Lecture 1: Introduction

AC215

Pavlos Protopapas
SEAS/Harvard

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).

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).
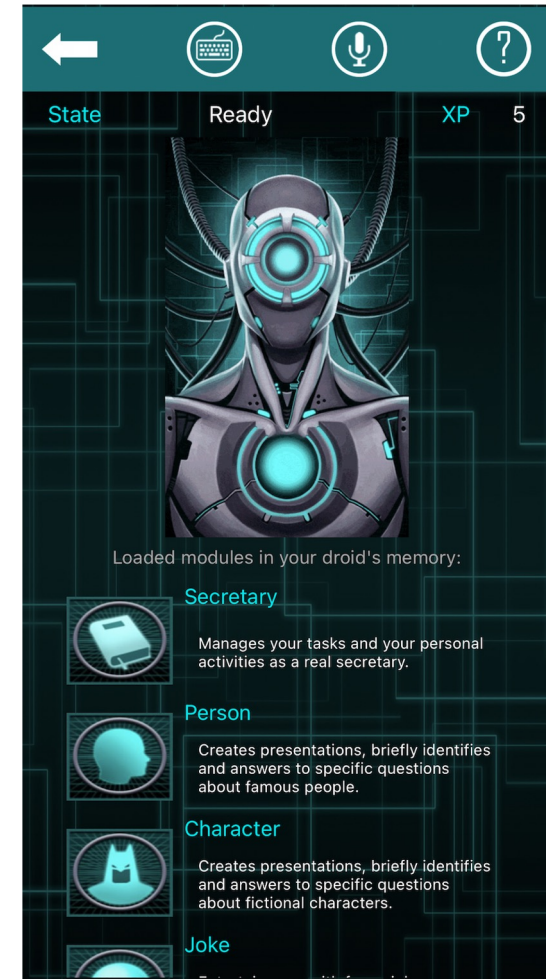
Projects

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 CS109A/B 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.[1]
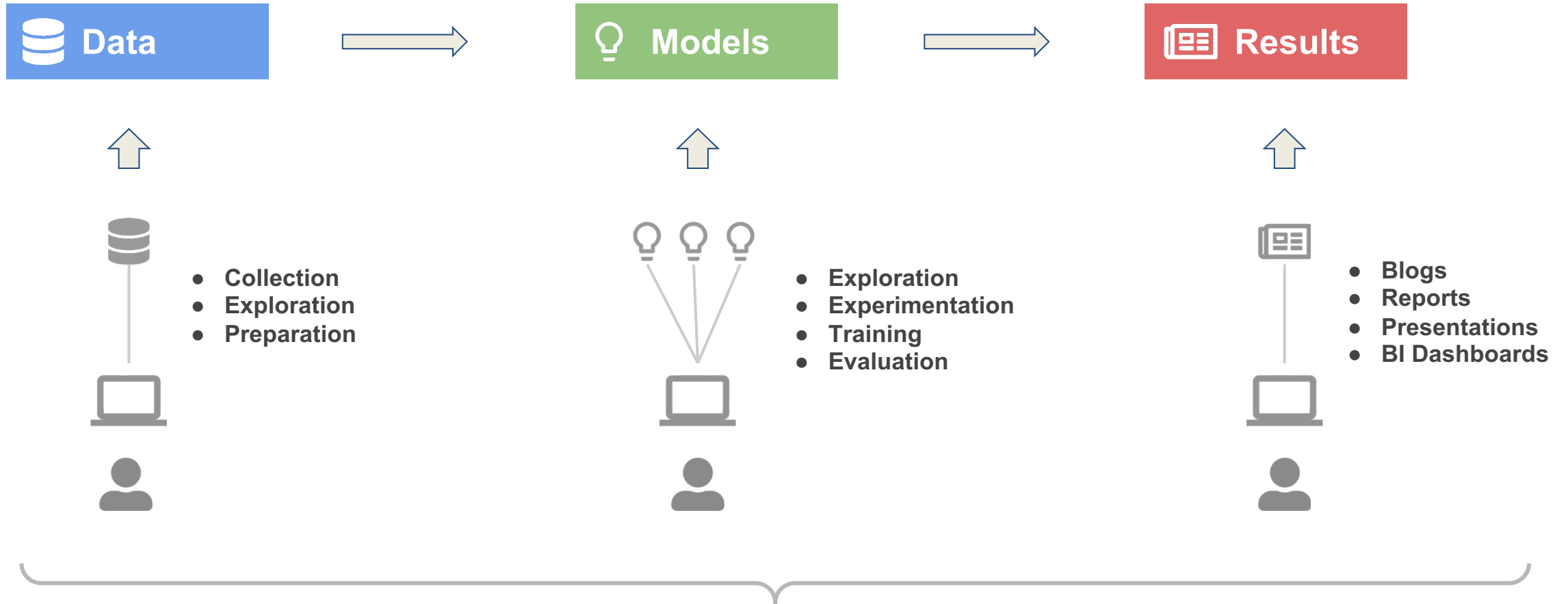
[1] https://arxiv.org/pdf/2011.09926.pdf

