

# Requirements Gathering and Architecture

Robert Clements

MSDS Program

University of San Francisco



# ML Systems Gone Wrong

## Racial bias found in widely used health care algorithm



An estimated 200 million people are affected each year by similar tools that are used in hospital networks

Nov. 6, 2019, 11:38 AM PST / Updated Nov. 7, 2019, 8:07 AM

By Quinn Gawronski

## Why everyone is mad about New York's AI hiring law

The law is a first step in regulating AI, but critics aren't happy

By Tate Ryan-Mosley

MIT  
Technology  
Review

July 10, 2023

KHARI JOHNSON

BUSINESS MAR 7, 2022 7:00 AM

WIRED

## How Wrongful Arrests Based on AI Derailed 3 Men's Lives

Robert Williams, Michael Oliver, and Nijeer Parks were misidentified by facial recognition software. The impact cast a long shadow.

# Cruise says it's not at fault for woman injured in San Francisco crash

By Zak Sos , Allie Rasmus and Andre Torrez | **Updated** October 3, 2023 8:10am PDT | San Francisco | KTVU FOX 2 | [↗](#)

BUSINESS

## California allows robo-taxis to expand and emergency responders aren't happy

AUGUST 10, 2023 · 4:44 PM ET



Dara Kerr

NEWS | BAY AREA & STATE

# Waymo driverless car set on fire, destroyed by San Francisco crowd

By **Katie Dowd**

Feb 11, 2024



# Who do you get requirements from?

- *All stakeholders:*

- Customers
  - Project owners/leadership
  - Operations teams
  - Regulators
  - Affected
- Project owners will change their minds often, and will be vague
  - Don't forget safety, responsibility, ethics (hopefully some of this comes from regulators)

# What requirements should we collect?

- Goals of the project
  - Measurable metrics for each goal
- Data sources
- Potential risks: this list is huge
  - Risk mitigation strategies
- Outline of architectural components
  - This may be flexible, depending on what's available to you

# What if we were asked to build a Smart Transportation System?

- The SFMTA wants to implement a smart transportation management system to optimize traffic flow, reduce congestion, and lower emissions. Using data from traffic cameras, vehicle sensors, public transit GPS, mobile apps, weather services, and scheduled events, the system would predict traffic patterns and adjust traffic light timings, suggest route alternatives, and manage MUNI schedules. The system should be able to adjust to unexpected events, such as collisions, and ensure equitable service across neighborhoods. It should be live 24/7 and demonstrate measurable improvements in commute times and air quality.



# What if we were asked to build a Smart Transportation System?

- Let's focus on just scheduling the lights on **one street** according to live traffic from traffic cameras and sensors. In Slack:

List ~~two~~<sup>3</sup> low-level goals? And what metrics can we use to measure our goals?



