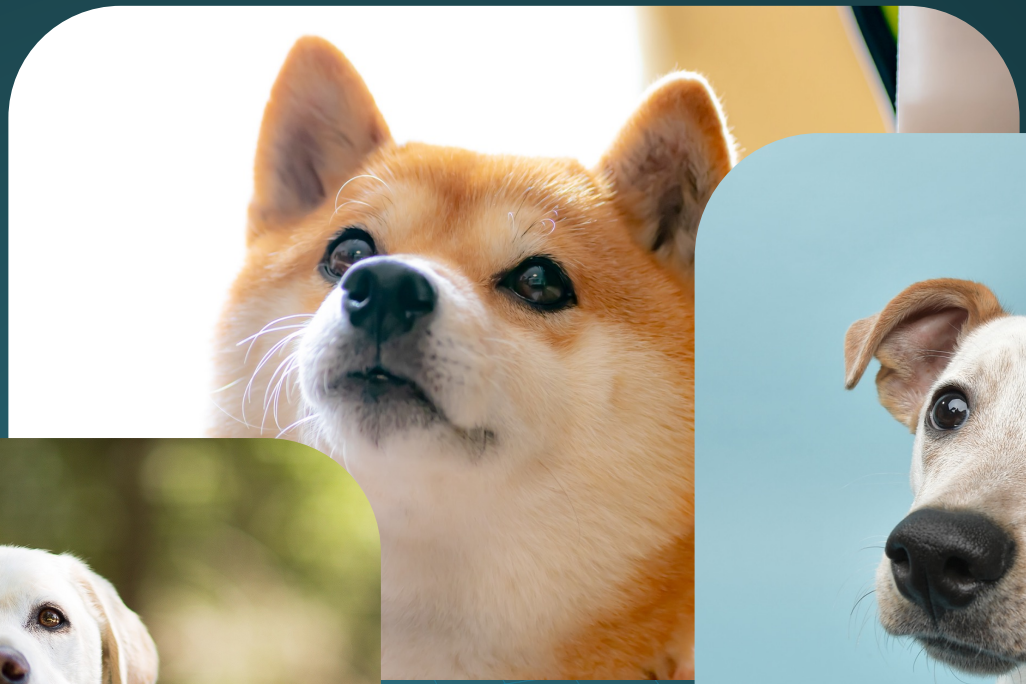
Solving MLOps from first principles

DataCamp, November ‘23, Dean Pleban








The 5 step process

Step 1: Define the problem

Step 2: Outline problem parameters

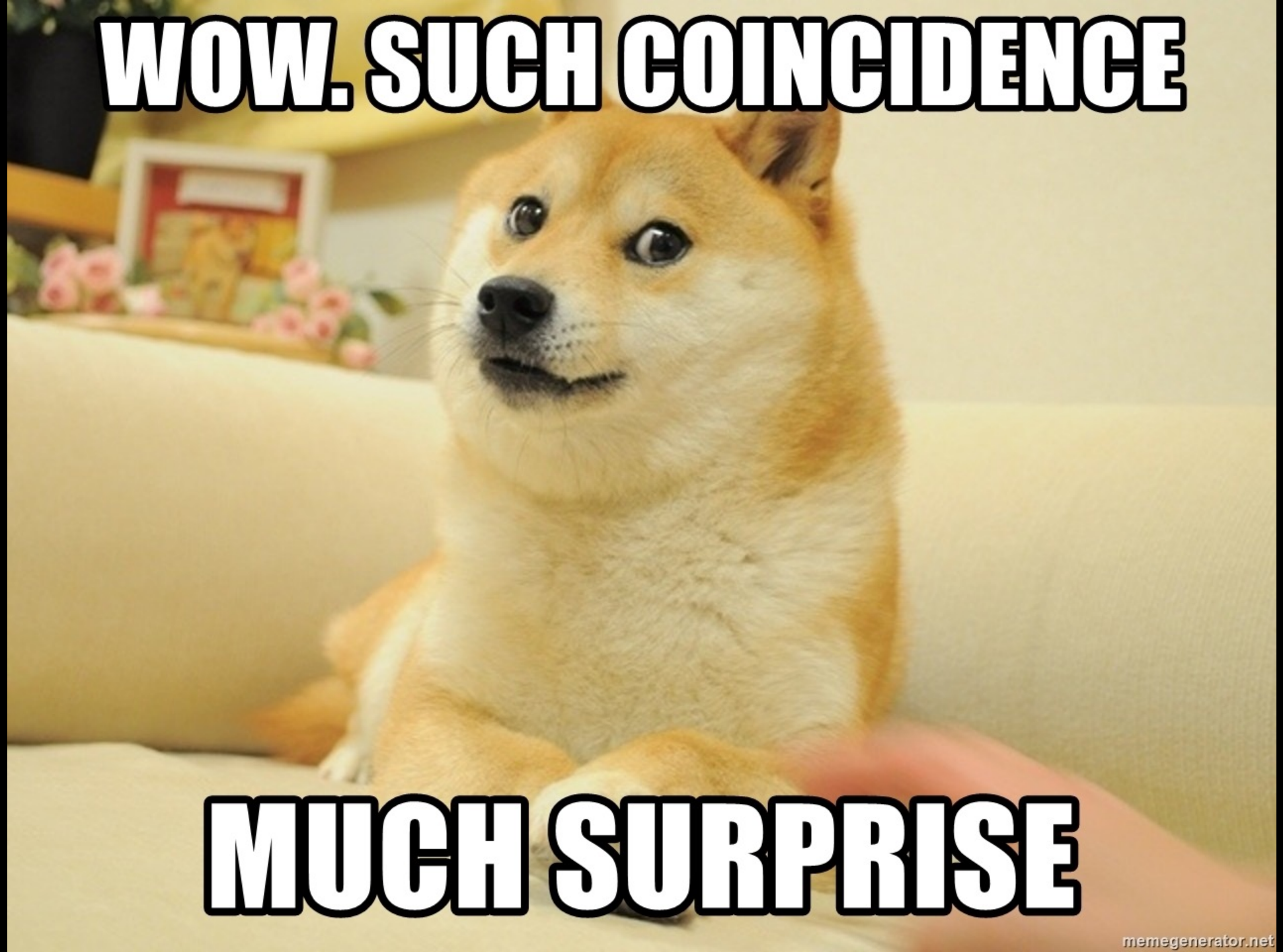
Step 3: Google / GPT the problem

Step 4: Evaluate solutions

Step 5: Integrate

The 5 step process

D O G E-I


WOW. SUCH COINCIDENCE
MUCH SURPRISE
memegenerator.net

About Me

Dean Pleban





Building tools for ML teamwork

Strongly believe in open source

DagsHub

The MLOps Podcast

Follow me:

