CS249r: Responsible AI

Nov 13

Course Logistics

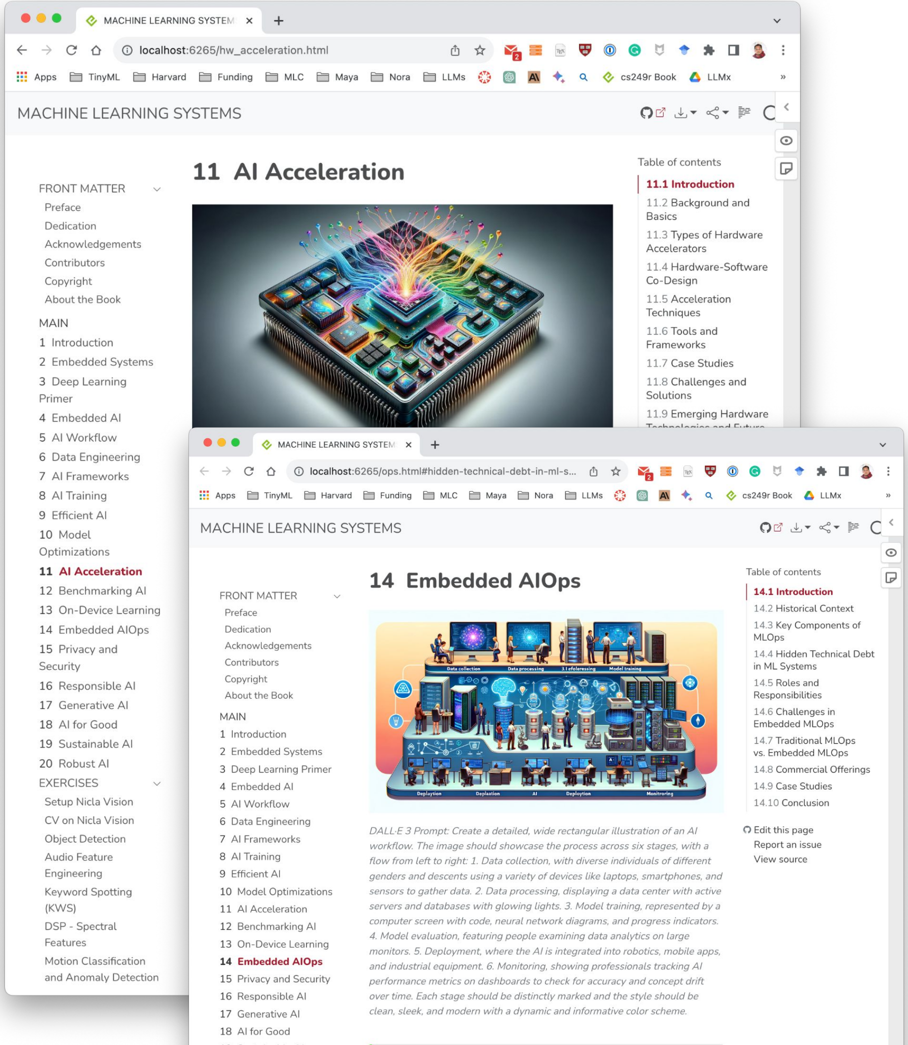
Assignment Schedule Updates

- Assignment 2
  - Due: October 23rd (Monday)
- Mid-Project Review
  - Due: October 30th (Monday)
- Assignment 3
  - Due: November 6th (Monday)
- ~~Assignment 4 Part 1~~
  - ~~Due: November 20th (Monday)~~
- ~~Assignment 4 Part 2~~
  - ~~Due: November 27th (Monday)~~
- Project Presentations
  - Due: December 4th (Monday)
- Final Report
  - Due: December 11th (Monday)

Part 1: Sustainability
&
Part 2: Responsible AI
—
Move to in-class activity!
(so no added HW)

Scribing

- Peer review - Generally, 1 detailed review
  - **MLOps** PR is ready
  - https://github.com/harvard-edge/cs249r_book/pull/57
  - Only 1 detailed review



MACHINE LEARNING SYSTEMS

FRONT MATTER
Preface
Dedication
Acknowledgements
Contributors
Copyright
About the Book
MAIN
1 Introduction
2 Embedded Systems
3 Deep Learning Primer
4 Embedded AI
5 AI Workflow
6 Data Engineering
7 AI Frameworks
8 AI Training
9 Efficient AI
10 Model Optimizations
11 AI Acceleration
12 Benchmarking AI
13 On-Device Learning
14 Embedded AIOps
15 Privacy and Security
16 Responsible AI
17 Generative AI
18 AI for Good
19 Sustainable AI
20 Robust AI
EXERCISES
Setup Nicla Vision
CV on Nicla Vision
Object Detection
Audio Feature Engineering
Keyword Spotting (KWS)
DSP - Spectral Features
Motion Classification and Anomaly Detection

