Vol.9 No.1 (2026)

Journal of Applied Learning & Teaching

ISSN: 2591-801X

Proudly owned and sponsored by Kaplan Business School, Australia

Content Available at: **https://jalt.open-publishing.org/index.php/jalt/index**

When AI Joins the Lab: Reflections on ChatGPT in Research and Graduate Education. An Interview with Professor Timothy Mitchison.

Timothy Mitchison[A] [A]*Harvard Medical School, USA*

Yulu Hou[B] [B]*Michigan State University, USA*

Keywords

ChatGPT;
doctoral supervision;
generative AI;
graduate education;
higher education practice;
teaching and learning.

Correspondence

houyulu@msu.edu[B]

Article Info

Received 23 September 2025
Received in revised form 19 November 2025
Accepted 20 November 2025
Available online 12 February 2026

DOI: https://doi.org/10.37074/jalt.2026.9.1.3

Abstract

This interview with Professor Timothy Mitchison from a leading U.S. medical school explores his evolving use of generative artificial intelligence (AI), particularly ChatGPT, in research and doctoral education. The conversation examines how he incorporates AI into literature searches, hypothesis generation, experimental design, coding support, and laboratory communication, while also noting persistent limitations such as inaccurate references and conceptual errors.

Beyond technical applications, TM reflects on how AI reshapes cognitive effort, writing practices, and the dynamics of graduate supervision. He highlights both the risks of cognitive offloading, where students may bypass deeper engagement with learning, and the potential for AI to support creativity, accelerate routine tasks, and foster collaborative dialogue between supervisors, students, and AI tools. The discussion also addresses broader questions of knowledge creation, mentorship, and the role of AI in shaping higher education. By situating AI adoption within the everyday practices of scientific research and graduate training, this interview offers valuable insights for academics, supervisors, and students navigating the promises and pitfalls of integrating AI into teaching and learning.

About the Conversation

Timothy Mitchison (TM) is a professor of systems biology at a leading U.S. medical school. His laboratory focuses on understanding fundamental mechanisms of cell biology, with a particular emphasis on microtubules, intracellular transport, and mitotic regulation. TM has contributed significantly to the fields of cell biology. Alongside his research, he plays an active role in mentoring PhD students and contributing to curriculum development in biomedical graduate education. In recent years, he has also developed a personal and professional interest in the use of generative AI tools in scientific research and education.

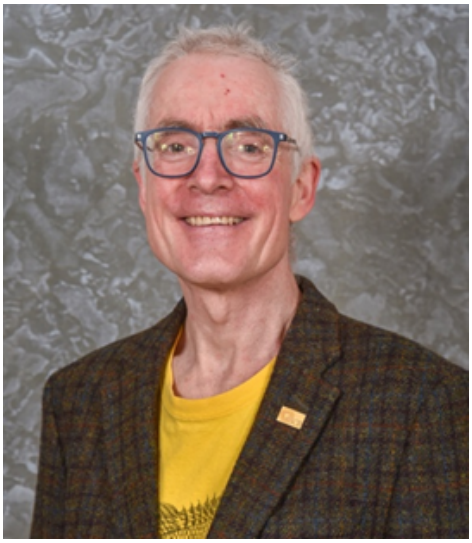

Figure 1: Professor Timothy Mitchison.

Yulu Hou (YH) is a doctoral student in higher and adult education. Her research explores the intersection of technology, identity, and knowledge creation in graduate education, with a particular interest in how scholars and students engage with AI tools. She is also involved in collaborative studies on doctoral education, international student experiences, and researcher development.

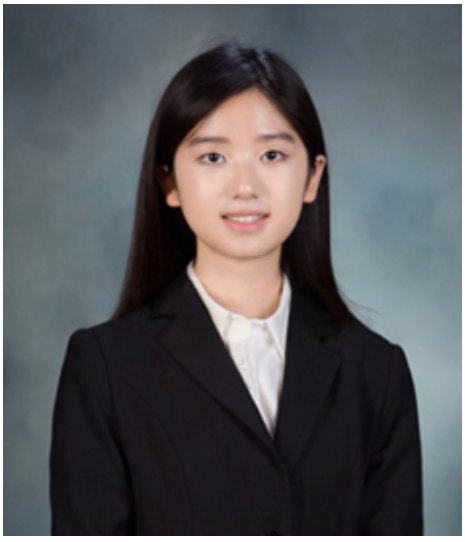

Figure 2: Yulu Hou.

