



Sharpen decisions in Financial Services
with High-Scale Production & Trusted AI

Deep learning from physics to financial services

Jeremie Abiteboul

AMLD – AI & Physics Track

January 28, 2020



Who are we ?



Where do we come from ?

Université Paris 7 Denis Diderot
UFR de Physique

Thèse en vue de l'obtention du diplôme de
docteur de l'université de Paris 7 en physique
Ecole doctorale 517, Particules-Noyaux-Cosmos

Etude théorique et expérimentale des corrections électrofaibles au
processus de production inclusive de jets
Développement de méthodes de détection de topologies extrêmes

Nicolas MERIC

Thèse réalisée au CEA, LPNHE, LPTHE
Sous la direction de :
Gavin SALAM et Philippe SCHWEMLING

Thèse soutenue le jeudi 19 septembre 2013.

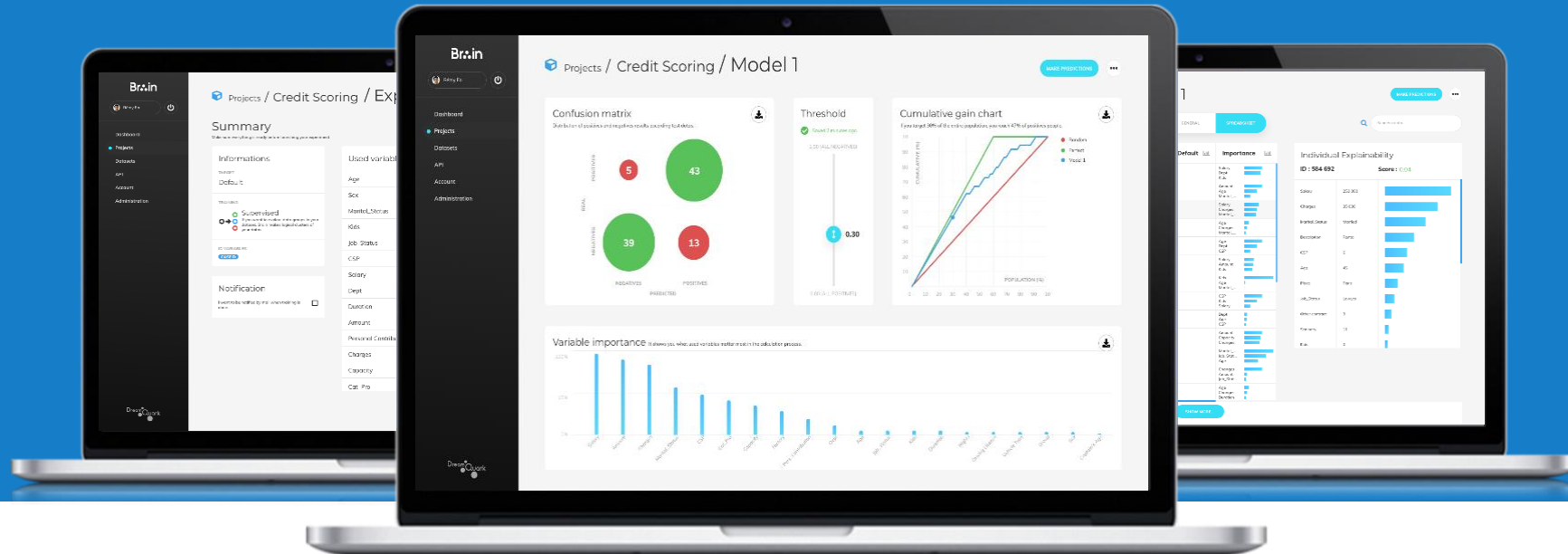
- Deep learning @ATLAS

- DreamQuark established 2014

Empower your action thanks to science

- My background: computational
plasma physics, turbulence modeling

=> originally applied ML to analyze
experimental data & simulation data



Fraud & Compliance



Fraud detection
AML
Compliance supervision

BANKS & INSURANCE



Marketing

Product upsell/cross-sell
Propensity
Customer segmentation
Satisfaction & anti-churn
Product recommendation



