DATA 605 Ethical & Legal Issues in Data Science

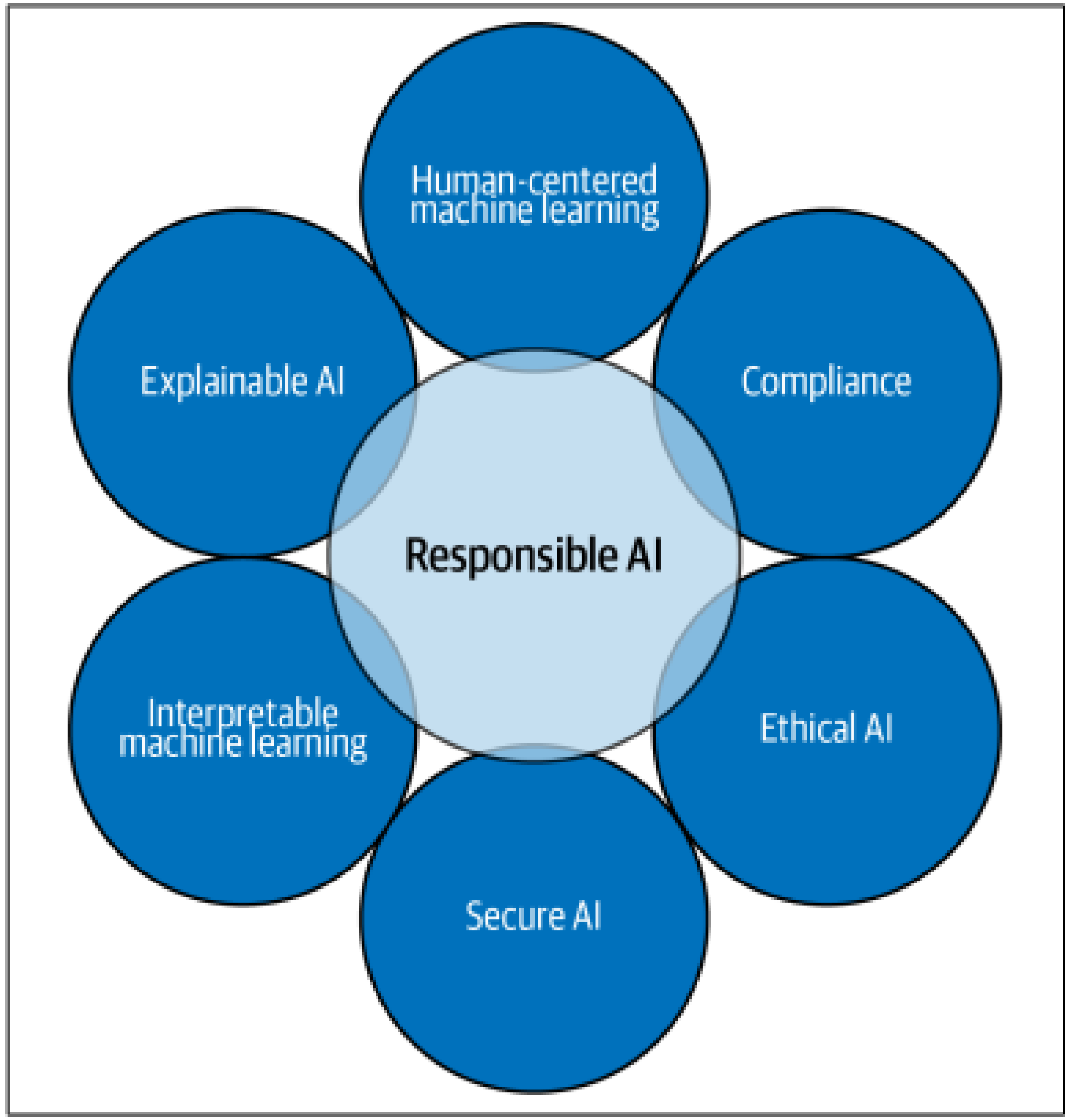
SPRING 2022

SUNELA THOMAS

APRIL 14, 2022

AGENDA

- Questions?
- NO “live” class on April 21st ; materials and online graded class discussion topic will be available on April 21st
- Responsible Machine Learning
- Breakout
- Team Assignments for Group Presentation



What is Responsible Machine Learning?

