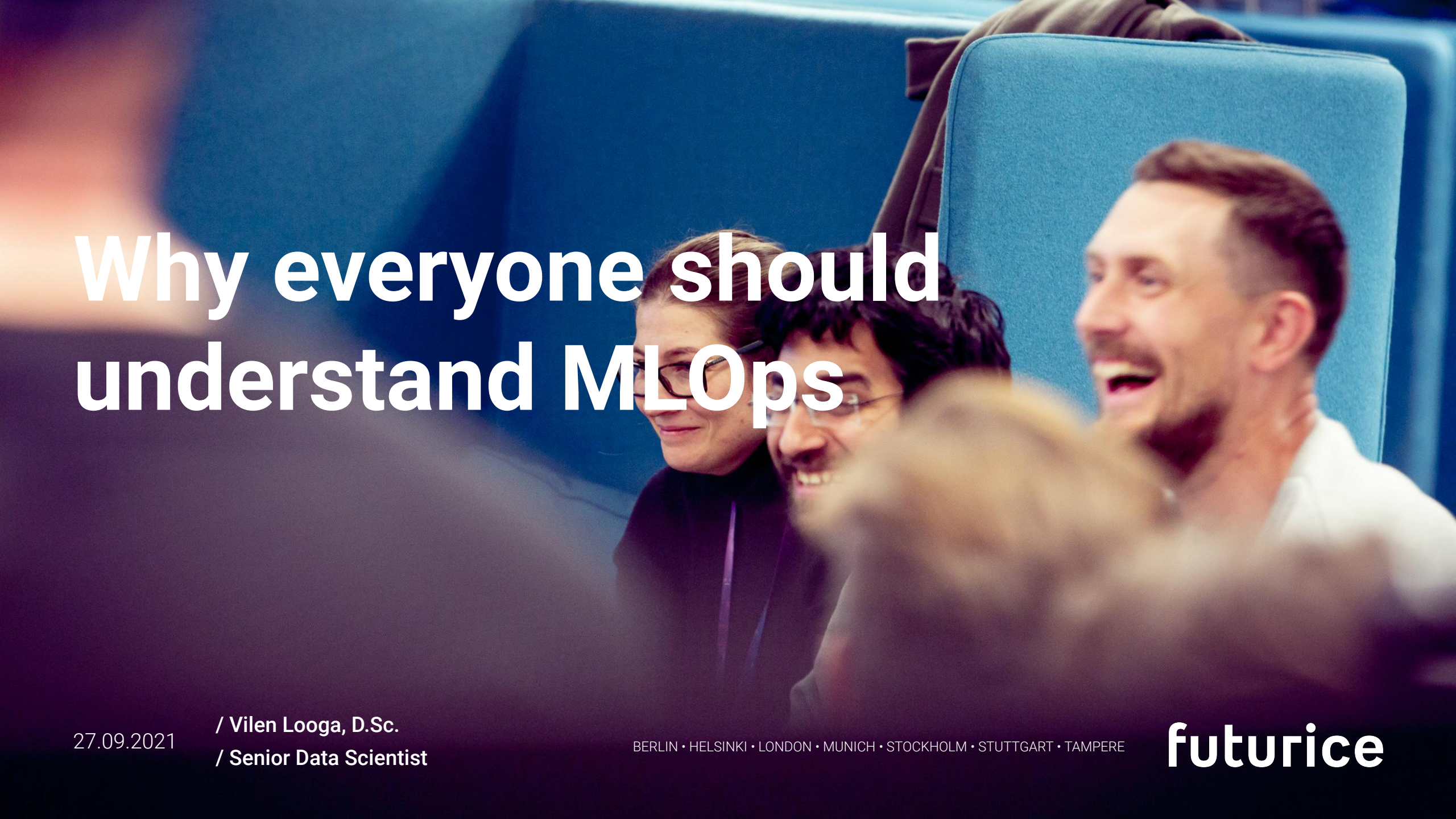




Why everyone should understand MLOps

27.09.2021

/ Vilen Looga, D.Sc.
/ Senior Data Scientist

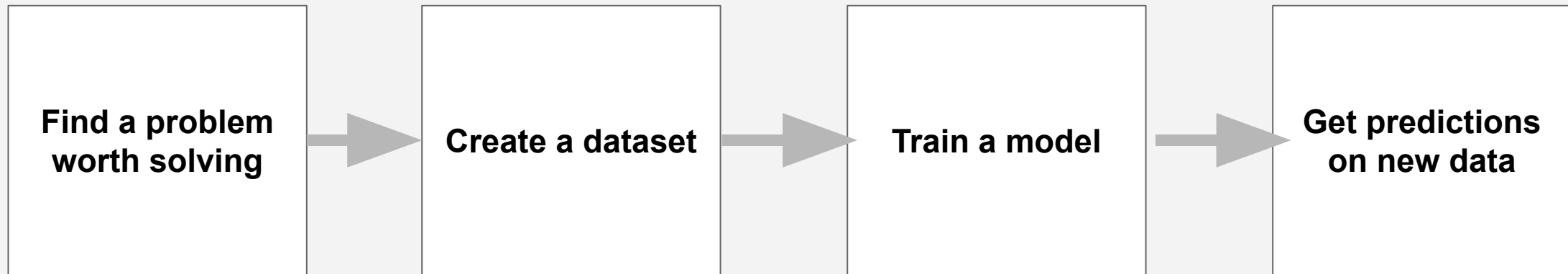
BERLIN • HELSINKI • LONDON • MUNICH • STOCKHOLM • STUTTGART • TAMPERE

futurice

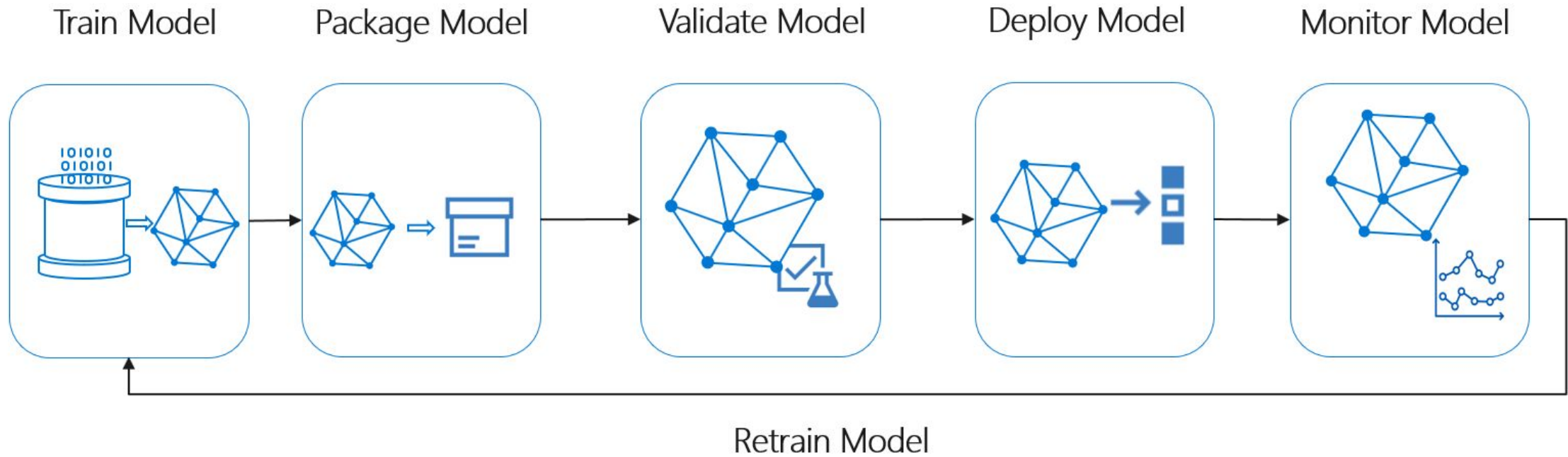
Who am I?

- Senior Data Scientist at Futurice Oy
- Mostly working on end-to-end machine learning solutions for clients in construction, healthcare, public sector etc.
- Previously:
 - Co-founded a computer vision startup
 - Doctor of Science from Aalto University, Computer Science department

