

# Lecture 1: Introduction

AI-5

Productionizing AI (MLOps)

Pavlos Protopapas, Shivas Jayaram

# Outline

---

1. Why should you take this class and why not?
  2. Who are we?
  3. Course structure and activities?
  4. Class organization (Workload, Logistics, Grades).
- 

## Projects

# Outline

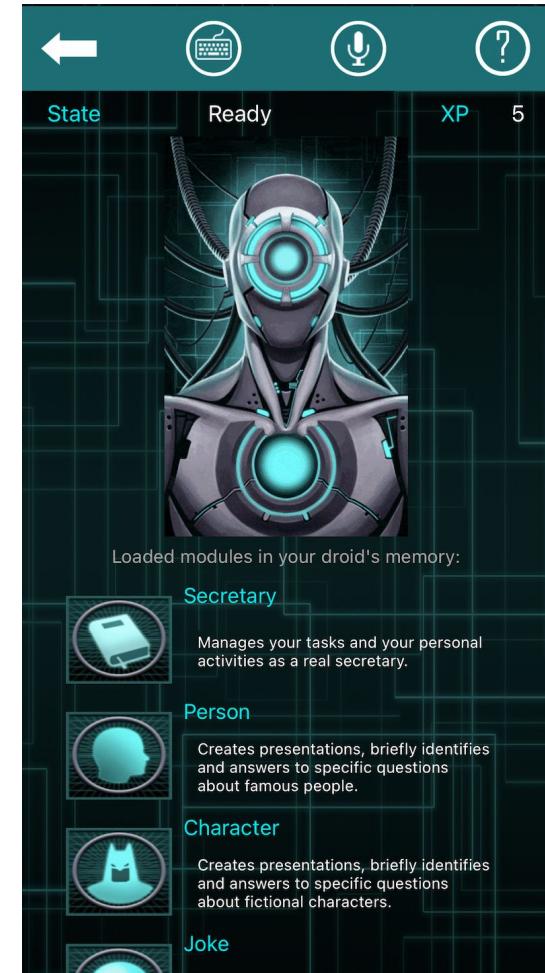
---

- 1. Why should you take this class and why not?**
  2. Who are we?
  3. Course structure and activities?
  4. Class organization (Workload, Logistics, Grades).
- 

## Projects

# Why you should take this class

So you can build awesome apps like this:



- <https://runwayml.com/>
- <https://www.databot-app.com/>

# Why you should take this class

---

Because you want to learn how to:

- Put your models in production
- Build an application using your models
- Take advantage of available models
- Integrate and orchestrate applications
- Deploy increasing amount of data

# Reasons This Course May Not Be For You

---

1. **Lack of Commitment:** This course demands full engagement with materials and projects.
2. **Expectation of a Traditional Lecture Format:** This is a project-based course, with a focus on hands-on learning.
3. **Limited Background in Prerequisites:** If you are unfamiliar with key concepts from AI-1,2,3 such as:
  - Basic Machine Learning
  - CNNs, RNNs, Autoencoders, GANs, etc.
  - Basic shell commands.

# Why you shouldn't take this class

---

4. **Unwillingness to Receive and Apply Feedback:** We provide detailed feedback on your projects, expecting you to implement changes in subsequent milestones.
5. **Seeking a "Easy A":** While project-based courses may seem easier, we hold rigorous evaluation standards.

# Motivation

---

Mckinsey Global Survey findings on Adoption of AI shows nearly 25% year over year increase in the use of AI. 50% of companies spend between 8 and 90 days deploying a single AI model, with 18% taking longer than 90 days. A report by IDC that surveyed 2,473 organizations and their experience with ML found that a significant portion of **attempted deployments fail**, quoting **lack of expertise**, as one of the key factors.<sup>[1]</sup>

