Responsible Deep Learning: Part 2

Danna Gurari
University of Colorado Boulder
Spring 2024

https://dannagurari.colorado.edu/course/neural-networks-and-deep-learning-spring-2024/

Review

- Last lecture:
  - Deep Learning that Discriminates
  - FAT (Fair, Accountable, & Transparent) Algorithms
  - Ethics in Deep Learning

- Assignments (Canvas):
  - Final project presentation due in 1.5 weeks

- Questions?

Today: 2-Stage Activity

- STAGE 1: Can you see it coming?
  - Choose a deep learning application or use case (can not yet exist)
  - Ask yourselves what could go wrong?
    - E.g., bad actors, vulnerabilities, bias, large-scale societal impacts?
    - This outcome could be either tomorrow or in the future (e.g., in 50 years).
  - What might be effects from this problem (e.g., to individuals, communities, society?)

- Each group should write the following:
  - Group member names
  - What is your application or use case?
  - What goes wrong?
  - What is the result?
  - A news headline based on what you came up with

Today: 2-Stage Activity

- STAGE 2: Can you get ahead of it?
  - Discuss the other assigned group's example of something that can go wrong.
  - Ask yourselves what could have prevented this?
    - e.g., technical methods, regulation social structures?
    - If prevention isn't feasible, what could be done now?
    - If it feels like a solution isn't possible, then address why.
    - Feel free to suggest multiple possible strategies or solutions!

- Each group should write the following:
  - What's your solution?


The End