AI4Business
Deploying AI Solutions

Roadmap AI4Business

Part I

Introduction to AI

Part II

Developing AI tools

Part III

Data and Value

Part IV

Deploying AI

Part V

Monitoring

Table of contents

1 Real Life AI

AI is everywhere

Alphabet
Google

Self-driving cars
Play several Atari games and Go
Voice interface to make phone calls & schedule your appointments

IBM

Man vs. machine competitions:
Deep Blue (chess)
Watson (tv quiz Jeopardy)
Debater (professional debates)

Alibaba

Predict what you buy
Generate product descriptions
Reduce traffic jams in smart cities
Monitor farming crops

amazon

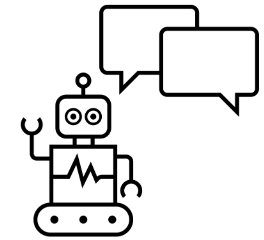
Digital assistant Alexa
Ship before you buy
Buy without checkout

NETFLIX

Content recommendations
Optimized streaming
Autogenerate personalized thumbnails

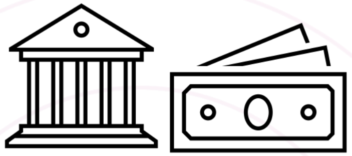
FaceID & Siri
Song recommendations
Navigation in Maps

Example: AI in the banking sector



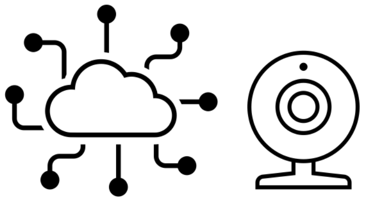
AI-powered chatbots

Mobile banking apps

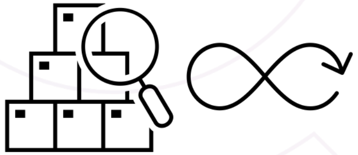


Loan default prediction

Fraud detection



Cyber-security



Product cross- or upselling

If AI systems are everywhere…

- *How easy is it to build an AI solution?*
- *What are the requirements to build an AI solution?*
- *Is building an AI solution the same as building any piece of software?*
- *What are the challenges to make my AI solution work?*
- *What kind of special tools do I need to build an AI solution?*

Why are AI solutions so difficult?

