

# Projects in Artificial Intelligence: Software 2.0

Nadia Ahmed



**Project Scope & Feasibility**



**Team**



**Planning**



**Data**



**Tooling and Infrastructure**



**Deployment**



# Project Scope & Feasibility

**"87% of data science projects never make it into production"**

**- Venture Beat**



# Project Scope & Feasibility

Why?

- Technically infeasible
- **Lack of Collaboration**
- Unclear Success Criteria
- “The Value of the Project Does Not Outweigh its Technical Complexity” (Tobin, 2022)



# Project Scope & Feasibility

Machine Learning: The High-Interest Credit Card of Technical Debt (Sculley, et al)

- Complex Models Erodes boundaries
- Data Dependencies Cost More than Code Dependencies
- Feedback Loops
- Anti-patterns
- Configurational Debt
- Subject to instability



# Project Scope & Feasibility

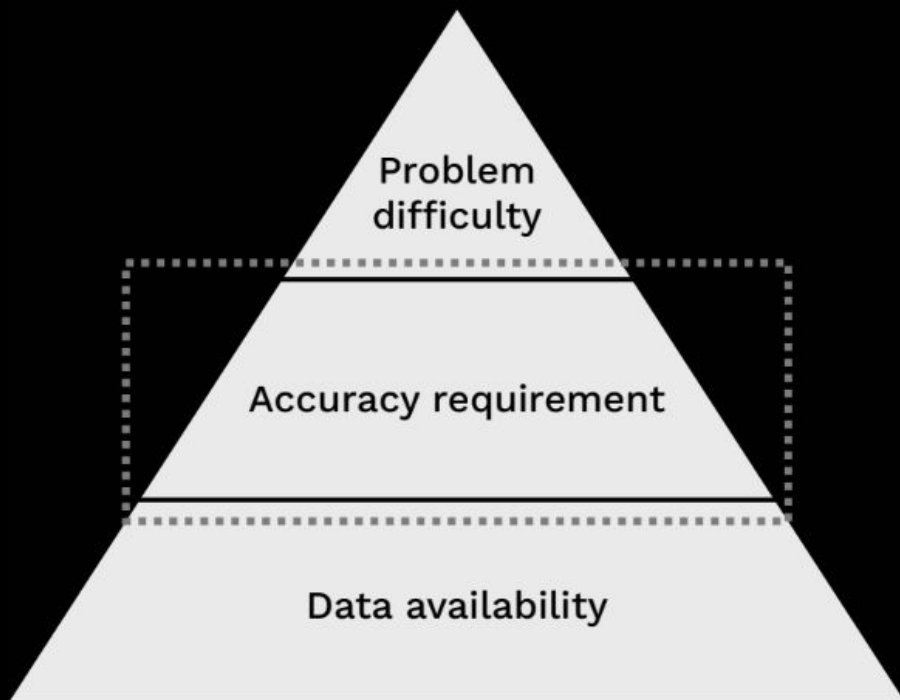
Can we identify **high** impact, **low-cost** problems that can benefit enough to outweigh the **complexity cost** that ML brings to development?



# Project Scope & Feasibility

## Cost drivers

---



## Main considerations

---

- Is the problem well-defined?
- Good published work on similar problems?  
(newer problems mean more risk & more technical effort)
- Compute requirements?
- Can a human do it?
- How costly are wrong predictions?
- How frequently does the system need to be right to be useful?
- Ethical implications?
- How hard is it to acquire data?
- How expensive is data labeling?
- How much data will be needed?
- How stable is the data?
- Data security requirements?