Risk

Risk model
Credit scoring
Fragile clients
Recovery



Transparency of decisions



Ability to do more business



Ease of use



Ease of deployment

THEY TRUST US

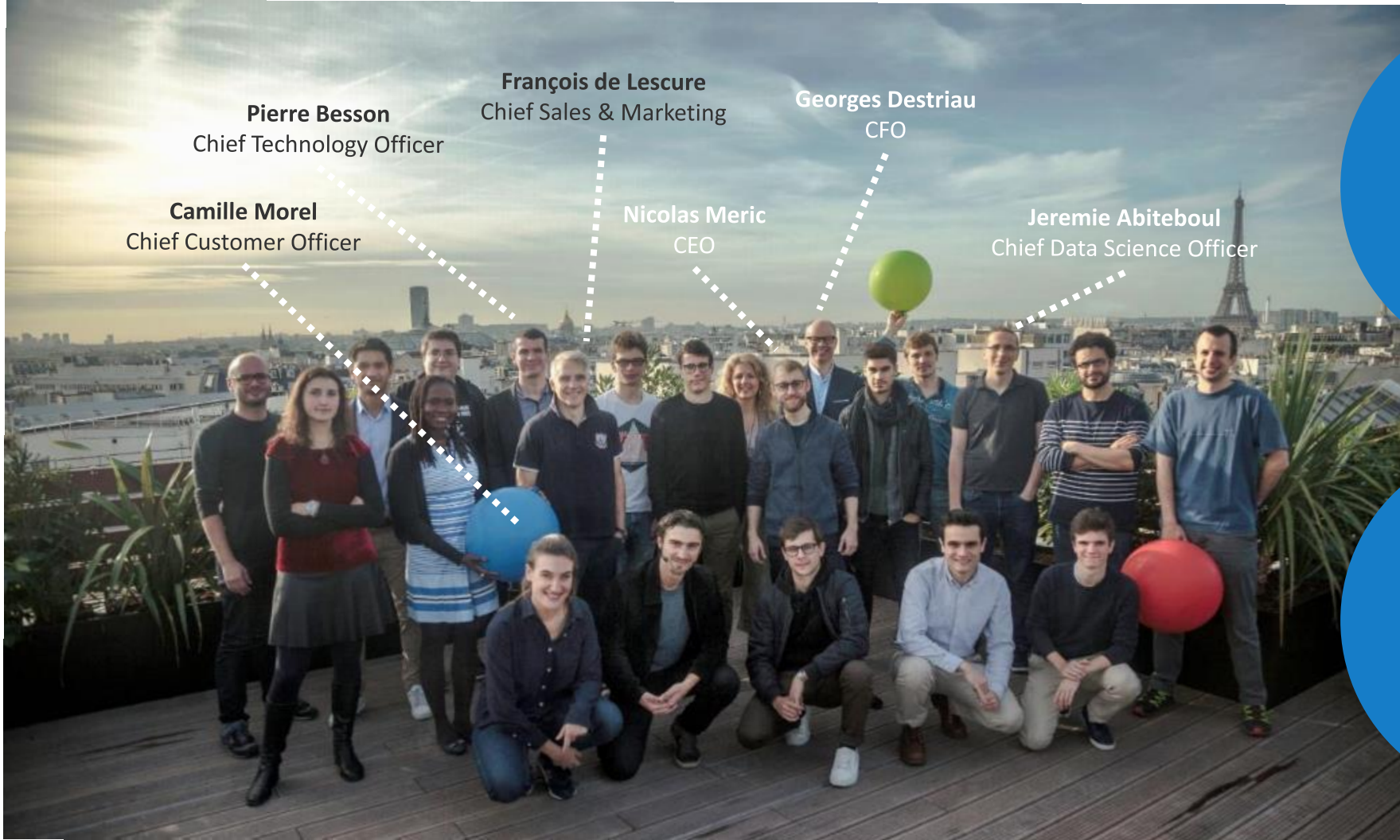


AG2R LA MONDIALE



RECENT AWARDS

The success of a team



Now a team
of
25

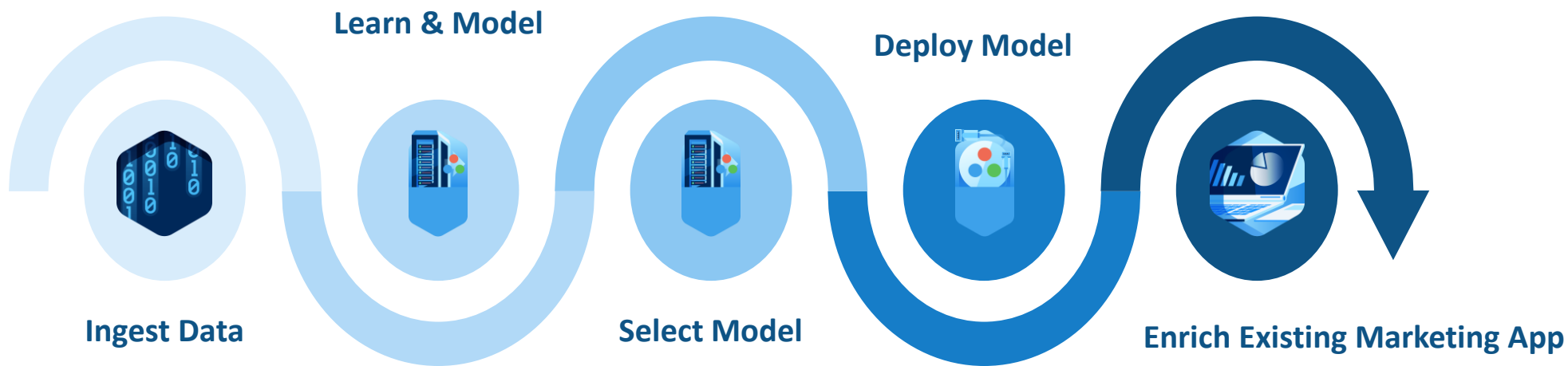
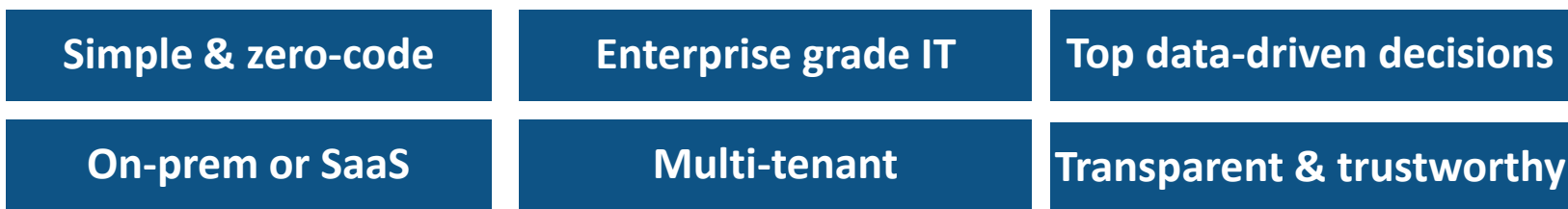
**Series-B
competed**

What do we do ?



Leverage data to reinvent the customer experience

To make it a reality in large scale production, DreamQuark created an AI platform for business experts:



Empower business analysts with predictive analytics => **Citizen Data Scientist**