11 AI Acceleration

Table of contents
11.1 Introduction
11.2 Background and Basics
11.3 Types of Hardware Accelerators
11.4 Hardware-Software Co-Design
11.5 Acceleration Techniques
11.6 Tools and Frameworks
11.7 Case Studies
11.8 Challenges and Solutions
11.9 Emerging Hardware Technologies and Future

MACHINE LEARNING SYSTEMS

FRONT MATTER
Preface
Dedication
Acknowledgements
Contributors
Copyright
About the Book
MAIN
1 Introduction
2 Embedded Systems
3 Deep Learning Primer
4 Embedded AI
5 AI Workflow
6 Data Engineering
7 AI Frameworks
8 AI Training
9 Efficient AI
10 Model Optimizations
11 AI Acceleration
12 Benchmarking AI
13 On-Device Learning
14 Embedded AIOps
15 Privacy and Security
16 Responsible AI
17 Generative AI
18 AI for Good

14 Embedded AIOps

DALL·E 3 Prompt: Create a detailed, wide rectangular illustration of an AI workflow. The image should showcase the process across six stages, with a flow from left to right: 1. Data collection, with diverse individuals of different genders and descents using a variety of devices like laptops, smartphones, and sensors to gather data. 2. Data processing, displaying a data center with active servers and databases with glowing lights. 3. Model training, represented by a computer screen with code, neural network diagrams, and progress indicators. 4. Model evaluation, featuring people examining data analytics on large monitors. 5. Deployment, where the AI is integrated into robotics, mobile apps, and industrial equipment. 6. Monitoring, showing professionals tracking AI performance metrics on dashboards to check for accuracy and concept drift over time. Each stage should be distinctly marked and the style should be clean, sleek, and modern with a dynamic and informative color scheme.

Table of contents
14.1 Introduction
14.2 Historical Context
14.3 Key Components of MLOps
14.4 Hidden Technical Debt in ML Systems
14.5 Roles and Responsibilities
14.6 Challenges in Embedded MLOps
14.7 Traditional MLOps vs. Embedded MLOps
14.8 Commercial Offerings
14.9 Case Studies
14.10 Conclusion

Edit this page
Report an issue
View source

Project Rubric

- Final Project Rubric - **45 points**
    - **December 4th - Presentation (~5-7 mins, ~3 min QnA): 10 points**
        - Reuse prior presentation and add in your new approach content

    - **December 11th - Write-up (4-5 pages max): 10 points**
        - Template for write-up here
        - Structure
            - Introduction
            - Background/Related work
            - Approach/Method
            - Insights/Findings
            - Contributions - who did what? **5 points**

    - **December 11th - Video (~2 mins max): 5 points**
        - Demo

    - **December 11th - Technical Deliverables - Code/Hardware: 15 points**

Project Expectations

Questions to answer

- Motivate your problem statement well, show **related work**
- Present your **analysis** - systematic thinking through approach/methodology and discussion of why it did or didn't work
- Completion vs. Work in progress
    - Technical challenges
    - Theoretical challenges

Project Check-Ins

Check-ins

- Not mandatory, but **strongly encourage** you to meet with the TAs to discuss the status of the project
- Check-in this week and/or next week

Who to check-in with

- TA: Meet with TAs for informal updates next week (nothing to prepare)
- VJ: Extended office hours on Monday from 3:30pm to 5pm

Check List

Scribing

- 14 people completed Peer Reviewing so far
- Check canvas Peer Reviewing assignment

Paper Group Leads

- Check canvas for your grades and adjust if necessary

Assignments

- 1 and 2 (done)
- 3 (in progress, EOD)
- ~~4~~

Final Projects

- Grading breakdown (slide 5)