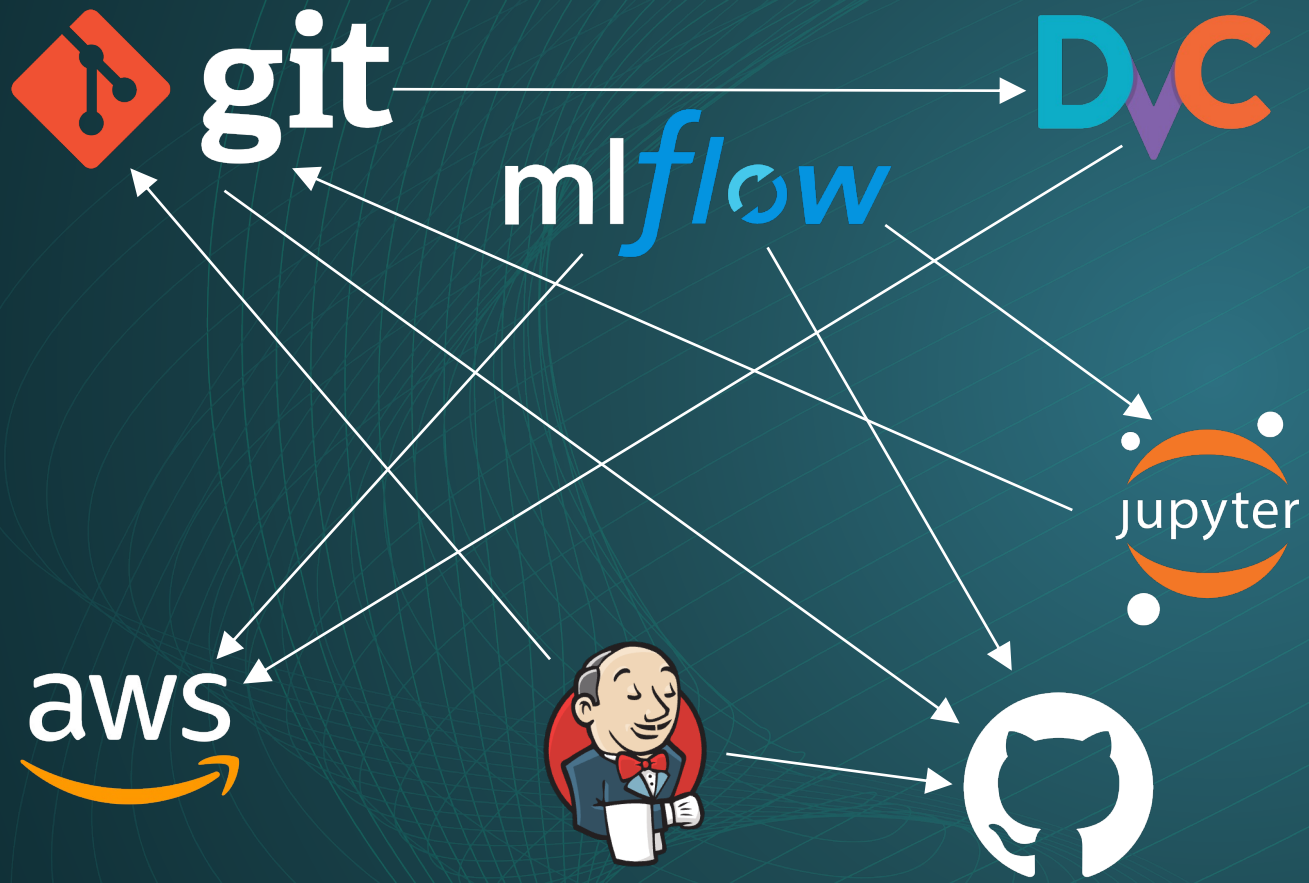
@DeanPlbn

/DeanPleban

First principles thinking



MLOps Fatigue – Too many tools, manually synchronized



300

MLOPS TOOLS TO CHOOSE FROM & INTEGRATE WITH

4 Stages of Grief



Everything is manual



Analysis Paralysis

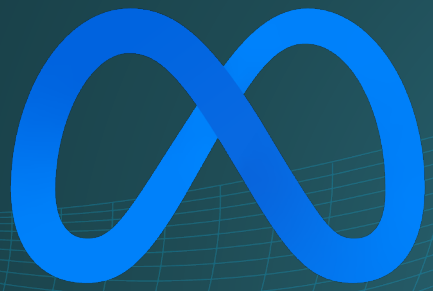


All or nothing



Building everything in-house

Assumptions about the PROBLEM







You are not Google / Facebook / Open AI



Existing MLOps tools will change

Assumptions about the PROBLEM

Production ready  **Save time / money**





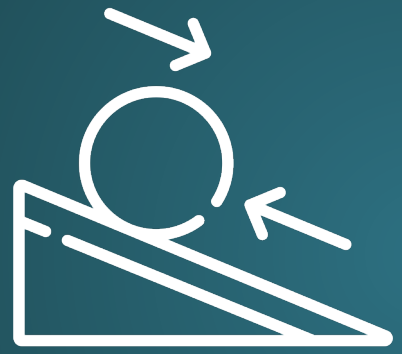
Future proof

It’s a matter of **tradeoffs**

Assumptions about the SOLUTION



Problem vs Feature focus



Find real friction points by going to prod



The ML Twist



