1) What classification scheme does Awad use to distinguish “types” of AI, and why do these types matter for scientific research?

Awad says we shouldn’t think of AI as one single thing. Instead, AI is more like a group of different tools that do different jobs in research like a “toolbox” with different kinds of tools inside it . She groups AI into three main categories: foundational paradigms, decision-oriented paradigms, and cross-cutting principles.

The foundational paradigms are the “core” kinds of AI, like:

- Descriptive AI (finding patterns in lots of data),
- Predictive AI (predicting future outcomes based on past data),
- Generative AI (creating new outputs like text, images, or even possible hypotheses)

Then there are decision-oriented paradigms, like optimization AI, which helps choose the best experimental setup when there are many options and constraints . Finally, there are cross-cutting principles like explainable AI, causal AI, and privacy-aware AI, which are meant to make AI results more understandable, reliable, and ethical .

Micro-claim 1: These categories matter because research happens in stages, and different stages need different kinds of AI support . For example, if a scientist is doing genetics research and wants to discover groups in messy data, descriptive AI is useful.

Micro-claim 2: Having different “types” of AI lets scientists use the right method for the right job like using predictive AI for modeling climate patterns, or using optimization AI to reduce time and cost in experiments.

2) Does Awad clearly distinguish AI as a tool vs AI as a scientific collaborator?

Yes. Awad makes a clear difference between AI as a **tool** and AI as more of a **collaborator**

When AI is a tool, it mainly helps humans do tasks faster like classifying data, finding patterns, or summarizing information. It’s like a powerful calculator or a helpful assistant that follows instructions. But Awad says there is a newer shift toward AI acting more like an “epistemic agent,” meaning it doesn’t just follow steps it can help decide what steps to take next

Awad suggests AI is starting to move from “doing tasks” to “helping shape the research process,” which can affect what scientists treat as evidence and how conclusions are made. For example, autonomous labs (like the one she mentions) can use reinforcement learning to improve experiments over time with less human direction.

Systems that can plan, adapt, and make choices based on results are closer to collaboration than tools, because they influence the direction of research. Awad also mentions multi-agent systems and LLM-based setups that can help generate hypotheses in scientific fields.

Is this a real shift or mostly an extension? Awad argues it's a deeper shift. Some uses (like older machine learning models) feel like an extension of existing methods, but "agentic" systems can change how science is done because they can generate ideas and decide next steps.

A skeptic could say: "This is still just advanced statistics and automation." But Awad's point is that when AI starts generating hypotheses or finding connections in ways humans can't easily explain, the process starts to feel different from normal human-led research.

3) What limitations/risks does Awad discuss? (Interpretability, bias, reproducibility, theory formation)

Awad talks about several risks of using AI in science, especially when the models become very complex.

One major issue is interpretability. Some AI models (like deep learning systems) can give strong results, but they can be hard to explain. Awad describes how these models can act like a "black box" when you see the output, but it's difficult to understand how the model got there.

If scientists can't explain the reasoning behind an AI's output, it becomes harder to defend results during peer review and harder for other researchers to trust the conclusions. This also affects reproducibility, because if the system is not transparent, repeating the result becomes harder.

Awad also highlights bias, especially in AI that works with human language (NLP). These systems can absorb stereotypes from training data, which can affect scientific conclusions in social science, medicine, and other areas.

Bias in AI can distort what counts as "evidence," because biased outputs can push research in unfair or inaccurate directions.

Finally, she discusses risks for theory formation, especially with generative AI. Models can produce confident-sounding text that is wrong. She mentions Galactica, which was withdrawn after it generated misleading claims and fake citations. If researchers rely on this kind of output, they might build explanations or theories on false information, which weakens trust in science.

4) Is AI more likely to accelerate discovery or reshape the scientific method?

Awad suggests AI will do both, but she leans toward the idea that AI could reshape how science is done, not just speed it up. I agree with that.

AI doesn't only make research faster; it can change the structure of research by turning science into a continuous feedback loop, like "closed-loop systems" where models and experiments keep updating each other . That's different from the traditional idea of science being a mostly human-driven process with separate steps.

When AI helps generate hypotheses, guide experiments, and influence what scientists test next, it starts affecting the method itself, not just the speed of results. For example, if AI chooses the next most useful experiment using Bayesian optimization, it's doing strategic decision-making that used to belong mostly to humans. That means the scientific method becomes more "machine-augmented," not just "human-led with tools."

What I learned

I learned that "agentic AI" means AI systems that can plan and make choices (not just follow instructions), which makes them closer to collaborators than simple tools.

What I'm still uncertain about

I'm not sure how scientists can create standard rules for verifying AI-based conclusions when even experts may not fully understand how some models produced their answers.

One question I'd ask the author

Since models like Galactica can "hallucinate" and sound confident while being wrong, should there be a special review process for AI-generated hypotheses before they get treated like real scientific claims?