Data Management

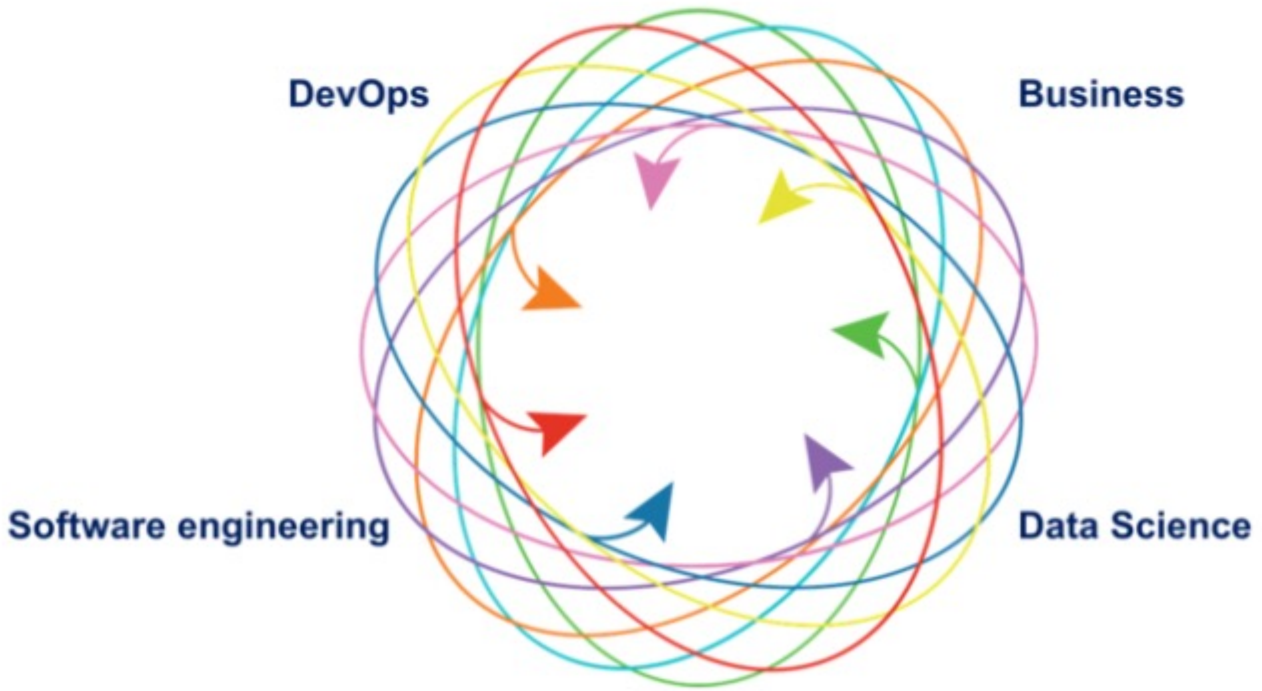
Configuration

Entanglement

Different expertise

Model Governance

AI requires high collaboration



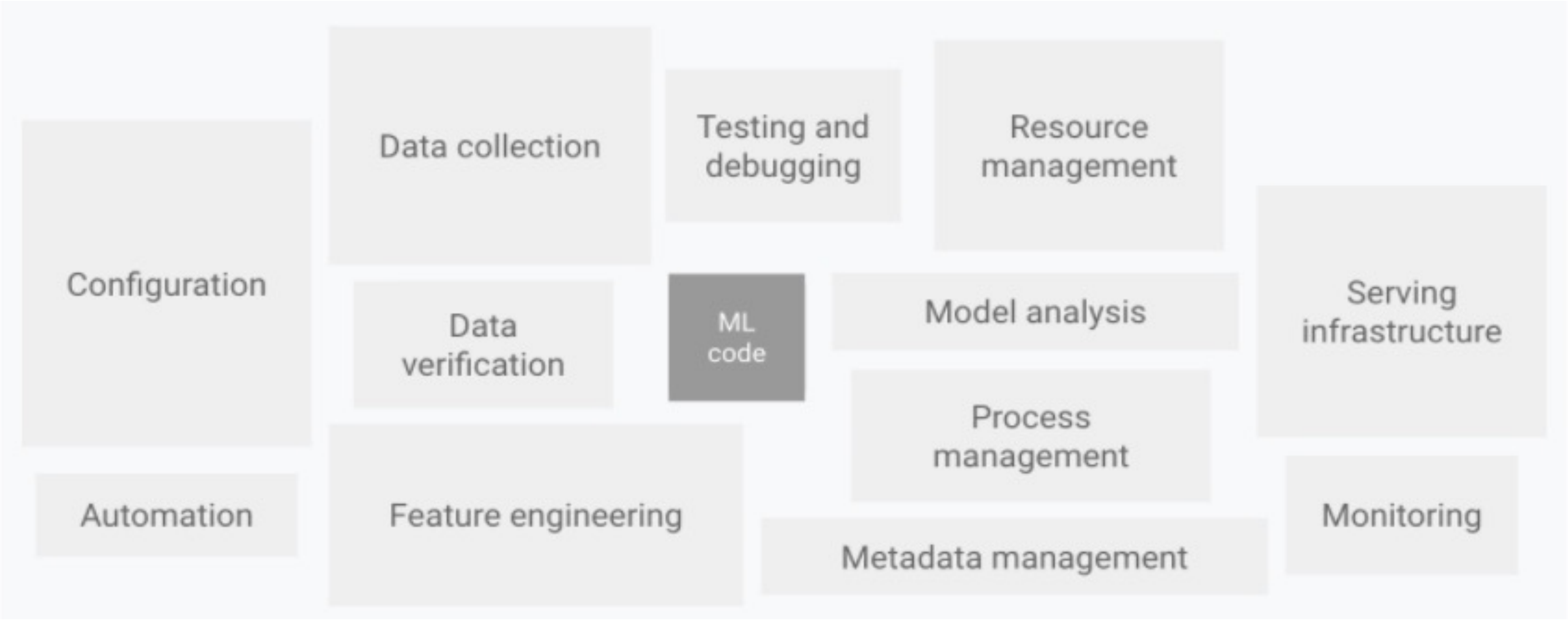
2 PoC to Production Gap

PoC versus Production

"All of AI, .., has a proof-of-concept-to-production gap. The full cycle of a machine learning project is not just modelling. It is finding the right data, deploying it, monitoring it, feeding data back [into the model], showing safety—doing all the things that need to be done [for a model] to be deployed. [That goes] beyond doing well on the test set, which fortunately or unfortunately is what we in machine learning are great at."

- Andrew Ng

The big picture



[Google - CD and automation pipelines in ML]

Basic ML building blocks

Data Management	Experimentation	Production
Process and govern the data used by models: • Usually large data sets • Should be of high quality • Should be compliant with legislation • Should be tracked	Build a model based on business requirements, after iteration of experimentation: • Workflow is iterative • Experiment should be tracked • Code should have standards • Accuracy metrics should be tracked • Retraining should be possible • Requires specific infrastructure	Integrate prediction into production and business processes: • Generate systematic predictions • Track performance across time • Follow best engineering practices

Moving to production is hard

(Not so) Fun fact

According to VentureBeat, roughly 1 out of 10 Machine Learning models actually makes it into production. But why?

The Set up is not right

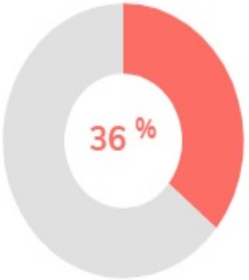
- Bad infrastructure
- Disconnect between the relevant parties
- Poor data management
- Leadership doesn't understand

ML has its own difficulties

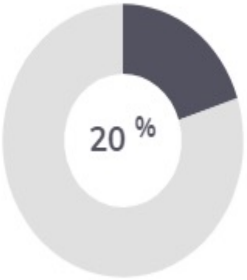
- Scaling is not easy
- Duplication is widespread
- Management not on board
- Lack of Reproducibility
- Support across technologies

Deploying models takes time

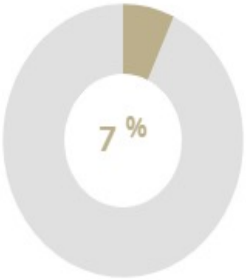
36% of survey participants said their data scientists spend **a quarter** of their time deploying ML models



36% of survey participants said their data scientists spend **a quarter to half** of their time deploying ML models



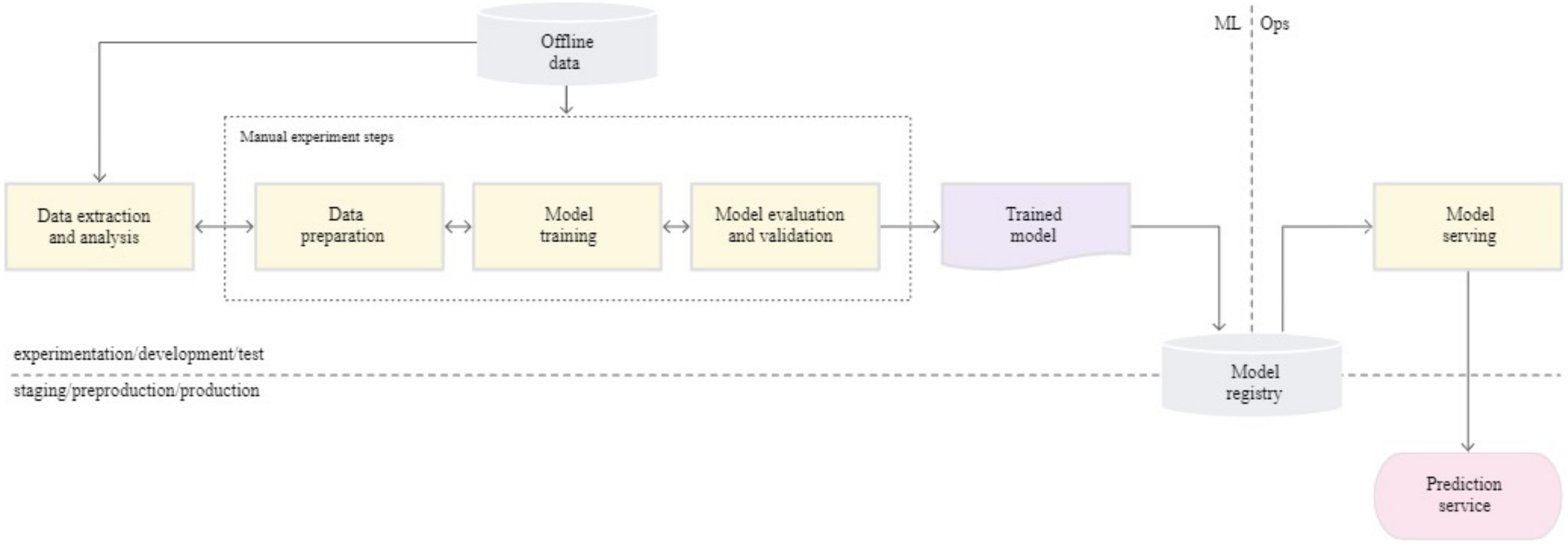
20% of survey participants said their data scientists spend **half to three-quarters** of their time deploying ML models



7% of survey participants said their data scientists spend **more than three-quarters** of their time deploying ML models