[1] <https://arxiv.org/pdf/2011.09926.pdf>

# Motivation

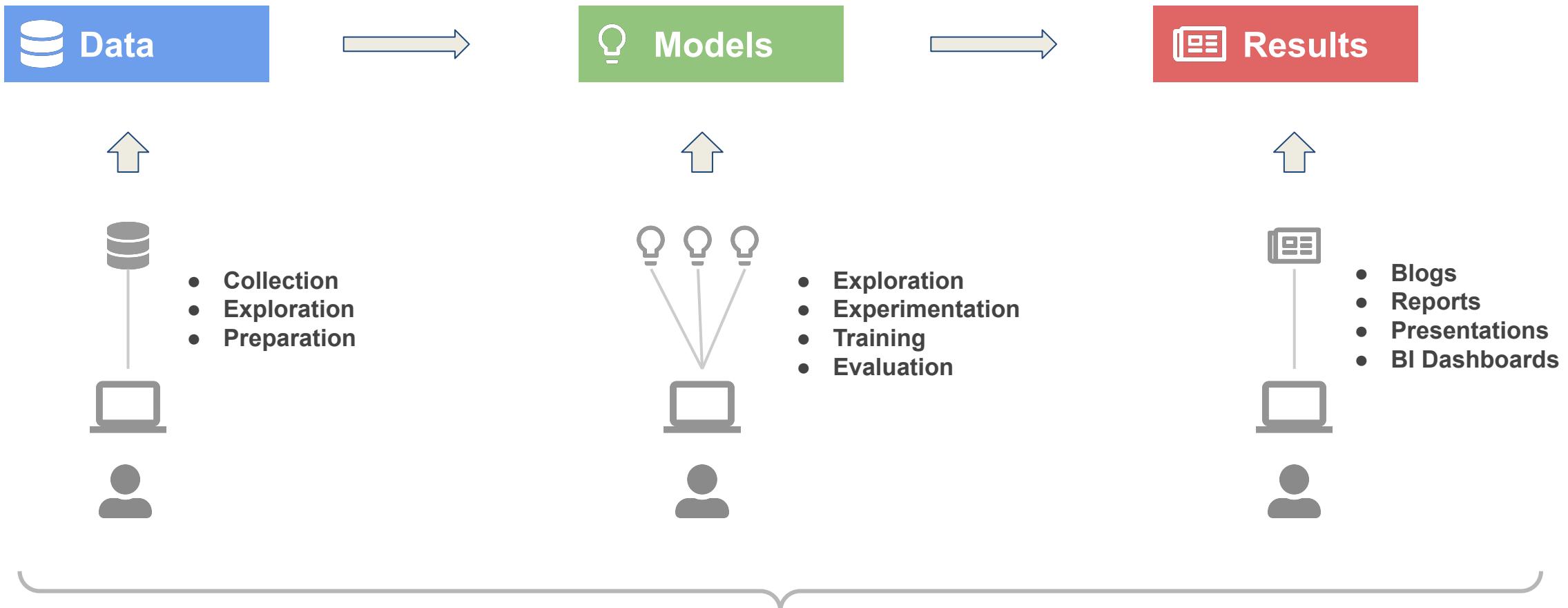
---

A recent International Data Corporation ([IDC](#)) survey of global organizations that are already using artificial intelligence (AI) solutions found only 25% have developed an enterprise-wide AI strategy. At the same time, half the organizations surveyed see AI as a priority and two thirds are emphasizing an "AI First" culture.

IDC: <https://www.idc.com/>

# Data Science Series to Real World

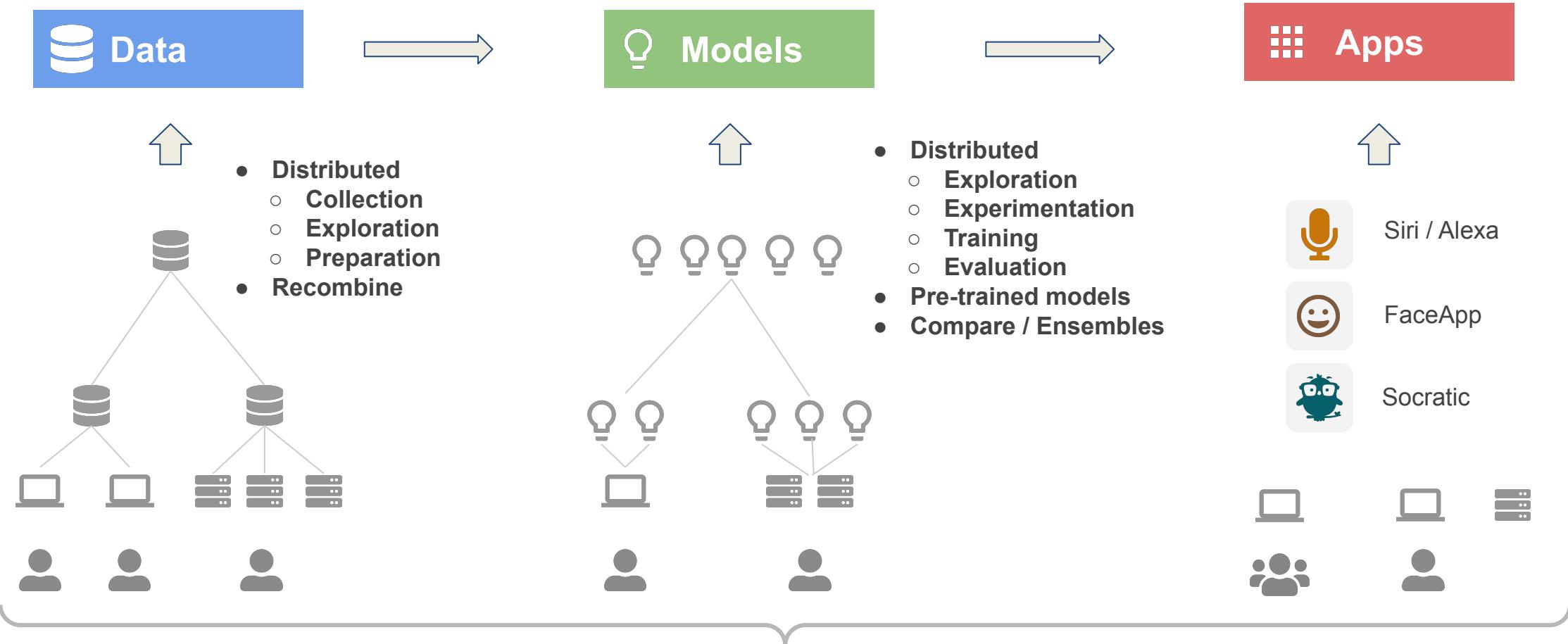
Data Science Series AI 1 - 4



Single developer on one computer. Projects are individual to 2-3 member team.