# Project Scope & Feasibility

## The Rise of Software 2.0: Data Driven Development

### 2.0 Developers (Data)

- Curation
- Updates & Growth
- Transforming
- Cleaning

### 1.0 Developers (Traditional)

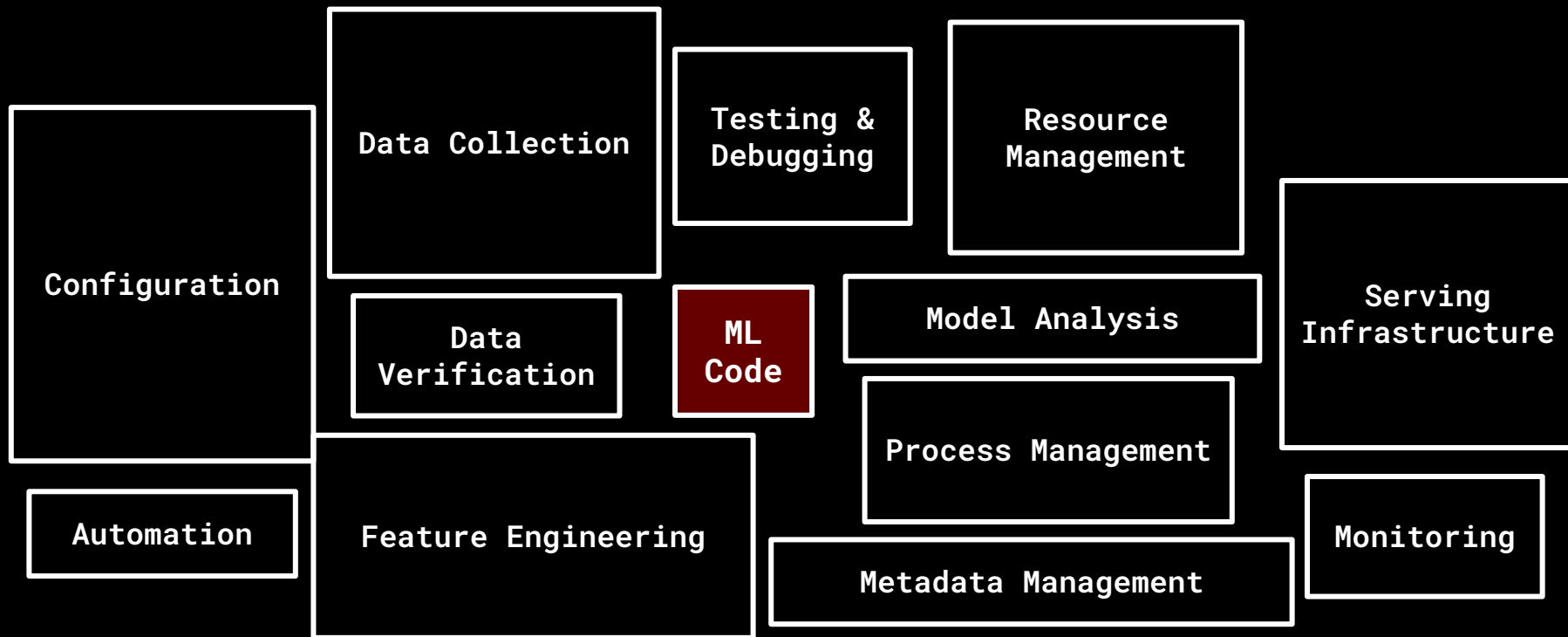
- Training Infrastructure  
Development
- Analytics
- Visualization
- Labeling

(Karpathy, 2017)