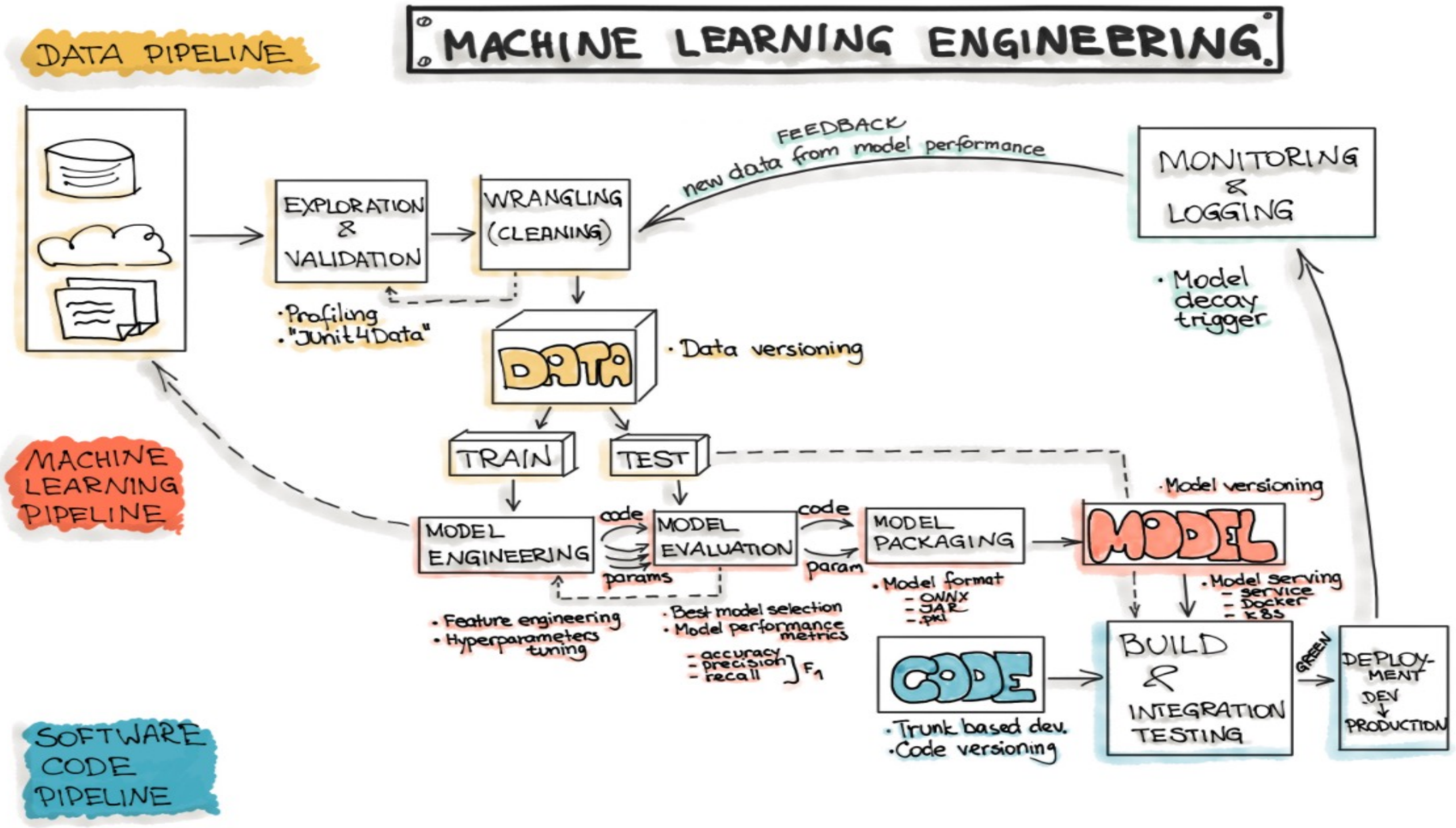
Algorithmia - State of Enterprise ML

Basic process for building a model



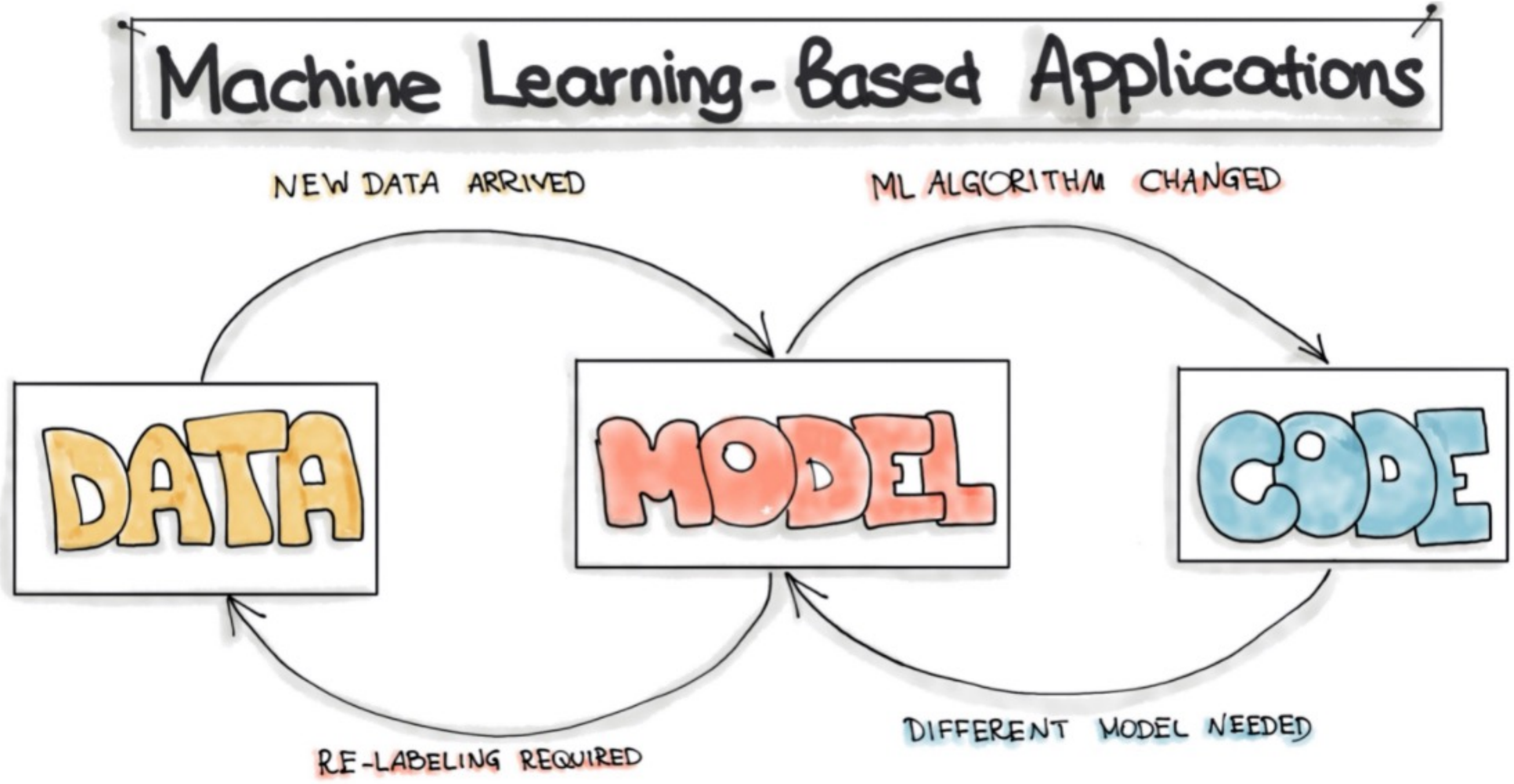
[Google - CD and automation pipelines in ML]

Real life is a bit more complicated



[MLOps.org - End-to-end workflow]

Changing anything changes all



[MLOps.org - Motivation for MLOps]

Hidden technical debt

Developing and deploying ML systems is relatively fast and cheap, but maintaining them over time is difficult and expensive. Some of the reasons for this are:

- Data dependencies cost more than code dependencies
- Feedback Loops
- ML-Systems anti-patterns
- Configuration debts
- Always changing external world
- Other ML related debt (e.g Data testing, Reproducibility debt)

Other production issues

- Data quality:
  - ML models reflect the data they are build on, so they are very dependent on its size and quality
- Model decay:
  - As times goes by, there might be changes in behavior that the original data would not necessarily reflect causing the quality of the model to drop
- Locality:
  - The quality of the performance of ML model does not always translates completely to production

3 Deployment Challenges

Challenges of building AI systems

Any software application comes with many challenges.