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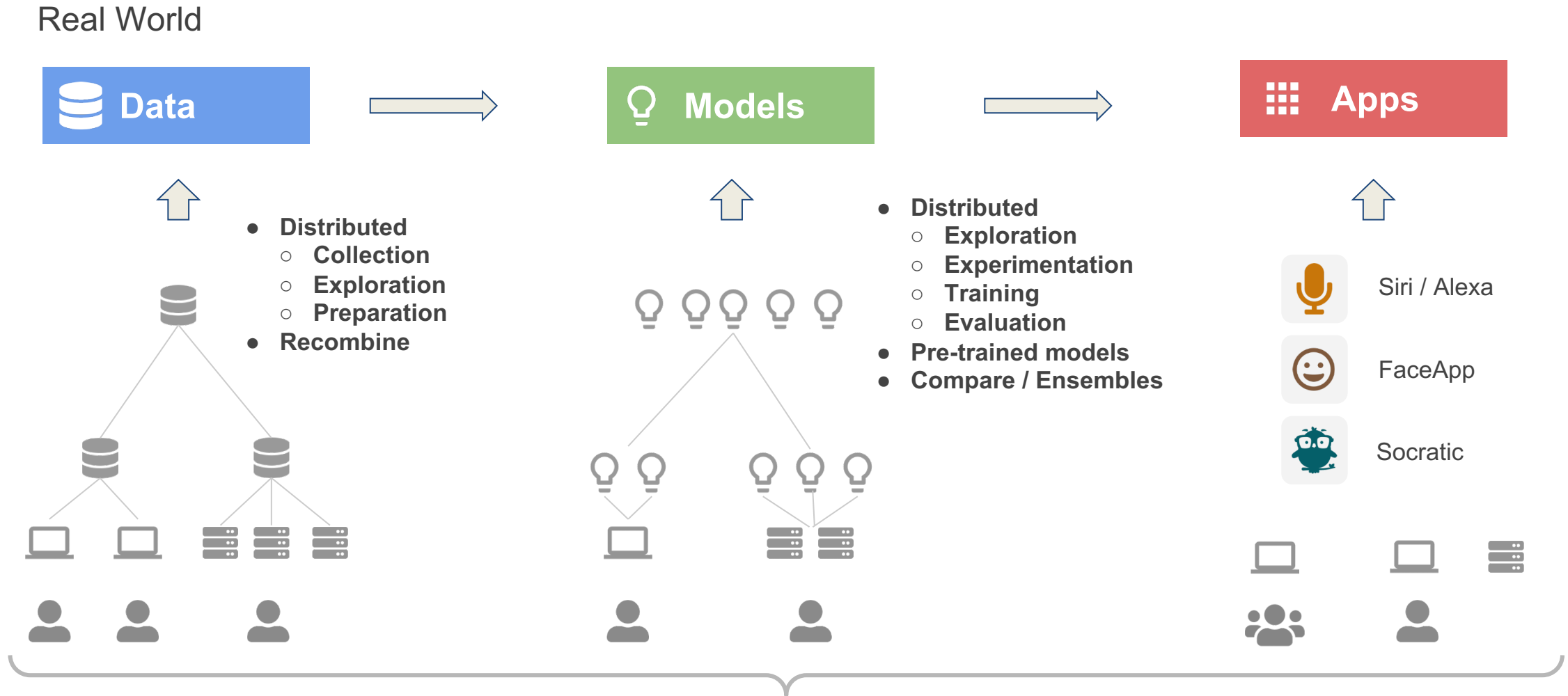
IDC: https://www.idc.com/