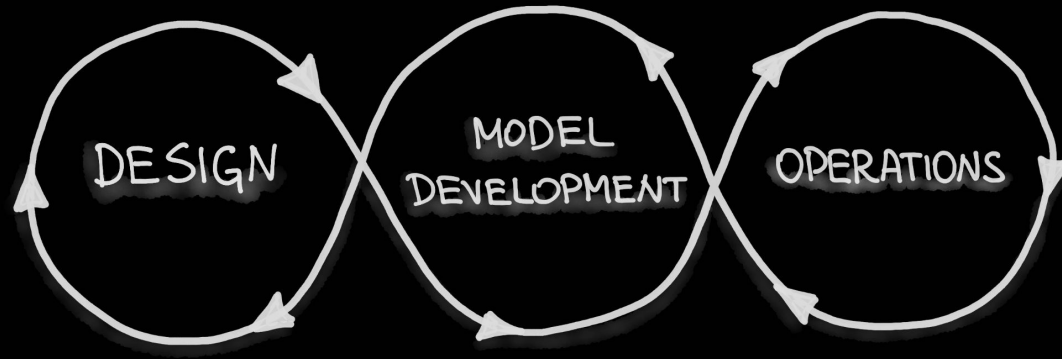
# Project Scope & Feasibility





# Project Scope & Feasibility

MLOps



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

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

- ML Model Deployment
- CI/CD Pipelines
- Monitoring & Triggering

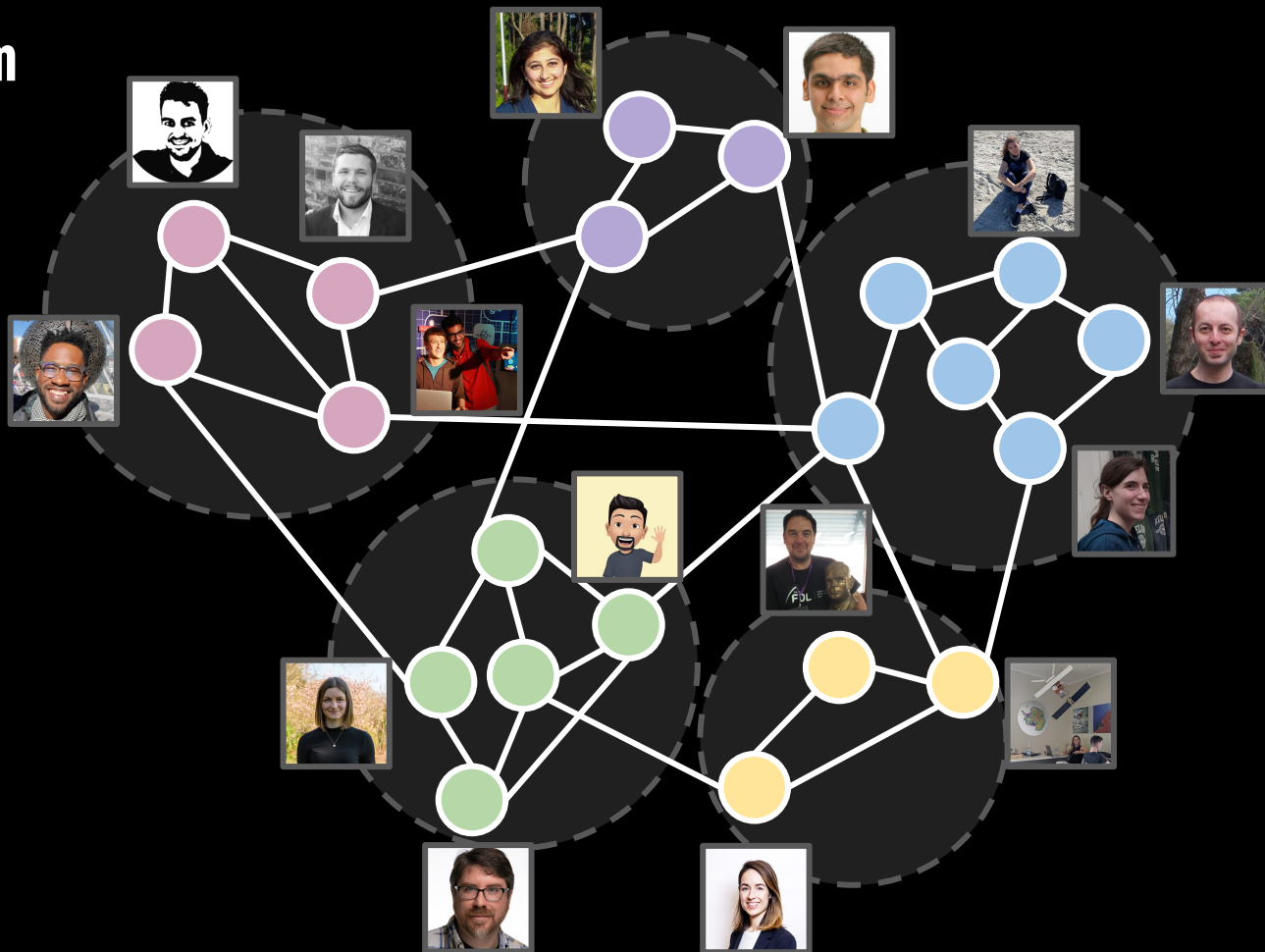