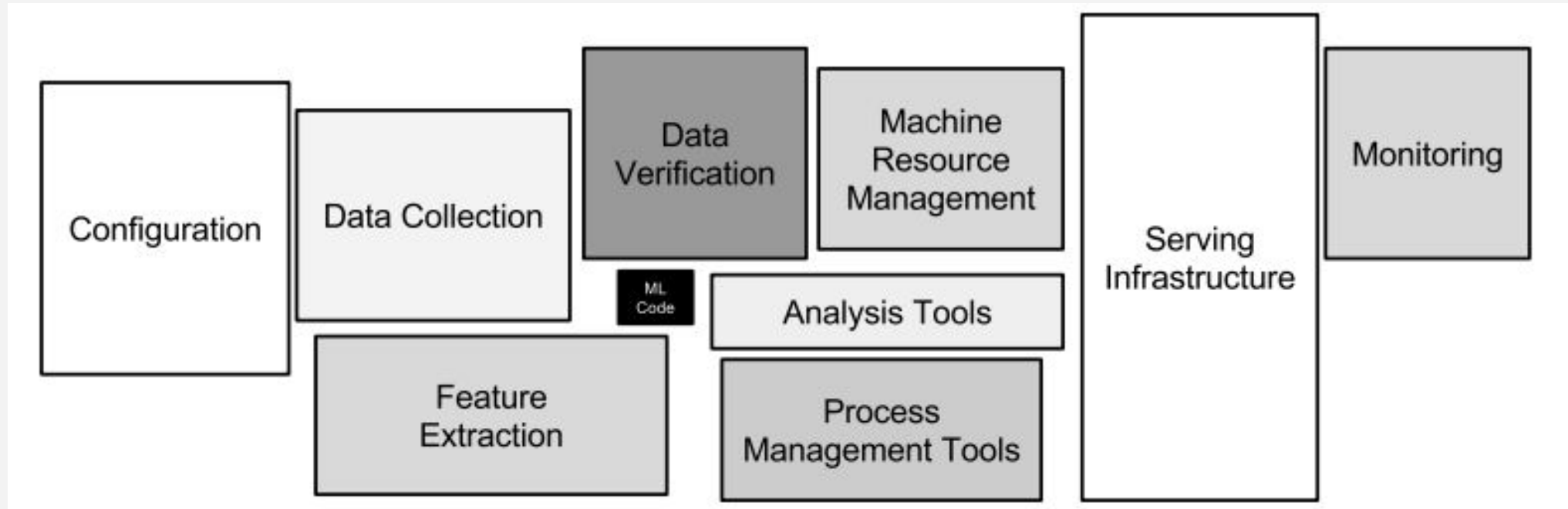
What do we need to do to create a useful ML model?



How can we deliver a useful model continuously?



What do we need to do to get the model to the end users?



How can Machine Learning software integrate with the rest?

Software Development

- Robust and long-existing quality practices
- Predictable work output
- Agile workflow across multiple disciplines
- Toolchains
- Deterministic output

Data Science

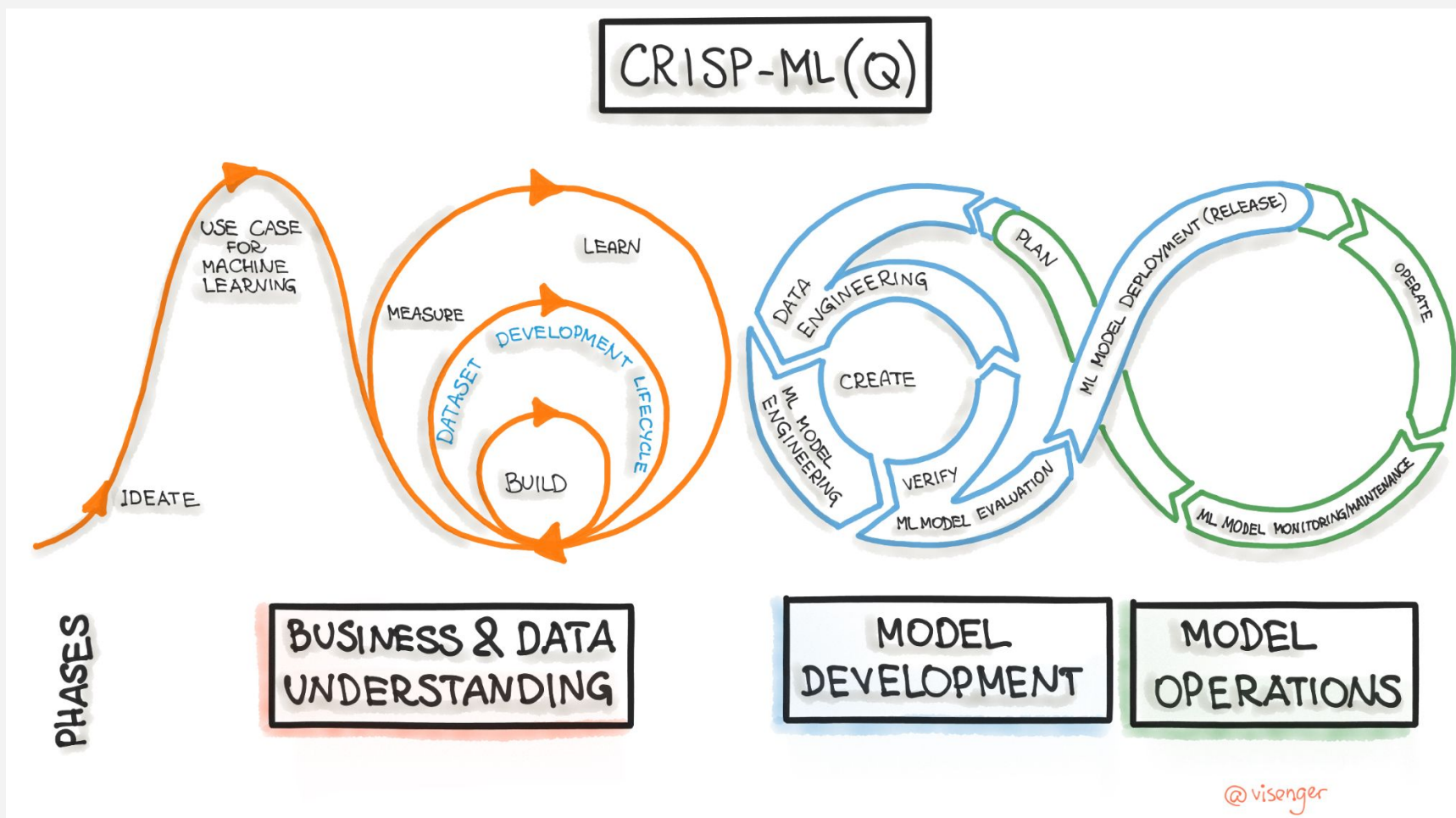
- Experiment driven
- Multiple experiments at the same time
- No guarantee of success
- Only recently got out the ivory tower
- Evolving and often overlapping toolchains
- Evolving quality practices
- Probabilistic output

