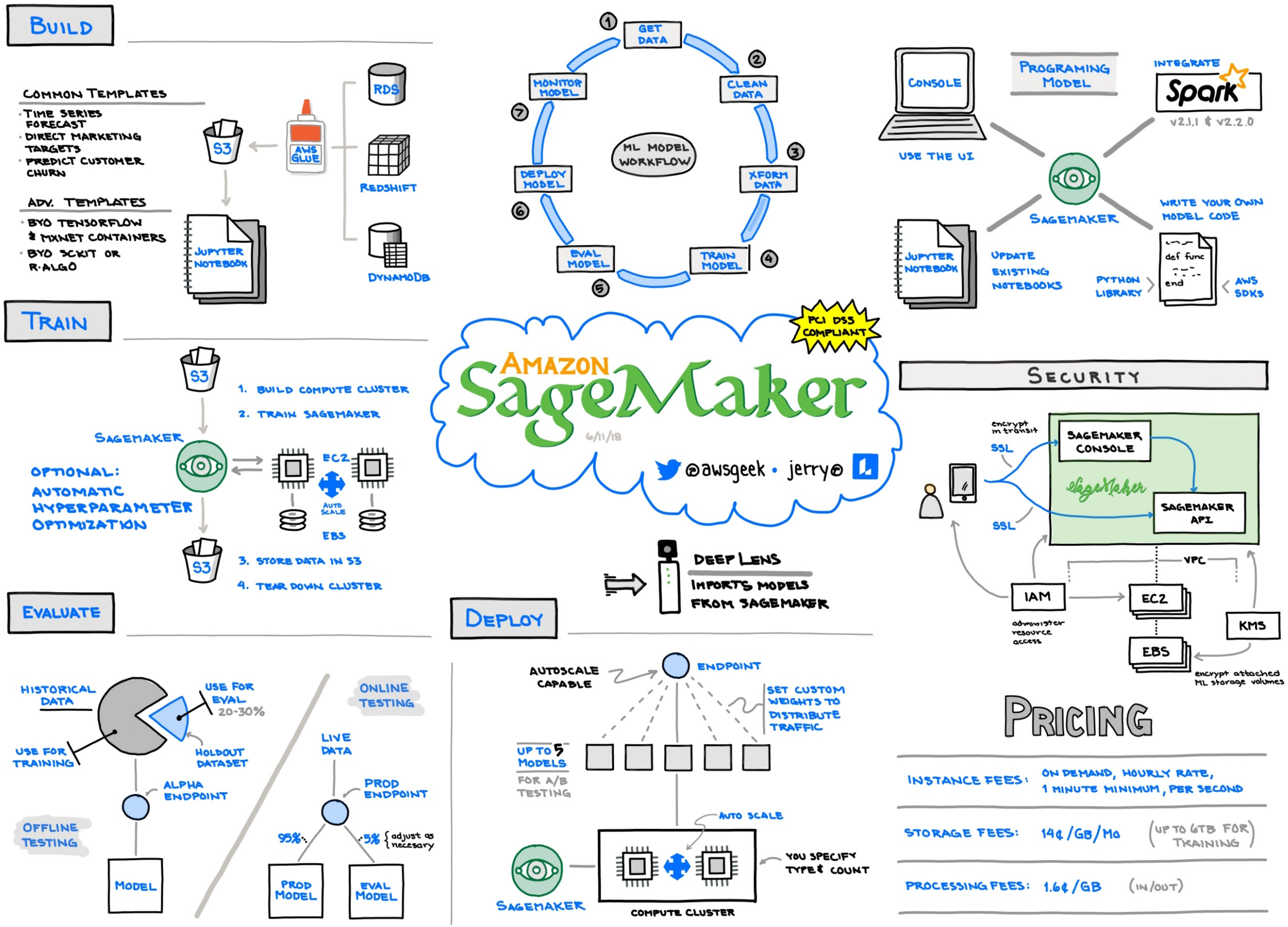
Industrial ML
(abridged)


MACHINE INTELLIGENCE 3.0
ENTERPRISE INTELLIGENCE
VISUAL
Orbital Insight
planet
clarifai
DEEP VISION
cortica
Algocian
SPACE_KNOW
Captricity
netra
deepomatic
AUDIO
Gridspace
TalkIQ
nexidia
twilio
CAPIO
Expect Labs
Clover
Mobvoi
Qurious.AI
popUP archive
SENSOR
PREDIX
IoT
MAANA
Sentenai
PLANET OS
UPTAKE
IMUBIT
Preferred Networks
thingworx
KONUX
Alluvium
INTERNAL DATA
PRIMER
IBM WATSON
Cycorp
Palantir
ARIMO
Alation
sapho
Outlier
Digital Reasoning
MARKET
mattermark
Quid
DataFox
PREMISE
Bottlenose
MOTIVA
enigma
CB INSIGHTS
Tracxn
predata
ENTERPRISE FUNCTIONS
CUSTOMER SUPPORT
DigitalGenius
Kasisto
ELOQUENT
Wise.io
ACTIONIQ
zendesk
Preact
CLARABRIDGE
SALES
collective[i]
6sense
fuse|machines
AVISO
salesforce
INSIDE SALES .COM
clari
Zensight
MARKETING
MINTIGO
Lattice
RADIUS
LiftIgniter
[PERSADO]
brightfunnel
retention science
COGNICOR
AIRPR
msg.ai
SECURITY
CYLANCE
DARKTRACE
ZIMPERIUM
deepinstinct
Sentinel
DEMISTO
graphistry
drawbridge
SignalSense
AppZen
RECRUITING
textio
entelo
Wade & Wendy
hiQ
unitive
SpringRole
GIGSTER
Hire Vue
AUTONOMOUS SYSTEMS
GROUND NAVIGATION
drive.ai
AdasWorks
ZOOX
MOBILEYE
UBER
Google
TESLA
nuTonomy
Auro Robotics
AERIAL
SKYDIO
SHIELD AI
Airware
dji
LILY
DroneDeploy
pilot.ai
SKYCATCH
INDUSTRIAL
JAYBRIDGE
OSARO
CLEARPATH ROBOTICS
fetch robotics
KINDRED
HARVEST AUTOMATION
rethink robotics
AGENTS
PERSONAL
amazon alexa
Cortana
Allo
facebook
Siri
Replika
PROFESSIONAL
butter.ai
pogo
SKIPFLAG
clara
x.ai
slack
talla
Zoom.ai
sudo
INDUSTRIES
AGRICULTURE
BLUE RIVER
mavrx
tule
TRACE GENOMICS
Pivot Bio
TerrAvion
AGRI-DATA
Descartes Labs
udio
abundant robotics
EDUCATION
KNEWTON
volley
gradescope
CTI
coursera
UDACITY
altschool
INVESTMENT
Bloomberg
sentient
iSENTIUM
KENSHO
alphasense
Dataminr
CEREBELLUM CAPITAL
Quandl
LEGAL
blue J
BEAGLE
Everlaw
RAVEL
seal
ROSS
LEGAL ROBOT
LOGISTICS
NAUTO
Acerta
PRETECKT
Routific
clearmetal
MARBLE
PITSTOP
TECHNOLOGY STACK
AGENT ENABLERS
OCTANE.AI
howdy.
Maluuba
KITT.AI
OpenAI Gym
Kasisto
AUTOMAT
semanticmachines
DATA SCIENCE
DOMINO
SPARKBEYOND
rapidminer
kaggle
DataRobot
yhat
AYASDI
data iku
seldon
yseop
bigml
MACHINE LEARNING
CognitiveScale
GoogleML
context relevant
Cycorp
HyperScience
naralogics
minds.ai
H2O.ai
SCALED INFERENCE
sparkcognition
loop
GEOMETRIC INTELLIGENCE
deepsense.io
reactive
skymind
bonsai
NATURAL LANGUAGE
agolo
AYLIEN
LEXALYTICS
Narrative Science
loop ai Labs
spaCy
LUMINOSO
cortical.io
MonkeyLearn
DEVELOPMENT
SIGOPT
HyperOpt
fuzzy.io
kite
rainforest
lobe
Anodot
Signifai
LAYER 6
bonsai
DATA CAPTURE
CrowdFlower
diffbot
CrowdAI
import.io
Paxata
DATASIFT
amazon mechanical turk
enigma
WorkFusion
DATALOGUE
TRIFACTA
parsehub
OPEN SOURCE LIBRARIES
Keras
Chainer
CNTK
TensorFlow
Caffe
H20
DEEPLEARNING4J
theano
torch
DSSTNE
Scikit-learn
AzureML
neon
MXNet
DMTK
Spark
PaddlePaddle
WEKA