Participation/Attendance - check canvas

Course Topics

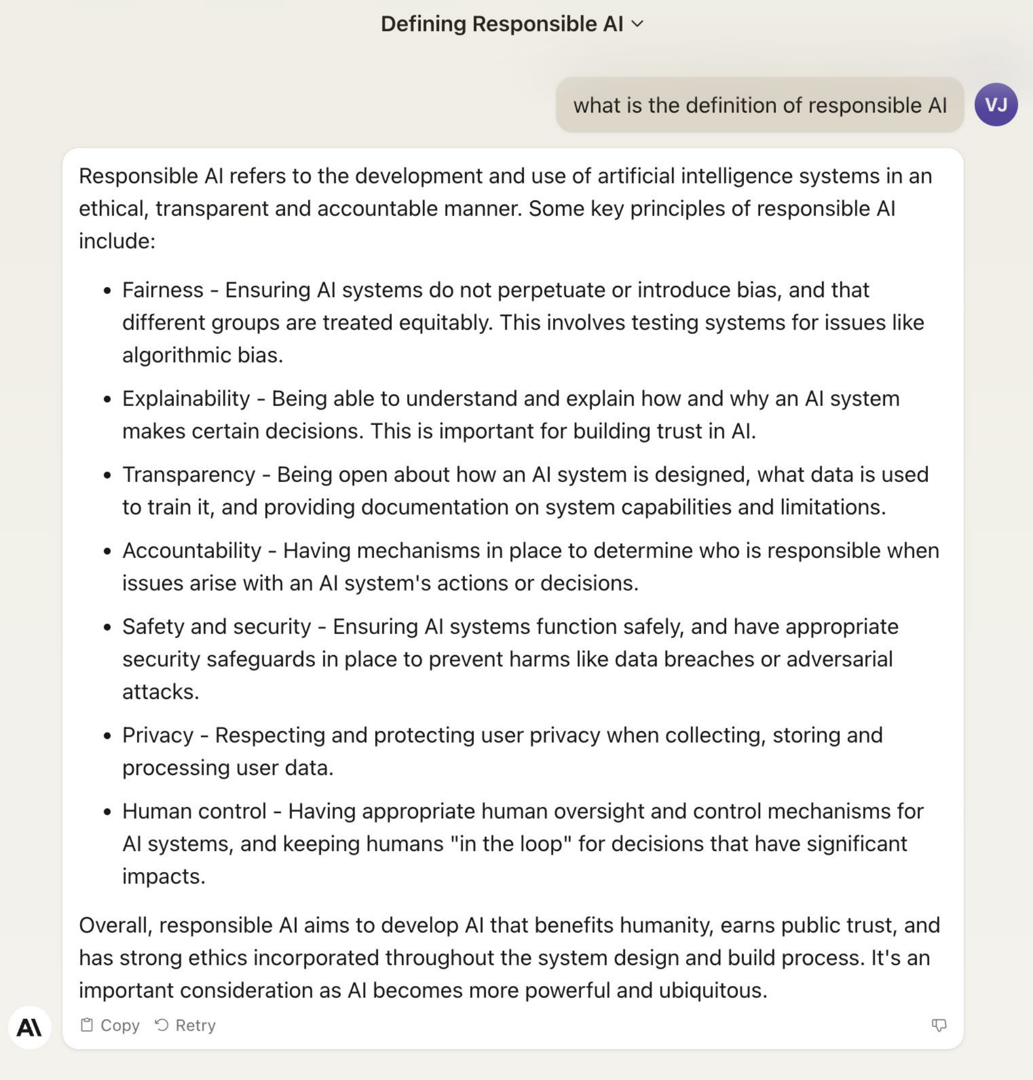
1. Overview and Introduction to Embedded Machine Learning
2. Data Engineering
3. Embedded Machine Learning Frameworks
4. Efficient Model Representation and Compression
5. ~~Performance Metrics and Benchmarking of ML Systems~~
6. Learning on the Edge
7. Hardware Acceleration for Edge ML: GPUs, TPUs and FPGAs
8. Embedded MLOps
9. Secure and Privacy-Preserving On-Device ML
10. **Responsible AI**
11. Sustainability at the Edge
12. Generative AI at the Edge

Responsible AI

Why are talking about responsible AI when we are in a systems class?

We will be talking about responsible AI in the context of generative AI systems (T2I image models)

Generative AI at the edge is coming, the next year or two will be dominated by edge AI



what is the definition of responsible AI

Responsible AI refers to the development and use of artificial intelligence systems in an ethical, transparent and accountable manner. Some key principles of responsible AI include:

- Fairness - Ensuring AI systems do not perpetuate or introduce bias, and that different groups are treated equitably. This involves testing systems for issues like algorithmic bias.
- Explainability - Being able to understand and explain how and why an AI system makes certain decisions. This is important for building trust in AI.
- Transparency - Being open about how an AI system is designed, what data is used to train it, and providing documentation on system capabilities and limitations.
- Accountability - Having mechanisms in place to determine who is responsible when issues arise with an AI system's actions or decisions.
- Safety and security - Ensuring AI systems function safely, and have appropriate security safeguards in place to prevent harms like data breaches or adversarial attacks.
- Privacy - Respecting and protecting user privacy when collecting, storing and processing user data.
- Human control - Having appropriate human oversight and control mechanisms for AI systems, and keeping humans "in the loop" for decisions that have significant impacts.

Overall, responsible AI aims to develop AI that benefits humanity, earns public trust, and has strong ethics incorporated throughout the system design and build process. It's an important consideration as AI becomes more powerful and ubiquitous.

Today’s Schedule

1245pm to 1pm

- Course logistics and setting up the context

1pm to 130pm

- Jessica will introduce gen AI, diffusion models etc. and setup MLC accounts in class with everyone

130pm to 2pm

- Paper discussion

2pm to 3pm

- Lora's lecture

3pm to 330pm

- In class nibler hacking

Lora Aroyo

Lora Aroyo is research scientist at Google Research, NYC where she works in the area of Responsible AI specifically focussing on research for Data Excellence, e.g. metrics and strategies to measure quality of human-labeled data in a reliable and transparent way.