**MLOps tries to bridge the gap between
experimentation oriented Data Science workflows
and the robust practices of Software Development
& Continuous Deployment**

MLOps Principles

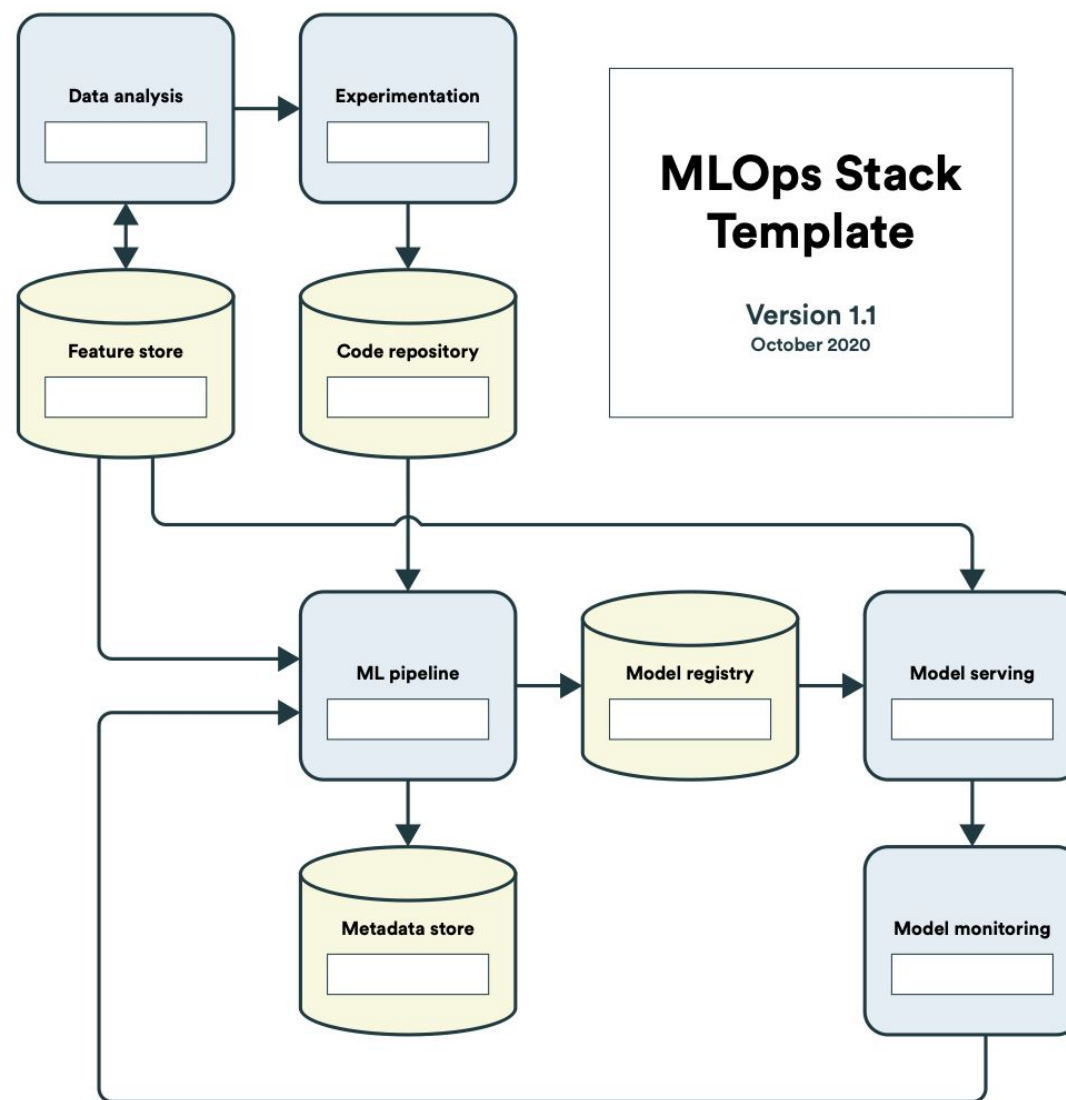
- Automation - source control, feature store, model registry, test&build&deploy services
- Continuous Everything (Integration/Delivery/Training/Monitoring)
- Versioning - datasets/scripts/models for rollback/reproducibility/governance
- Experiment Tracking - running multiple experiments requires tracking of all relevant metadata
- Testing - data/features/model development/ML infrastructure
- Monitoring - dependency changes in data schema/input & prediction distributions/staleness
- Reproducibility - data collection/feature engineering/model training/deployment
- Modularity - generally weaker boundaries between components so good structure is important