Building on OSS

Example: Data Versioning

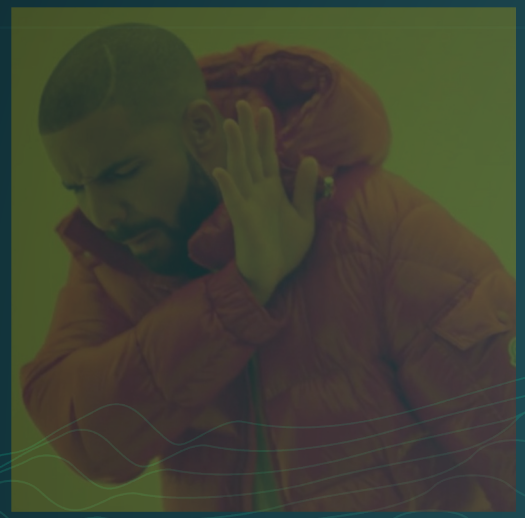


I want to version my data



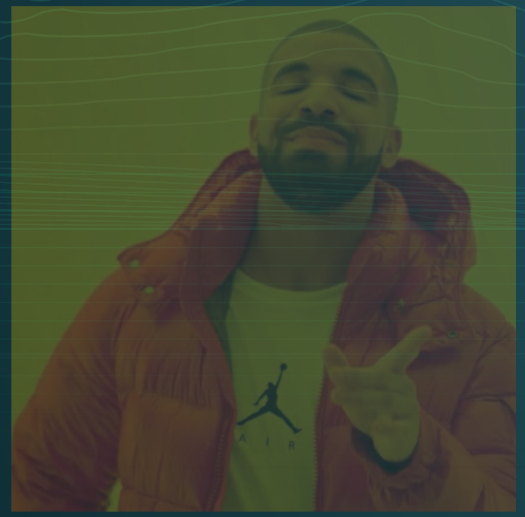
My data is regularly changing and I want to revert back to an older version for disaster recovery

Example: Data Versioning



I want to version my data

[Step 1: Define the problem]



My data is regularly changing and I want to revert back to an older version for disaster recovery

Step 1: Define the problem



Revert in case of bug



Find where the data is used



Edit data without breaking

Step 1: Define the problem

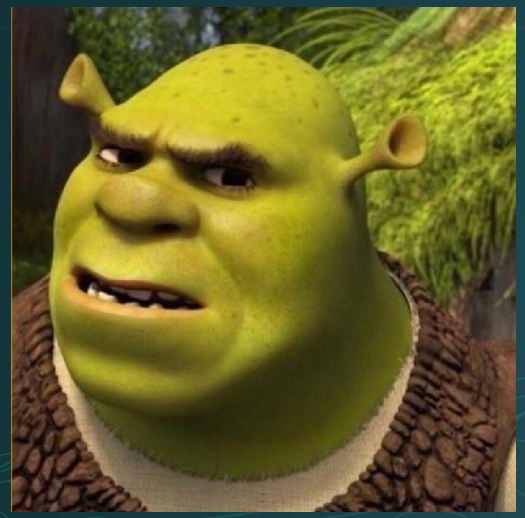


Ad-hoc definition



“All of the above”

Example: Data Versioning



We need to get a flexible tool that supports data versioning, so that we can revert in case of an emergency



