Main Goal of Interview:

Understand ML Interpretability <u>Needs</u>, <u>Challenges</u> and <u>Practices</u> in <u>Industry</u>

Outline

1. Your projects and goals
2. Your team
3. Interpretability (when/why)
4. Interpretability (what)
5. Interpretability (who)
6. Interpretability (how)
7. Wishlist (dreaming big)

Estimated time: approximately 60 mins
Interview is recorded, ok?

Projects and Goals

Projects + goals

- What problems are you trying to solve?

Data

- What kind of data? (type, volume, dimensionality)

Models

- What kind of models (NNs vs. traditional)?

Team

Size of your team/roles/expertise
Communication/workflow

Interpretability (When/Why)

What role does it play in your projects?
When/why does interpretability matters?

What do you aim to achieve?
What are the biggest challenges?

Examples of specific projects with interpretability needs? (discuss 2-3 if possible)

Interpretability (Purpose)

Model comparison/debugging/performance
E.g., Verification model with domain experts

Model auditing
E.g., Ethical model, bias detection, legal compliances

Improving products
E.g., Providing explanations to end-users

... what else?

Interpretability (What Type)

What type of interpretability do you use/need in your workplace?

- Global vs. local
- Black-box (NN) vs. white-box (decision tree)
- Inherently interpretable models

Accuracy vs. interpretability

When/why isn't accuracy enough?

Interpretability (Who)

Who needs interpretability (for what reason)?

- Data scientists
- Domain experts
- Managers
- End-users
- External regulators
- … who else?

How do these people communicate? Challenges?

Interpretability (How)

How do you currently deal with interpretability needs using current methods, tools:

- Existing methods (LIME, SHAP, etc.)
- Surrogate models
- Intrinsically transparent models
- … What else?

Challenges?

Wishlist

What kind of solutions would solve the problem you have?

If you could "dream big", what kind of solutions would you like to have?

Anything else?

Before we conclude, is there anything else you would like to share with us?

Is there anything you think we should investigate which we did not discuss?

Thank you!