ENTERPRISE FUNCTIONS
CUSTOMER SUPPORT
DigitalGenius Kasisto
ELOQUENT Wise.io
ACTIONIQ zendesk
Preact CLARABRIDGE
SALES
collective[i] 6sense
fuse|machines AVISO
salesforce INSIDE SALES .COM clari
Zensight
MARKETING
MINTIGO Lattice RADIUS
Liftigniter [PERSADO]
brightfunnel retention science
COGNICOR AIRPR msg.ai
SECURITY
CYLANCE DARKTRACE
ZIMPERIUM deepinstinct
Sentinel DEMISTO
graphistry drawbridge
SignalSense AppZen
RECRUITING
textio entelo
Wade & Wendy hiQ
unitive SpringRole
GIGSTER Hire Vue
AUTONOMOUS SYSTEMS
GROUND NAVIGATION
drive.ai AdasWorks
ZOOX MOBILEYE
UBER Google TESLA
nuTonomy Auro Robotics
AERIAL
SKYDIO SHIELD AI
Airware dji LILY
DroneDeploy
pilot.ai SKYCATCH
INDUSTRIAL
JAYBRIDGE OSARO
CLEARPATH ROBOTICS fetch robotics
KINDRED
HARVEST AUTOMATION rethink robotics
AGENTS
PERSONAL
amazon alexa
Cortana Allo
facebook
Siri Replika
PROFESSIONAL
butter.ai pogo SKIPFLAG
clara x.ai slack
talla Zoom.ai sudo
INDUSTRIES
AGRICULTURE
BLUE RIVER mavrx
tule TRACE GENOMICS Pivot Bio
TerrAvion AGRI-DATA
Descartes Labs udio abundant robotics
EDUCATION
KNEWTON volley
gradescope
CTI coursera
UDACITY altschool
INVESTMENT
Bloomberg sentient
iSENTIUM KENSHO
alphasense Dataminr
CEREBELLUM CAPITAL Quandl
LEGAL
blue J BEAGLE
Everlaw RAVEL
seal ROSS
LEGAL ROBOT
LOGISTICS
NAUTO Acerta
PRETECKT clearmetal
Routific
MARBLE PITSTOP
INDUSTRIES CONT'D
MATERIALS
zymergen Citrine
Eigen Innovations
SIGHT MACHINE
GINKGO BIOWORKS nanotronics
CALCULARIO
RETAIL FINANCE
TALA zest finance
Lendo earnest
Affirm MIRADOR
wealthfront Betterment
HEALTHCARE
PATIENT
PULSE CareSkore
ZEPHYR HEALTH IBM Watson Health
Oncora SENTRIAN
Atomwise Numerate
IMAGE
BUTTERFLY 3SCAN
ARTERYS enlitic
BAYLABS imagia
Google DeepMind
BIOLOGICAL
iCarbonX color GRAIL
deep genomics RECURSION
LUMINIST Numerate
Atomwise verily WHOLE BIOME
data iku seldon yseop bigml
MACHINE LEARNING
CognitiveScale GoogleML context relevant
Cycorp HyperScience naralogics minds.ai H2O.ai
SCALED INFERENCE sparkcognition loop GEOMETRIC INTELLIGENCE
deepsense.io reactive skymind bonsai
NATURAL LANGUAGE
agolo AYLIEN LEXALYTICS
Narrative Science loop ai Labs spaCy LUMINOSO
cortical.io MonkeyLearn
DEVELOPMENT
SIGOPT HyperOpt fuzzy.io kite
rainforest lobe Anodot
Signifai LAYER 6 AI bonsai
DATA CAPTURE
CrowdFlower diffbot CrowdAI import.io
Paxata DATASIFT amazon mechanical turk enigma
WorkFusion DATALOGUE TRIFACTA parsehub
OPEN SOURCE LIBRARIES
Keras Chainer CNTK TensorFlow Caffe
H2O DEEPLEARNING4J theano torch
DSSTNE Scikit-learn AzureML neon
MXNet DMTK Spark PaddlePaddle WEKA
HARDWARE
KNUPATH TENSTORRENT Cirrascale
NVIDIA intel nervana Movidius
tensilica GoogleTPU 10^26 Labs Qualcomm
Cerebras Isosemi
RESEARCH
OpenAI nnaisense ELEMENT AI vicarious
KNOGGIN Numenta Kimera Systems Cogitai
shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA

Stages of Machine-Learning study in industry





http://bit.ly/2JAaa5h

Evolution of HEP x ML Engineering

Layer	HEP (Circa 2013)	HEP (Circa 2018)	Industry
Data Layer	ROOT Files	ROOT Files	DB / HDFS etc.
Loading Layer	Ad hoc ROOT ETL logic	Numpy / HDF5 Converters / Loaders	Numpy / HDF5 Converters / Loaders
Training Layer	TMVA	Keras, TensorFlow, PyTorch, XGBoost, scikit-learn, …	Keras, TensorFlow, PyTorch, XGBoost, scikit-learn, …
Serving Layer	Deployment Target (TMVA)	Deployment Target (lwtnn, TensorFlow, TMVA wrappers)	Deployment Target (TensorFlow Serving, SageMaker, etc.)

11

Luke De Oliveira, 2018

Notable ML-infrastructure trends

Automation

- Reuse, reproducibility
- Workflows (Amazon SageMaker)
- Access to data and computing resources
- Continious offline testing