PROJECT WM PE pre-Webinar > EXPERIMENT Web PE 01 > MODEL model_e45c

ENABLE MODEL API

Information

Target: **Has PE**

Training: supervised

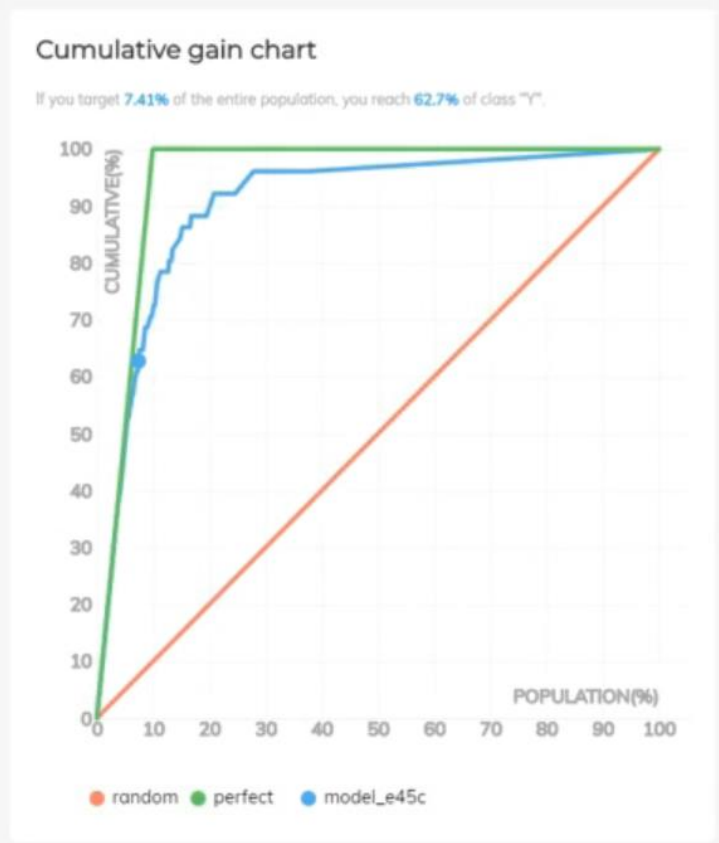
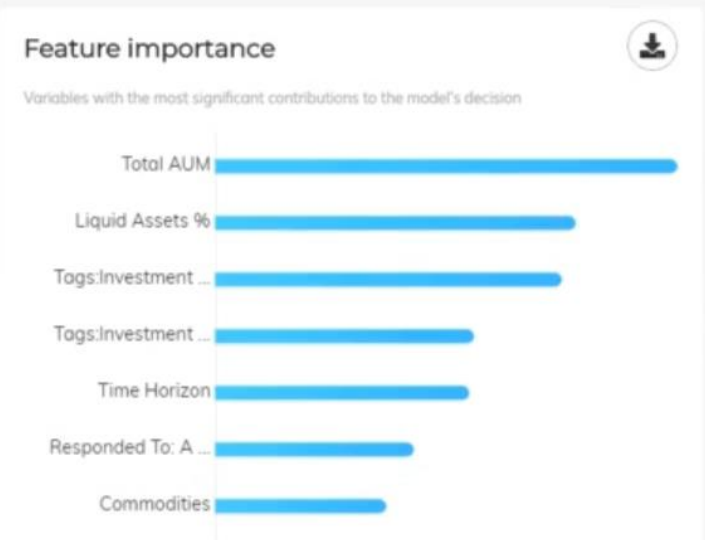
AUC

0.958 TEST

0.959 VALID

The plots shown in this page are computed with test data which was not used to train the model.

Model PSI: 0.054



Individual Scores

Search by line number...

CALL	PREDICTED SCORE	ALERT	VARIABLE IMPACT	VALUE	IMPACT
0	0.009	All Clear!	Tags: Investment Themes... Commodities Discretionary Age Bracket Advisory	1 23632 Yes 2 Yes	
1	0.159	All Clear!	Tags: Investment Themes... Liquid Assets Age Bracket Responded To: Alternatives Number Of Touchpoints L...	1 66280.71 1 Low 2	
2	0.027	All Clear!	Tags: Investment Themes... Property Responded To: Alternatives Age Bracket Liquid Assets %	1 10848.66 Low 1 24	
3	0.015	All Clear!	Tags: Investment Themes... Tags: Investment Theme - ... Discretionary Total AUM Number Of Touchpoints L...	0 0 No 895959.2281 57	
4	0.766	All Clear!	Liquid Assets % Commodities Total AUM Tags: Investment Theme - ... Age Bracket	23 53262.65 1278610.415 1 3	
5	0.000	All Clear!	Tags: Investment Theme - ... Tags: Investment Product... Property Tags: Investment Product... Commodities	0 0 333.6586 0 333.6586	
6	0.818	All Clear!	Commodities Tags: Investment Themes... Liquid Assets % Discretionary Age Bracket	67072.25 1 24 No 2	
7	0.005	All Clear!	Tags: Investment Themes... Tags: Investment Theme - ... Number Of Touchpoints L... Age Bracket Discretionary	1 0 16 2 No	
8	0.000	All Clear!	Tags: Investment Theme - ... Tags: Investment Product... Property	0 0 156035.15	



EDIT

ACTIONS +

CONTACT

Mrs. Anne ALPHABET

CLIENT

AGE 59 yo

MARITAL STATUS Single

GENDER Female

NATIONALITY Swiss

COMMUNICATION

EMAIL annealphabetbq@gmail.com

MOBILE PHONE 07 06 05 04 03

BUSINESS PHONE 032 900 80 70

CONTACT METHOD Mobile message

VALUE & RISK

LIFETIME VALUE 100 k€

CUSTOM VALUE €€€€

RISK PROFILE Good

NET PROMOTER SCORE 9

CHURN SCORE 10%

MENU

- Overview
- Household
- Portfolios →
- Documents
- Claims
- Life Events
- Complaints
- Opportunities
- Client Activity
- Addresses

Mrs. Anne ALPHABET

Client for 3 years
Joe CARRY (Insurance Agent)

104,200

4 Products & Services

Household Insurance - Policy #01234

PROPOSED

2,500 2 month ago

VIEW DETAILS

▼

Car Insurance - Policy C 31024

ACTIVE

500 1 year ago

VIEW DETAILS

▼

Saving - Policy L 20456

ACTIVE

100,000 2 years ago

VIEW DETAILS

▼

Health Insurance - Policy B 52103

ACTIVE

1,200 2 years ago

VIEW DETAILS

▼

ALPHABET CORP.

Client for 3 years
Joe CARRY (Insurance Agent)

15,000

3 Products & Services

Recommendations

Powered by DreamQuark

Pet Insurance

96%

Excellent match!

