

Machine Learning Operations

02457 Machine Learning Operations

Nicki Skafte Detlefsen,

Postdoc

DTU Compute

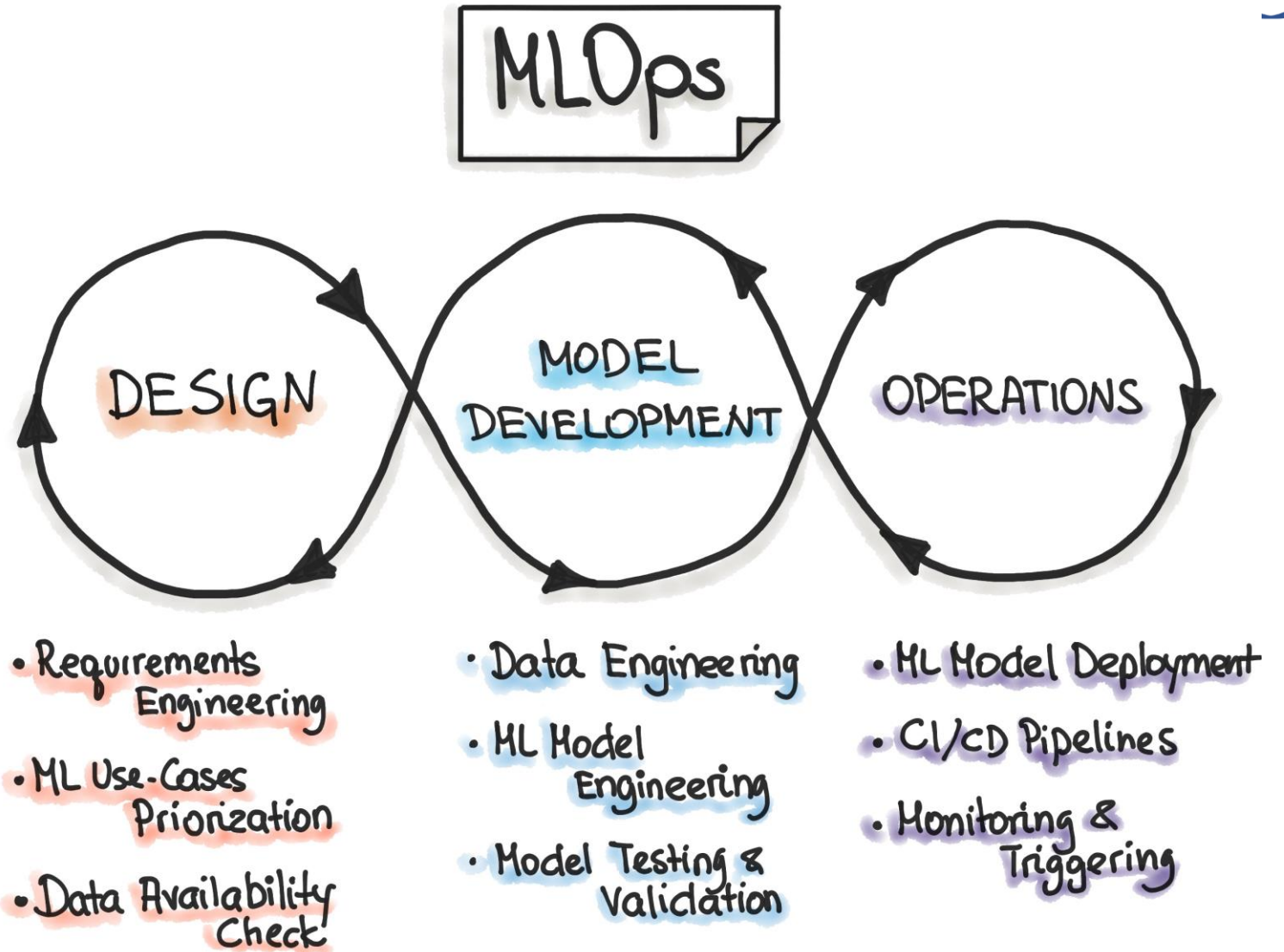
Loosely based on <https://www.youtube.com/watch?v=VU5Em1qkWDU>