# Data Science Series to Real World

Real World



# Data Science Series to Real World (cont)

---

## Challenges:

- Required Installations for Specific Operating Systems
- Guidelines for Code Collaboration
- Methods for Sharing Datasets and Models
- Requirements for Multi-GPU Usage or Extended Training Sessions
- Automation of Data Gathering and Model Training
- Onboarding Procedures for New Team Members
- Resolving "It Works on My Machine" Issues 

# Ops for Machine / Deep Learning

---

## **Development Operations (DevOps):**

Unifies software development (Dev) & operations (Ops) for efficiency.

## **Machine / Deep Learning Operations (MLOps):**

Integrates ML/DL model development with app development & operations.

## Machine / Deep Learning:

- Data collection & exploration
- Model exploration & selection
- Training & evaluation
- Distillation & compression

## Application Development:

- APIs / Model serving
- ML integration
- Web & mobile apps
- Edge device apps
- Automation scripts

## Operations:

- Provisioning and managing deployment servers, on-demand GPU servers
- Maintain 100% uptime of app / apis
- CI/CD: Continuous Integration / Deployment
- Continuous Data Collection / Model Training
- Model/data monitoring
- Model/data versioning
- ML Workflow Management

# MLOps - Tech Stack



Data



Models



Development



Operations



Data Engineers

Data Scientists

Software Engineers

Systems Engineers

# MLOps - Tech Stack



Data



Models



Development



Operations

- ★ Spark
- 🐶 Hadoop
- 🐘 Kafka
- ⚡ Dask
- ↗ Airflow
- 🐘 Pachyderm
- 🐶 DagsHub
- 🔍 Elastic
- .neo Neo4j
- >Weaviate