In this interview, YH invited TM to reflect on his evolving use of ChatGPT in his scientific and supervisory work. The conversation, held in person in July 2025, explored how TM integrates generative AI into literature searches, experimental design, mentorship, and academic writing, while also raising broader questions about cognition, reasoning, and co-creation. Drawing on TM's disciplinary background in biology and YH's perspective from education research, the conversation also probed the boundaries of AI assistance and human–machine collaboration in graduate training. The transcript has been edited for clarity and flow.

Professional Background

YH: Could you please briefly describe your roles at your university?

TM: Absolutely. I'm a professor at our medical school. Historically, the school was focused on teaching medical students, but over the years, faculty like me have been doing less and less of that. In fact, I'm not involved in medical education at all anymore. That's now handled by dedicated medical education professionals.

My job today primarily revolves around supervising a research program, which is mostly funded by external government grants. I also play a voluntary role in PhD education. The most important part of that is mentoring PhD students who are doing their thesis research in my lab. Beyond that, I've participated in PhD program leadership, helping to organize programs and directing courses in the biomedical area. So overall, I would say my work is about 95% research and 5% teaching.

YH: Could you remind me when you joined this university?

TM: I came here in 1997, so it's been roughly 25 years now. Before that, I was a professor at UCSF.

Meeting ChatGPT for the First Time

YH: When did you first start using generative AI tools like ChatGPT?

TM: It's funny—my first realization of the power of deep learning came from a board game. I'm a fan of Go (called Weiqi in Chinese), and I remember the famous 2016 match where DeepMind's AI beat the top Korean player. That was a turning point for me. Go had always seemed like something the human brain could do but computers couldn't. So seeing an AI win was a signal that everything was about to change.

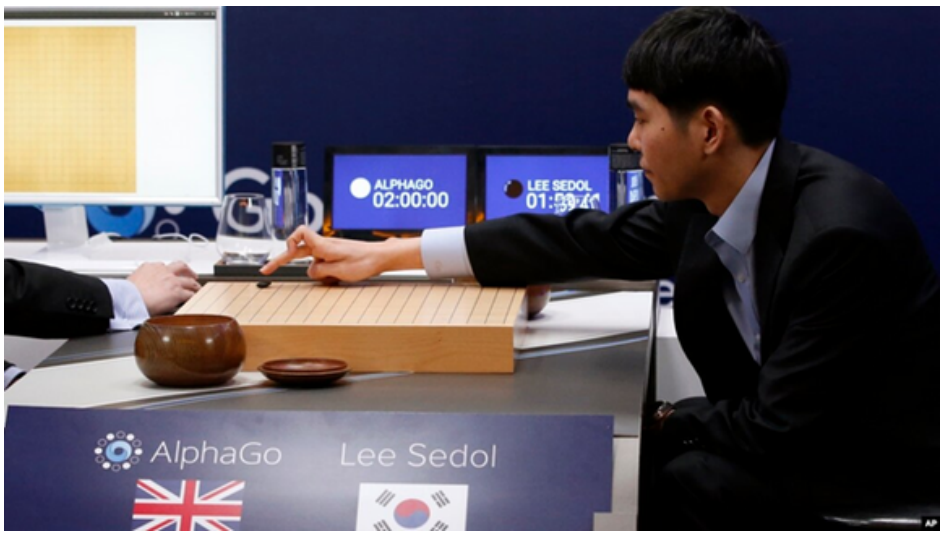

Figure 3: South Korean Go champion Lee Sedol, right, places the first stone against Google's AI program AlphaGo (VOA, 2016).

But I didn't start using generative AI myself until much later, when it became available to the public. That must have been in 2022, when GPT-3.5 launched and was free to use. Like many people, I went to the website and tried it out with some research-related questions, and I was immediately amazed. It was extraordinary: being able to have a conversation with a computer that felt like it was thinking.

Like most professors, one of the first things I did was ask it to generate literature citations. In research, one of the main ways we use the internet is to find out what's already known about a topic. We're always asking: What drugs have people studied? What do we know about this pathway?

So I asked ChatGPT to provide citations—and it did. But they were fake. They looked real, with plausible-sounding j-

-ournal names and author lists, often with Chinese names, but when I tried to look them up, they didn't exist. That was shocking. It was especially unsettling because the rest of the conversation was actually helpful, so the fake citations were hard to detect.