↓  
ML/AI  
Devices  
Infra  
Team

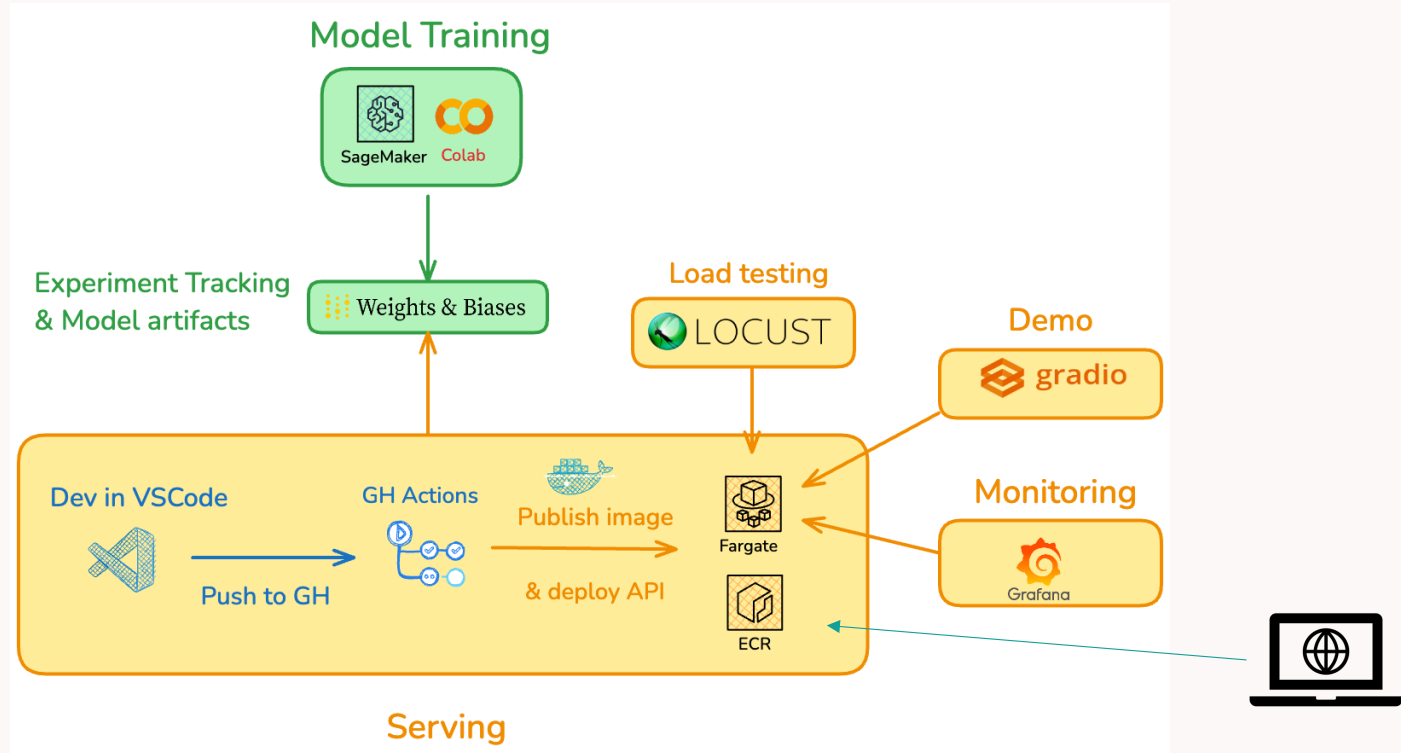


# ML System Architecture

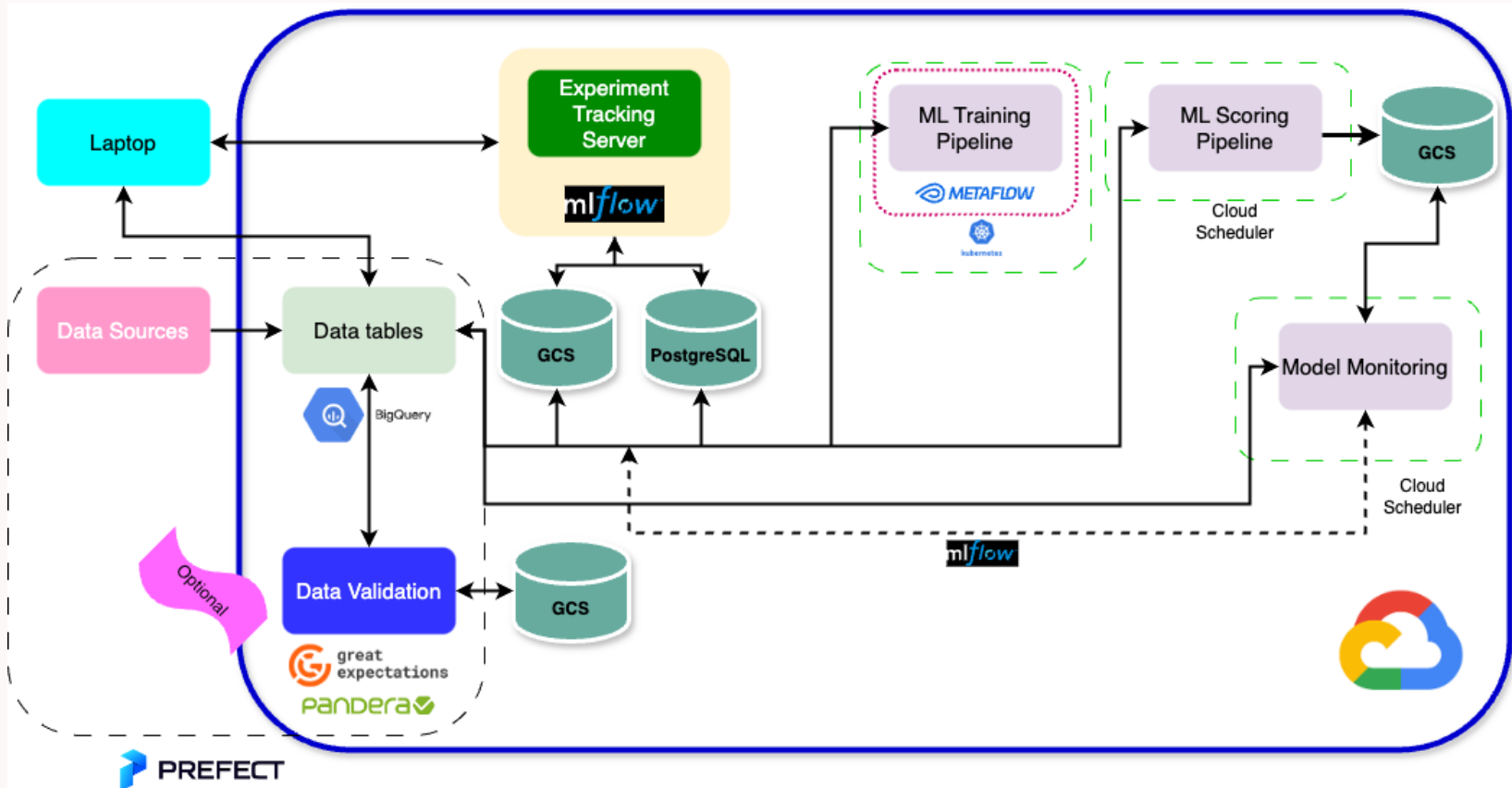


Let's first zoom in on the MLOps  
architecture

# What pieces is the system made of?



# What pieces is the system made of?



# Environments

## DEV

- Dev data
  - Exploration
  - Model dev
- Mirror of prod

- Dev cluster
- Github: use dev branch

SANDBOX  
ENV

Dev. models  
EDA  
Track exps

# Environments

## STAGE

- Stage data
- Mirror of prod

- Stg. cluster (compute)
- Github: merge w/stg branch
- Trigger "build"
- Trigger unit/integration/regression tests

# Environments

- Prod data
- Feature store (optional)

## PROD

- Prod cluster (compute)
- Github = merge w/main branch
- Deploy model
- Monitor model (inputs, outputs, performance, resources)