AI/ML brings around a couple of extra ones:

- Data is difficult to manage and resource consuming
- Iteration is necessary but slow
- The expertise needed is abundant and diverse
- Scaling quickly becomes an issue
- Maintenance becomes particularly difficult
- Selecting the right tool is not always so easy

Data is an investment

Having easily available and high quality data is expensive

Why invest in data?

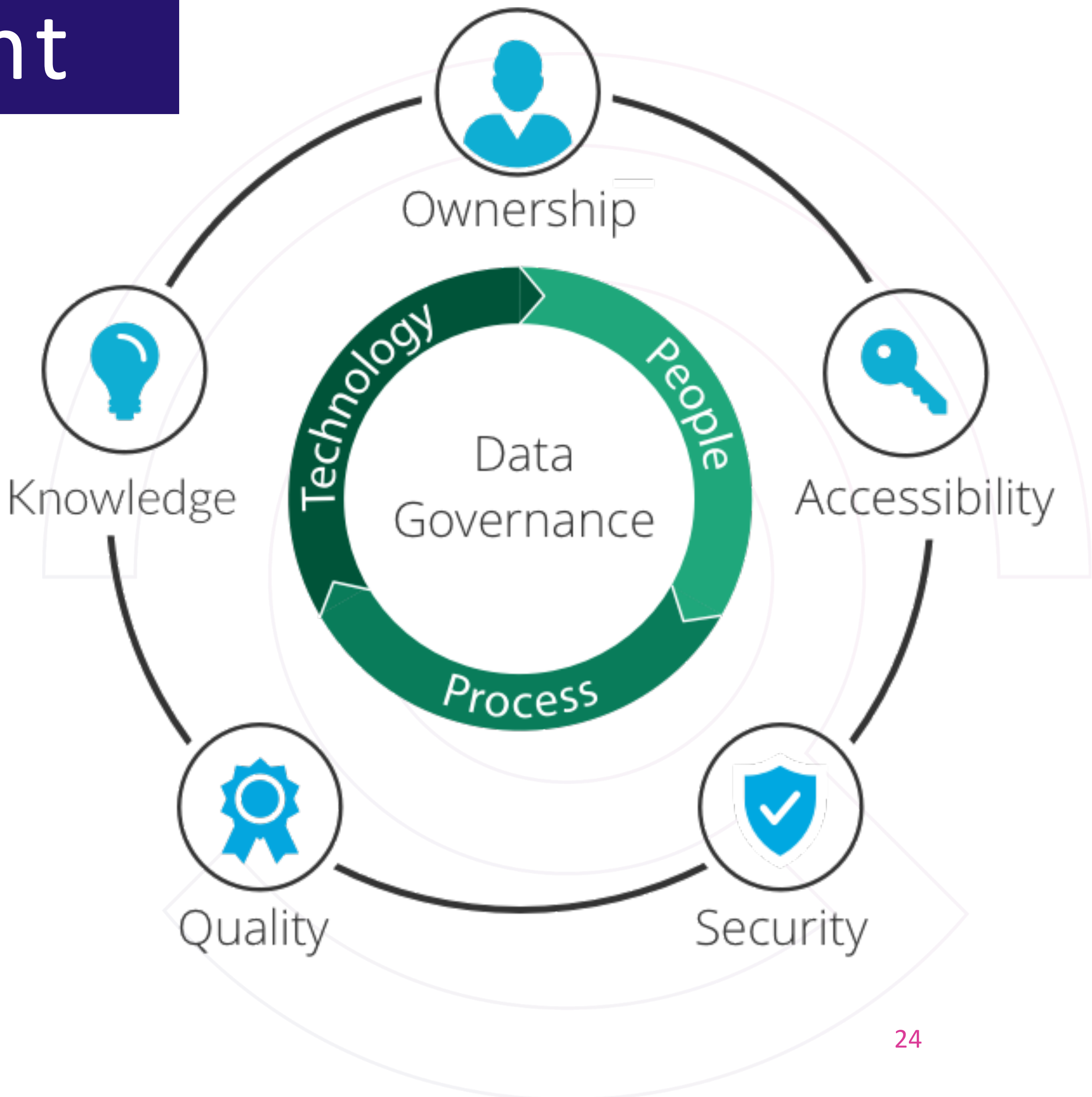
- Model quality depends on data quality
- Data is needed after deployment
- Data is worthless if not usable
- Data is at the core of the AI system

- Bad data increases complexity → need easy access to high quality data

Data is an investment

Data governance is key

- Being data driven is more than just buying expensive data
- Having the processes is as valuable as having the right tool
- Having a cohesive data strategy is the key to success.
- Data governance is not the same as Data management



[Inperva- Data Governance]