What is machine learning operations



*Is a set of tools,
processes, and mindset
that aim to make ML
Lifecycle **reproducible**,
trackable, **testable** and
maintainable*

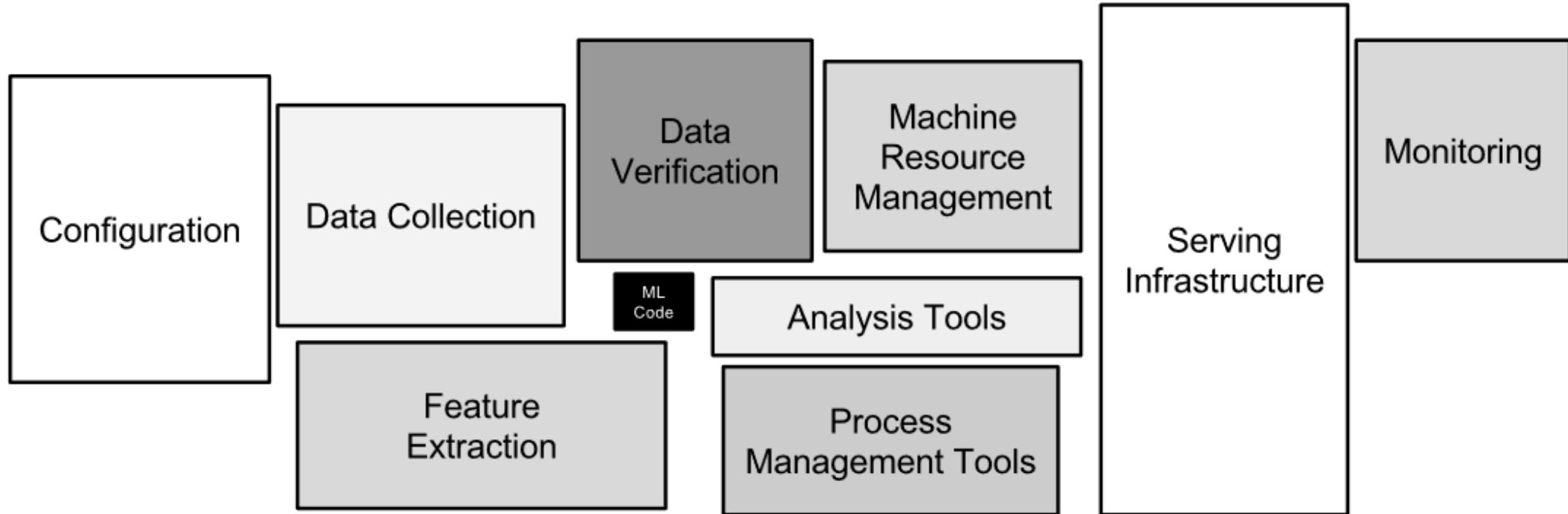
Notice: ITS A CYCLE!



Why should you care?



