Suggest having a brief “4.0 Introduction” that introduces and lays out the arc of the module

4.1 Explainable ML

All the examples are nice and straightforward, everything from explanations for ads to the concrete approaches to model explanation, but I would just suggest ensuring that this discussion of explainability is explicitly linked to cybersecurity. Right now—and I may be misinterpreting, since this is based on your slides not your actual presentation—the initial framing is that explainability is a desideratum because they’ll accountable for explanations due to the GDPR, the press, etc.

One question would be whether it’s possible to make more explicit (and, forgive me, you may plan to do this in your presentation) the considerations that drive the need for such accountability (for example that there are regulations in place and one may be pilloried in the press for good reasons, some of which may have to do with cybersecurity or just good practices in machine learning etc.).

Slide 4, Potential Audiences: small note, I’m thinking the “third parties” may be the most interesting category to this audience—not journalists maybe, which is the example you give, but their bosses and the lawyers and risk mitigation folks within their organizations

Would “Attackers” be a potential audience to include here (not an intended audience, but one they would need to consider when crafting explanations)

Slide 8, Models That Are Intuitively Explainable: not spelled out yet but I’m assuming the distinction is going to be between e.g. linear regression and models that function as a black box?

4.2 Fairness

Slide 13, What Would You Do: Love these business applications/what would you do slides in your decks. How are you planning on approaching this? Talking through the considerations that you would have? We could e.g ask students to post their thoughts to a discussion forum, or to come prepared to discuss in the live sessions

4.3 Privacy

Have a similar question about framing as in 4.1—can this discussion about privacy be initiated by orienting it in the context of cybersecurity? What are the stakes involved, or to put another way, why should this audience care?

4.4 Cleaning Missing Data

The arc of the module is a little unclear just reviewing the PowerPoints (again, you may plan to flesh it out orally). The movement from Explainability🡪Fairness🡪privacy🡪Cleaning data etc. could use some explanation. Also, would it make sense for 4.5 (data collection) to proceed cleaning missing data?

Any examples here, real or hypothetical, would be great if possible.

Would it make sense for slides 5f (“Ideal Process”) to precede the “questions to ask” (slides 2f)?

I could see this one lending itself to a great hands-on exercise

4.5 Data Collection

Slide 9, Human-Subject Data for Cybersecurity ML: Just noting, these seem like harder and more general questions than in previous decks, which are introduced with a hypothetical. Might introduce these things to consider, and then apply more specific questions to a hypothetical.

I’m just thinking about ways to help students move from the concrete examples that you provide, to a general framework that they can use themselves, to practicing that framework on an example.

4.6 Anonymization

I believe these failures all relate to cases where data that was anonymized proved less anonymous than presumed. For this crowd, might be useful to discuss in some detail what this might mean for their models. I.e., are these failures of anonymization only a concern if the underlying data may be released? Or should they be thinking about these considerations regardless (e.g., if they’re just training and deploying a model, with no intent to release the data)?

Slides 7-10: Is there a way to present in synopsis what these different anonymization techniques are, and perhaps where the tradeoffs are or in what use cases each may be most applicable (understand that it may not be possible, if this will always vary by dataset and problem)

Slides 14-18: As noted during the connect, these images are somewhat pixelated.

Slide 20, Anonymity’s Interaction With Cybersecurity ML: same thought as end of 4.5—might consider giving these considerations, and then giving a hypothetical scenario in which they can be entertained in more concrete terms

4.7 ML Data Lifecyles

Slides 16-19: Are there best practices, rules of thumb, general standards, etc. that can be given for some of these things?

To whatever extent feasible, it would be great to tie together and summarize the different parts in this module—either briefly at the end of this last video, or in another video.