Overall, my biggest takeaway was that ethical decision making shows up in the gray areas for example, how you frame your question, how you decide what counts as evidence, and how you respond when results are messy or inconvenient. I also took the idea that professionalism and ethics are tied together because the way you communicate, document, and collaborate affects more than just you, it can also affects your advisor, your lab, and the researchers who might rely on your results later.

This can tie directly into my thesis on a unified benchmark, where the whole point is improving reproducible benchmarking for multi agent reinforcement learning in multi robot settings. Even though, my thesis is computational and simulation based with no human subjects, but the ethical stakes are still real because benchmarking can quietly shape what other researchers and labs believes is "reliable" or "state of the art." If I cherry pick my metrics, hide failure cases, overtune to a single task, or leave out important configuration details, I am not just making my thesis weaker, I am feeding the exact reproducibility gap my thesis aims to fix. Thus, I want to treat transparency as a core deliverable, not just like an afterthought. In doing so, I plan to lock in evaluation metrics early, keep clean experiment logs across seeds, and publish configurations and code in a way that researchers outside my lab could actually rerun and replicate the results. I also want to be more intentional about how I describe extra factors like communication reliability and observation limits, since the framework is designed to vary those conditions in a controlled way. Additionally, if my thesis is trying to give guidance on which methods hold up under realistic constraints, I need to make sure my process matches that standard.

I also noted that ethics is not only about avoiding obvious misconduct, it is about designing your process so you do not accidentally fool yourself. In benchmarking, that means committing to

evaluation choices early and sticking to them, even when the results are not what I hoped for. My proposal already lays out the direction I want to follow, performance metrics like episodic return and success rate, plus safety metrics like collisions, and then robustness plots as communication reliability and observation consistency change. To make that ethical commitment real, I am going to treat those metrics as the contract, run multiple random seeds, and report uncertainty instead of only showing the best run, documenting everything I do with clear commented code. I also want my final deliverables to reflect that same standard by documenting task configurations, hyperparameters, dataset generation procedures, and evaluation scripts so someone else can reproduce what I did without guessing.

Furthermore, despite the fact that my thesis is mostly software and simulation, there are still plenty of moments where it would be easy to blur lines, like what counts as a meaningful improvement, how to present negative results, or how to divide credit when code and ideas are shared across a lab. Since my benchmark is meant to be modular and eventually released publicly with documented task configurations, algorithm settings, dataset generation procedures, and evaluation scripts, I want to treat that release as a responsibility, not just a nice portfolio add on. That means keeping my Git history clean, writing clear experiment notes, and being upfront with my chair and committee about what is working, what is not, and what assumptions I am making. It also means building in simple safeguards like unit tests for tasks, consistent evaluation pipelines, and repeatable configuration files, because those are the boring details that prevent accidental mistakes from turning into misleading claims later.