Development & production integration

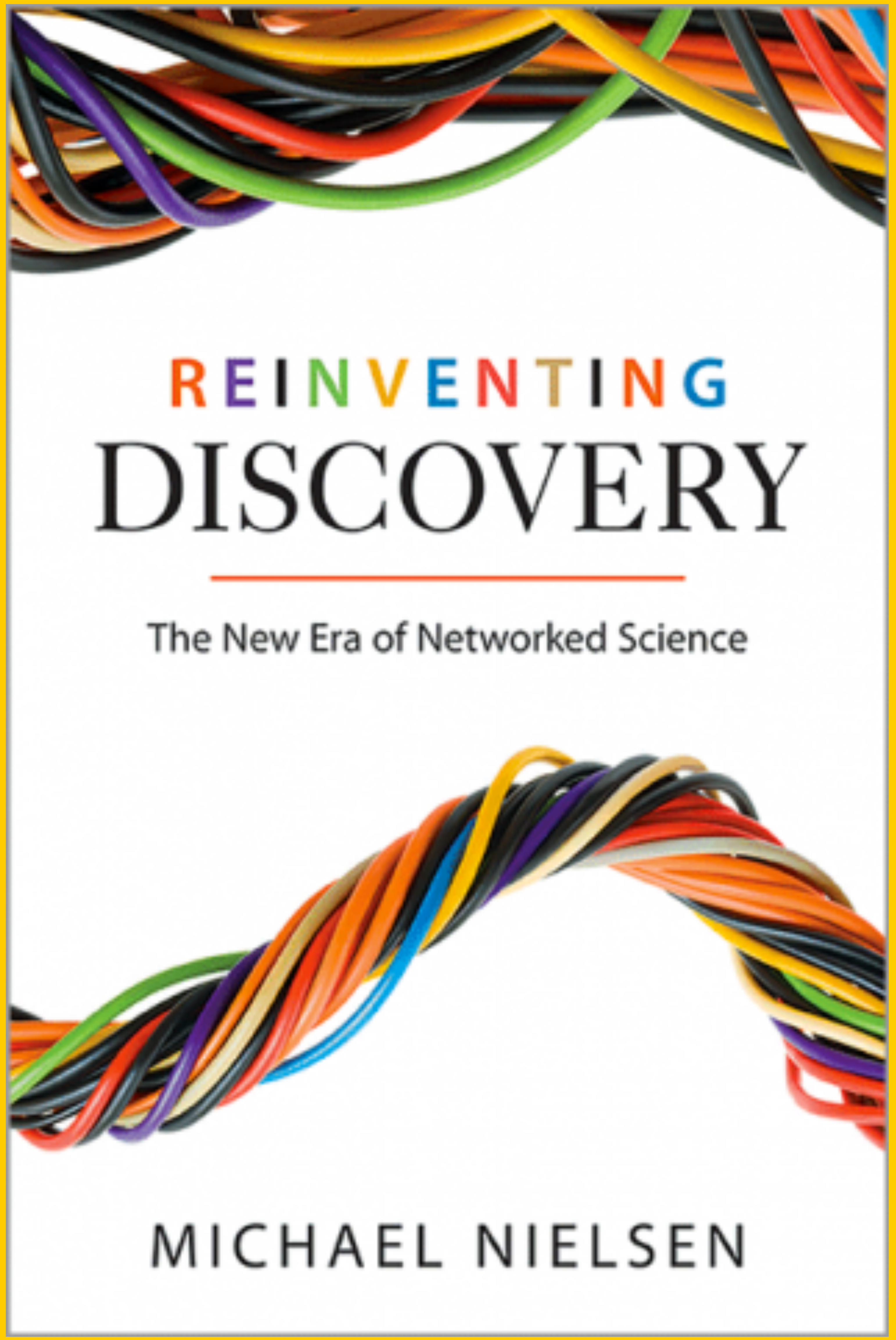
ML as a service

- Algorithm design, optimization (studio.ml, modelgym)
- Deployment

Interoperability

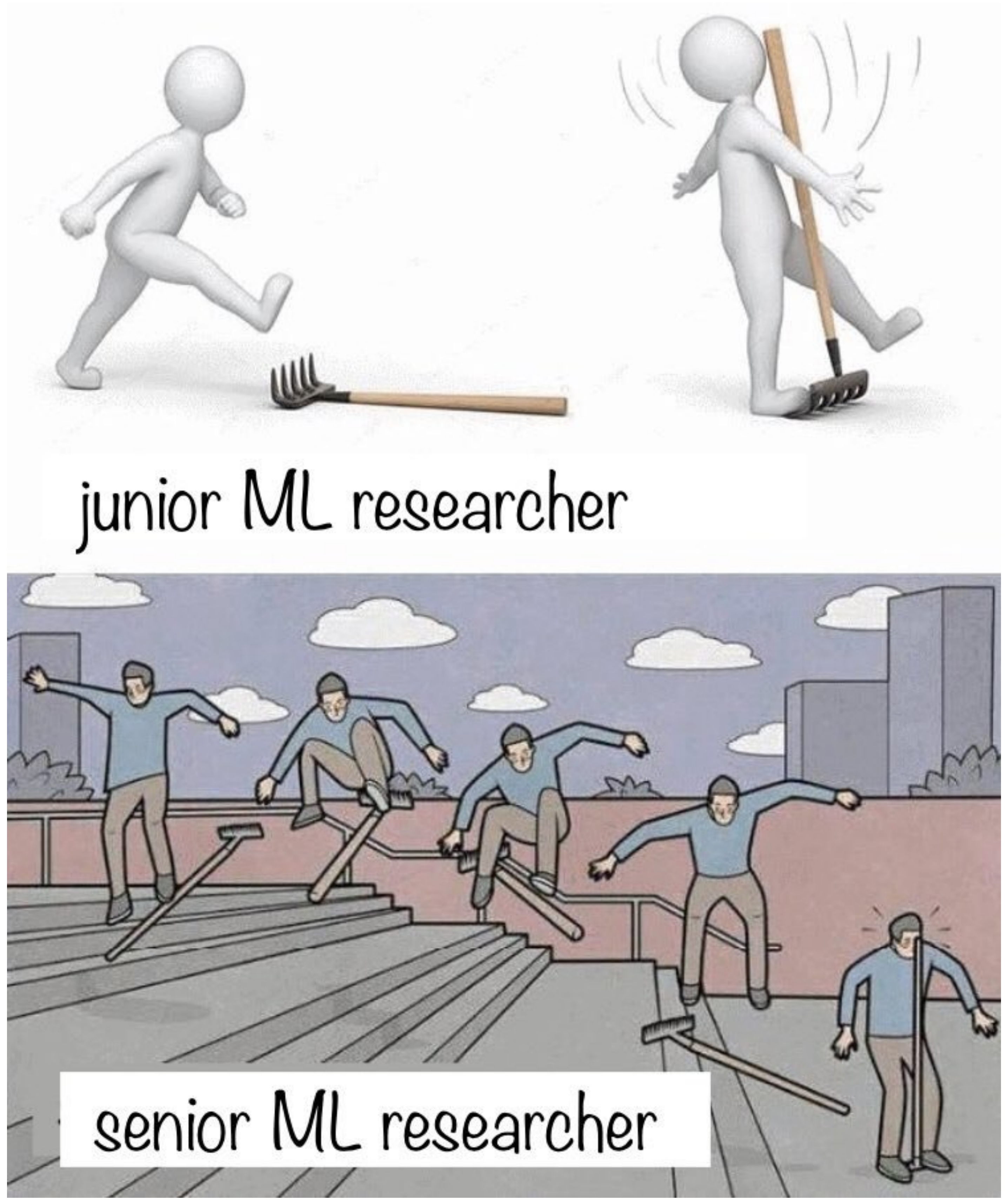
- Open Neural Network Exchange, https://onnx.ai/

One more trend



Collaboration with Data Science (DS)

- Domain science researches do not necessarily have required skills and background to properly adapt those methods (High Energy Physics, Astro Physics, Neuroscience)
- Industry or Academic data scientists are eager to help, but sometimes it is difficult for them to cope with domain specificity



DataScience competition: Netflix Prize

- Netflix prize – prediction of DVD titles renting (1M USD)
  - training data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies
  - Each training rating is a quadruplet of the form <user, movie, date of grade, grade>
  - The user and movie fields are integer IDs, while grades are from 1 to 5 (integral) stars
  - The qualifying data set contains over 2,817,131 triplets of the form <user, movie, date of grade>, with grades known only to the jury
  - A participating team's algorithm must predict grades on the entire qualifying set, but they are only informed of the score for half of the data, the **quiz** set of 1,408,342 ratings. The other half is the **test** set of 1,408,789, used to find winners.
  - Submitted predictions are scored against the true grades in terms of root mean squared error (RMSE), and the goal is to reduce this error

https://wiki2.org/en/Netflix_Prize

Netflix Prize timeline

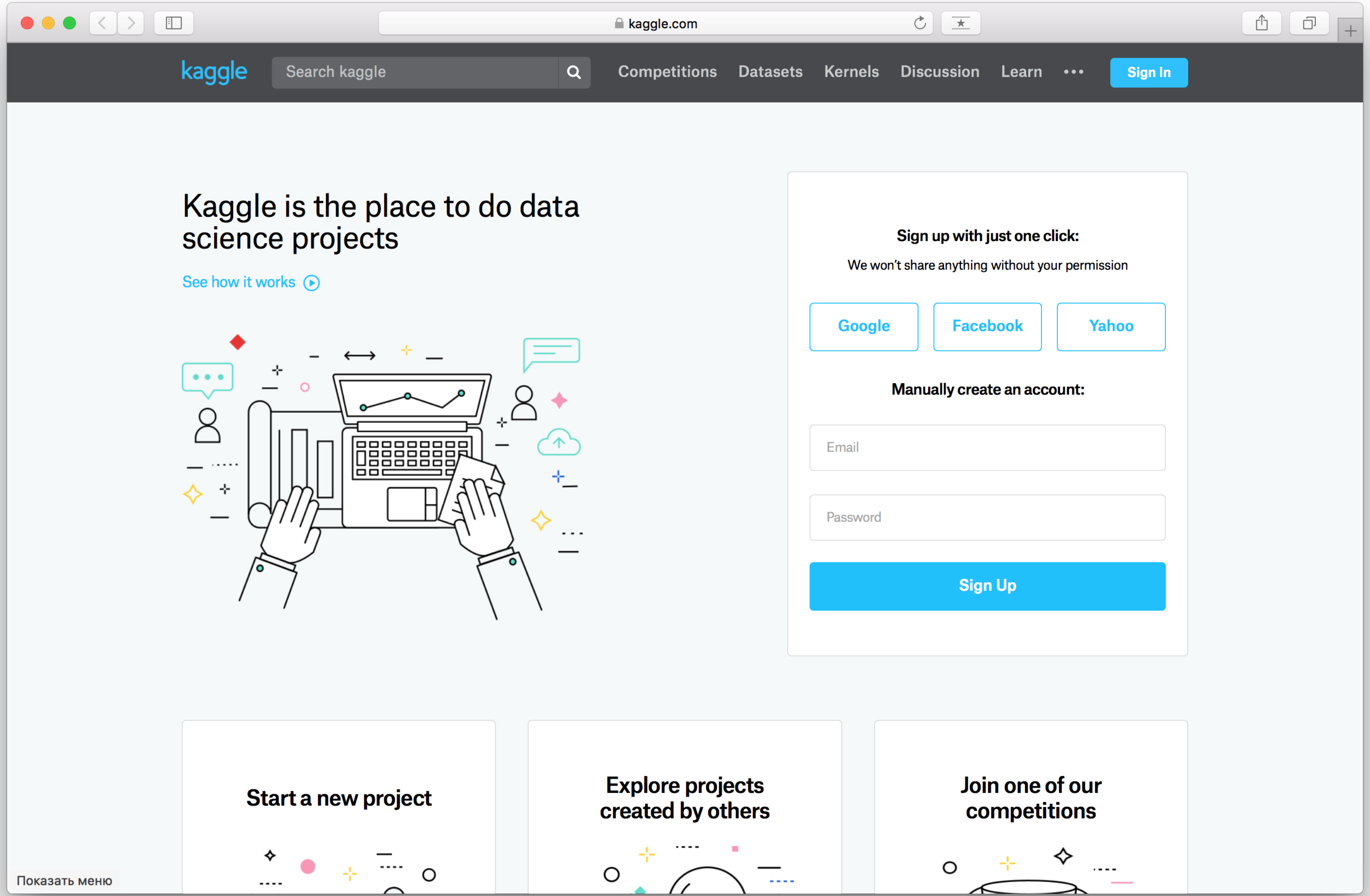
Netflix prize – prediction of DVD titles renting (1M USD for improving baseline by 10%)