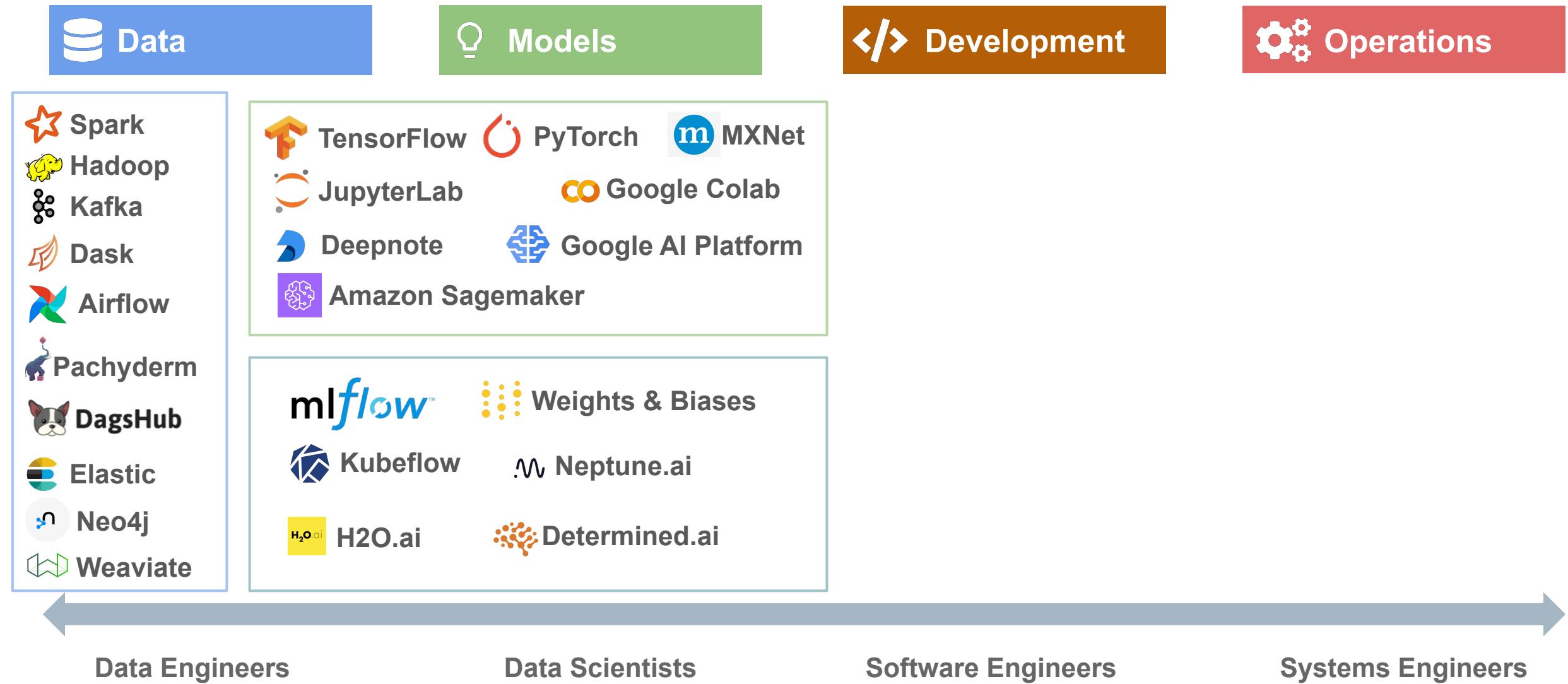
Data Engineers

Data Scientists

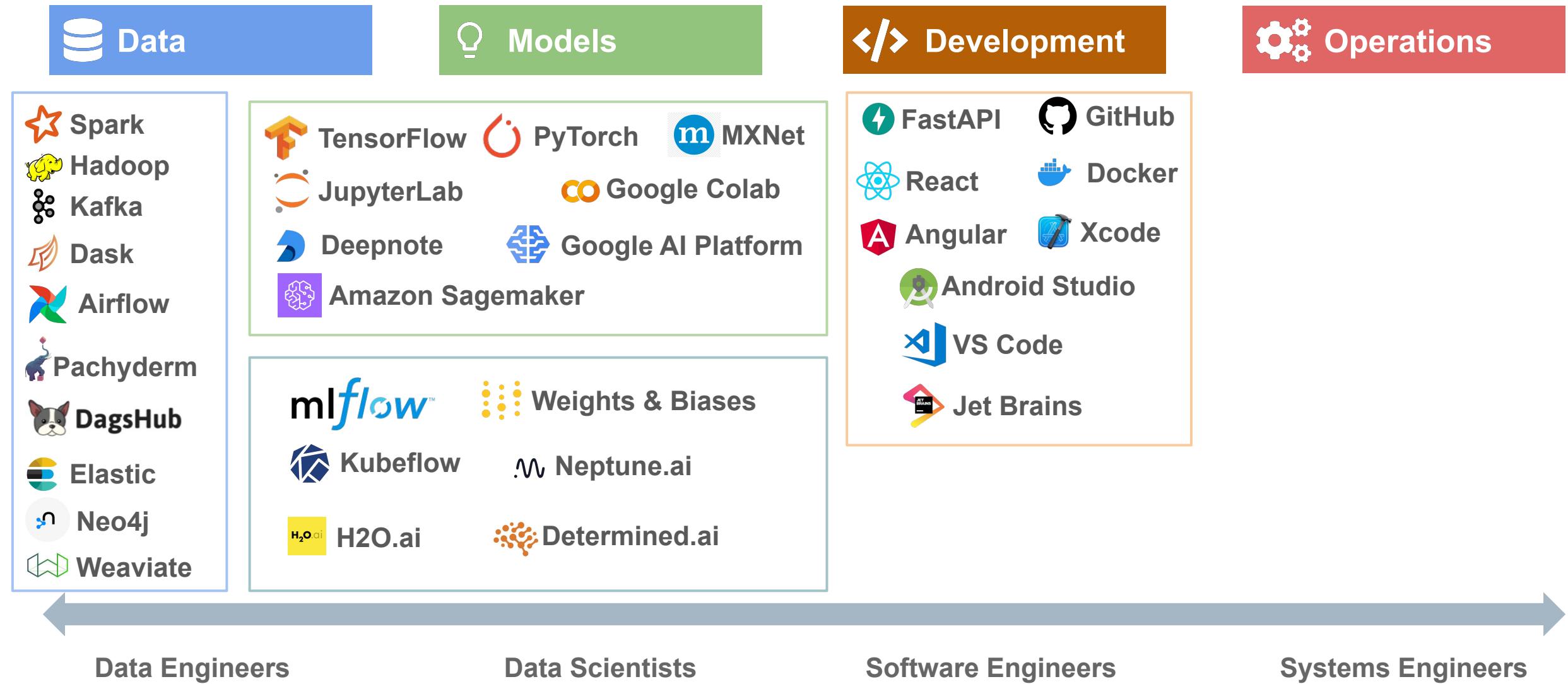
Software Engineers

Systems Engineers

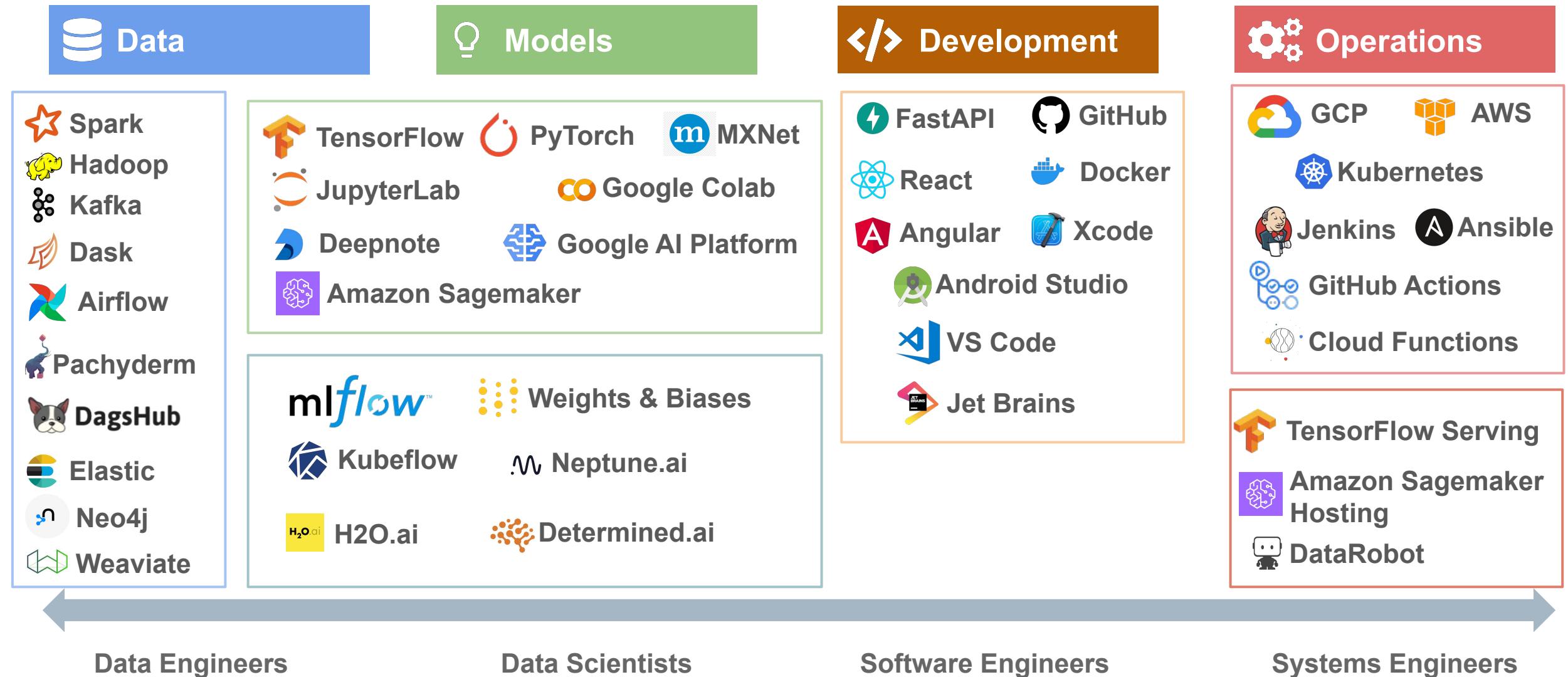
# MLOps - Tech Stack



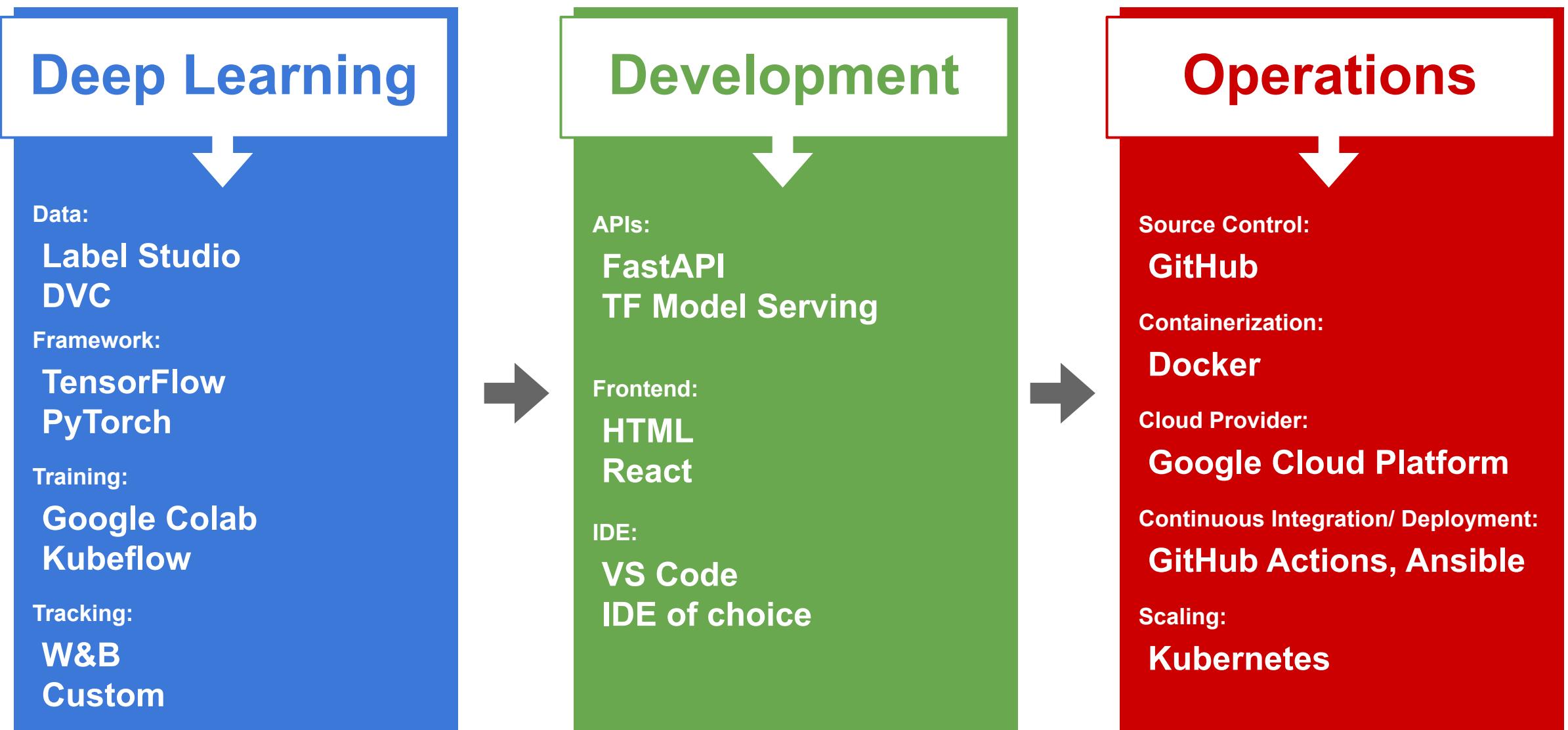
# MLOps - Tech Stack



# MLOps - Tech Stack



# MLOps - Tech Stack



# Outline

---

1. Why should you take this class and why not?
  2. Who are we?
  3. Course structure and activities?
  4. Class organization (Workload, Logistics, Grades).
- 

## Projects

# Who? The Astro-AI-Statistician Who Rocks The Kitchen!

---

## Roles:

- The Science Wizard: Scientific Director of IACS.
- Course Maestro: Instructs CS109a, CS109b, and AC215 like a boss.
- Astro-Guru: Stoked about the next-gen telescopes that are about to revolutionize our view of the universe!