Data Science Series to Real World

Data Science Series CS109 A/B

Data		Models		Results

Data:
- Collection
- Exploration
- Preparation

Models:
- Exploration
- Experimentation
- Training
- Evaluation

Results:
- Blogs
- Reports
- Presentations
- BI Dashboards

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

Data Science Series to Real World

Real World

Data → **Models** → **Apps**

- **Distributed**
  - **Collection**
  - **Exploration**
  - **Preparation**
- **Recombine**

- **Distributed**
  - **Exploration**
  - **Experimentation**
  - **Training**
  - **Evaluation**
- **Pre-trained models**
- **Compare / Ensembles**

Siri / Alexa

FaceApp

Socratic

Team of developers on single/ multi node clusters on a Cloud Platform. Projects are 5+ member teams

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 ¯_(ツ)_/¯

Ops for Machine / Deep Learning

Development Operations (DevOps):

DevOps is a practice that brings together software development (Dev) and operations (Ops) to streamline the process for better productivity and shorten development life cycle.

Machine / Deep Learning Operations (MLOps):

MLOps is a practice that brings together machine learning or deep learning model development, application development, and operations together to streamline the interaction between the three and simplify the machine learning life cycle.

MLOps - Tasks

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**

16

MLOps - Tech Stack

Data

Models

Development

Operations

- ⭐ **Spark**
- 🐘 **Hadoop**
- **Kafka**
- 🍂 **Dask**
- 🎡 **Airflow**
- 🐘 **Pachyderm**
- 🐶 **DagsHub**
- **Elastic**
- **Neo4j**
- **Weaviate**

Data Engineers   **Data Scientists**   **Software Engineers**   **Systems Engineers**

17

MLOps - Tech Stack

Data

Models

Development

Operations

- Spark
- Hadoop
- Kafka
- Dask
- Airflow
- Pachyderm
- DagsHub
- Elastic
- Neo4j
- Weaviate

TensorFlow | PyTorch | MXNet
JupyterLab | Google Colab
Deepnote | Google AI Platform
Amazon Sagemaker

mlflow | Weights & Biases
Kubeflow | Neptune.ai
H2O.ai | Determined.ai

Data Engineers

Data Scientists

Software Engineers

Systems Engineers

18

MLOps - Tech Stack

Data

- Spark
- Hadoop
- Kafka
- Dask
- Airflow
- Pachyderm
- DagsHub
- Elastic
- Neo4j
- Weaviate

Models

- TensorFlow
- PyTorch
- MXNet
- JupyterLab
- Google Colab
- Deepnote
- Google AI Platform
- Amazon Sagemaker

- mlflow
- Weights & Biases
- Kubeflow
- Neptune.ai
- H2O.ai
- Determined.ai

Development

- FastAPI
- GitHub
- React
- Docker
- Angular
- Xcode
- Android Studio
- VS Code
- Jet Brains

Operations

Data Engineers   Data Scientists   Software Engineers   Systems Engineers

19

MLOps - Tech Stack

Data

- Spark
- Hadoop
- Kafka
- Dask
- Airflow
- Pachyderm
- DagsHub
- Elastic
- Neo4j
- Weaviate

Models

- TensorFlow
- PyTorch
- MXNet
- JupyterLab
- Google Colab
- Deepnote
- Google AI Platform
- Amazon Sagemaker

- mlflow
- Weights & Biases
- Kubeflow
- Neptune.ai
- H2O.ai
- Determined.ai

Development

- FastAPI
- GitHub
- React
- Docker
- Angular
- Xcode
- Android Studio
- VS Code
- Jet Brains

Operations

- GCP
- AWS
- Kubernetes
- Jenkins
- Ansible
- GitHub Actions
- Cloud Functions

- TensorFlow Serving
- Amazon Sagemaker Hosting
- DataRobot

Data Engineers   **Data Scientists**   **Software Engineers**   **Systems Engineers**