Teeny tiny part is actual ML code, the rest is operations

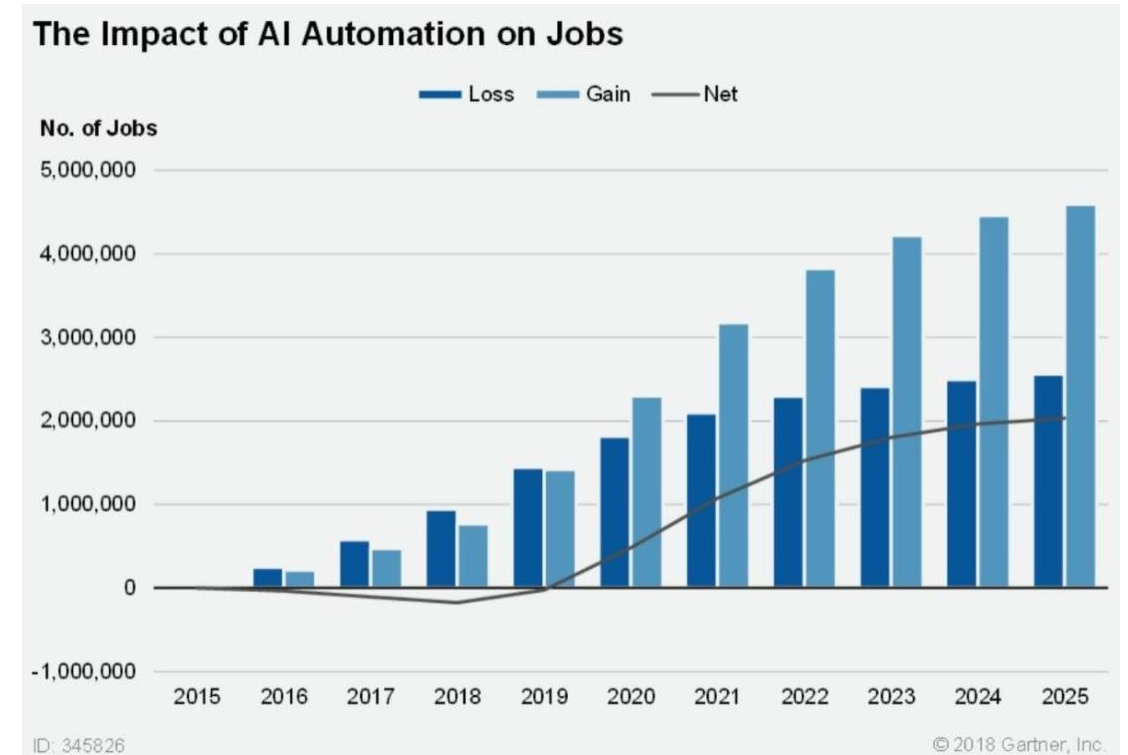


D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. 2015. **Hidden technical debt in Machine learning systems**. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15)*. MIT Press, Cambridge, MA, USA, 2503–2511.

Why does companies care



- ML automatization is going to increase over the years
- Examples:
 - Which stocks to buy or sell?
 - Where is the tumor in the picture
 - What should be the price of a banana today?

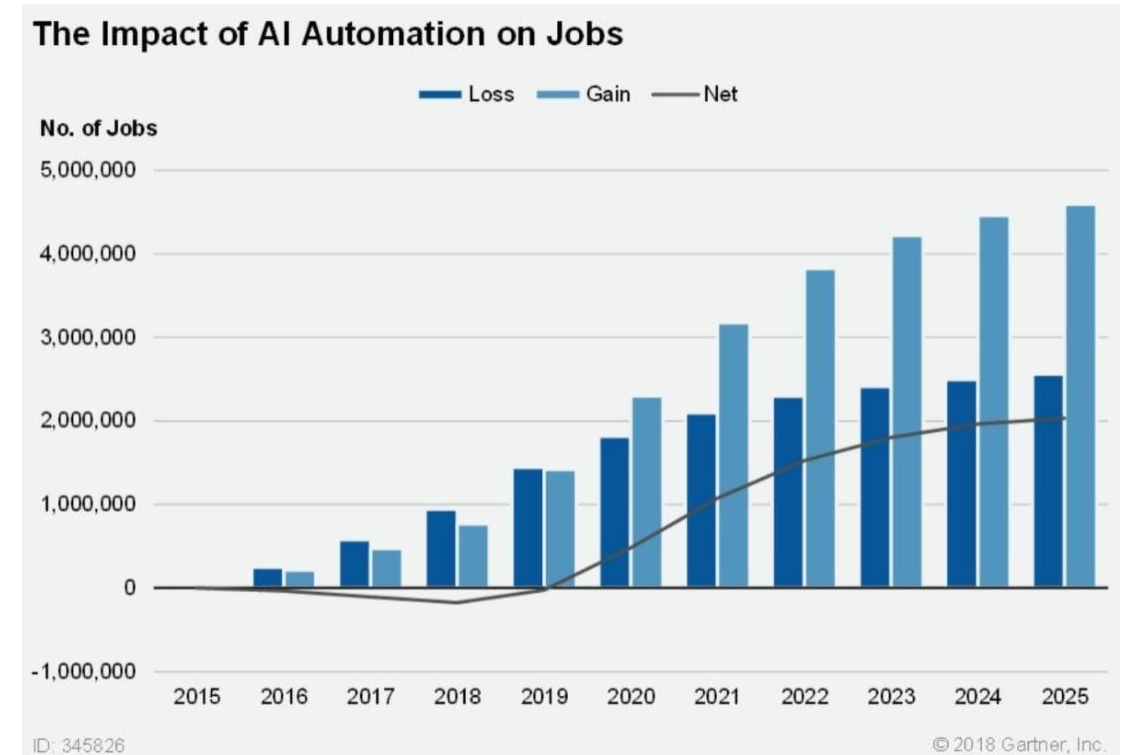


Why does companies care



- Having automated model deployed with errors can cost ALOT of money:

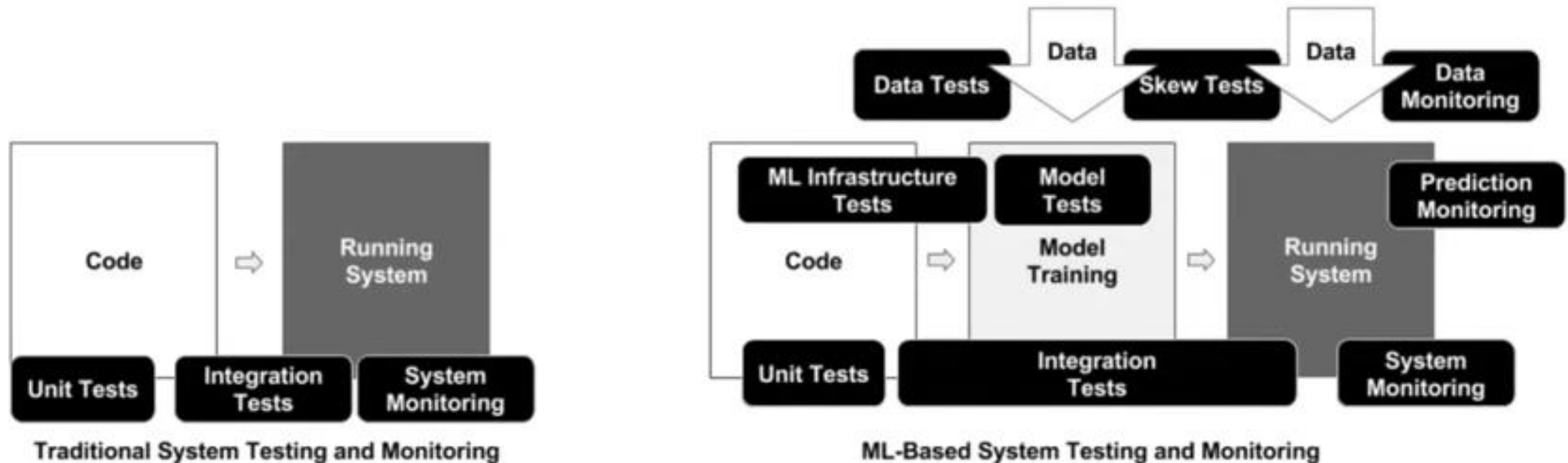
"A famous example of the dangers here was Knight Capital's system losing \$465 millions in 45 minutes, apparently because of unexpected behavior from obsolete experimental codepaths" – Hidden Technical depth in Machine Learning Systems



Why is MLOps harder than DevOps



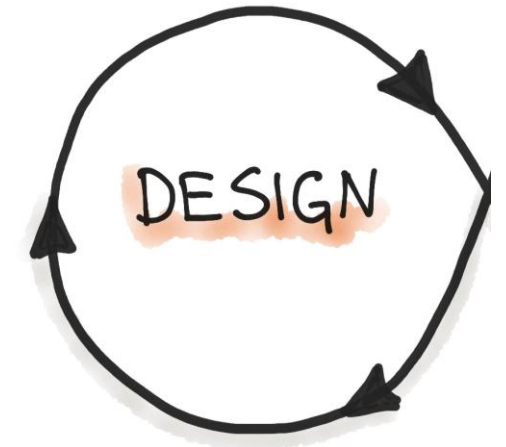
- It involves a freaking lot of testing



Design



- This is the main part we train you at DTU
 - Analyze a problem
 - Look in literature for references
 - Check if you have access to data for investigating this