MLOps - How do we do it?



OK, how about the tools?

- Currently there is an explosion of tools and stack to support MLOps
- All the big cloud players have their own (opinionated) tech stack to do MLOps
- There is a lot of open-source tools, of varying quality.
- There are a lot of small and big players that take care of parts or almost end-to-end MLOps

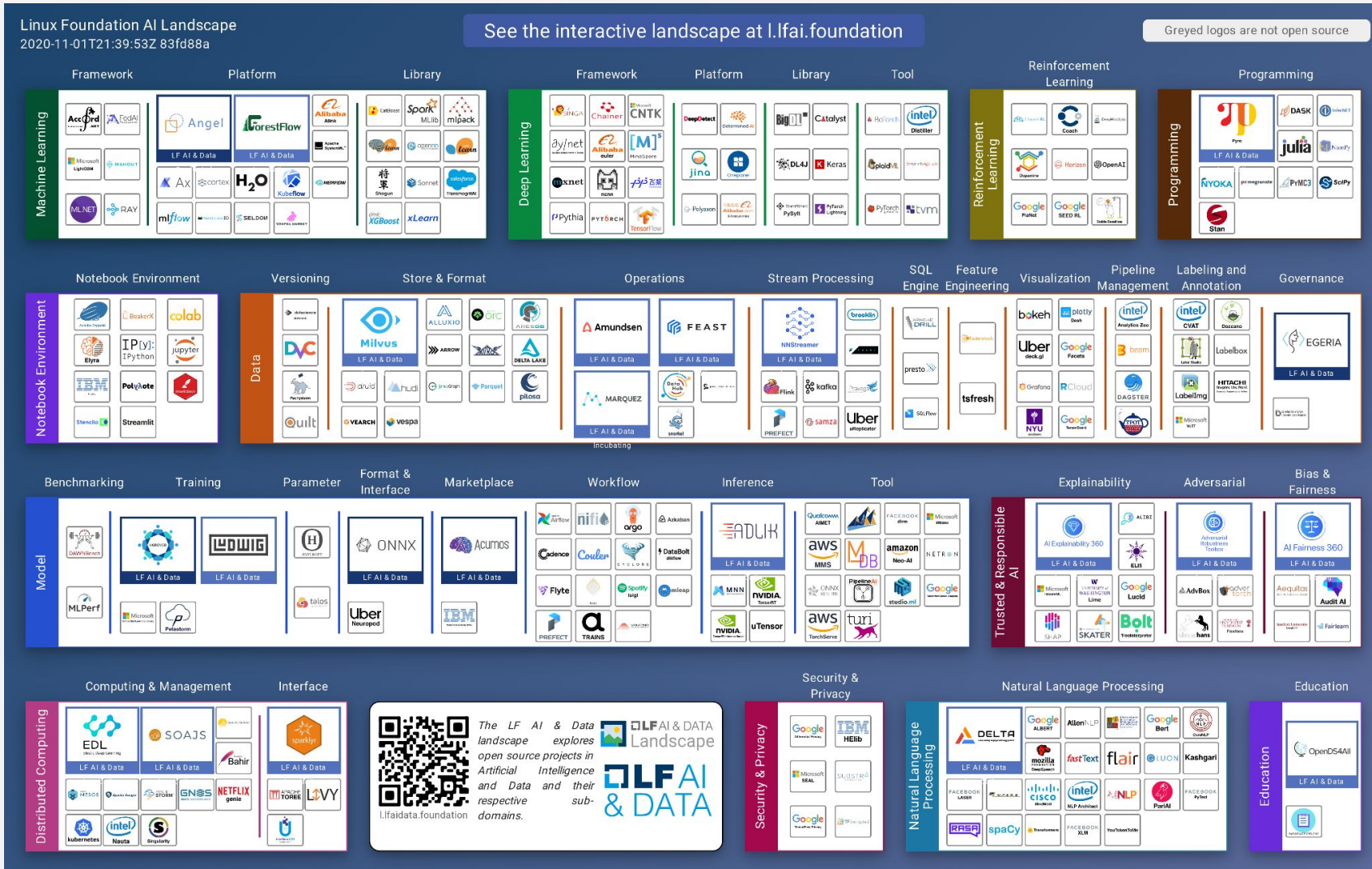


MLOps Stack Template