# Team



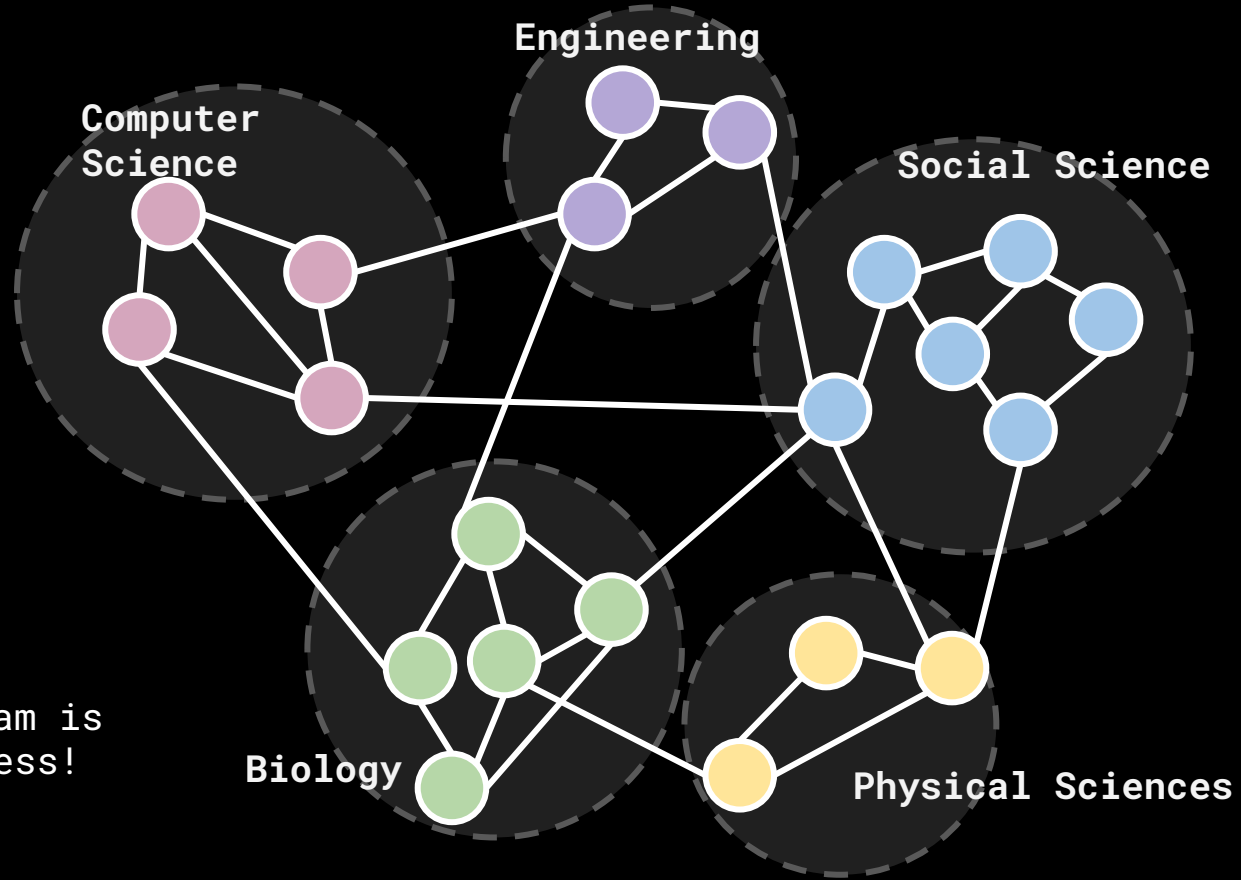


# Team





# Team



**Diversity** of Team is  
the key to success!



# Team

- Diversity of Background
- Openness to learning
- Strong Subject Matter Expertise
- Strong Computer Science Skills
- Collaboration



# Planning

From Prototype to Production