That experience put me off a little. Some of my colleagues never used generative AI again after that. They felt burned. And I've read that this was a common experience. Even now, GPT-4 still struggles with citations.

But I stayed interested, thanks in part to a junior colleague of mine, a former PhD student who's now in another lab. He's very enthusiastic about generative AI and would regularly update me on ways to work around the citation problem and new use cases people were discovering. That helped keep me engaged.

There was also a memorable faculty meeting where a colleague presented a project he had done in Israel: he asked ChatGPT to write scientific papers using only existing literature (Ifargan et al., 2025). He's a serious scientist, and I was really impressed by how far the AI could go, formulating hypotheses, structuring papers. It made a big impression.

Since then, generative AI has become more integrated into my work. I don't want to sound like an advertisement for a specific tool, but I've mainly used GPT-4, based on my colleague's recommendation. I've been paying for a subscription ever since. I tried a few other tools but didn't find them as satisfying, though that might just be because I've gotten used to how GPT works. Like any tool, it takes time to learn how to use it well.

Now, I use GPT-4 almost every day, mostly for research. Sometimes I use it on its own, and sometimes alongside other search tools. It's like having a colleague who is extremely broadly knowledgeable, but whose knowledge in any specific area is shallow and not always reliable. Still, no human has that breadth. So, it's incredible. And despite its flaws, like occasionally being wrong or resistant to correction, I've had many experiences where it made interesting connections or surfaced valuable insights.

YH: Thanks for sharing your experiences. You mentioned earlier that GPT still struggles with generating accurate references. Have you found any reliable workarounds?

TM: Yes, the fake citations are still a major flaw. One workaround is to use it in combination with a proper literature search engine. I sometimes have it formulate the search query, then summarize the results. That forces the citations to be real, though the synthesis can become less coherent.

Other times, I ask GPT for citations knowing they'll be fake, and I use those fake references as search terms in PubMed to find what it was probably referring to. That's kind of ridiculous, but it still saves time.

YH: Besides literature searches, what other specific tasks do you assign to GPT in your research?

TM: I've had all sorts of experiences using GPT for research, some frustrating, some really illuminating. One of the most compelling examples happened recently. I was studying a very obscure protein. There wasn't much known about it, but we observed it getting phosphorylated in response to one of the drugs we're testing. What struck me was that this protein had two completely different sets of annotations in the literature. One suggested it was involved in a cell death pathway; the other described it as a GDP exchange factor for a particular GEF protein. These two roles seemed totally unconnected.

So, I asked GPT-4: Why might a protein involved in a cell death pathway also act as a GEF? It's a highly advanced question. GPT-4 responded with several mechanistic possibilities. As usual, some of the citations it gave were incorrect, but one of the hypotheses really stood out. It pointed me to a very niche area of the literature I wasn't familiar with—an unexpected corner that, I think, opened up a genuinely interesting line of inquiry. That, to me, is a key strength of generative AI in research: it prompts new ways of thinking. It helps you frame better questions.

I also use it cautiously in tutorials, particularly when we're discussing topics that fall slightly outside my core expertise. In those cases, I really enjoy three-way conversations: me, a student or colleague, and the large language

model. These interactions are often stimulating, thought-provoking, and productive in unexpected ways. There's something very useful about having that kind of triadic dynamic: it pushes the conversation into new territory.

Knowledge or Retrieval?

YH: It sounds like you treat GPT-4 as a knowledgeable colleague, but one that can make mistakes. This aligns with how some in the literature describe generative AI as a "more knowledgeable other (Stojanov, 2023)."You also mentioned earlier that it helped you make novel connections. In that protein example, would you consider that a form of knowledge creation by GPT-4? Or was it just surfacing information that already existed?

TM: It's not that GPT-4 synthesized new knowledge. It was more like it highlighted existing knowledge I hadn't seen. Maybe I could have found that information through traditional searches—PubMed, for example—but it would have taken longer, and I might not have thought to ask that specific question in the first place.

I haven't yet encountered an example where ChatGPT truly creates novel knowledge—something unsupported by any current literature—that I would trust. As you dig deeper into a topic, GPT-4 will start offering more synthetic suggestions, but those are often unreliable. The deeper you go, the more likely it is to suggest something imaginative but false.