- Baseline algorithm – Cinematch (linear model)
- Aug 2007 – international conference, announcement
- Oct 2007 – BellKor FTW – 8.43% improvement! (among 20k teams)
- Oct 2008 – Big Chaos took lead
- Late Oct 2008 – BellKor + Big Chaos – 9.43% impovement
- June 2009 – BellKor's Pragmatic Chaos – 10.05%
- 26 July 2009 18:18:28 – BellKor's Pragmatic Chaos – 10.09%
- 26 July 2009 18:38:22 – Ensemble – 10.10%

Got same result on final test! The prize was awarded to BellKor's Pragmatic Chaos.

Second challenge was cancelled due to privacy concerns.

https://wiki2.org/en/Netflix_Prize



O(100) datasets
O(1000) competitions
O(10000) users
$O(10^8)$ submissions

http://bit.ly/2JDMo8j

Competitions on Particle Physics at Kaggle

Higgs Boson Search

- “Discovery” of Xgboost algorith

Flavour of Physics ($\tau \rightarrow 3\mu$)

- Data doping – nice approach for training on mixture of MC & data

TrackML

- ...in progress until end of October (co-hosted by Codalab)

Private challenges (YSDA):

- Tracking, PID, EM-showers, for MLHEP schools and local courses

Restrictions:

- Single metric, difficult to check solutions

Caveats

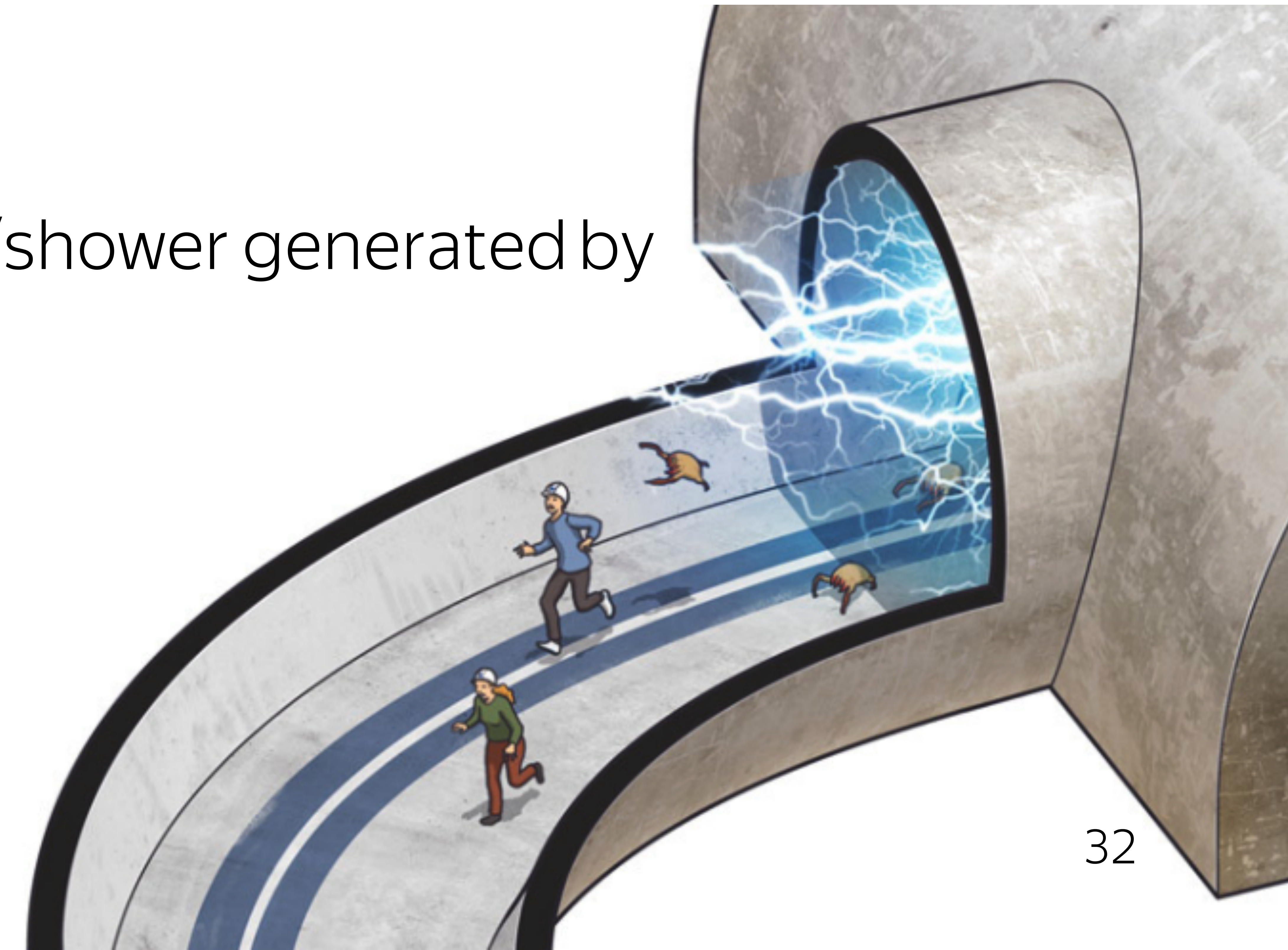
Domain-specific barriers

- Developed terminology and mindset
- Structured and semantically-rich data
- Weird constraints (“systematics”, “calibration”) due to the fact that ML part is just a step of a bigger picture
- Enormous data flows
- No obvious metrics for ’sanity’ checks (is a jet/shower generated by NN looks realistic enough?)

Reproducibility/traceability of results

Peer review?

Motivation for DS people?



Research Coopetition* Platform Candidates

- Github (belongs to Microsoft)
  - No reward mechanism, too generic
- Kaggle (belongs to Google)
  - No micro-reward motivation, no reward for popular contribution, single metric from pre-defined list
- CodaLab
  - No micro-reward motivation, single metric, no means of publishing / reuse / peer review

*) Coopetition or co-opetition (sometimes spelled "coopertition" or "co-opertition") is a neologism coined to describe cooperative competition.

Research Coopetition Platform Features

- Target audience
  - DS-intensive courses / Universities
- Built on top of existing services
  - GitHub, CodaLab, Jupyter, etc
  - Data storage
- Support for complex workflows
- Media for dialogue between domains

(more details in backup section)

Research Coopetition Platform Benefits

- Motivation for students:
  - Mini-grants to participants for computing access
  - Motivation through social dynamics of published code (likes/claps/forks)
  - Mini-grants for participants meeting evaluation criteria
  - Automatic reward system through smart-contract evaluation
- Motivation for problem owners:
  - Many students may eventually improve well-formulated problems
- Motivation for advanced ML/DS:
  - New datasets / settings that they can develop own method for
- Motivation for universities:
  - Keep student’s contribution, more adequate grading

Personal Experience

- 3 projects in that format during 2018/2019 across several universities:
  - YSDA, HSE, MIPT, CSC
- 5 summer student projects in collaboration across
  - HSE, YSDA, MIT, DCU, MIPT
- One of the master-student projects has been advanced to the level of mature publication in applied Computer Science

Conclusion & Focus points

- Trends in data processing and analysis:
  - algorithms get more and more sophisticated thanks to Machine Learning
  - new sub-domains emerge in Data Science
  - instrumentation gets more powerful (workflows, virtualization, testing, experiments)
- ML will continue to play an increasingly important role in science.

How can we scale and cooperate?