INFLUENCING FACTORS

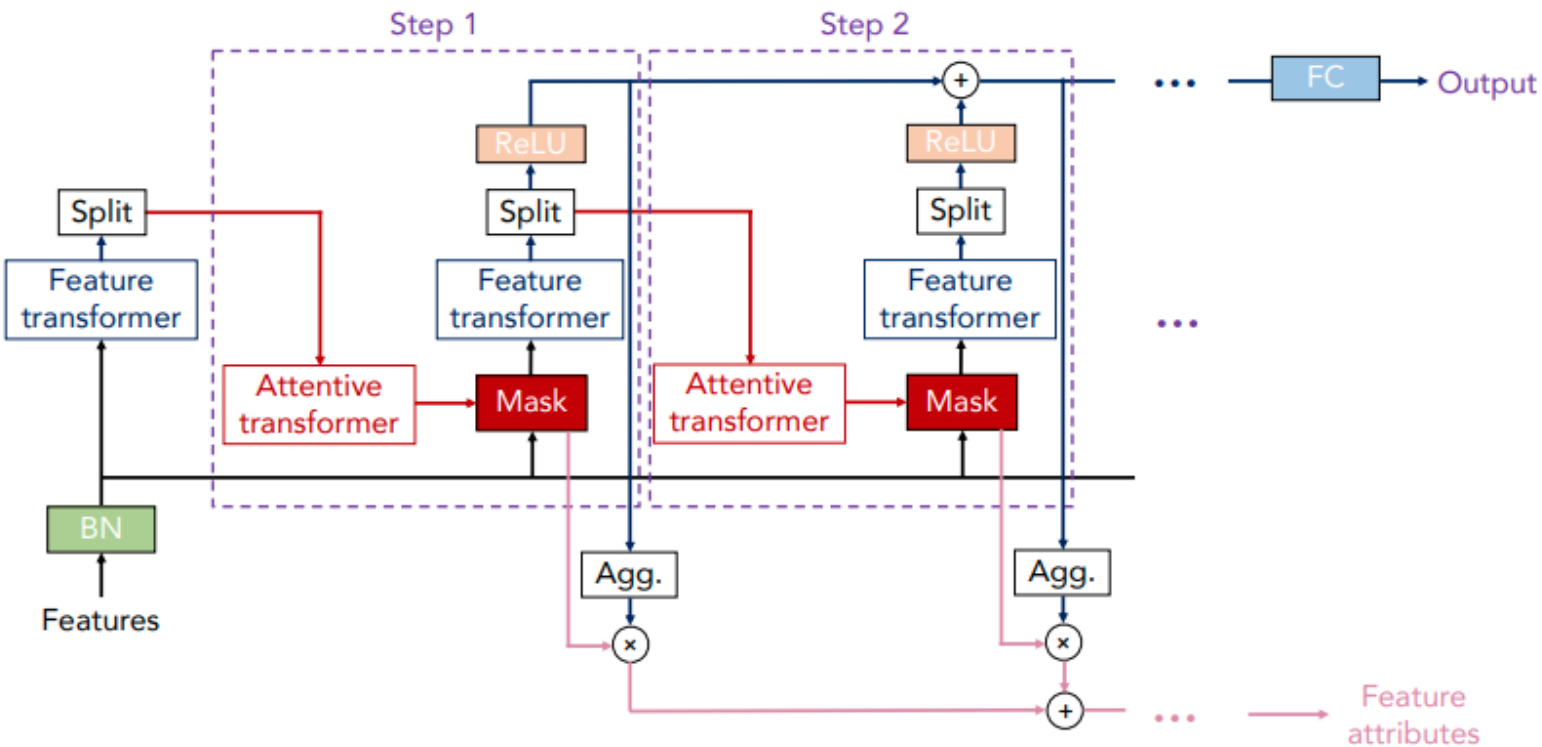
Age	59	<div style="width: 90%; height: 10px; background-color: white;"></div>
Marital Status	Single	<div style="width: 80%; height: 10px; background-color: white;"></div>
Sex	Female	<div style="width: 60%; height: 10px; background-color: white;"></div>
House Type	apartment	<div style="width: 40%; height: 10px; background-color: white;"></div>

- Automatically training neural networks on tabular data
- Explainable AI: “opening the black-box”
- Fairness of AI systems
- Robustness of AI systems
- Challenges for AI systems in production

Training neural networks on tabular data



- Most popular applications of neural networks focus on “unstructured” data (images, text, audio, video)
- Tabular data is actually less “structured” than e.g. images
- Most of the AI community still uses standard ML methods (boosting algorithms, random forests, etc) for tabular data
- We’ve found that neural networks are competitive if
 - Design suitable architectures (not necessarily very deep)
 - Find a reasonable grid for hyperparameter search [patent pending]
- Bonus: ensembling of neural networks and standard ML







- Sparse *learned* feature selection
- Sequential multi-step architecture (each step contributes to a portion of the decision)
- Attentive transformer ensures different steps learn from different features
- Non-linear processing of selected features (=> model capacity)

- Original publication by Google Cloud AI team: <https://arxiv.org/abs/1908.07442>
- Open-source (pytorch) implementation by DreamQuark: <https://github.com/dreamquark-ai/tabnet>