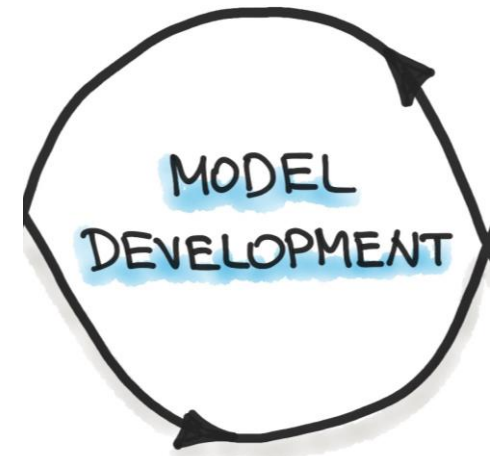


- Requirements Engineering
- ML Use-Cases Priorization
- Data Availability Check

Development



- This is somewhat covered in other courses
 - Going from ideas to practical implementation
 - How should data be formatted to guide the development
 - How should model be validated and tested
- This course will introduce tools to be more organised in this phase



- Data Engineering
- ML Model Engineering
- Model Testing & Validation

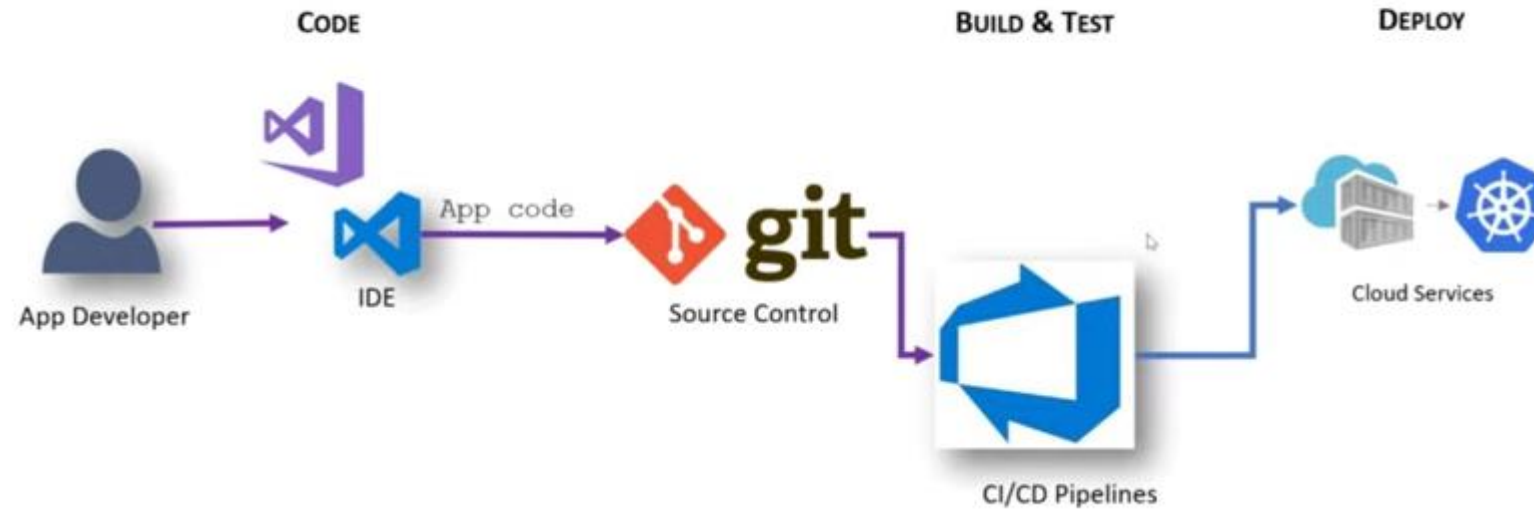
Operations (The new kid)



- To my knowledge, is not taught at DTU
- Operations = How to make sure models do not break
 - My hope is that you will get a feeling of this topic
 - Specifically we will touch upon deployment and CI