## Research Lab:

- StellarDNN Maverick: Tackles cosmic mysteries where astronomy, ML, and stats collide!
- His Interests? Cracking open differential equations with deep neural networks, being a detective in deep neural network inference, and teaching NLP techniques to chat with stars (well, in astronomical time series analysis, but let's keep it jazzy!).

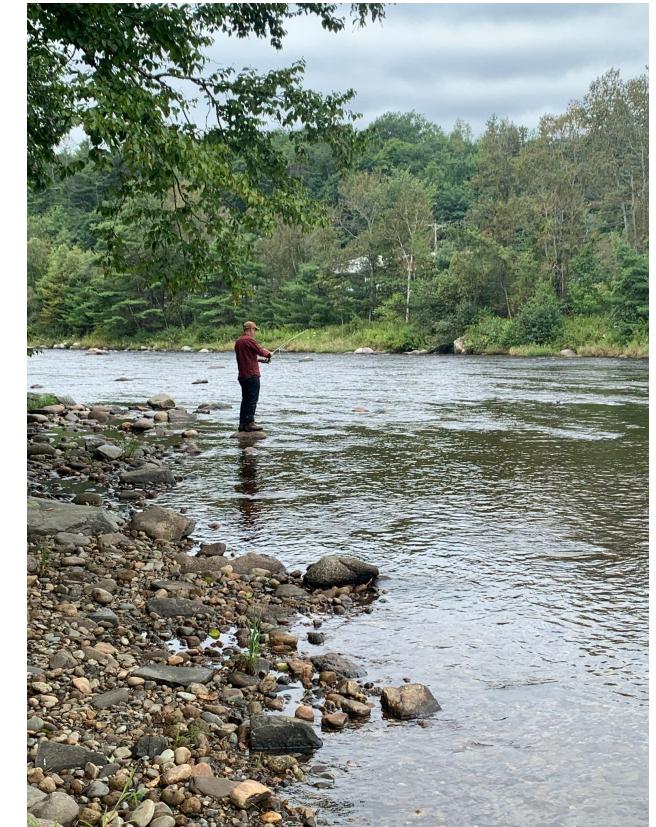


# Who? The Astro-Statistician Who Rocks The Kitchen!

---

## Fun Facts:

- Musical Soul: Classical tunes and opera are his jam. The Boston Symphony Orchestra is his second home!
- Culinary Artist: Holds a cooking badge of honor from Le Cordon Bleu and enjoys both whipping up a storm and devouring the results.
- Adventure Junkie: From biking up mountains to skiing down them, from kayaking to hooking fish mid-air—this professor is always on the go!



# Who ?

---



Shivas Jayaram



Ignacio Becker



Rahul Dave

# Who ?

---



Anshika Gupta



Aamir Ahmad  
Ansari



Shibani  
Budhraja



Vishnu M

# Outline

---

1. Why should you take this class and why not?
  2. Who are we?
  3. **Course structure and activities?**
  4. Class organization (Workload, Logistics, Grades).
- 

## Projects

# Course Structure and Activities

---

- One lectures/tutorials per week.
- Projects

# Lectures

---

## Two lectures per week - What to expect



There will be one reading suggestion per week

# Topics

---

- Containers
- Data Pipelines, Dask & Cloud Storage
- Data Parallelization
- Data Versioning
- Advanced Training Workflows
- Advanced Inference Workflows
- Pipeline
- App Design, Setup, and Code
- APIs & Frontend
- Scaling (k8)

# Team Projects: Crafting Your Own AI Solutions

---

**Objective:** Hands-on project development in AI & MLOps; transform your idea into a functional app.

**Guidance:** Weekly demonstrations from Pavlos' project provide practical insights and a reference point.

**Milestones:** Assess project evolution and grasp of MLOps concepts; crucial for grading.

**Creativity:** Open platform for start-up ideas, research, or personal hobbies.

**Assessment:** Milestones are key to **grades** and holistic development.

**Group Formation:** Starts today, aiming for teams of 3-5.

# Projects Milestones

Milestone	Description	Due Date	Points
MS1	Project Proposals & Team Formation	30th Dec	5
MS2	MLOps & Advanced Training	20th Jan	10
MS3	Scalable Computing	10th Feb	15
MS4	Midterm Presentation	24th Feb	25
MS5	Full-Stack Development	30th Mar	10
MS6	Final Presentation and Deliverables	TBD	35

# Outline

---

1. Why should you take this class and why not?
  2. Who are we?
  3. Course structure and activities?
  4. **Class organization (Workload, Logistics, Grades).**
- 

## Projects

# Workload

---

- 1 hour *Reading*
- 2.5 hours *Lectures*
- 1 hour *Office Hour/meet with your TF*
- 7.5 hours *Project Milestones*
- ~ 12 hours/ week

# Expectations: Attendance

---

Attendance in class is not merely a requirement; it is a criterion I heavily weigh when considering academic and professional endorsements. Please understand that consistent presence and engagement in the classroom are highly valued in this course.