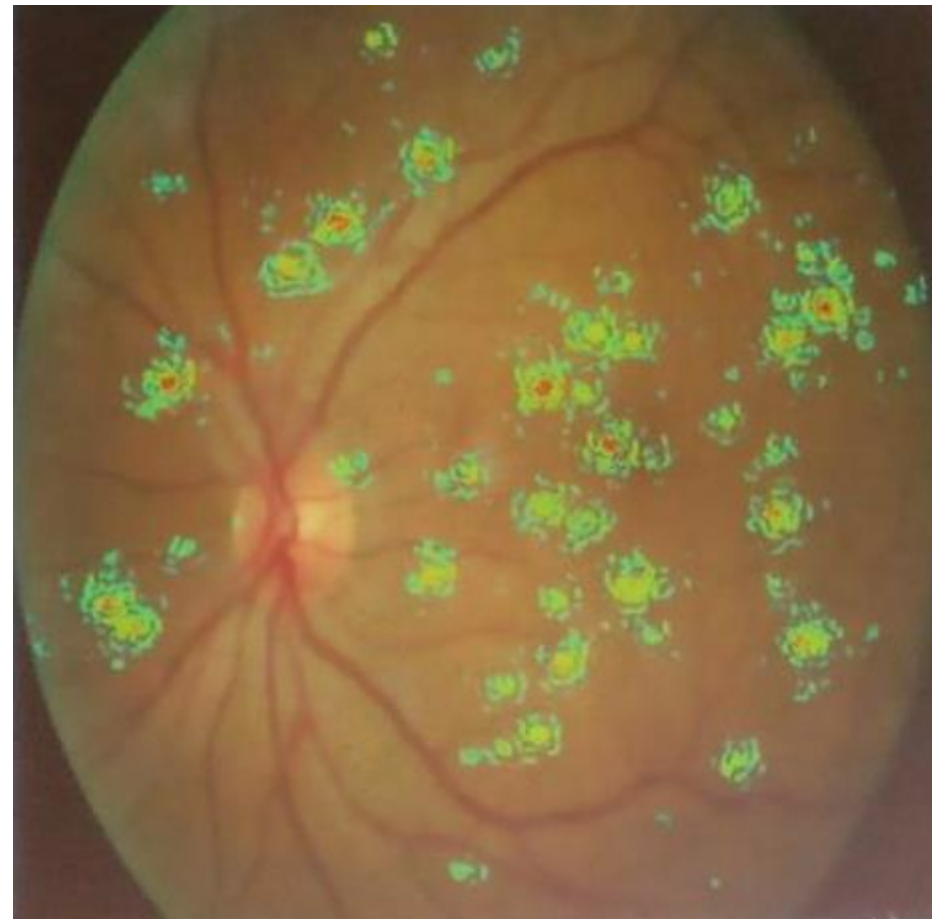
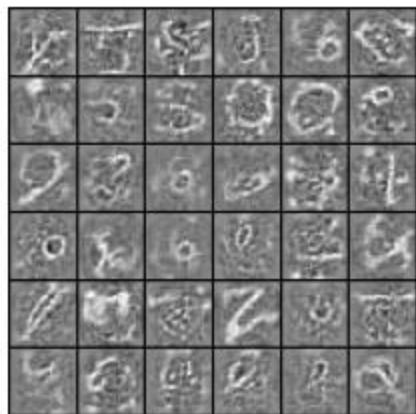
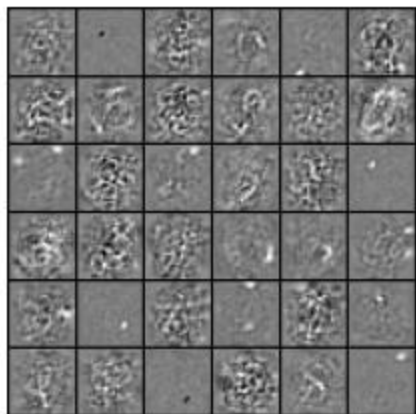
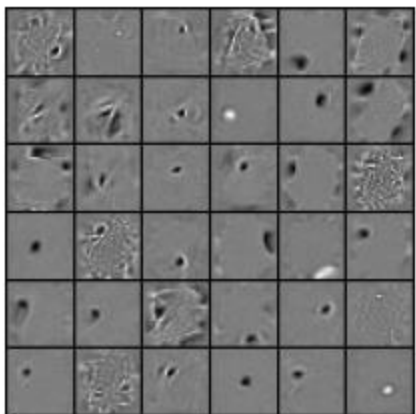
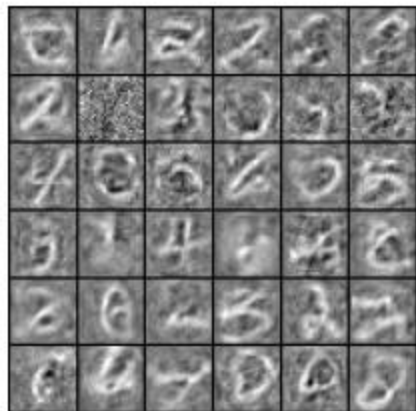
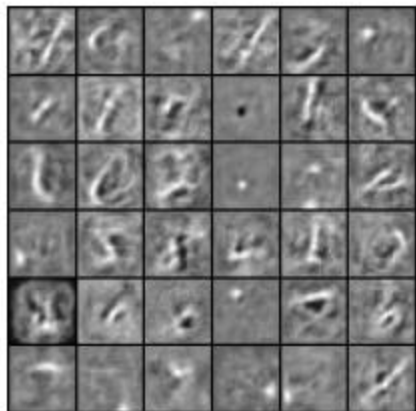
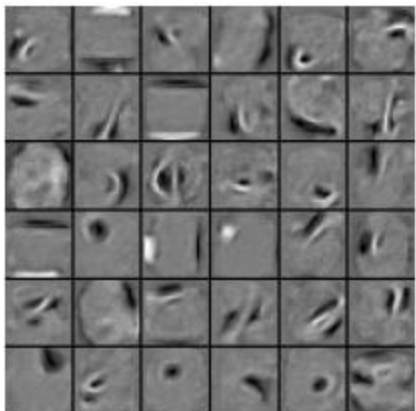
Explainable AI



- Understanding model decisions is a requirement for our users
 - Business users must have **confidence** in models before deploying them
 - **Regulations** require that decisions based on algorithms be made “understandable by non-experts”
 - For marketing use cases, explanations also make decisions **actionable**

CALL	PREDICTED SCORE	ALERT	VARIABLE IMPACT	VALUE	IMPACT
	Threshold: 0.32				
0	0.639  Y	All Clear!	Tags:Investment Theme... Property Commodities Total AUM Tags:Investment Theme...	1 99193.84 123992.3 2730231.85 0	
1	0.648  Y	All Clear!	Property Tags:Investment Theme... Total AUM Fixed Income Tags:Investment Theme...	88167.32 1 2075721.9433 154292.81 1	

- Also much academic interest around this research topic
 - DARPA XAI program: <https://www.darpa.mil/program/explainable-artificial-intelligence>



Erhan, Bengio et al. 2013

DreamQuark 2016

Cell sensitive to position in line:

```
The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.
```

Cell that turns on inside quotes:

```
"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.
```

```
Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."
```

Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
    siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!(current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

A large portion of cells are not easily interpretable. Here is a typical example:

```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
}
```

Cell that turns on inside comments and quotes:

```
/* Duplicate LSM field information. The lsm_rule is opaque, so
 * re-initialized. */
static inline int audit_dupe_lsm_field(struct audit_field *df,
    struct audit_field *sf)
{
    int ret = 0;
    char *lsm_str;
    /* our own copy of lsm_str */
    lsm_str = kstrdup(sf->lsm_str, GFP_KERNEL);
    if (unlikely(!lsm_str))
        return -ENOMEM;
    df->lsm_str = lsm_str;
    /* our own (refreshed) copy of lsm_rule */
    ret = security_audit_rule_init(df->type, df->op, df->lsm_str,
        (void **) &df->lsm_rule);
    /* Keep currently invalid fields around in case they
     * become valid after a policy reload. */
    if (ret == -EINVAL) {
        pr_warn("audit rule for LSM '%s' is invalid\n",
            df->lsm_str);
        ret = 0;
    }
    return ret;
}
```

Cell that is sensitive to the depth of an expression:

```
#ifdef CONFIG_AUDITSYSCALL
static inline int audit_match_class_bits(int class, u32 *mask)
{
    int i;
    if (classes[class]) {
        for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
            if (mask[i] & classes[class][i])
                return 0;
    }
    return 1;
}
```

Cell that might be helpful in predicting a new line. Note that it only turns on for some "":

```
char *audit_unpack_string(void **bufp, size_t *remain, si
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
    if (len > PATH_MAX)
        return ERR_PTR(-ENAMETOOLONG);
    str = kmalloc(len + 1, GFP_KERNEL);
    if (unlikely(!str))
        return ERR_PTR(-ENOMEM);
    memcpy(str, *bufp, len);
    str[len] = 0;
    *bufp += len;
    *remain -= len;
    return str;
}
```