Iteration is a must

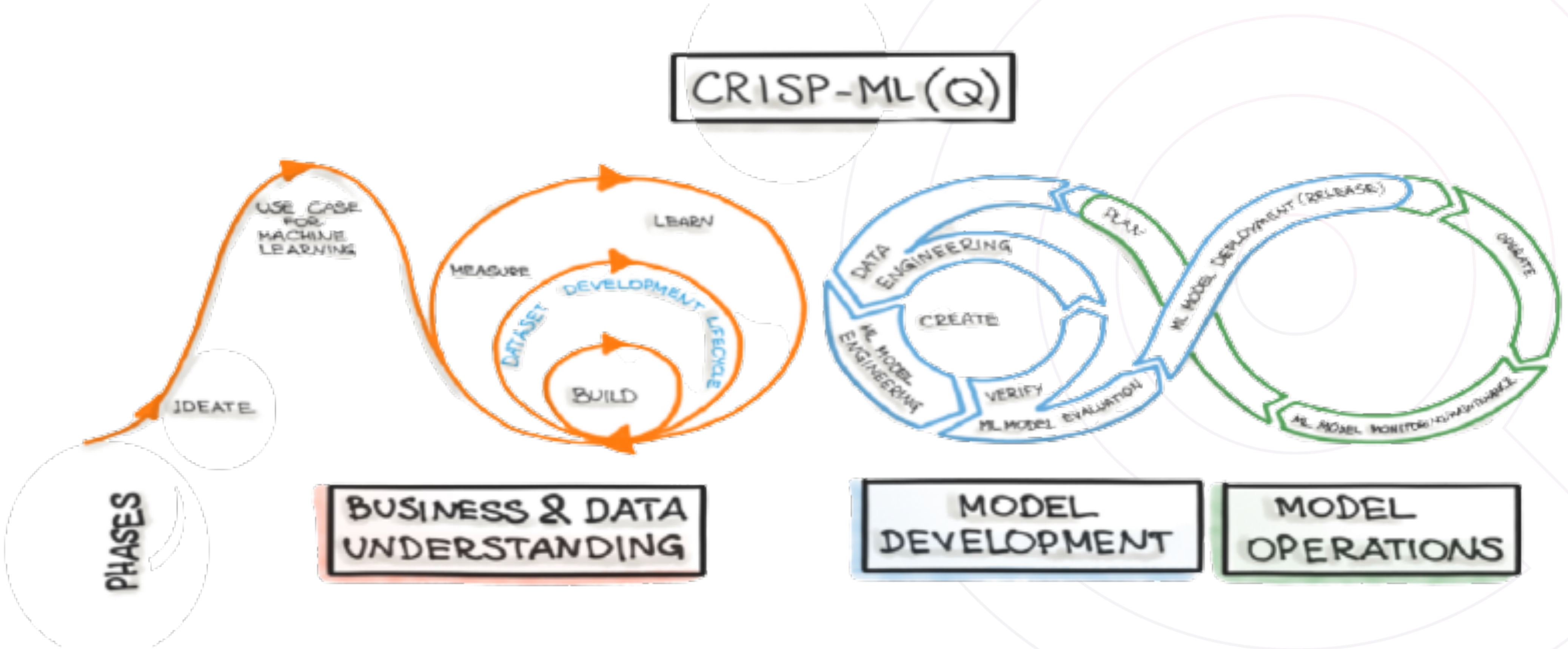
Be ready to iterate over:

- What does the business need?
- Do we have data as needed?
- Is data ready for modelling?
- What model should we build?
- Is our model good enough?
- How do we make results available?

CRISP-DM

Iteration is a must

As complexity of the environment increases, so does the workflow



MLOps.org - CRISP-ML(Q)

Business and Technical Leaders

Aligning business and technical leaders is not always easy.

But it is necessary to bridge the gap:

Business leaders

- Update their Data/AI literacy
- Understand the uncertainty around AI systems

Technical leader

- Set right expectations ahead of time
- Plan resources efficiently

Business and Technical Leaders

If deploying AI for the first time:

- Start small
- Look for low hanging fruits
- Look for problems with visible value

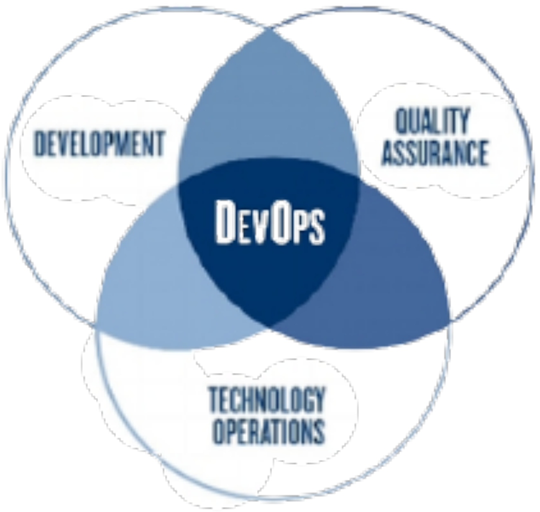
AI is not going to replace managers, but managers that use AI will replace those that do not

Remember: AI is not bulletproof, but when used correctly can be an extremely powerful tool

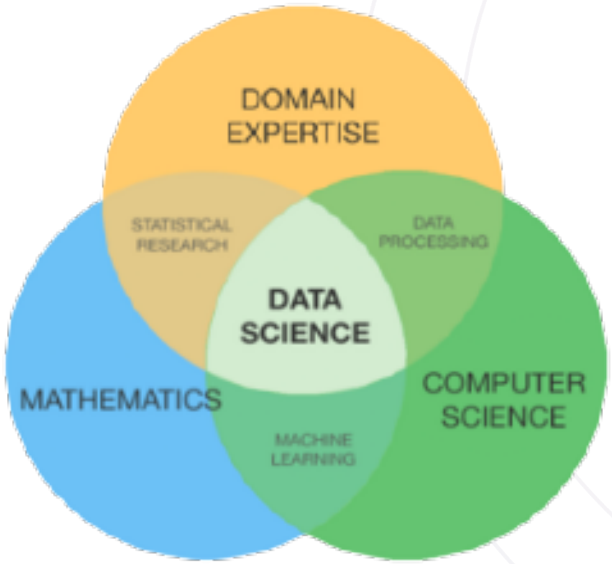
Technical teams working together

DevOps, IT and Data Scientist often organized in silos at organizations.

These silos must be connected* for AI. **Unless you found a unicorn that can do everything*