There are areas where it's just flat-out wrong. That's especially problematic when it reinforces misconceptions—things that sound intuitive but are actually incorrect. This happens often in chemistry and physics, where the correct answer is counterintuitive. I can give you an example. There's a simple problem in statistical mechanics: imagine particles randomly bouncing around in a solution near a hard wall. How does their concentration change as you approach the wall? I assumed the drop-off would be shallow. My colleague told me it actually drops very steeply, which is the correct answer. When I asked ChatGPT, it gave me the same wrong answer I had. It echoed the same mistaken entropy-based reasoning I was using. And even when I challenged it repeatedly, it couldn't be convinced otherwise. That was disconcerting.

But in other cases, especially when doing rough, back-of-the-envelope calculations in biology, it can be quite helpful. Say I want to estimate how many molecules are in a 1-micron box at 1 molar concentration. I used to do that by hand. Now I ask GPT-4, and it usually does a decent job, though sometimes it makes order-of-magnitude errors.

What's interesting is that if I say, "That doesn't sound right. Can you check it?" It often revises its answer and gets it right the second time. It's capable of self-correction when prompted, which is amazing. Not always, but often enough that I keep using it.

Expanding Use Cases

TM: I tried using it to learn Python from scratch, but I didn't get far because my starting knowledge was zero. Still, I now use it to generate command-line expressions in Excel for tasks that would be too confusing to Google. Excel can do fairly complex operations, and GPT-4 can usually produce a working command that I just paste in. That's incredibly useful.

Professional coders use it in a similar way, for writing and debugging code. But I've noticed that if you know nothing, GPT-4 doesn't help much. It doesn't substitute for foundational knowledge. That applies to literature searching too. I can evaluate its suggestions because I know the field. I can quickly spot which of the five responses are worth pursuing. So right now, it's most helpful for experts: it makes us faster, and sometimes more creative.

I've also started using it for lab protocols. If I need to express a protein, I'll ask for a protocol. It's usually decent—perhaps no less reliable than what you'd find in a published paper. So, for simple tasks, it works quite well.

And here's another unexpected use: in biology, we often work with lists of proteins and want to categorize them. For example, if I have 100 proteins and want to know which ones bind microtubules, GPT-4 can help. If I give it the list and ask which ones likely bind microtubules, it does an okay job—not great. But if I ask it to generate a list of 10-

-0 microtubule-binding proteins, it does a very good job. That task is officially handled by gene ontology databases, but those aren't always reliable or convenient. GPT-4 ends up being faster and sometimes more useful.

YH: It sounds like, as a senior researcher, GPT-4 has helped you with many routine tasks, convenient and time-saving, even if it makes mistakes occasionally. What about more advanced tasks in research?

TM: The protein example I mentioned earlier is probably the most advanced use I've had. In that case, I was looking at two datasets from two different students, published in separate papers. I was trying to synthesize across them—looking for a connection. That kind of synthesis is where GPT-4 has helped me most. I asked GPT-4 to explore potential links between two seemingly unrelated findings. That's not something you'd easily get by asking a human expert or using traditional search tools. It pushed me to think beyond obvious explanations and consider alternative mechanisms.

So yes, generating hypotheses across datasets, helping me find unexpected relationships, that's probably the most intellectually advanced thing I've used it for. And I don't really do anything in my work that's more complex than that kind of conceptual synthesis. GPT-4 is great for those quick, exploratory conversations. I've found that longer sessions don't improve the output much—you hit the limits quickly—but short, focused exchanges can be incredibly helpful.

Writing and Cognitive Effort

YH: What about writing?

TM: I've tried using it for writing, but I really dislike the style. I enjoy writing. It's one of the pleasures of my job, actually. I care about every word, especially in scientific writing. And I believe there is such a thing as style in science writing—but it's subtle. So, when I'm writing a paper, it's an intense process. Every sentence matters. There are no shortcuts. And frankly, the figures are harder than the text anyway.

I also do a lot of communication in bullet points—PowerPoint slides, shared docs within the lab. I've found GPT-4 pretty good at taking a paper and turning it into bullet points. It's less good at turning bullet points into polished paragraphs, in my opinion.

YH: You mentioned that you don't want to delegate writing tasks to GPT-4 because they're integral to your thinking. In EdTech research, we call that cognitive offloading, using a tool to take over part of the mental labor (e.g. Zou & Zhao, 2025). For you, writing isn't something you want to offload.