- Several approaches stand out in the community:
 - **LIME** [Ribeiro et al, 2016 & <https://github.com/marcotcr/lime>]
 - Model agnostic, generates and scores random points around an example
 - Trains a simple (e.g. linear) model to reproduce this local behavior
 - **SHAP** [Strumbelj et al, 2014 & <https://github.com/slundberg/shap>]
 - Game theory approach: “connects optimal credit allocation with local explanations using the classic Shapley values from game theory”
 - **Permutation** Feature Importance
- DreamQuark method for individual explainability [patent pending]
 - For each ("local") prediction, use the neural network's underlying architecture to determine (locally) the most important variables for the model's decision
 - Decorrelate input variables in order to present more consistent explanations
- Definitely not the end of the story here, research is ongoing

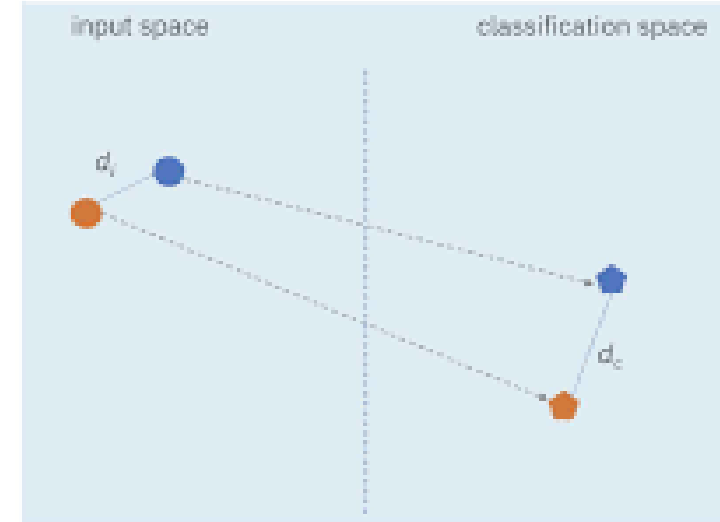
Challenges for real-life AI systems



- What constitutes fairness ?
 - *Demographic blindness* (training data is not correlated with demographics)
 - *Demographic parity* (outcomes are proportionally equal for all classes)
 - *Equal opportunity* (true-positive rates are equal for all classes)
 - *Equal odds* (true-positive and false-positive rates are equal for all classes)
- “impossibility theorems” of fairness, see e.g. [Zhao&Gordon, 2019]
- *Equality of Opportunity in Supervised Learning* [Hardt et al, 2016]
 - Pre-define sensitive groups before training a model
 - => ensure no discrimination against protected groups in the learned model

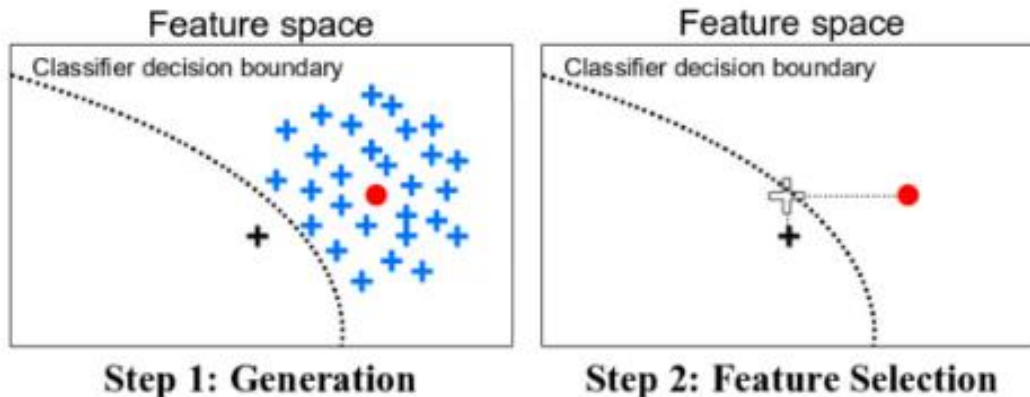
- Definition of individual fairness [Dwork et al, 2011]
=> “similar individuals are treated similarly”
- In other words: a fair model is a robust model
- Difficulty: how do we define *similar* inputs ?

Fairness Through Awareness



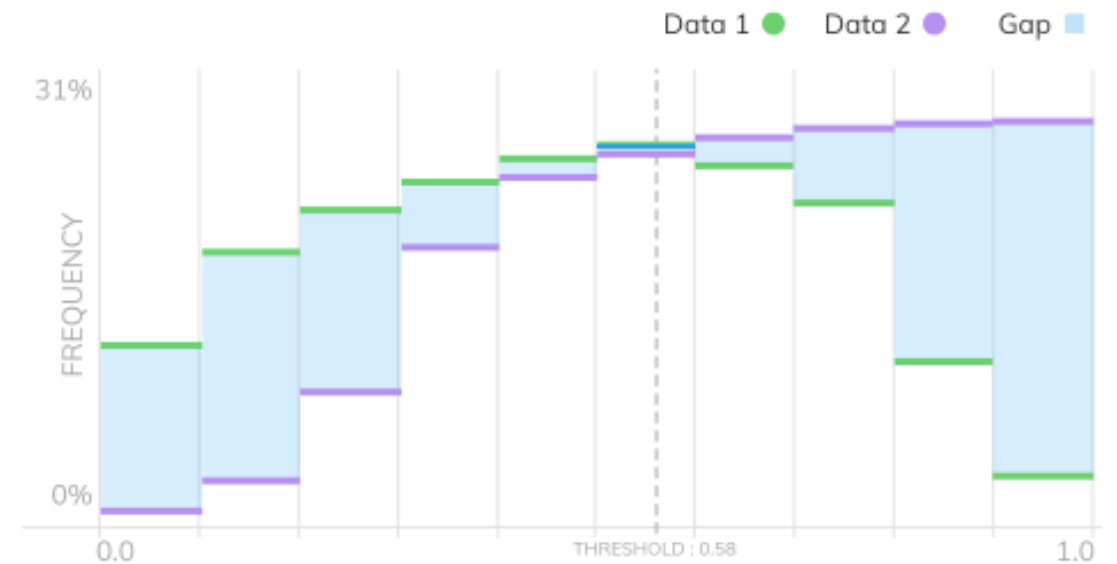
Dwork et al. Fairness Through Awareness, 2011.