# Expectations: Projects

---

This course is project-based, and your entire grade will depend on the quality of your project. We will place significant emphasis on both the effort you put into the project and the quality of your completion of its milestones.

While project-based courses generally make it easier to achieve a high grade, we will thoroughly assess each component and provide comprehensive feedback. We expect you to incorporate this feedback into your work for subsequent milestones.

# Course Components

## Course web page

The screenshot shows the AI-5 course web page. The navigation bar includes links for Schedule and Calendar, Projects, Readings, Staff / Contact, and FAQs. The main content area features a large heading "Productionizing AI (MLOps): AI-5." and a "TABLE OF CONTENTS" section with the following items:

- 1 Course Introduction
- 2 Course Topics Overview
- 3 Prerequisites
- 4 Course Components
- 5 Grade Distribution
- 6 Course Policies
- 7 Accessibility:
- 8 Inclusion and Belonging Statement

<https://easternempire.github.io/AI-5/>

## Discord

The screenshot shows the AI-5 Discord server interface. The left sidebar displays the "Get Started" channel, which contains sections for Server Guide, 1 Event, Channels & Roles (NEW), Members, INFORMATION, and TEXT CHANNELS. The right side of the screen shows a general chat window with a message from shibani: "Next: introduce yourself!". Below it, there's a "New to Forum Channels? Check out our guide!" section and a post from shibani: "Using this discussion board! :point\_right:". The top right corner of the interface includes standard Discord controls like search and settings.

# Grades

Assignment	Final Grade Weight
Milestone 1	5%
Milestone 2	10%
Milestone 3	15%
Milestone 4	25%
Milestone 5	10%
Milestone 6	35%
<b>Total</b>	<b>100%</b>

# Final Details

---

- We will be using discord for discussions, announcements and surveys

Submissions for project milestones and projects will be using GitHub – details will follow soon

# Outline

---

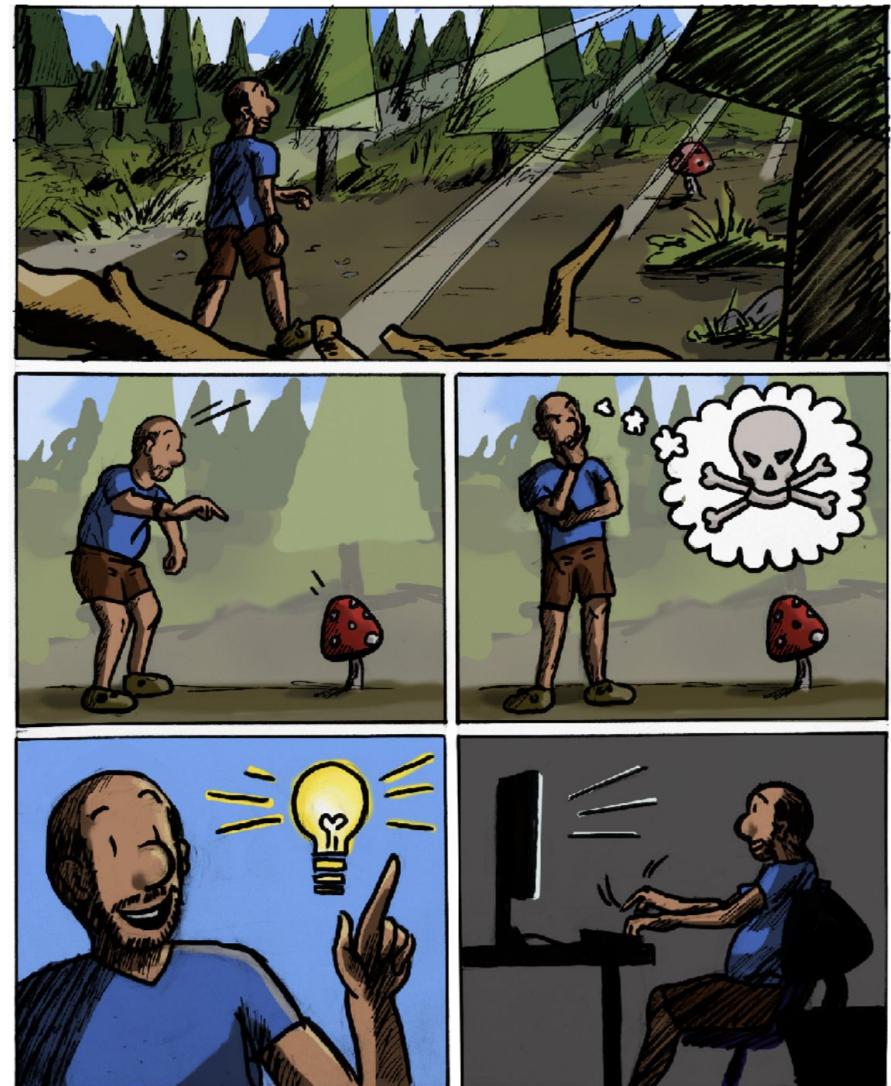
1. Why should you take this class and why not?
  2. Who are we?
  3. Course structure and activities?
  4. Class organization (Workload, Logistics, Grades).
- 

## Projects

## In Class Demo Mushroom Identification App

# Project Idea

- Pavlos likes to go the forest for mushroom picking
- Some mushrooms can be poisonous
- Help build an app to identify mushroom type and if poisonous or not
- [Project Summary](#)



Credit: Nikolas Protopapas

# Problem Definition

---

Pavlos enjoys visiting the woods to collect mushrooms, finding the endeavor both enjoyable and beneficial since some mushrooms can be eaten. However, the forest where he gathers mushrooms is also home to numerous toxic varieties. While some are easy to identify, there are others for which he needs assistance in recognizing.