TM: Right. I don't want to offload that part of the process. There are a few reasons for that. First, writing isn't the bottleneck in my productivity, so I don't gain much from speeding it up. Second, I actually enjoy writing. It's part of what makes my work satisfying. And third, it stimulates my internal thinking. If I delegate that to a machine, I'm not engaging my brain in the same way. Maybe GPT-4 could technically do a better job, but it's my job to think about these things.

That's why I'm concerned about AI use in education. If we let students offload too much cognitive effort, will they ever truly learn? But I also understand the counterpoint. Students today are under immense pressure, with limited time. Maybe AI can help them cover more ground. I don't know the answer yet, but I hope our class experiment will shed some light.

That said, I'm planning to run an experiment next semester in a new master's-level paper reading class. We'll ask half the students to read papers and write notes the traditional way, and the other half to input the paper into GPT-4 first, get a bullet-point summary, and then read it. We'll alternate groups and compare the experiences. I'm genuinely curious whether the AI-assisted reading helps students, especially with papers that are outside their core training. For instance, a biochemist trying to understand a medicinal chemistry paper. It might help bridge that gap.

YH: You're also pointing to a broader question: when does using AI support learning, and when does it short-circuit it?

TM: Exactly. One of my colleagues thinks if students use AI up front, they'll never really understand the paper. Another colleague—much younger—is convinced this is the future. I'm in the middle. That's why I want to try it out and see how students respond.

What you said earlier about cognitive offloading really resonates with me. I don't want to offload writing because it brings me joy, and because it stimulates other kinds of thinking. But that's not true for everyone. Some students might benefit from offloading certain tasks if it frees them to focus on other areas. It might come down to individual preferences and where people want to invest their effort.

Pedagogy and Mentorship

YH: That's a thoughtful approach for the classroom. I'm also curious about your supervisory role. How do you guide PhD and master's students when it comes to using generative AI tools like ChatGPT?

TM: For PhD students, my approach is that they're junior colleagues—not trainees. I don't tell them how to study or how to do their research. They're responsible for discovery and innovation, and how they get there is up to them. Of course, I'm happy to share how I work. And I learn from them too. That's a big part of why I get along well with students—we treat each other as equals.

With master's students, the relationship is a bit different. There's more of an expectation that we're teaching them something. We've had serious debates in our master's program team about how much to embrace AI. Should we tell students to use it for everything because it's the future? Or should we hold back because they still need to build foundational skills? We're not sure yet.

YH: That's fair. A lot of higher education faculty are still figuring this out. And when it comes to doctoral education, the landscape is even murkier.

TM: Exactly. For PhD students, what matters is the result—the discovery. That's the assessment: publication, peer review, contribution to knowledge. We all know AI is having an impact on assessment. Some educators have reacted by banning devices, insisting on in-class exams, leaving cell phones at the door. Others say: I don't care how you get there; I just want to see the answer.

In our program, we've noticed that if we assign quantitative take-home problems, students will almost certainly use AI tools. So, we've moved away from that kind of assignment and toward short-answer questions or in-class assessments. There's no perfect solution.

It's a huge topic, and I'm not the right person to tackle it deeply. But I am aware that assessment shapes how students learn. I just tend to see things more from the research side: how students engage with uncertainty and make new knowledge. That's where my interest lies.

YH: That's understandable. AI and Assessment is a major topic in undergraduate education, but we know far less about how it plays out at the doctoral level. You mentioned earlier that the core goal for PhD students is discovery. Do you think GPT can support that goal?

TM: Yes, I do. Both my doctoral students and I have been using GPT in ways that make us more efficient. Someone uses it to code; I use it to design experimental protocols. That's not cognitive offloading; it's just a faster way to get information we already know how to process.

But when you're using it to connect ideas across datasets, that's a more advanced kind of use. I think GPT-4 can support cognitive enhancement in that space. It's like having a colleague who helps you maximize your mental output.

The kind of student pursuing a PhD has also changed. When I started, people pursued PhDs to push the boundaries of knowledge. That's how I approached it. It's not just about the tools we use—it's about the mindset. But these day-

-s, many students see the PhD as a stepping stone to a good job. They want to do the minimum. They're less interested in being stretched. And they want more formal education: more coursework, more structure.

Maybe they're right to want that. But I've had to adjust my expectations. I'm still very elitist in my scientific thinking. I've always worked at the top places, and I've always believed that if you don't know something, you go figure it out. In that mindset, GPT is just a new tool for inquiry, like the library used to be.