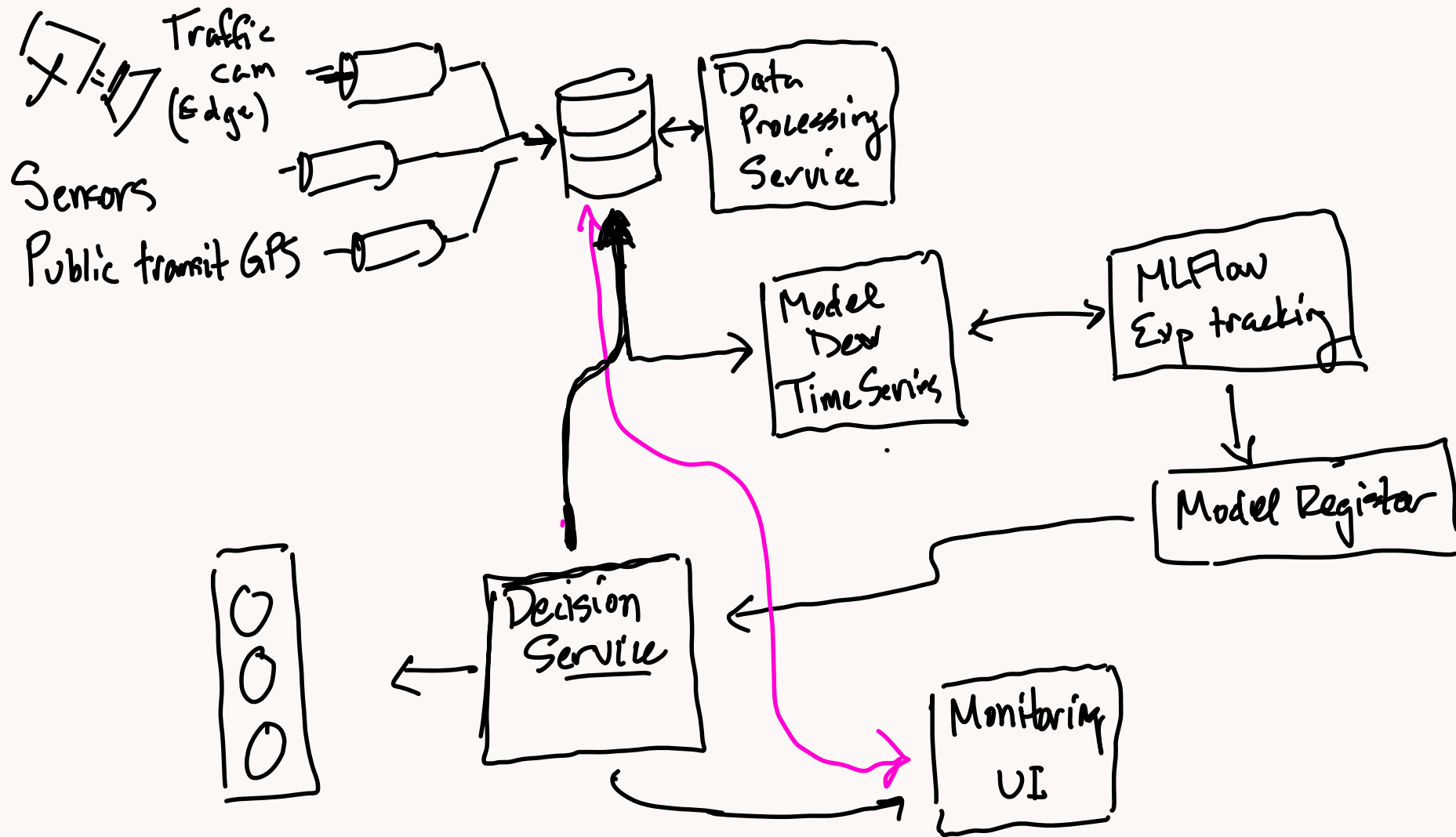
Now let's zoom out and  
consider the entire system

# What if we were asked to build a Smart Transportation System?

- Let's focus on just scheduling the lights on ***one street*** according to ***estimated*** (not live) traffic from traffic cameras and sensors.







# How about these other use cases?

This is next word prediction

More considerations:

- Cost
- Privacy
- Fairness
- Accuracy
- Model size
- Explainability
- Feedback loop

# How about these other use cases?



## More considerations:

- Cost
- Privacy
- Fairness
- Accuracy
- Model size
- Explainability
- Feedback loop

# How about these other use cases?

The screenshot shows a Google search interface. The search bar contains the text "How much energy is used by AI". Below the search bar, the navigation tabs include "All", "Images", "News", "Videos", "Short videos", "Shopping", "Web", and "More". The "All" tab is selected. The search results are displayed in a list format. The first result is titled "AI Overview" and includes a summary of AI energy consumption. A "Show more" button is located at the bottom of the first result. To the right of the main results, there is a "Learn more" link and a list of related search suggestions.

Google

How much energy is used by AI

× |

All Images News Videos Short videos Shopping Web : More Tools

◆ AI Overview [Learn more](#) ⋮

AI, particularly generative models, consume a significant amount of energy, with training a model like GPT-3 estimated to use **nearly 1,300 megawatt-hours (MWh) of electricity**, equivalent to the annual power consumption of 130 U.S. homes.

Here's a more detailed breakdown of AI's energy usage:

**Training Energy:**

- Training a large language model like GPT-3 is estimated to consume nearly 1,300 MWh of electricity.
- This is roughly equivalent to the annual power consumption of 130 U.S. homes.

[Show more](#) ▾

How much energy will AI really consume? The good, the bad and ...  
Mar 5, 2025 — On average, according to their latest results, generating an image from a text prompt consumes about 0.5...  
 Nature ⋮

How much electricity do AI generators consume? - The Verge  
Feb 16, 2024  
 The Verge ⋮