Responsible AI

Ethical AI Sociological fairness in ML predictions (i.e., whether one category of person is being weighed unequally or unfavorably)

Explainable AI The ability to explain a model after it has been developed

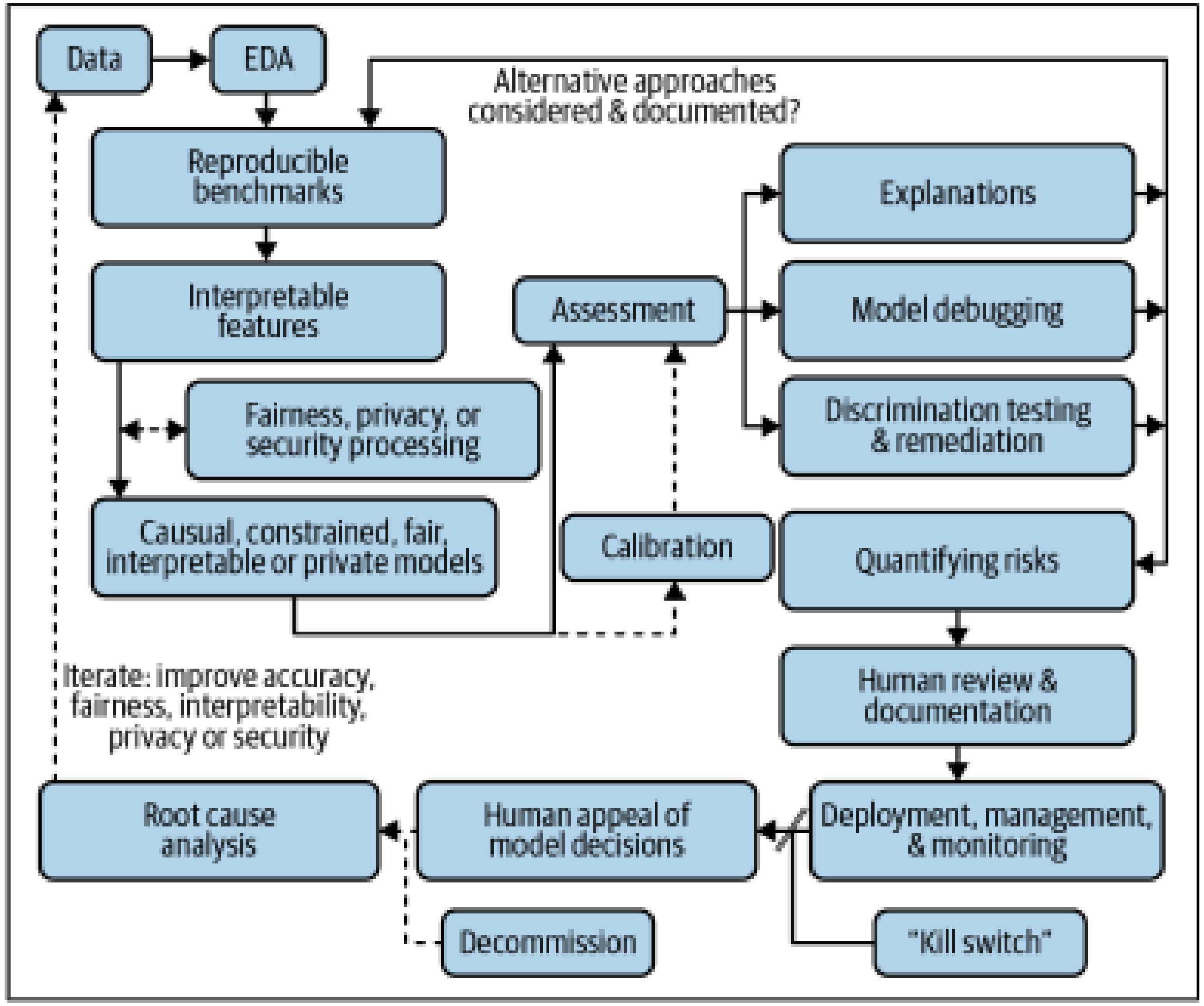
Human-centered machine learning Meaningful user interactions with AI and ML systems

Interpretable machine learning Transparent model architectures and increasing how intuitive and comprehensible ML models can be

Secure AI Debugging and deploying ML models with similar counter measures against insider and cyber threats, as seen in traditional software

Compliance Aligning your ML systems with leading compliance guidance such as the EU GDPR, the Equal Credit Opportunity Act (ECOA), or the US Federal Reserve's SR 11-7 guidance on model governance

Responsible Machine Learning Workflow - Sample



Principles for Responsible ML

- Human Augmentation
- Bias Evaluation
- Explainability by Justification
- Reproducible Operations
- Displacement Strategies
- Practical Accuracy
- Trust by Privacy
- Data Risk Awareness

Responsible Machine Learning Culture

Accountability

Dogfooding

Demographic and Professional Diversity

Cultural Effective Challenge

Going Fast and Breaking Things

Get in the Loop

Human Audit of ML Systems

Domain Expertise

User Interactions with ML

User Appeal and Operator Override

Kill Switches

Accountability

Who tracks the way ML is developed and used at my organization?

Who is responsible for auditing our ML systems?

Do we have AI incident response plans?