- **Problem Refinement**
- Collect data
- Clean and preprocess data
- Explore the data (EDA)
- Feature and model selection
- Train model
- Evaluate results
- Deploy



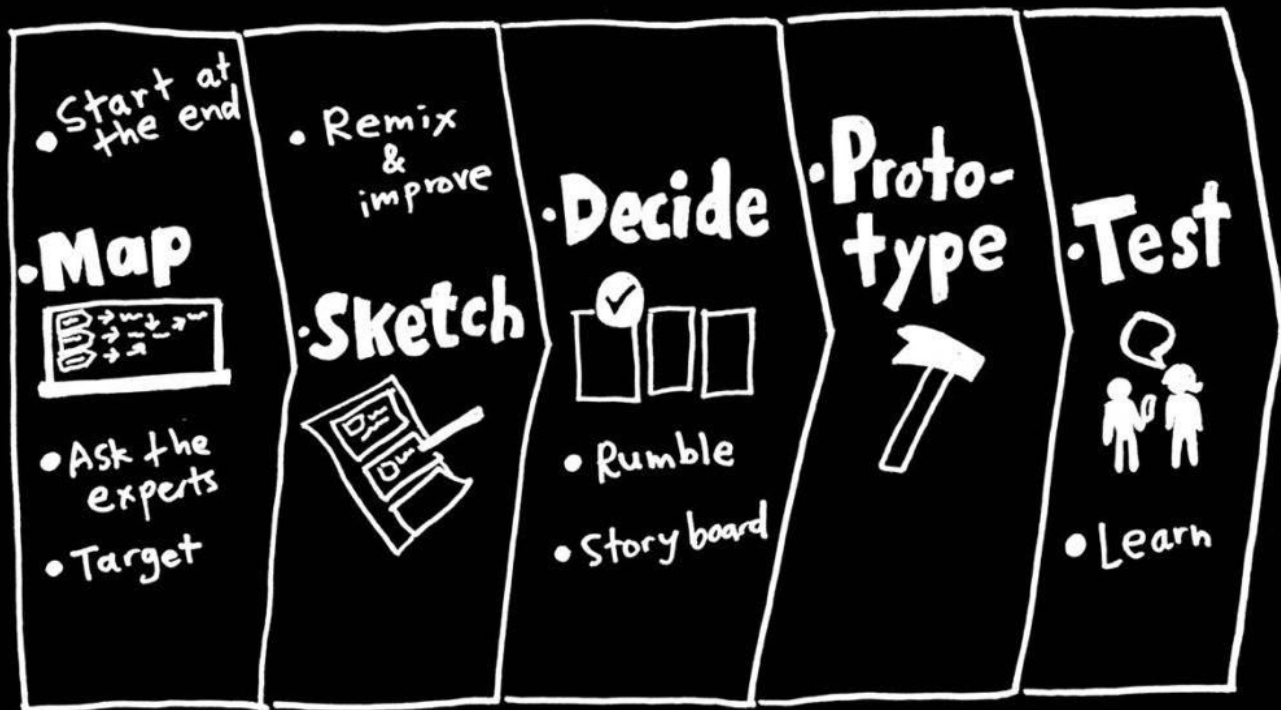
# Planning

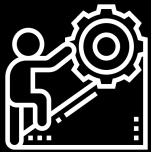
How might we use **Agile techniques** to plan our project?