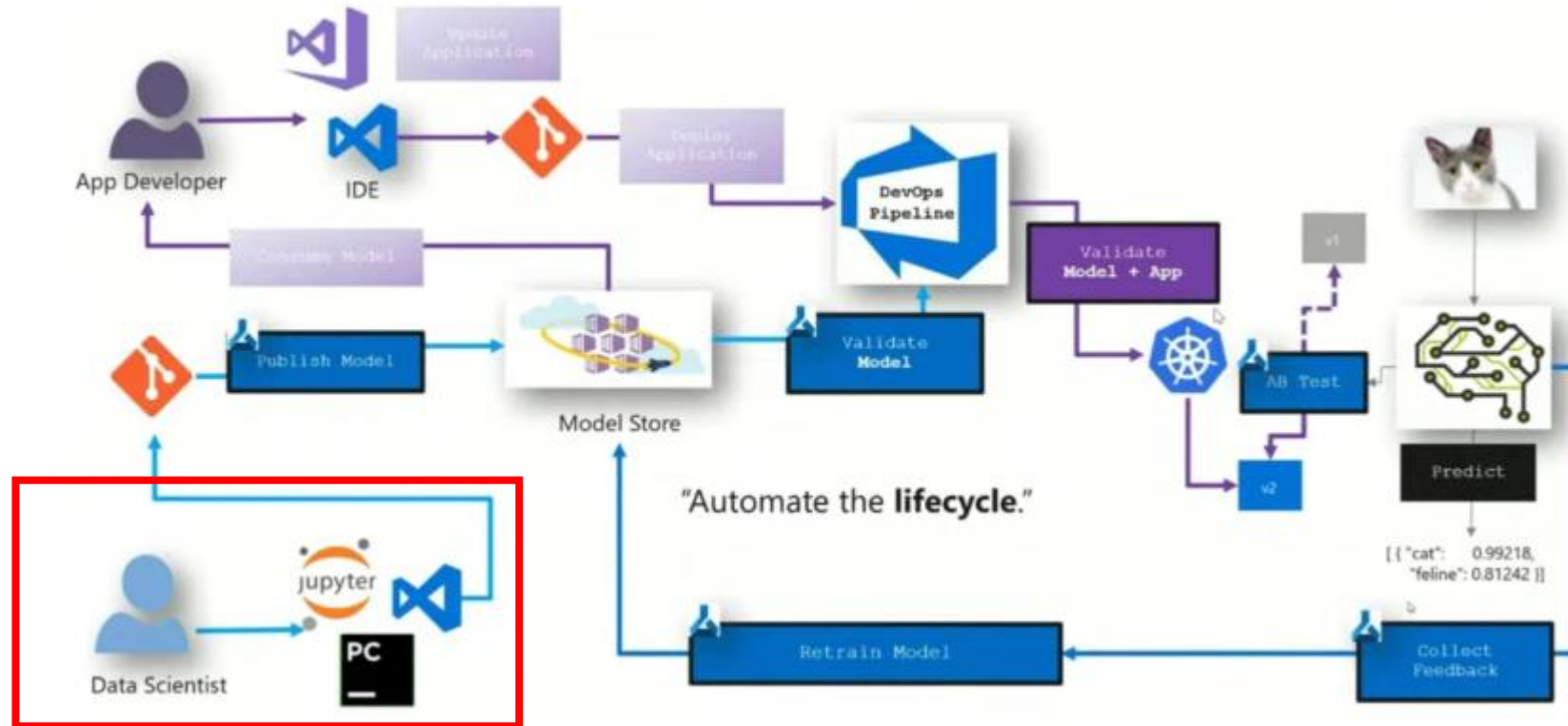
The workflow of standard DevOps



The workflow of MLOps



DevOps on steroids



The big difference is MLOps requires domain knowledge

MLOps at a high level



1. Optimizing workflows
 - Getting organized cost time initially but will save you time down the line
2. Versioning
 - Keep track of code changes, trained models etc. so everything can be backtracked
3. Automatization and Continuous X
 - Make sure that new changes automatically gets tested, deployed etc.
4. Reusability
 - Why rewrite the same code for a new project if you can reuse
5. Reproducibility
 - Make sure that your results can be redon by others

The first step of MLOps: Getting organised



Today's exercises are all about organising your workflow.

Note that

- While organization is maybe not that big of a deal on personal projects, it is an essential factor when working on large scale projects

Meme of the day