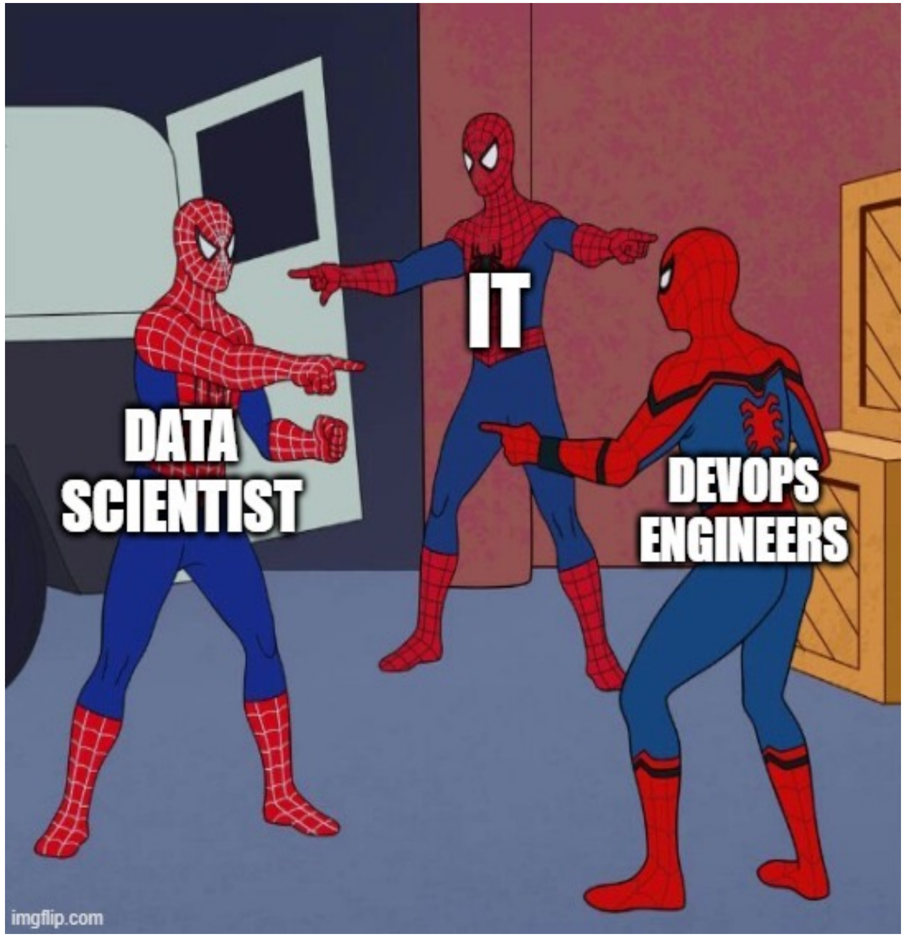


+



=

Technical teams working together

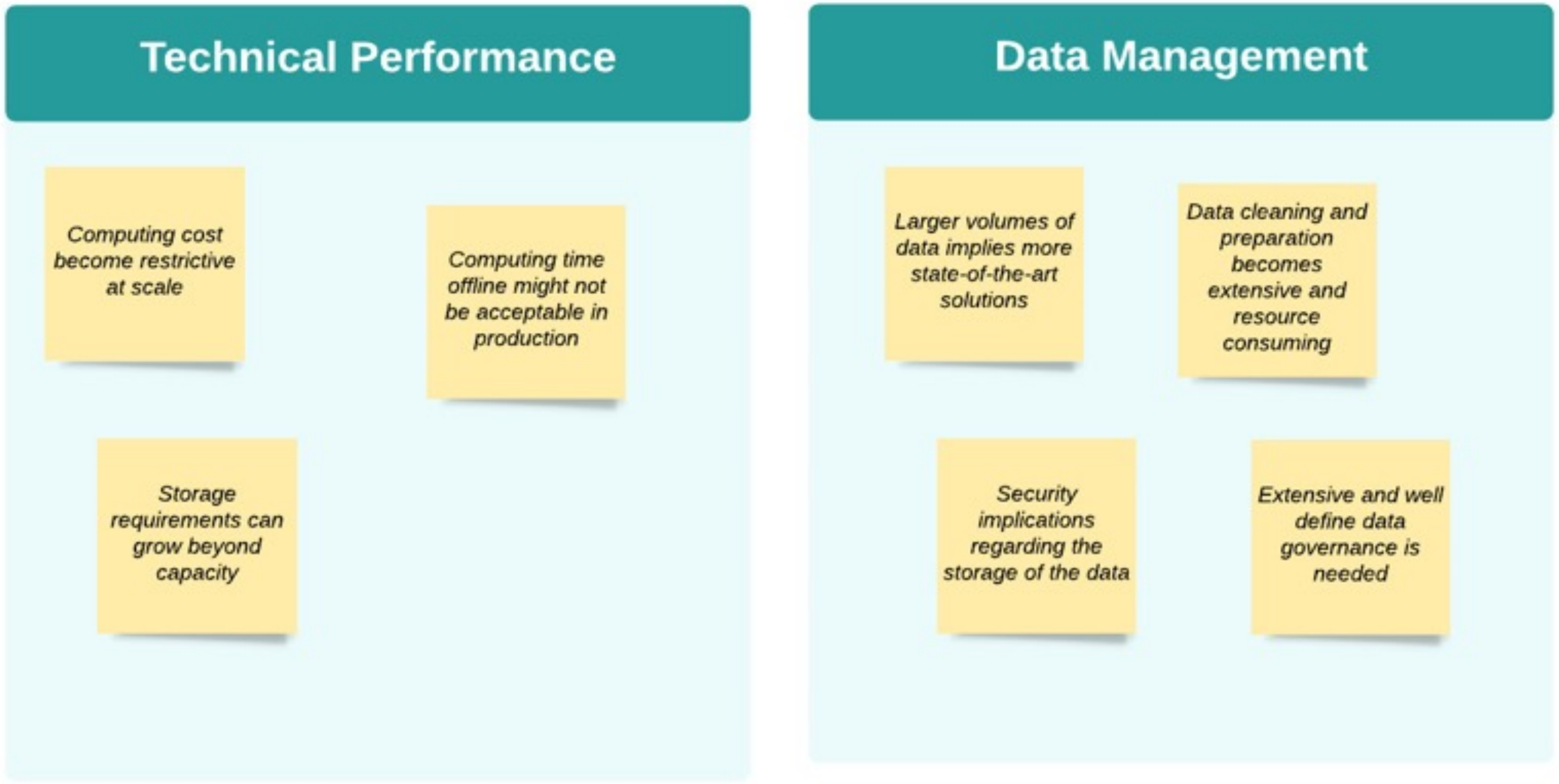


When working in silos:

- Impossible to have a high-level overview of the solution
- Constant blaming across teams
- Can’t tackle complex problems (e.g. real time applications)
- Maintenance rapidly becomes a nightmare