Typically answer????

if an organization assumes everyone is accountable for ML risk and AI incidents, the reality is that no one is accountable.

Dogfooding

Is a term from software engineering that refers to an organization using its own software, i.e., “eating your own dog food.”

Brings an additional layer of alpha or pre-alpha testing that is often neglected in the mad dash to profit from a perceived ML gold rush.

If an organization has developed an ML system that operates in a manner that, say, violates their own privacy policies, or is meant to be deceptive or manipulative, employees engaging in dogfooding might find this objectionable and raise concerns.

Diversity

Demographic Diversity

Professional Diversity

Developing teams with deep cross-disciplinary professional experience can be invaluable as you look to deploy ML

Involving oversight professionals from the beginning is a great way to assess and mitigate the risks

Cultural Effective Challenge

When building complex ML systems, effective challenge roughly says that one of the best ways to guarantee good results is to actively challenge and question steps in the ML development process

A culture that encourages serious questioning of ML design choices will be more likely to catch problems before they balloon into AI incidents

Going Fast and Breaking Things

The mindsets of many top engineers and data scientists

Practitioners must recognize the implications and downstream risks of their work instead of racing towards results for an outdated maxim

Get in the Loop

Concrete steps practitioners or managers can take to get more control over ML systems

Human’s detailed review of ML systems

Staple for model governance – inventories and documentation

Without domain expertise, ML systems can be trained on incorrect data, results can be misinterpreted, audits are less meaningful, and data or programming errors may explode into full-blown AI incidents

Human Audit of ML Systems

Google has put forward a framework for ML model audits

Sample documentation for models and data

What can you and your organization do to promote human audits of ML systems?

- Create an inventory of ML systems
- Nominate accountable executive(s)
- Instate executive and technical review of documented ML systems
- Require technical and executive sign off before deploying ML systems
- Carefully document, validate, and monitor all ML systems

Domain Expertise

Real-world success in ML almost always requires some input from humans with a deep understanding of the problem domain

Experts can also serve as a sanity check mechanism

For instance, if you're developing a medical ML system, you should consult physicians and other medical professionals.

User Interactions with Machine Learning

For maximum impact, nontechnical and decisionmaker users need to understand and act on ML system results

When constructing ML systems, it is wise to con- sider the different types of users and personas who will need to interact with the system

User Appeal and Operator Override

What if a computer unjustly kept you in prison?

What if a computer erroneously accused you of a crime?

What if a computer kept you or a loved one out of the college of your dreams?

Steps you can take to prevent your organization's ML systems from making unappealable, and potentially illegal, black-box decisions:

- Use of interpretable ML models or reliable post-hoc explanation techniques (preferably both)
- Proper documentation of the processes used in these systems
- Meticulous testing of ML system interpretability features before deployment

Kill Switches

If your ML system goes seriously wrong, you will want to be able to turn it off fast

ML systems should be monitored for multiple kinds of problems, including inaccuracy, instability, discrimination, leakage of private data, and security vulnerabilities.

Breakout

BIAS IN IMAGE DATA

Additional Resources

https://berryvilleiml.com/interactive/

https://towardsdatascience.com/responsible-machine-learning-with-error-analysis-a7553f649915

https://www.weforum.org/agenda/2021/03/responsible-machine-learning-that-protects-intellectual-property/

GROUP PRESENTATION

- Groups are assigned (refer to next slide)
- Criteria:
  - Select your own topic for presentation – something that has an ethical issue in data science (e.g., can ads be banned in a browser, can genetic data be shared for analysis, does ethics differ in cultures, etc.)
  - Formal presentation – 10 minutes
  - Everyone in the team participates
  - Presentation to include:
    - Cover page – title and team members listed
    - Problem Statement/Summary
    - Ethical Issues & relation to the theories learned
    - Proposed Solution
    - References, if any
  - Copy of the presentation will be due to me on May 11th by 11:00pm ET
  - Live presentation to the class on May 12th

GROUP PRESENTATION – Team Assignments

Team #1	Team #2	Team #3	Team #4
Soumya Kasireddy	Carol Kingori	Sai Sridhar Nenavath	Jael Kruthi Battana
Pavan Chinthakunta	Daniel Rimdans	Sravani Ravulaparthi	Tahereh Hematian Pour Fard
Sai Gangadhar Veeramreddy	Chanakya Polisetty	Sahithi Veeranki	Showri Yeruva
Sai Krishna Jakkampudi			

Team #5	Team #6	Team #7	Team #8
Nidhishree Sanam	Shiridinath Konduru	Lavanya Telapudi	Tarun Eswar Reddy Vuyyuru
Saketh Reddy	Yaswanth Reddy Annapureddy	Harshini Akkapally	Lokesh Katuri
Jaspreet Singh Bhatia	Victoria Borsetti	Prashanthi Ponakalla	Chandralekha Bhaviri