- Alternative approach: counterfactual explanations
=> Find the “closest” neighbor which the model predicts differently



[Lauger et al, 2017]
Method of « growing spheres »
for selecting sparse counterfactuals

- Model performance (e.g. AUC, accuracy), even on *test data*, does not guarantee robustness and results in production
- => models must be monitored in production, and eventually retrained
- **Model** stability (e.g. PSI)
- **Data** stability / data shifting
 - Does the distribution of datapoints in production match training data ?
 - Train a simple model to distinguish test data vs train data
 - If successful, your data has shifted



DreamQuark 2020

- Possible added benefit: robustness of AI system against adversarial attacks

Take-aways



- Physicists can successfully transfer their skills and experience
- Innovation is possible in the industry, even as a small company
- As a private company: use open source software, develop proprietary code and file patents, but also publish open-source code
- Deep learning on tabular data is an emerging topic in AI community
- Beyond accuracy of models, challenges for both academia and industry
 - Explainable AI
 - Fairness
 - Robustness
 - Model stability and data stability



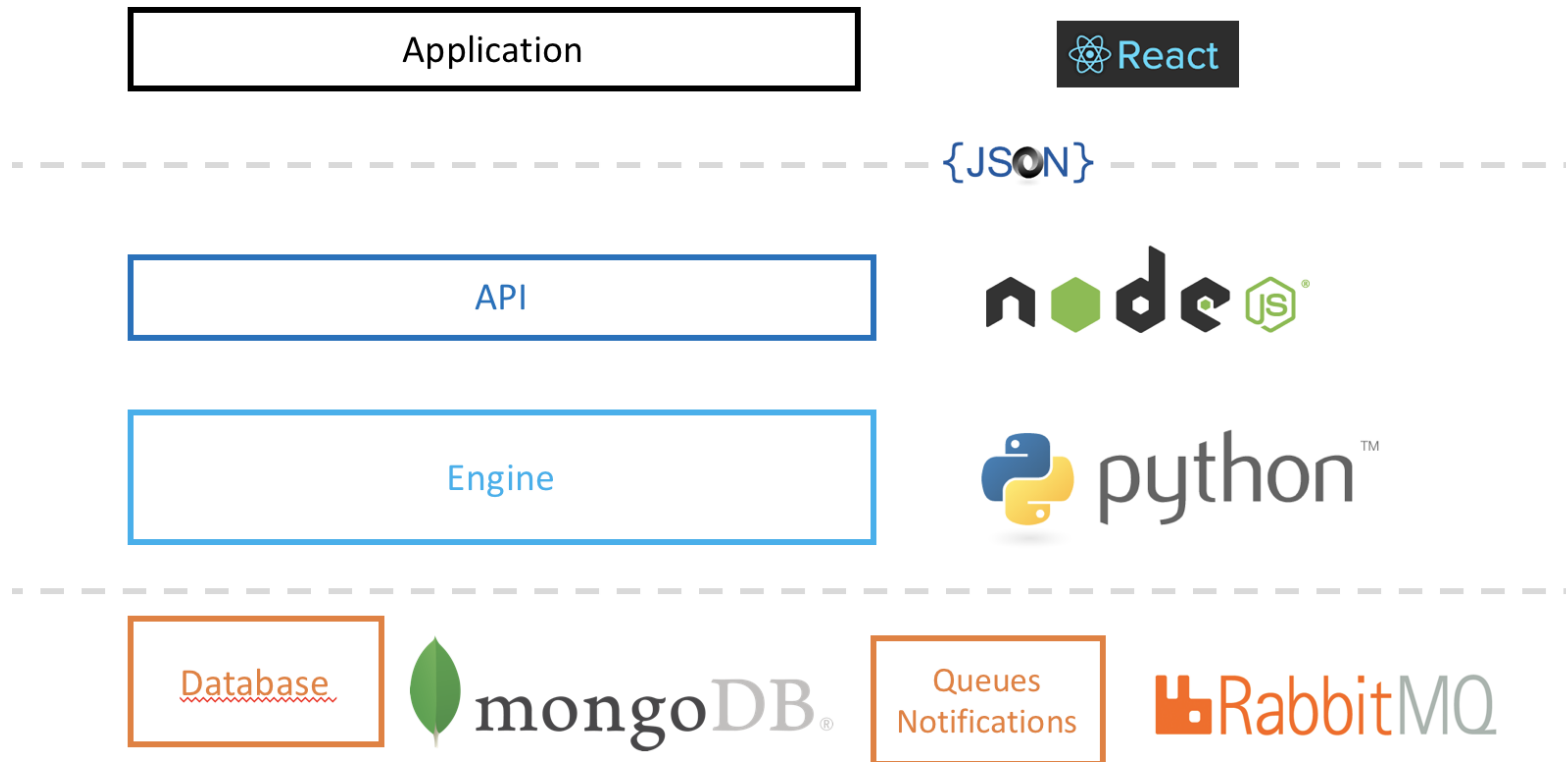
Jeremie Abiteboul
Chief Data Science Officer
jeremie.Abiteboul@dreamquark.com