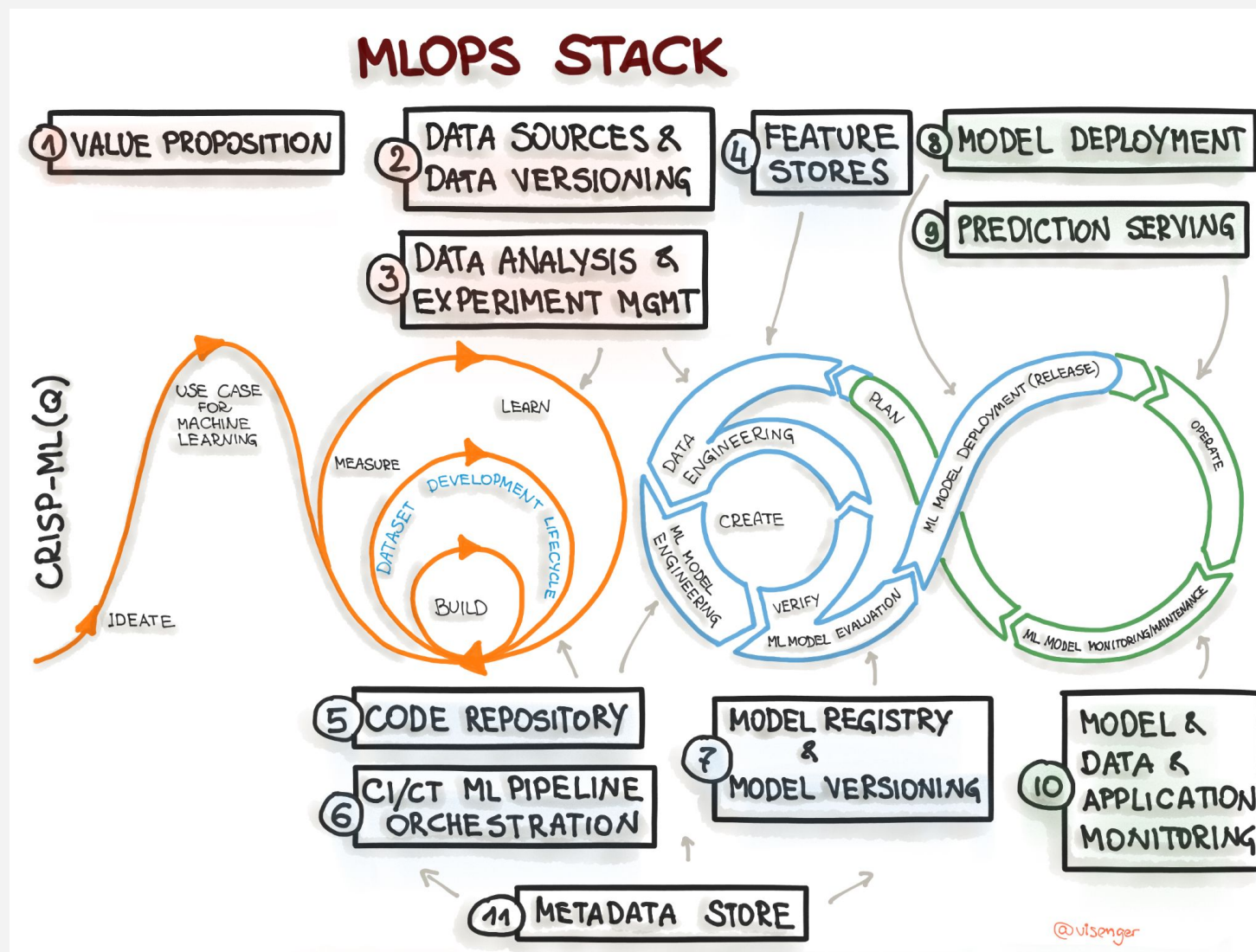
Version 1.1
October 2020

Component	Requirements	Tooling
Data analysis		
Experimentation		
Feature store		

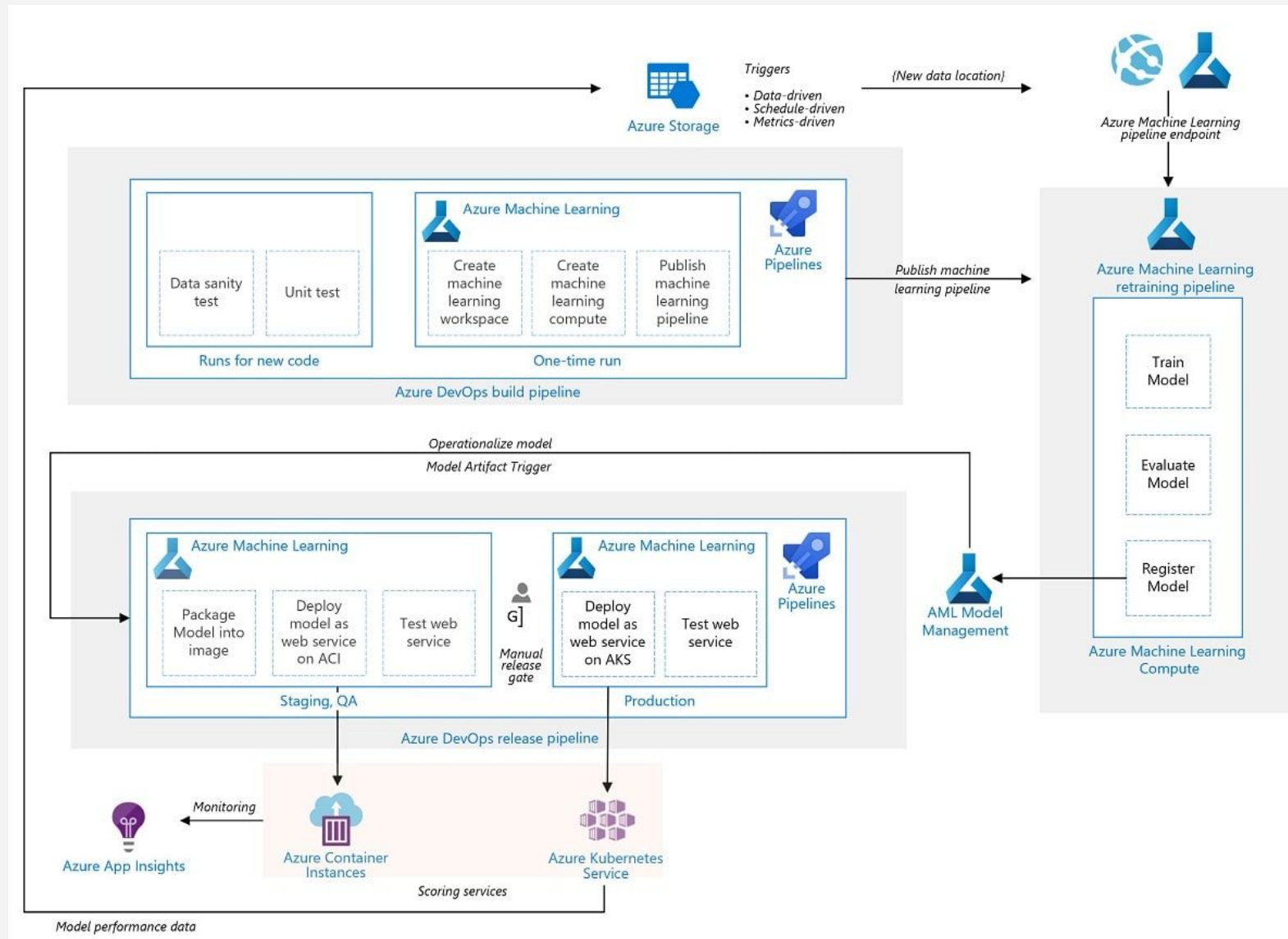
Are there really that many options? Oh, yes ..