We need a tool for tabular data that supports add-only changes, works with MySQL databases, and can support 100GB-10TB of data. We will only need to revert every few weeks.

Example: Data Versioning



Step 2: Outline problem parameters

We need to get a flexible tool that supports data versioning, so that we can revert in case of an emergency

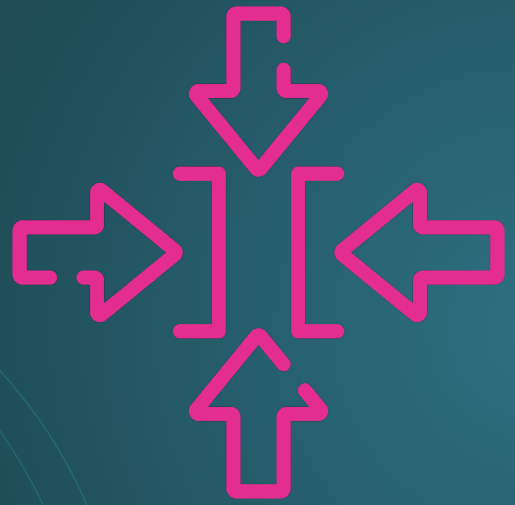


We need a tool for tabular data that supports add-only changes, works with MySQL databases, and can support 100GB-10TB of data. We will only need to revert every few weeks.

Step 2: Define the problem parameters



Problem-specific requirements



Organizational constraints



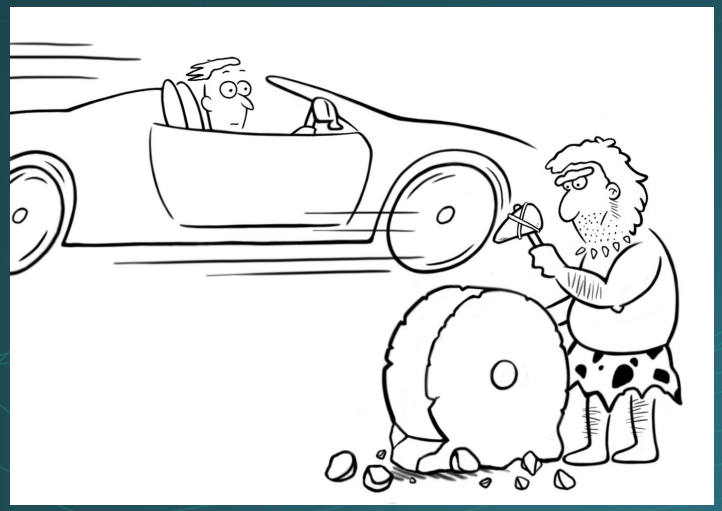
Who am I?