  - learn to speak ML-language (reasoning, dictionaries, intuition)
  - adoption of best industrial practices
  - cooperation with academic ML teams
  - inclusion of PhD curricula of ML courses from ML academic partners
  - tools flexibility, reproducibility
- Research Coopetition Platform (global, diverse, blockchain)

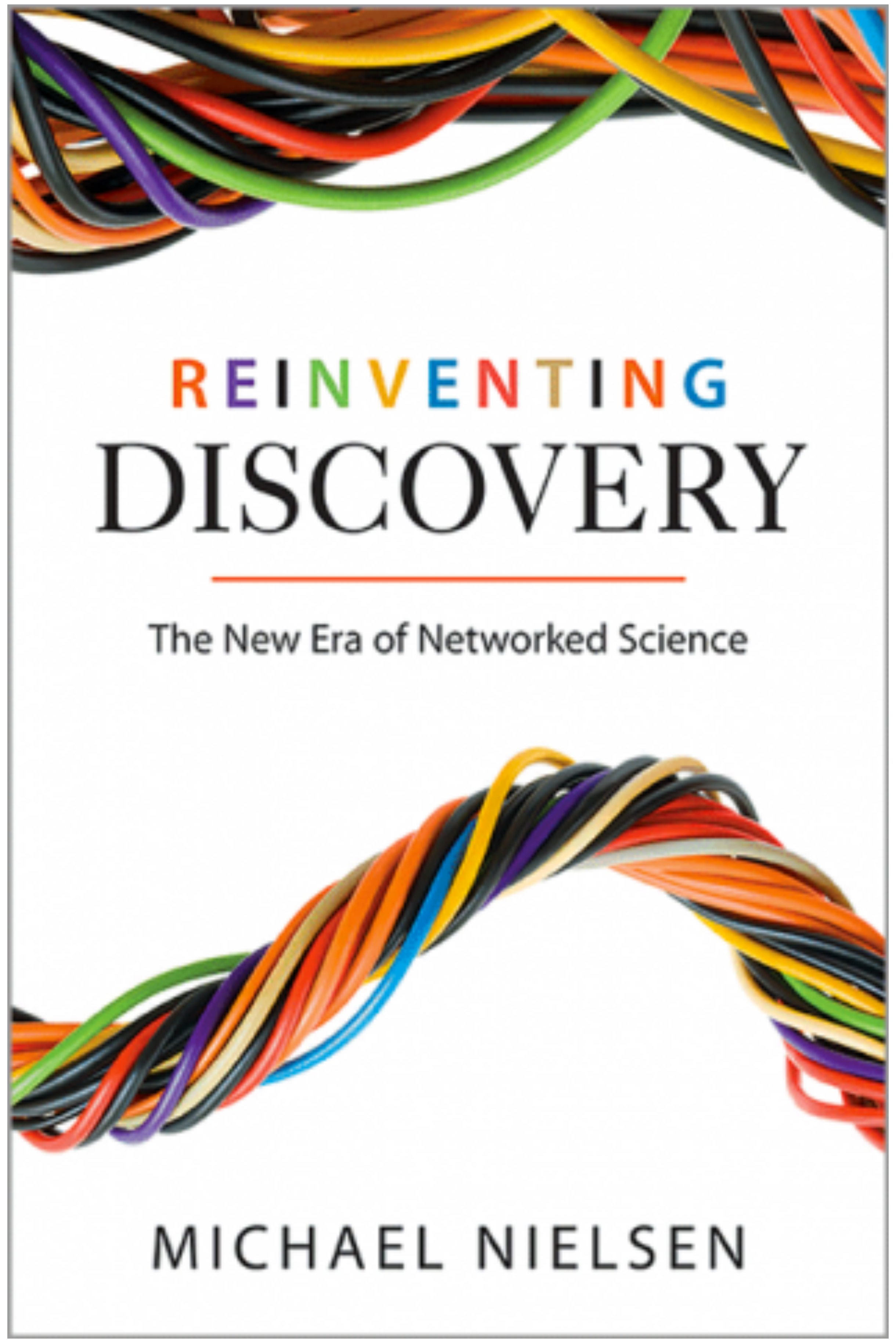
Andrey Ustyuzhanin

anaderiRu@twitter
austyuzhanin@hse.ru

Backup

Successful Citizen-Science project check list

- Clear goals, context and ambitions
  - marketing
- If you want to eat an elephant do it one bite a time
  - Split big goal in feasible steps
- Participant’s motivation even for weakly involved ones
  Specialist attention focus at percise moments
  - Progress announcemnts
  - Short contribution check cycle
- Check or reuse artifacts created by other participants
  Explanatory materials, methodological manifest, convention



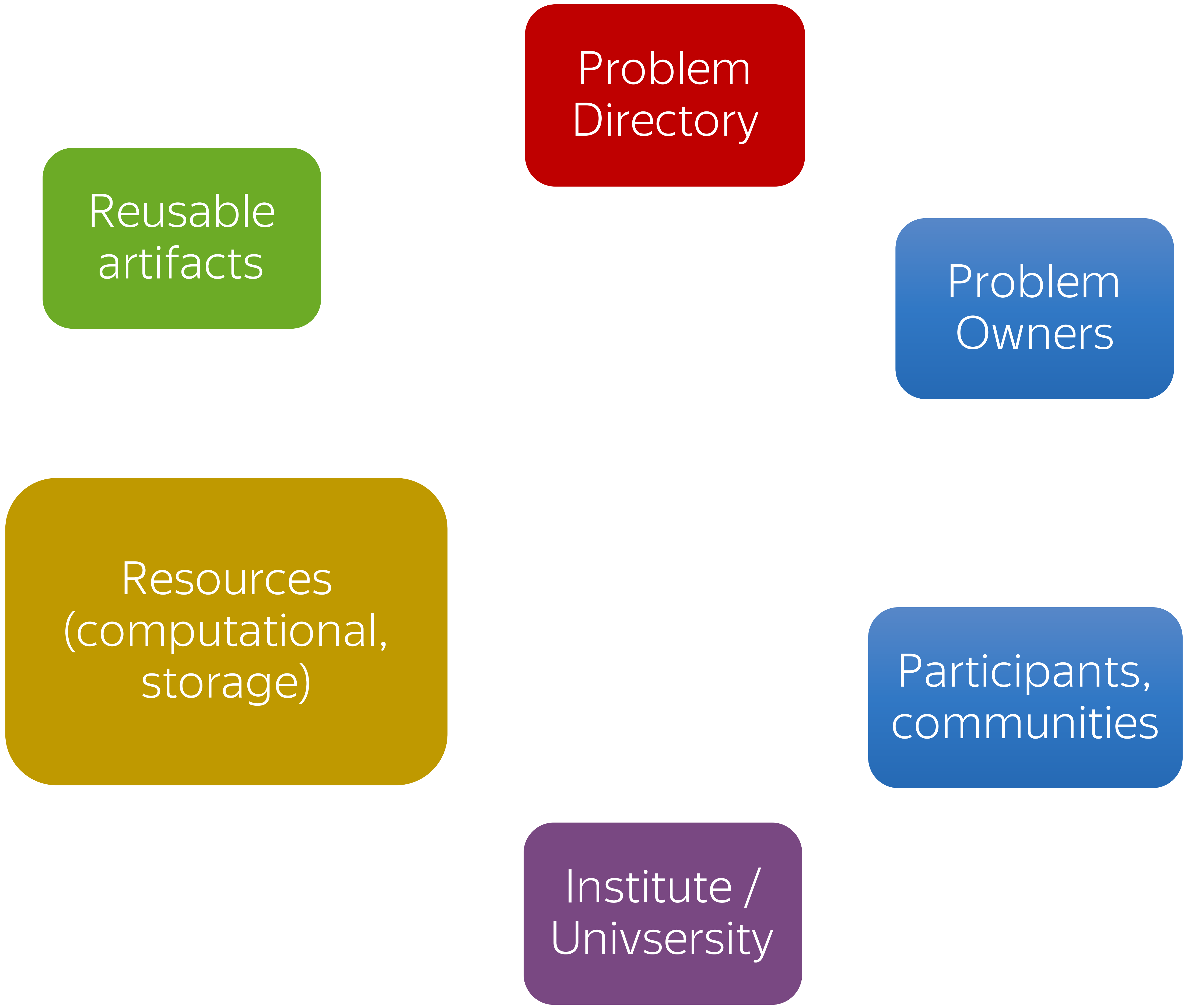
Michael Nielsen, Reinventing the Discovery, 2014