# Planning





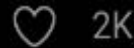
# Data



**François Chollet** ✓ @fchollet · Jan 24



ML researchers work with fixed benchmark datasets, and spend all of their time searching over the knobs they do control: architecture & optimization. In applied ML, you're likely to spend most of your time on data collection and annotation – where your investment will pay off.

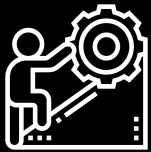


## Machine Learning Needs Data.



# Data

- **What prediction problem would you like to solve?**
  - Classification
  - Regression
- **What kind of data is needed to solve the problem?**
  - Time series data
  - Image data
  - Text data
  - Sound
  - Tabular



# Data

## Open Source Datasets

- [Kaggle](#)
- [UCI Machine Learning Repository](#)
- [US Government Data](#)
- [Open Data on AWS](#)
- [Google Cloud Public Datasets](#)
- Or create your own...



# Tooling & Infrastructure

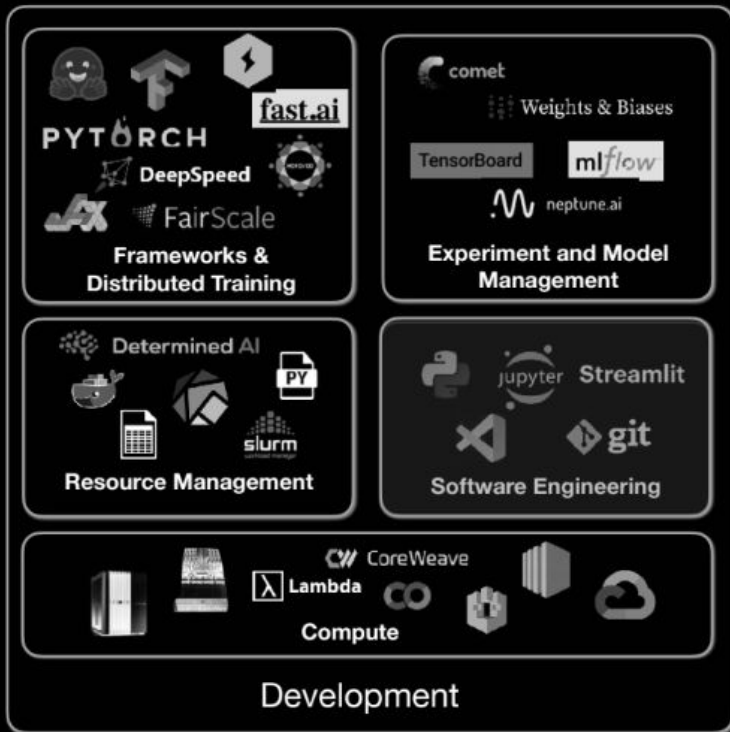
“All-in-one”



Amazon SageMaker



gradient<sup>o</sup>  
by Paperspace



(Karayev, et al 2019)