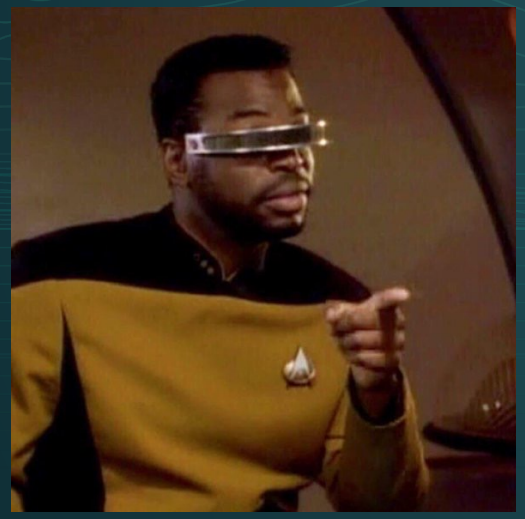


Define a “user story”

Example: Data Versioning







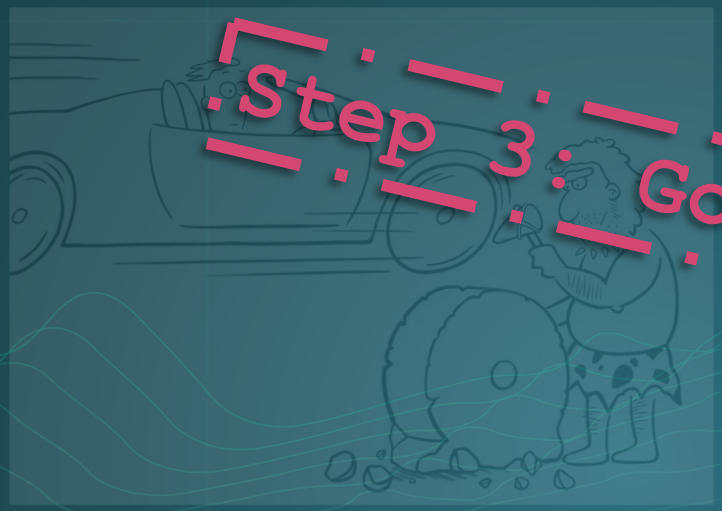
Google

ChatGPT

Example: Data Versioning







Step 3: Google / GPT the problem

Google

ChatGPT

Step 3: Google / GPT the problem



Budget time



Build an info page

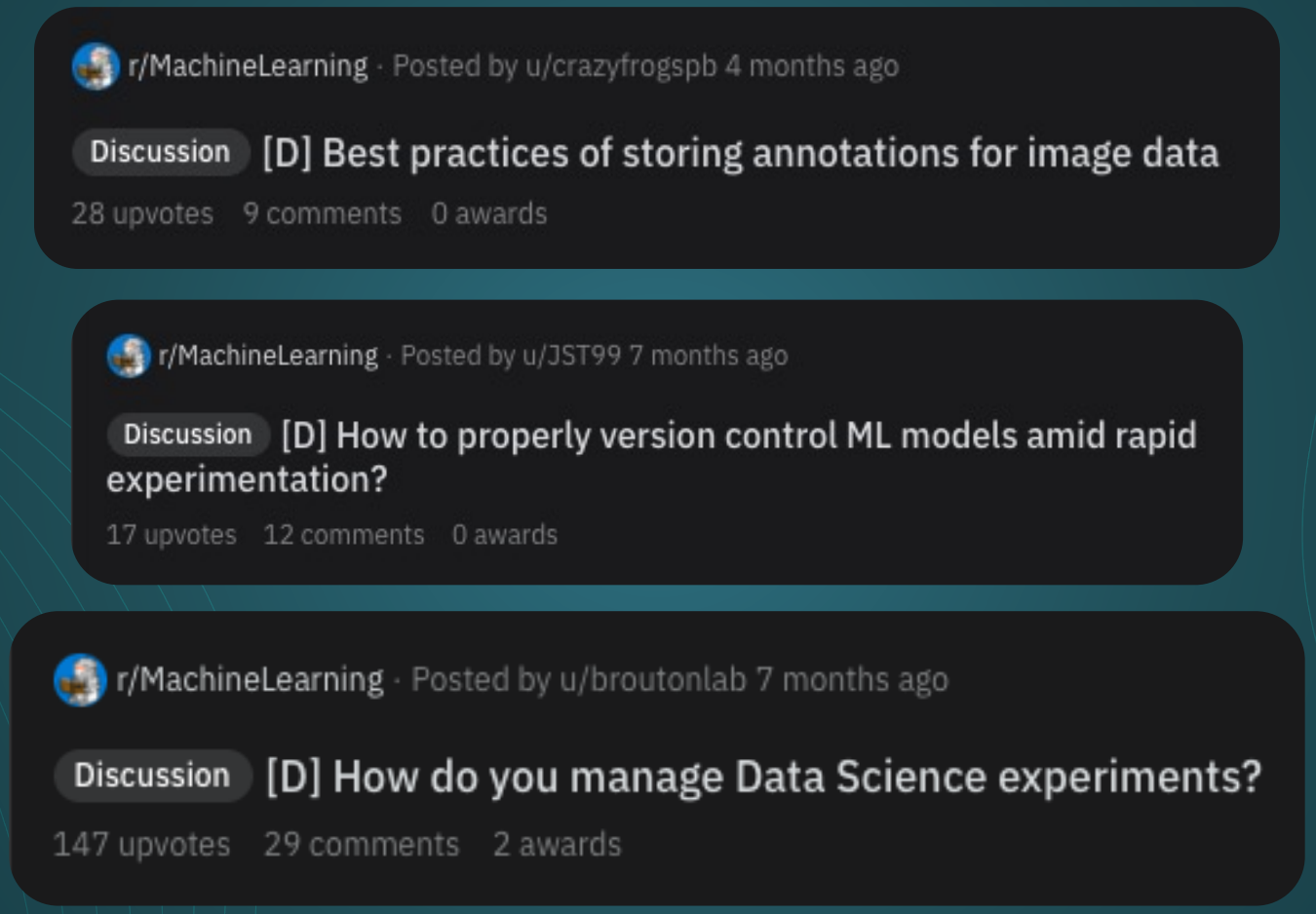


Discover the wrong words



Ask about the problem

Step 3: Google / GPT the problem



Reddit example

Step 3: Google / GPT the problem



Example: Data Versioning



Do 700 hello world tutorials



Methodically evaluate 2-3 tools

Example: Data Versioning



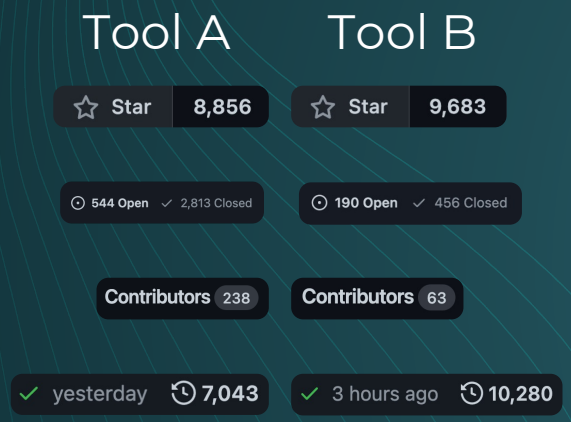
Do 700 hello world tutorials



Methodically evaluate 2-3 tools

Step 4: Evaluate solutions

Step 4: Evaluate solutions



"face value" Comparison



Operating principles deep dive



Get started tutorial

Step 4: Evaluate solutions – Stress Test



Adding a 10K image dataset

Example: Data Versioning

Step 5: Integrate





Going all in / drowning in success criteria



Start simple



Define success criteria, **or don't**



Review

The 5 step process

Step 1: Define the problem

Step 2: Outline problem parameters

Step 3: Google / GPT the problem

Step 4: Evaluate solutions

Step 5: Integrate

The 5 step process

D O G E-I

DagsHub

Thank You!



Follow me:

@DeanPlbn

/DeanPleban

Follow DagsHub:

@TheRealDAGsHub

/company/dagshub/