MLOps - Tech Stack
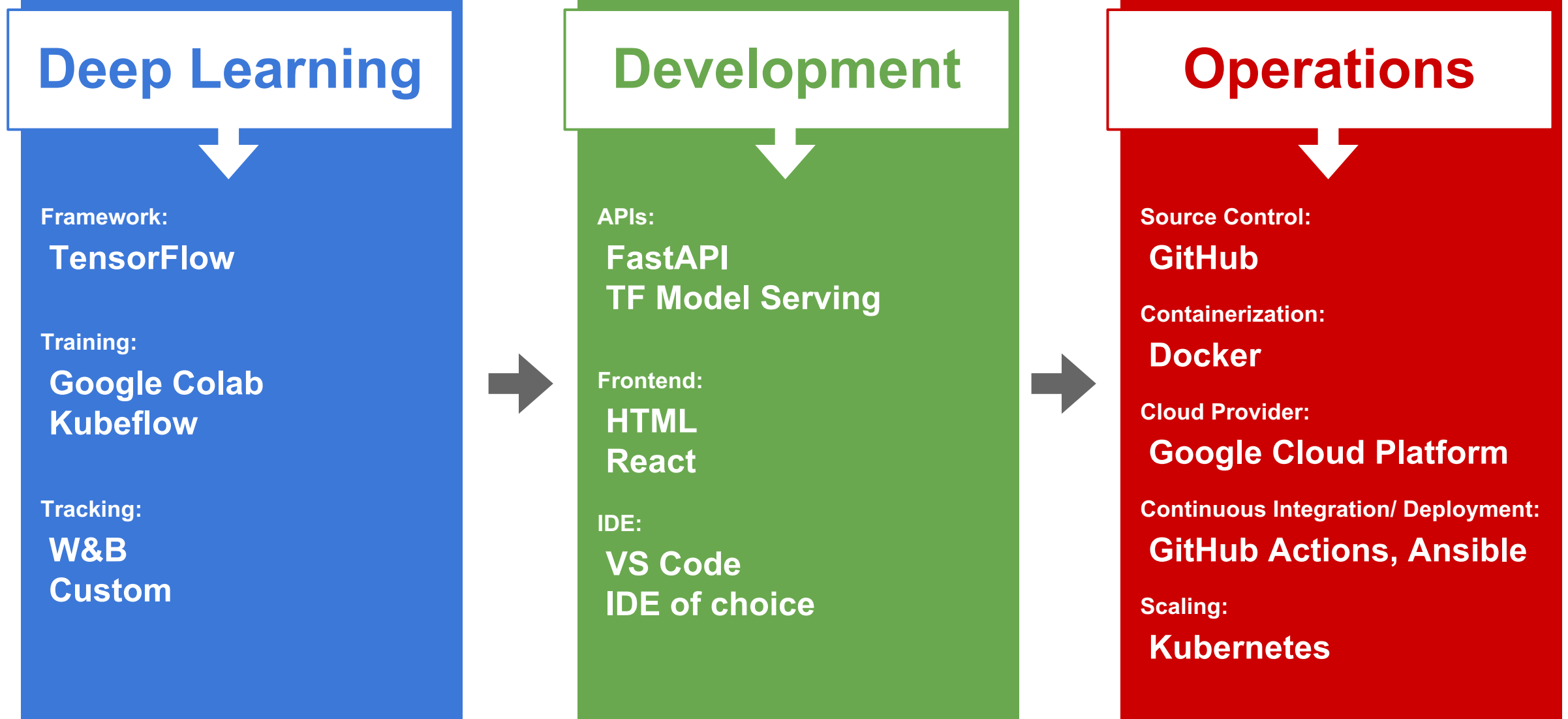
Deep Learning

Framework:

TensorFlow

Training:

Google Colab
Kubeflow

Tracking:

W&B
Custom

Development

APIs:

FastAPI
TF Model Serving

Frontend:

HTML
React

IDE:

VS Code
IDE of choice

Operations

Source Control:

GitHub

Containerization:

Docker

Cloud Provider:

Google Cloud Platform

Continuous Integration/ Deployment:

GitHub Actions, Ansible

Scaling:

Kubernetes

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-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 ?

Rashmi Banthia

TF for many Data Science
classes here at Harvard
including CS109A/B.

Fun Fact: Enjoys kaggle
competitions

Andrew Smith

Passionate about using machines
to model and assist the human
creative process

Fun Fact: Has produced concerts
on five different continents

Connor Capitolo

Machine Learning Engineer

Graduated from Master's in DS
program in May 2022

Fun Fact: Loves to go fly fishing

Who ?

Shivas Jayaram

Deep Learning Researcher, Educator and Practitioner

Working on medical-pharma knowledge platform startup

Fun Fact: He is a dog lover

Tale Lokvenec

Tale is a Founding Machine Learning Engineer at Zenlytic, specializing in natural language processing (NLP) and ML system design.

Fun Fact: He is a super basketball player and until the age of 16 he was in the national team of Macedonia.