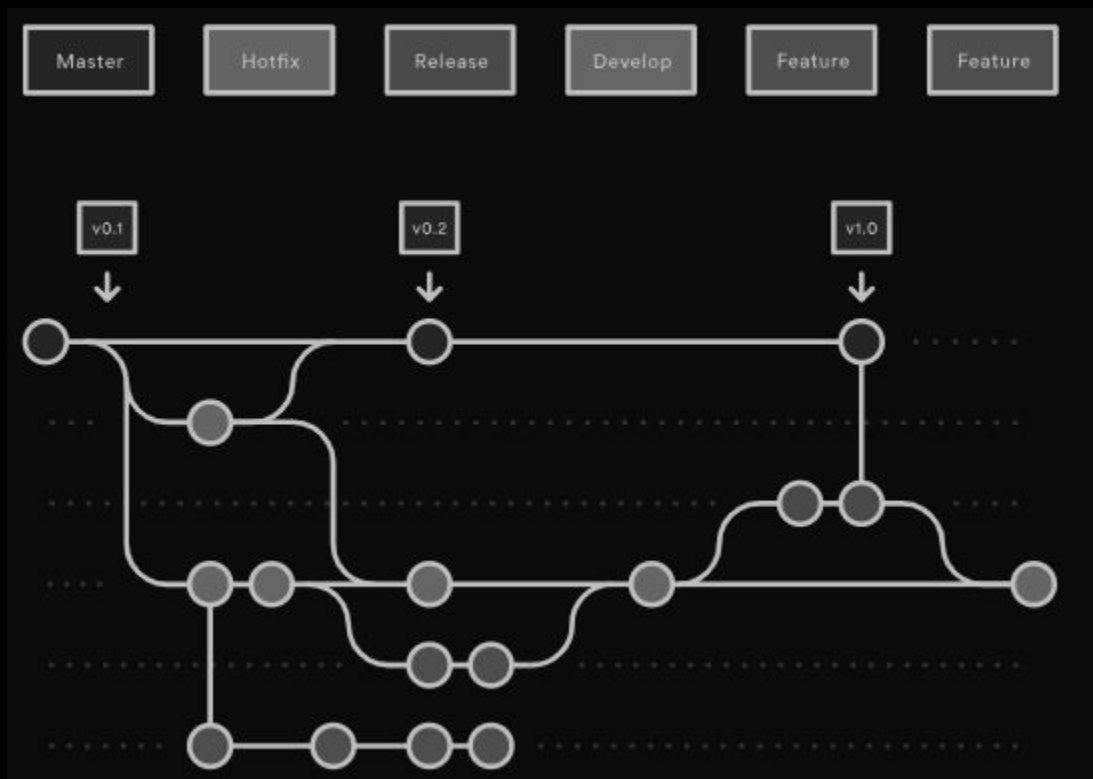
# Tooling & Infrastructure

## Continuous Integration and Development

- Data versioning
  - landing (raw)
  - staging (ML ready)
- Model versioning
- Results/Feature Deployment

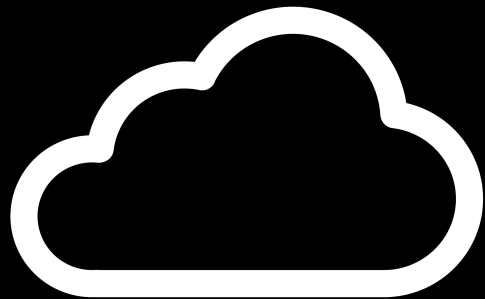


# Tooling & Infrastructure





# Deployment



Cloud



Payload











Mobile





Selected				2020-11-10 AOI01 CUSTOM-2021-03-
Feature:				12T13:53:21.991856_2020-11-10

Select event date(s) 01/01/2021 → 03/12/2021

- |   |            |                                |
|---|------------|--------------------------------|
|  | 2021-01-06 | AI01 Flood in Albania          |
|  | 2021-02-23 | AI01 Flood in Southern Ireland |
|  | 2021-02-23 | AI02 Flood in Southern Ireland |
|  | 2021-02-23 | AI03 Flood in Southern Ireland |
|  | 2021-02-23 | AI04 Flood in Southern Ireland |
|  | 2021-02-23 | AI05 Flood in Southern Ireland |
|  | 2021-02-23 | AI06 Flood in Southern Ireland |
|  | 2021-02-23 | AI07 Flood in Southern Ireland |

Redraw Specify event date: 11/10/2020 Add

# Thank You!