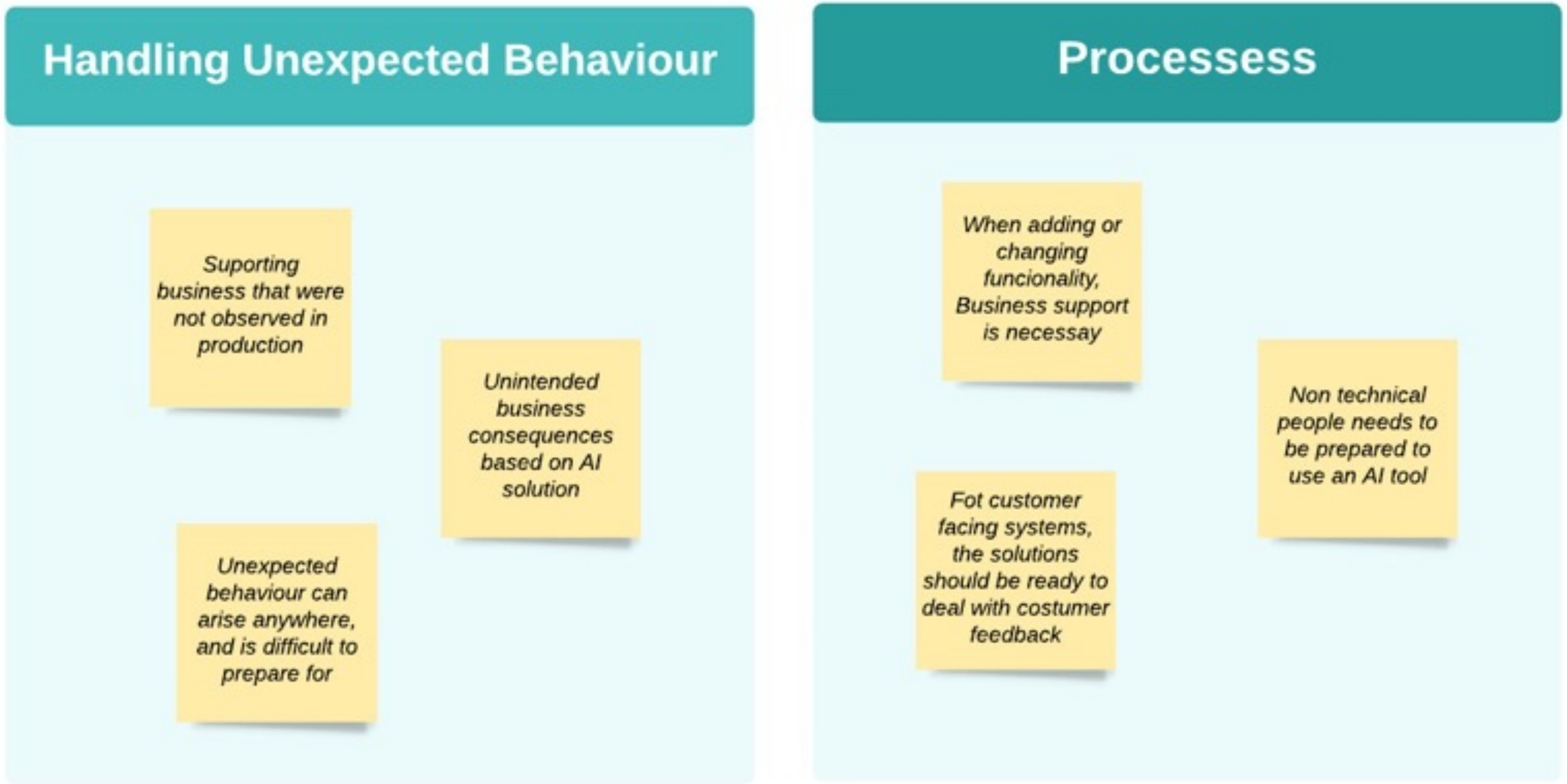
Think about scaling

Scaling AI solutions is not easy nor cheap



Think about scaling

Not only about money or resources but also about processes and behaviour



Data can significantly change

Data distributions can shift.

Assumption that past data is representative of future data is broken.

Data Drift

- Distribution of the features or target changes
- Past performance does not guarantee future results
- Models are not ever lasting but speed of decay increases

Concept Drift

- Occurs when patterns learned by the model no longer holds
- It might happen over time or suddenly
- Is more difficult to correct as is related to fundamentals

Maintaining AI solutions

"As the machine learning (ML) community continues to accumulate years of experience with live systems, a wide-spread and uncomfortable trend has emerged: developing and deploying ML systems is relatively fast and cheap, but maintaining them over time is difficult and expensive."

Sculley et al.

Maintaining AI solutions

Selecting right technology

Selecting the right tool for the problem at hand is not always simple, as the technology supporting AI is

- Diverse
- Fast growing
- Tailored

Remember: Don’t marry yourself to a tool. Tools are just means to an end.

Selecting right technology

Some general tips on selecting the right technology

Integration should be easy

- You are already on an ecosystem, new tools need to be easily integrable.

Flexibility is key

- Tools should easy to use and flexible to customization

Scalability is your friend

- Not all tools scale well for all problems

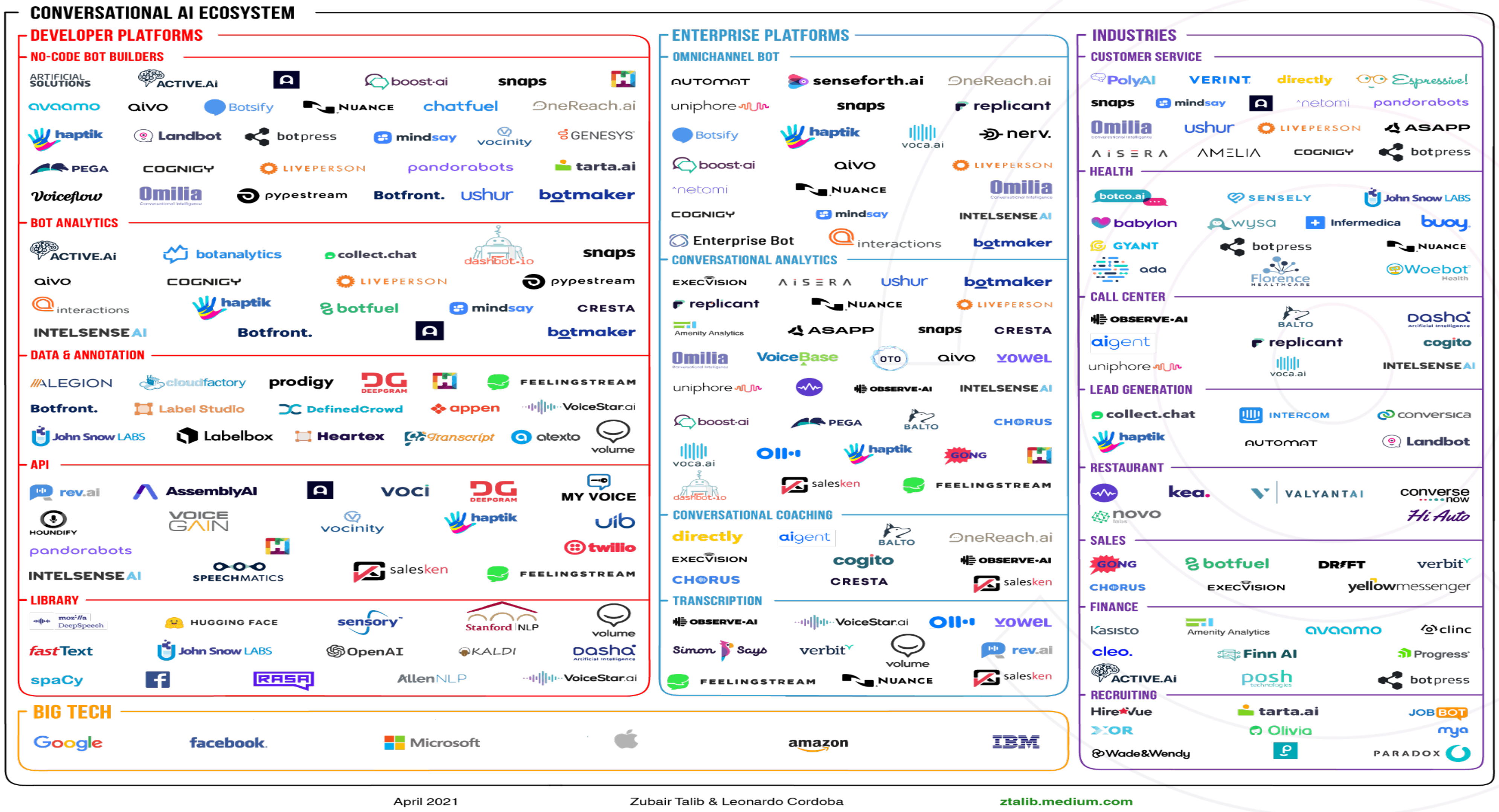
Right tool for the right job

- There are many tools available, no need to overspend

Support is important

- Having good support and a large community behind is key

AI Ecosystem is crowded



4 Data-Centric AI

The next big step

"The model and the code for many applications are basically a solved problem. Now that the models have advanced to a certain point, we got to make the data work as well."