Target Audience

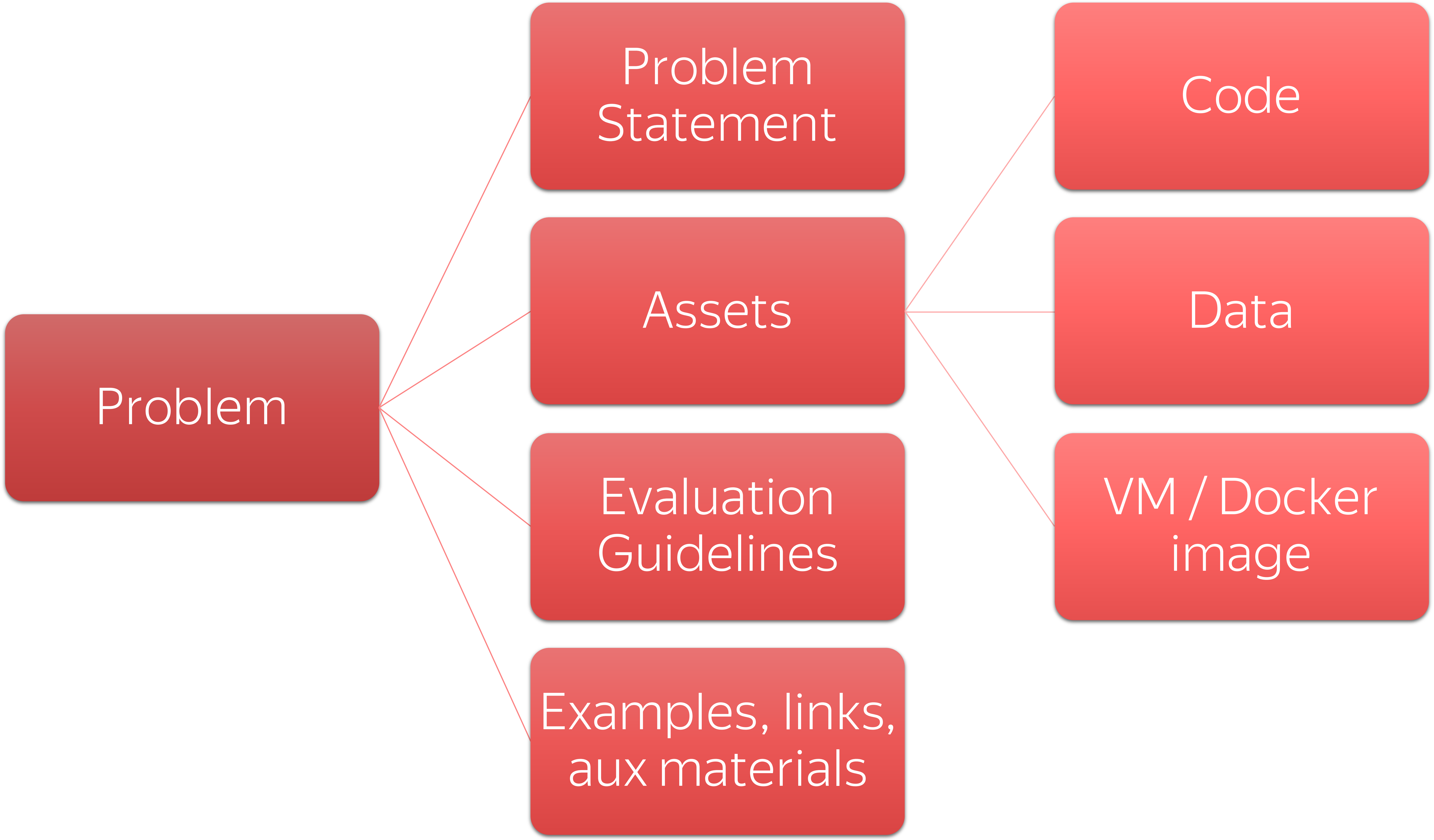
There are numerous people passing online ML courses, looking for decent problems to test their skills on

- Low-responsibility contribution
- Need for computational resources
- No time/resources for deep problem understanding
- Hungry for scoring records

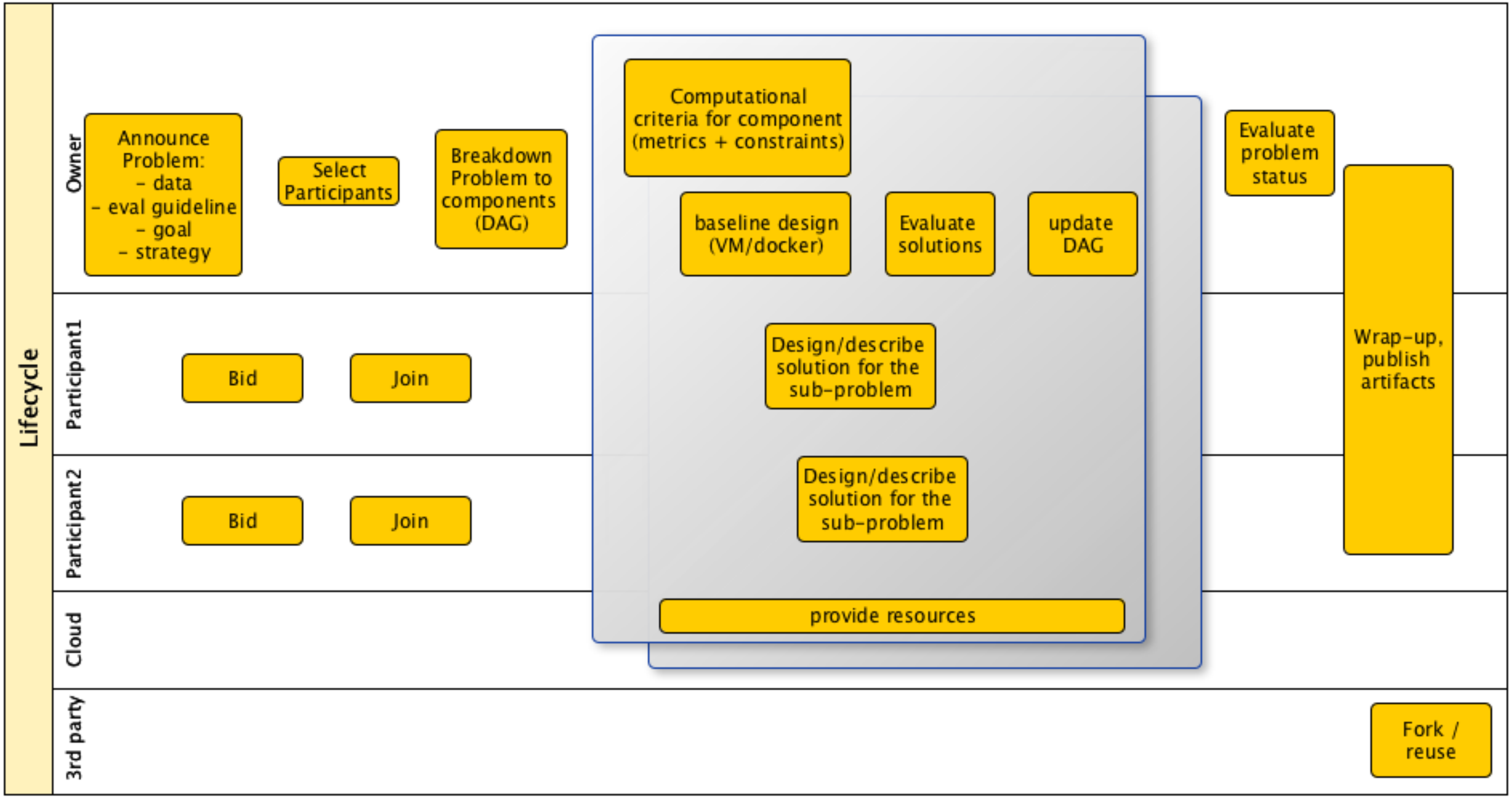
High-level platform Components



Problem Structure



Collaboration Lifecycle



Collaboration Highights

Preparation-stage

- Define the case goal(s), make it as independednt as possible
- Specify reasoning model, make it as clear as possible
- Produce dataset(s), describe the structure
- Produce evaluation baseline

Research-iterations

- Describe Figures of Merit (FOM) and constraints clearly
- Be comfortable with FOM evolution, repeat in cycles (sprints)
- Cycles are time-boxed
- For solution preparation and evaluation external resources are needed