Appendix



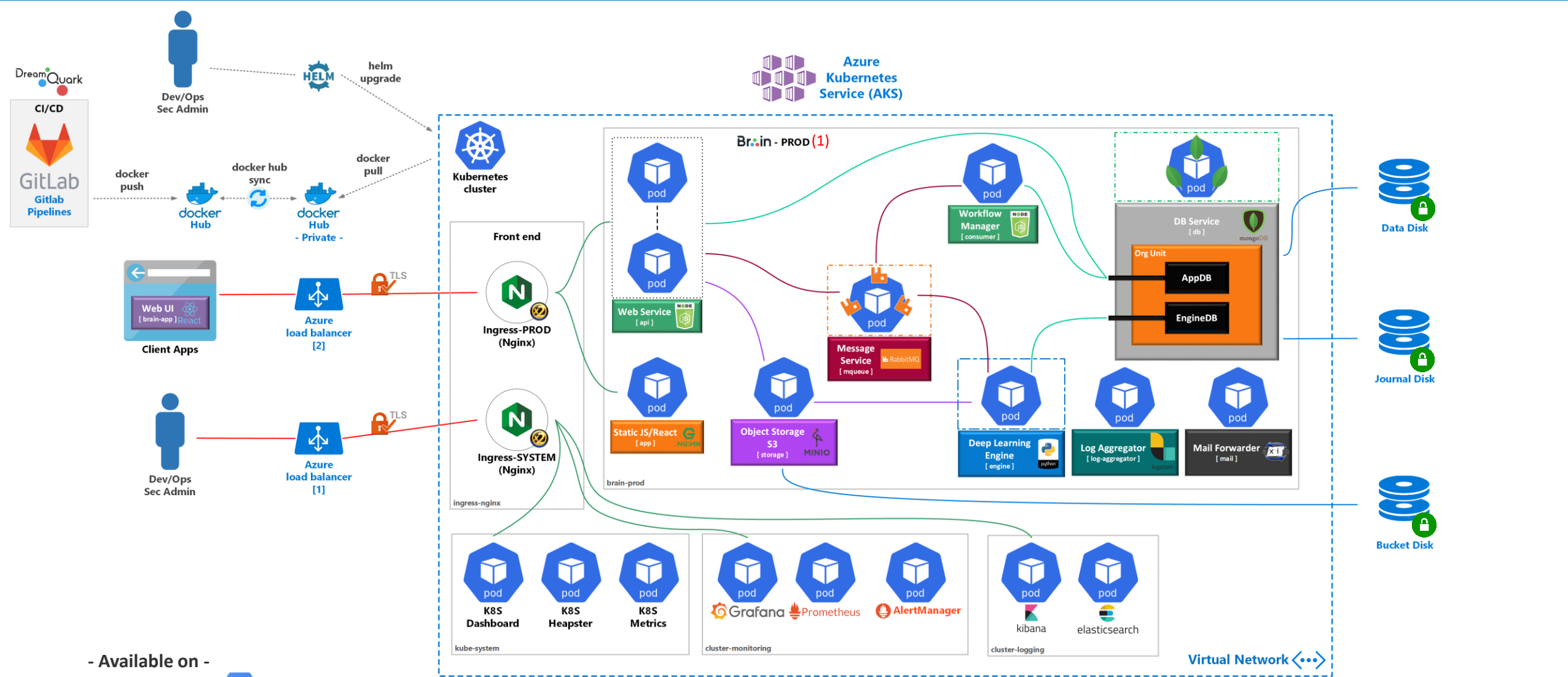
- Initial problem to solve:
 - Neural networks are a "universal approximator", but typically require significant manual effort for each new dataset
 - Many hyper-parameters: architecture, learning rate, regularization, etc
=> many-dimensional space for optimization, often carried out more or less manually
- Our solution to this problem, in a nutshell
 - Carry out training of a large number of models on a reduced dataset and/or for a small number of epochs, in order to gain "insights" on the dataset
 - => automatically deduce a reduced search space for hyperparameters
 - Optimize on this reduced search space using Bayesian methods



- + systematic use of Docker + Kubernetes for deployment (both cloud and on-premise)

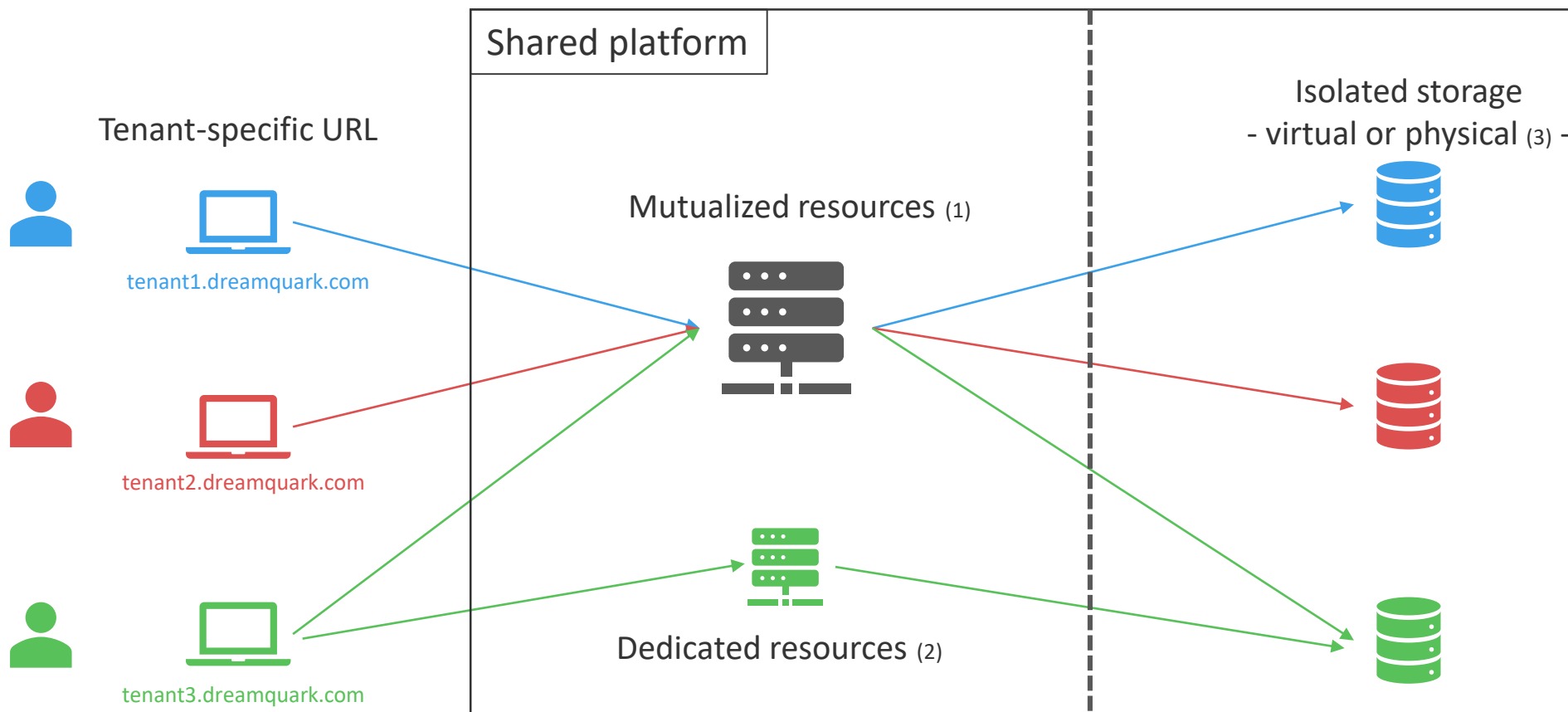
- Kubernetes: industry standard for container orchestration
- Deployment compatible with **multiple cloud providers & on-premise**
- Automated delivery of new releases
- Monitoring tools deployed as part of the product
- Easily manage many (independently scalable) micro-services





- Available on -





1. Instances of Brain micro-services (API or engine)
2. Only instances of engine micro-services
3. MongoDB: separated instances or separated databases

S3 Object Storage:
Separate Buckets or isolated "folders" in the same Bucket

Storage can be external and managed by the administrator