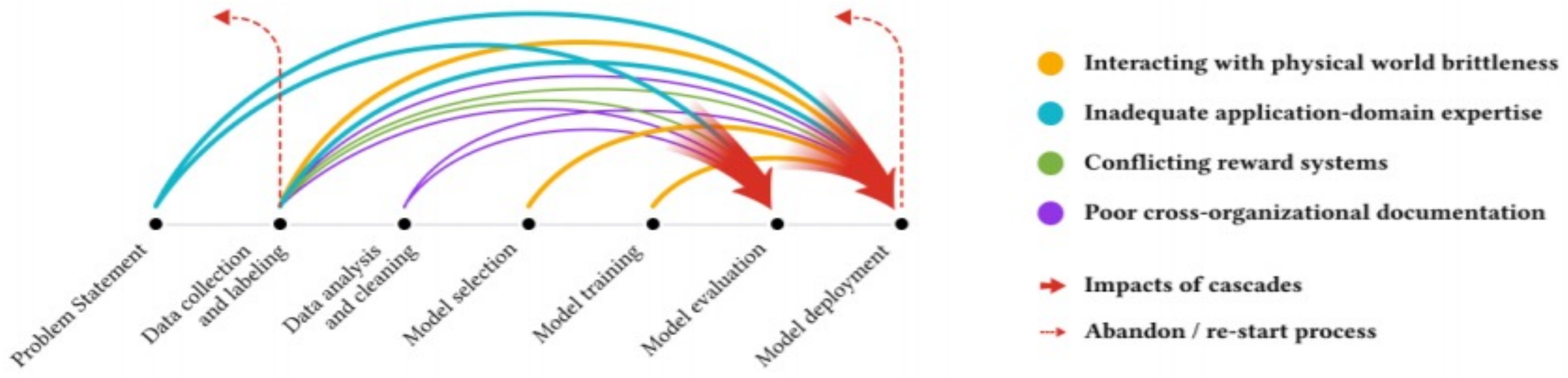
"Many data scientists have their own ways to clean data but what we don't have is a systematic mental framework for doing it"

"Just like the rise of deep learning a decade ago spawned tons of new jobs, I hope that data-centric AI development will spawn tons of new jobs in many industries."

-Andrew Ng

Data Cascades

Data Cascades are compounding events causing negative, downstream effects from data issues, that result in technical debt over time



Sambasivan et al, 2021

From Big data to Good data

The rise of Big data allows companies to extract value from AI.

Models improved, but what if we focus on improving the data instead?

What is good data?

- Defined consistently
- Cover of important cases
- Has timely feedback from production data
- Sized appropriately

Model-Centric status quo

Since the Big Data revolution, Data Scientists focused on improving models and code as much as possible, rather than the data.

This is expected

- Industry follows academic development closely
- Benchmark data sets are used so development focus on models
- Open source culture has made model improvement accessible

This is not necessarily a bad thing

- Following this approach AI has made tremendous progress

Data-Centric approach

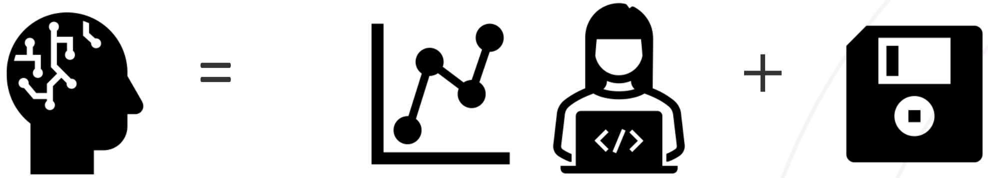
Aim of the data-centric approach is to make data quality a systematic issue

Why?

- Improving the data can greatly improv the quality
- Maintainability
- Ease of deployment

Model-Centric vs Data-Centric

AI = Code (and Model) + Data



Model-Centric

- Collect all the data you can
- Develop a model good enough to deal with that data
- Hold data fixed and iteratively improve the model

Data-Centric

- Consistency and quality of the data is paramount.
- Allows multiple models to do well
- Hold the code fixed and improve the data

Model-Centric vs Data-Centric

	Steel defect detection	Solar panel	Surface inspection
Baseline	76.2%	75.68%	85.05%
Model-centric	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
Data-centric	+16.9% (93.1%)	+3.06% (78.74%)	+0.4% (85.45%)

Takeaways

Systematic improvements

- For this approach to work, it should be and efficient and systematic process

High quality data

- Important to guarantee high quality data in the whole project life cycle

Is time to improve our approach

- The model-centric approach has taken us very far but we can still go further

Investing in data pays off

- Yet another reason why having a solid data strategy helps to improve results

A new systematic framework

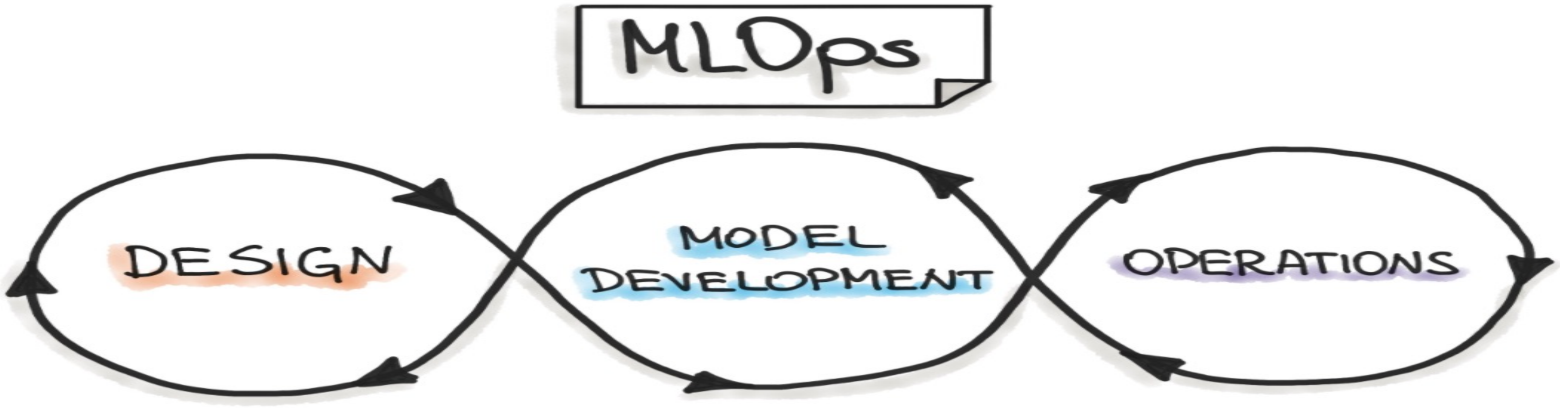
We know

Making data quality systematic improves general performance

But…What if?

We use the same principle for the whole project lifecycle

Enters…



PROPHECY LABS

ctg

nannyML

AI4Business

ESF
INVESTEERT IN
JOUW TOEKOMST

Europese Unie

Vlaanderen
is werk