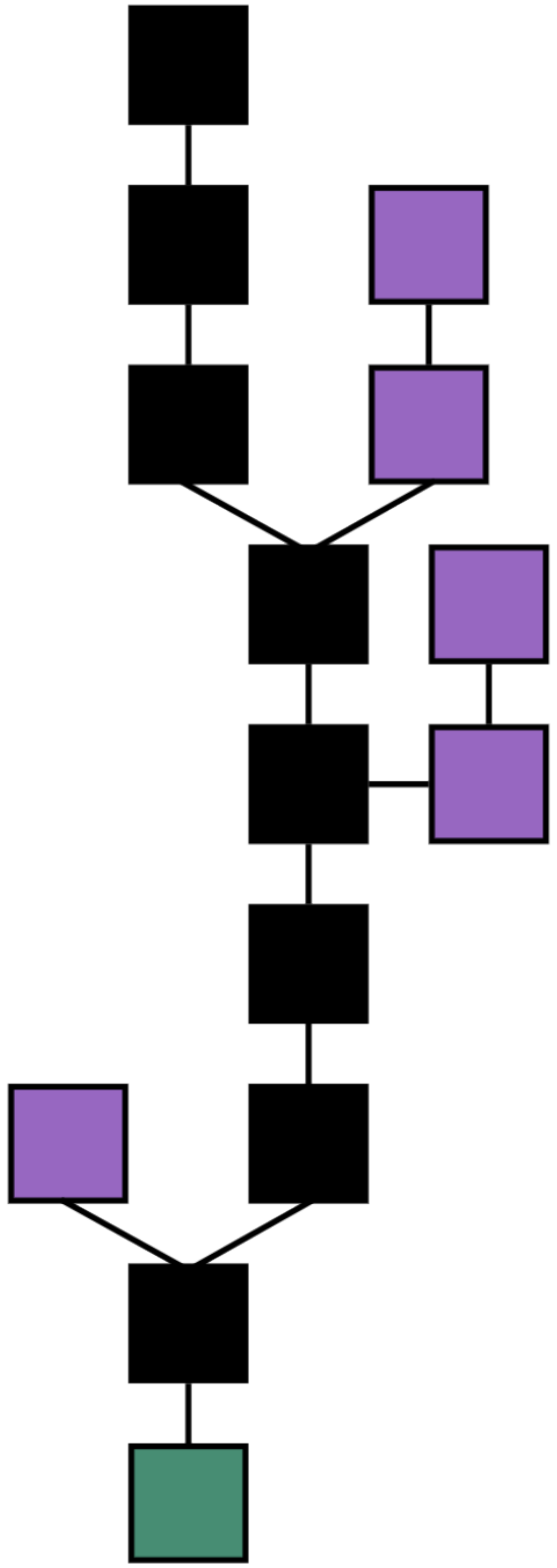
Wrap-up stage

- Publish reusable artifacts + result communication
- Generate track record for *each participant*, estimate impact of each contribution

What about diversity?

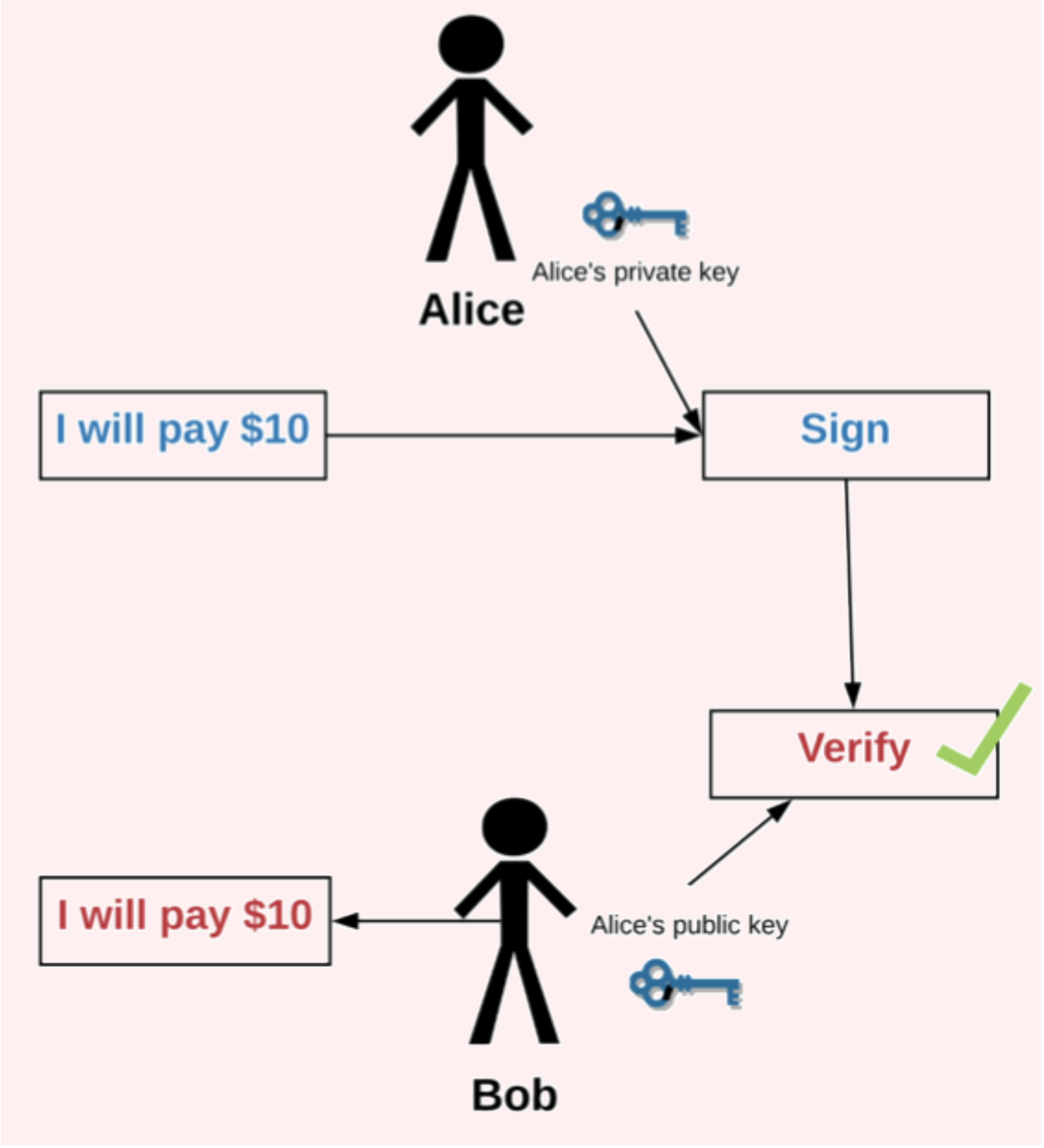
Blockchain - A Distributed Ledger Technology

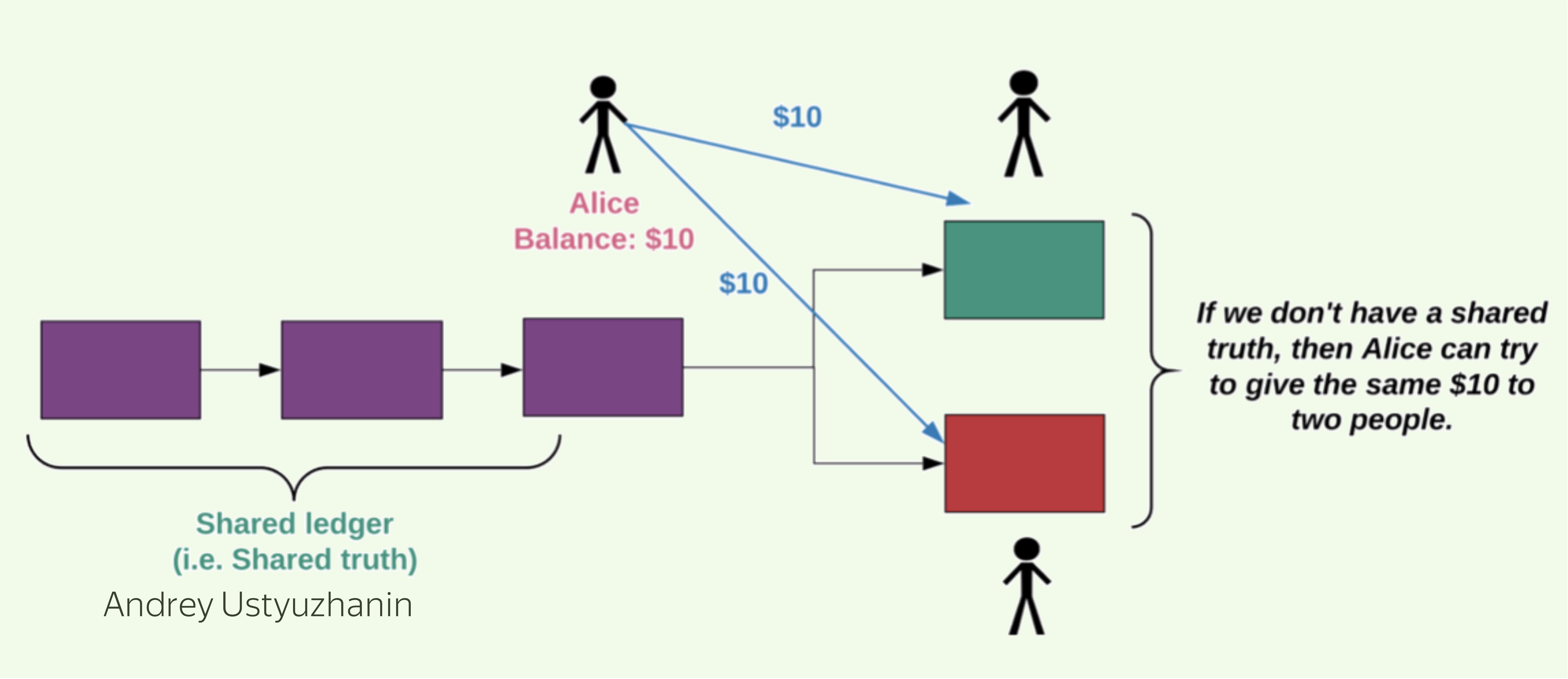
- A blockchain is a linked list where each node is connected to its predecessor by a cryptographic hash
  - All pointing back to the "genesis" block (right, in green) which may contain defining information about the rules for the blockchain protocol
  - In this way a blockchain comprises a verifiable public ledger
- Each node of the linked may contain additional transaction data (verifiable)
- Typically it's the longest contiguous chain (right, in black) which is considered valid (purple are orphaned blocks)
  - However it's up to the developers who define the protocol to determine the rules for consensus and evolution of the chain
- A variety of blockchains exist today, some exploring alternative architectures to test multiple aspects of scalability