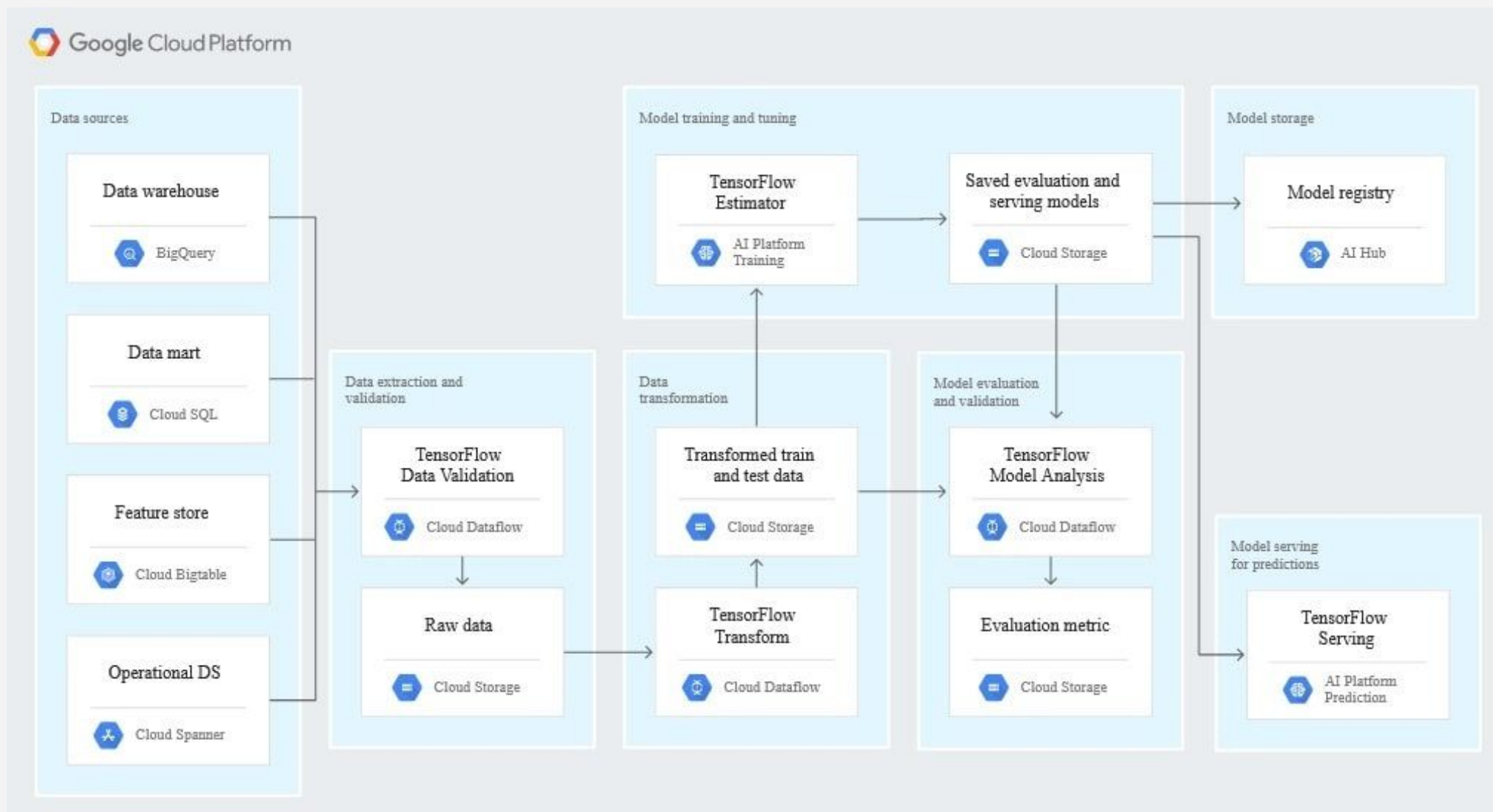
How do the tools support the process?



Example MLOps architecture on Azure



Example MLOps architecture on GCP



Wait, why should I personally care about MLOps at all?

- As a Data Scientist:
 - you want your work to scale and continue to deliver value. A model starts losing its value as soon as you finish training it.
 - your work process (and you) needs to integrate with the rest of the software development happening in the organization. No more ivory towers.
- As a Software Engineer:
 - There is a huge unmet need to bring good software practices (DevOps) to the organization that builds ML solutions. Full-stack JS is cool, but delivering models at a scale is cooler.
- For all: there is money, fame and careers to be made, if you understand MLOps :)

futurice

Demo time!

References

- Hidden Technical Debt in Machine Learning - <https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf>
- MLOps for Azure - <https://github.com/Microsoft/MLOpsPython>
- AzureML Pipeline, AutoML, AKS Deployment for Sentence Similarity - https://github.com/microsoft/nlp-recipes/blob/master/examples/sentence_similarity/automl_with_pipelines_deployment_aks.ipynb
- CRISP-ML(Q) The ML Lifecycle Process - <https://ml-ops.org/content/crisp-ml>
- MLOps Principles - <https://ml-ops.org/content/mlops-principles>
- Example MLOps architectures on AWS, Azure, GCP - <https://www.thecloudnavigator.com/cloud-based-mlops-architecture-aws-google-azure>
-

What is Futurice?

- Probably the best Nordic innovation consultancy, that actually builds cool stuff.
- We deliver solutions for our customers that blend design, technology and data.
- And we are hiring! Finland, Germany, Sweden - including fully remote positions!
- Check it out -> <https://futurice.com/careers/open-positions>

Thank you!
Kiitos!
Danke!
Tack!



Vilen Looga
Senior Data Scientist

vilen.looga@futurice.com

Or find me on LinkedIn!

futurice

Co-creating a resilient future