Matthew Nazari

Deep Learning Researcher and Student. Developing online language learning spaces for the Northeastern Neo-Aramaic dialects

Fun Fact: You can concatenate ChatGPT onto a morphological parser to perform zero-shot translation on Neo-Aramaic

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

- Two lectures/tutorials per week.

- Projects

Lectures

Two lectures per week - What to expect

<div>

Pre-session Assignment (Reading)	Lecture & Tutorials

</div>
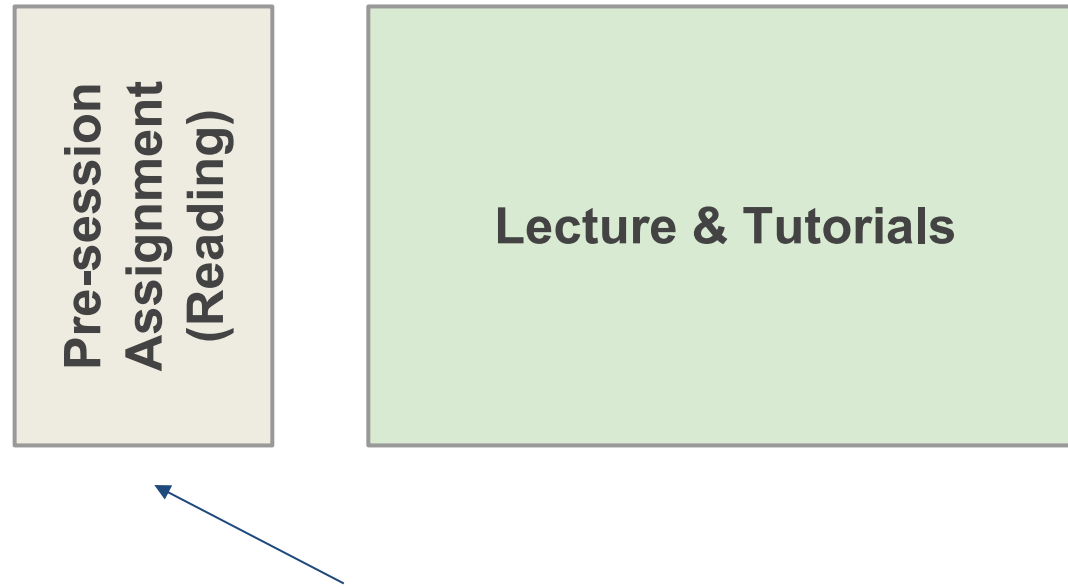
There will be one reading suggestion per week

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	09/14	5
MS2	MLOps & Advanced Training	09/26	10
MS3	Scalable Computing	10/05	15
MS4	Midterm Presentation	10/24	25
MS5	Full-Stack Development	11/14	10
MS6	Final Presentation and Deliverables	12/12	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 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

ED Stem

https://harvard-iacs.github.io/2023-AC215/

https://edstem.org/us/courses/42775/discussion/

Grades

Assignment	Final Grade Weight
Milestone 1	5%
Milestone 2	10%
Milestone 3	15%
Milestone 4	25%
Milestone 5	10%
Milestone 6	35%
Total	**100%**

Final Details

- We will be using ED for discussions, announcements and surveys
- Projects: Group

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

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 like to go to the forest to do mushroom picking. It is a fun activity and also rewarding as some mushrooms are edible. The problem is in the forest where Pavlos goes to pick mushrooms there are many varieties of poisonous mushrooms. Some of the mushrooms are obvious but there are some which he requires help in identification.

Proposed Solution

Pavlos will have his phone with him when he is in the forest. What if he could just take a picture of the mushrooms and and app could tell him what type of mushroom it is and weather it is poisonous or not

Project Execution Steps

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

How to Scope your Project
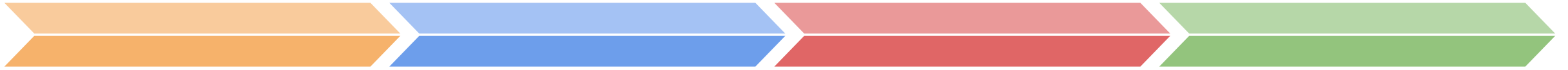
Proof Of Concept (POC)

- 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)

- Scrape mushroom data
- Verify images
- Experiment on some baseline models
- Verify new unseen mushrooms are predicted by the model(s)
- Visualize model activations to analyse 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

THANK YOU