Blockchain - A Distributed Ledger Technology

- Original purpose of the blockchain:
  - Keep shared (consensus) state of the "truth"
  - For example balance on each participant's account





Blockchain – Smart Contract

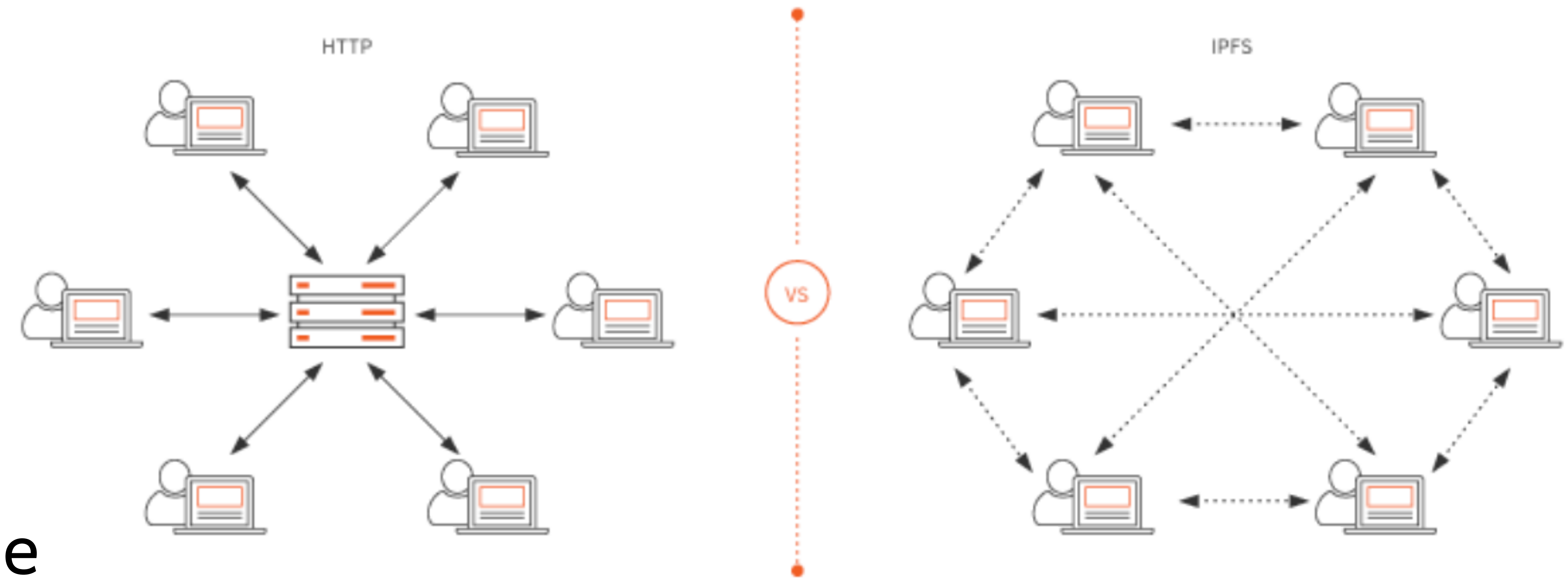
- Newer blockchains, Ethereum for instance, implement virtual machines that can execute byte code
- Smart contracts, implemented in this code allow binding between blockchain addresses and actions that are taken by the code
  - Typically the same code gets executed by all nodes in the network (extension of Nakamoto consensus)
- This can be used to implement a huge range of tasks
  - sub-currencies
  - timed payments
  - running of mathematical proofs
- Limited by blockchain transaction speed

```
pragma solidity ^0.4.21;

contract Coin {
    // The keyword "public" makes those variables
    // readable from outside.
    address public minter;
    mapping (address => uint) public balances;

    // Events allow light clients to react on
    // changes efficiently.
    event Sent(address from, address to, uint amount);

    // This is the constructor whose code is
    // run only when the contract is created.
    function Coin() public {
        minter = msg.sender;
    }

    function mint(address receiver, uint amount) public {
        if (msg.sender != minter) return;
        balances[receiver] += amount;
    }

    function send(address receiver, uint amount) public {
        if (balances[msg.sender] < amount) return;
        balances[msg.sender] -= amount;
        balances[receiver] += amount;
        emit Sent(msg.sender, receiver, amount);
    }
}
```

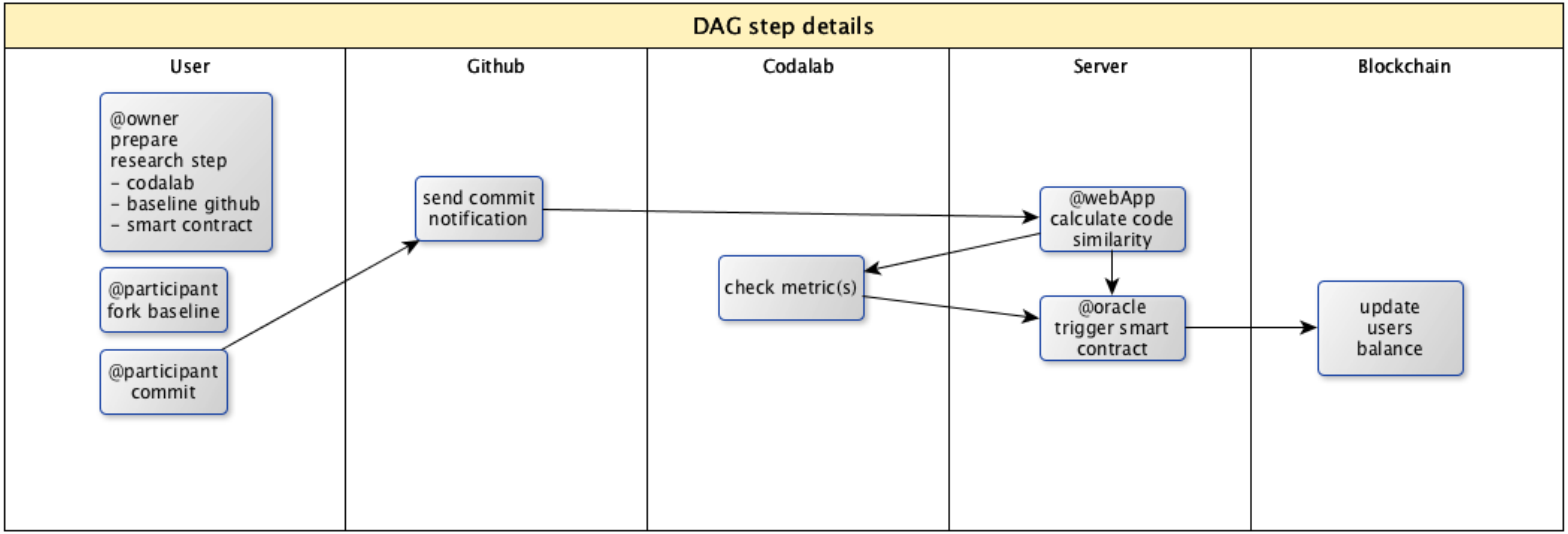
A simple example of a derived currency

Blockchain application

- Based on existing crypto-token
- Stores artifacts
- Manages computational resources allocation
- Records and rewards micro-contributions
  - Commit
  - Forks
- Remove bottle-neck and single vendor lock



DAG step (problem component) details



Blockchain application. Challenges

- Bootstrapping
  - Organizational (institutes / online education systems)
  - Marketing
- Should there be feedback loop from solution running in production?