# Proposed Solution

---

When Pavlos is in the forest, he always has his phone on hand. Imagine if he could simply snap a photo of the mushrooms, and an app could identify the species and indicate whether they are poisonous or safe to eat.

# Project Execution Steps

---

- Project Ideation / Requirements
- Data Exploration
- Model Exploration
- Prototyping
- Model Serving
- Product Development
- ML Integration
- Deployment

# Previous Year's Projects

<https://medium.com/institute-for-applied-computational-science/ac215-fall2021-projects/home>



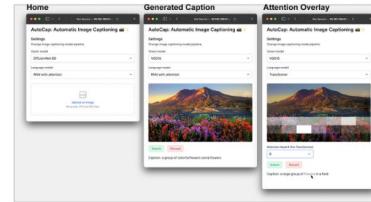
## Improved StyleCLIP for Text-Based Image Editing

This article was produced as part of the final project for Harvard's AC215 Fall 2021 course.



Michael Fein

Dec 13, 2021 · 7 min read



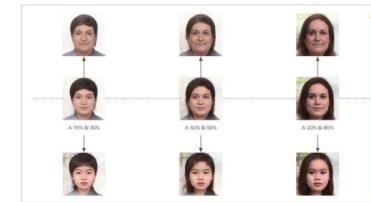
## AutoCap—Image Captioning

This article was produced as part of the final project for Harvard's AC215 Fall 2021 course.



James Parker

Dec 30, 2021 · 11 min read



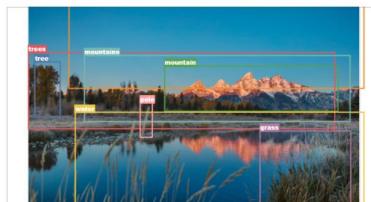
## Multifaceted face generator

Harvard AC215: Advanced Practical Data Science Author: Dongyun Kim, Vasco Meerman



Dongyun Kim

Dec 14, 2021 · 8 min read



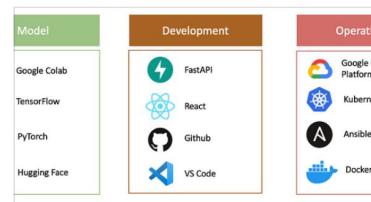
## ExBERT Image Captioning

Authors: Yuanbiao Wang, Shuyi Chen, Chunyi Li and Xinyi Li



Xinyi Li

Dec 30, 2021 · 6 min read



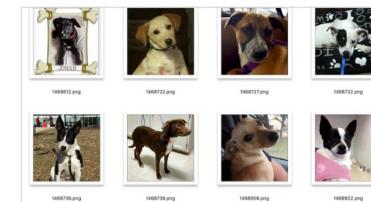
## Project Garble: Long Audio to Short Text Summarizer

We're going to build an app that takes a long audio file and generates a paragraph long summary—with all the main ideas included!



Joseph Kim

Dec 14, 2021 · 9 min read



## An End-to-End Approach Leveraging Computer Vision, NLP to Enable Better Pet...

This article was produced as part of the final project for Harvard's AC215 Fall 2021.



Guanhua Shu

Aug 29 · 7 min read

# How to Scope your Project



## Proof Of Concept (POC)

- Experiment potential ideas
- Check feasibility of the idea
- Use a subset of data to make experiments simpler to run
- E.g.: Verify if our language task can be performed by transfer learning using a transformer model
- **Users:** Internal team
- **Duration:** Days to few weeks

## Prototype

- A mockup or functional product that can showcase your ideas
- E.g.: A mockup web app to show user experience and flow
- **Users:** Internal team
- **Duration:** Weeks

## Pilot

- A usable and functional product of your solution
- Used to test out the product with real users and performing real use cases
- E.g.: An api endpoint of a model for prediction, a simple one page app to showcase a model's prediction capability
- **Users:** Internal / External
- **Duration:** Weeks

## Minimum Viable Product (MVP)

- Expanding on the Pilot to build something that real users can use
- E.g.: Production deployed app that can predict if a mushroom is poisonous or not
- **Users:** External
- **Duration:** Months

# Project Scope (Mushroom App)



## Proof Of Concept (POC)

- Scrap mushroom data
- Verify images
- Experiment on some baseline models
- Verify new unseen mushrooms are predicted by the model(s)
- Visualize model activations to analyze what the model is seeing

## Prototype

- Create a mockup of screens to see how the app could look like
- Deploy one model to Fast API to service model predictions as an API

## Minimum Viable Product (MVP)

- Create App to identify Mushrooms
- API Server for uploading images and predicting using best model

# Logistics

---

- Survey (will be shared on Discord)
- Make project groups
- Setup & Installations

**THANK YOU**