Live Session 6: Safety, Security and Accountability

Session will begin at 10:00

Dr Zak Varty

MLDS ETHICS PART I

Welcome!

This Session:

1. Review and Teasers (5-10 mins)

2. Group Activity - Writing Model Cards & Feedback from Groups (15 mins + 10 mins)

3. Comparison to Example and Templates & Feedback from Groups (15 mins + 10 mins)

Review and Teasers

Review

When putting models into prodution there are a lot of challenges to overcome:

- **User error**
- **Model generalisation**
- **Model misuse**
- **Model dating**
- **Model comparison**

Review

We can learn a lot from best practices in software engineering.

- **User error** Documentation, input checks, fail loudly, unit tests.
- **Model misuse** Model Cards, documentation of training data.
- **Model dating** Document dependencies, control charts, covariate shift / concept drift.
- **Model comparison & generalisation** Shadow deployment, A/B testing, adaptive experimental design.

Can you think of any more?

Writing model cards

Learn by doing - write your own

A lot of problems can be solved by writing documentation, but writing good docs is hard.

Activity

Working as a group, construct a model card for the following ML projects.

Focus on the structure, make up any details you need. (15 mins)

1. Anomaly detection in medical images.

2. Real-time anomaly detection to identify financial fraud.

3. Personalised product recommendations for an online retailer.

4. A ChatBot on the hompeage of an online education company.

5. Predictive maintenance of mechanical components in a factory.

Model card review

What did you include in your model card and why?

1. Anomaly detection in medical images.
2. Real-time anomaly detection to identify financial fraud.
3. Personalised product recommendations for an online retailer.
4. A ChatBot on the hompeage of an online education company.
5. Predictive maintenance of mechanical components in a factory.

Reviewing Model Cards

Reviewing Model Cards

Activity

Compare your model card & the example data sheets provided to those included in Mitchell et al (2019).

- What does each do well?
- What would you change about both?

Extension

Compare the structure of your model cards to the templates used by huggingface.

- Model Card Template / Data Sheet Template
- What does each do well?
- Why is a template helpful?

Feedback from Groups

Comparison to Mitchell et al and HuggingFace templates.

What does each do well? What would you change?

1. Anomaly detection in medical images.
2. Real-time anomaly detection to identify financial fraud.
3. Personalised product recommendations for an online retailer.
4. A ChatBot on the hompeage of an online education company.
5. Predictive maintenance of mechanical components in a factory.