I also found GPT surprisingly useful when mentoring PhD students, especially when they come to me for career advice. I've used GPT-4 as a kind of brainstorming tool for career direction.

AI Co-Creation and Human Thinking

YH: Earlier you described GPT as an efficient assistant, but I wonder, do you ever feel that you're actually co-creating with it? That it's more than just a tool?

TM: It certainly simulates agency very well. But I think we need to be careful about attributing human-like qualities to it. Sometimes I've had this conversation with colleagues: Is GPT thinking? How is it different from us? I tend to take a relatively skeptical—or maybe just humble—view of human reasoning. I actually don't think humans are as impressive as we like to believe. We all have deep blind spots. I can look at a chemical structure and intuitively reason about its behavior. It feels like reasoning, but it probably isn't really. That's training, not innate genius. When a physicist sees the same thing, it's just scribbles to them. So, I'm not convinced that our reasoning is as advanced or as distinct from machine reasoning as we might think. We don't fully understand how humans reason. Large language models generate the next word based on probabilities. But maybe that's not so far from how we operate.

So yes, maybe these machines are reasoning in ways that are different, but not necessarily lesser. I haven't tried this myself, but I've heard of researchers who use multiple AI agents, like two instances of GPT-4, where one plays the role of a creative thinker and the other a critic. They pose questions to each other and iterate (Yang et al., 2025). Apparently, this setup has produced ideas that combine novelty with rigor. That kind of synthetic back-and-forth is interesting.

YH: That's fascinating. In our field, we care a lot about how something was created. It affects how we value the findings. But maybe in science and engineering, it's more about the result?

TM: In science, it usually doesn't matter how you got there, as long as the result is valid, reproducible, and insightful. There's a lot of interest right now in building predictive models of biological systems using machine reasoning. For example, you might ask: If I give this patient two compounds, what will happen to their cancer? That kind of model would rely on probabilistic reasoning rather than causal chains, which are hard to establish in complex biological systems.

Some researchers are also trying to reverse-engineer how AI makes its predictions, with the hope that understanding the model's internal logic might reveal new mechanisms in biology. That's a whole new frontier, where machines might not just help us reason, but actually reason in ways we're not capable of. That's probably not the kind of research I'll personally pursue, but I think others will. And they should. These tools are going to be a big part of the future.

Emotion and Connection

YH: You mentioned that you're skeptical about human reasoning, and I noticed you use the word "feel" in describing chemical structure and its behavior: "it feels like reasoning." Do you think GPT can actually feel? Or perhaps simulate aspects of feeling?

TM: Personally, I don't find its emotional tone very appealing. As a biologist, I have a mechanistic understanding of feelings. We think using electrical signals at synapses, but we feel through chemical signals that diffuse and affect many neurons at once. That's what gives us mood and emotion. So, by that definition, computers can't feel. They d-

-on't work that way.

That said, humans are deeply emotional creatures. Our feelings are a source of both strength and vulnerability. Feelings drive us to tackle hard problems, but they also mislead us. They can cloud judgment, create bias, or lead to self-sabotage. That duality is part of being human—and it's something machines don't have. At least, not yet.

But I know these systems are designed to emulate human language, including emotional tones. And in some cases, that can be helpful. I've asked GPT-4 for recommendations—on finding a plumber, buying clothes—and it's worked out fine. But I can't imagine turning to it for relationship advice. That feels taboo to me, maybe a bit old-fashioned.

YH: Some people said ChatGPT gave them a sense of companionship during the lonely journey of their research. They described it as reassuring to have someone to "talk to" 24/7.

TM: I can see that. I'm privileged: I still have a lab group and wonderful students to talk to. But I'm 67 this year, and I do think about aging in science. I've seen different models. One colleague next door is still going strong in his lab. My father, also a scientist, aged more gradually. When it got too cognitively demanding to be in the lab, he still tried to do research from home. It kept him happy. He might have really benefited from something like ChatGPT—something to engage with intellectually when social interaction became more difficult. As we age, it becomes harder to stay socially and mentally active, but that's exactly why we need it more. It's like exercise: harder as you get older, but more necessary.

There's also fascinating work in robotics. I visited the MIT Museum recently and saw early social robots like Kismet from the 1980s. It had exaggerated facial features, like big eyes, expressive lips, and it was surprisingly effective at connecting with people, especially children with autism.

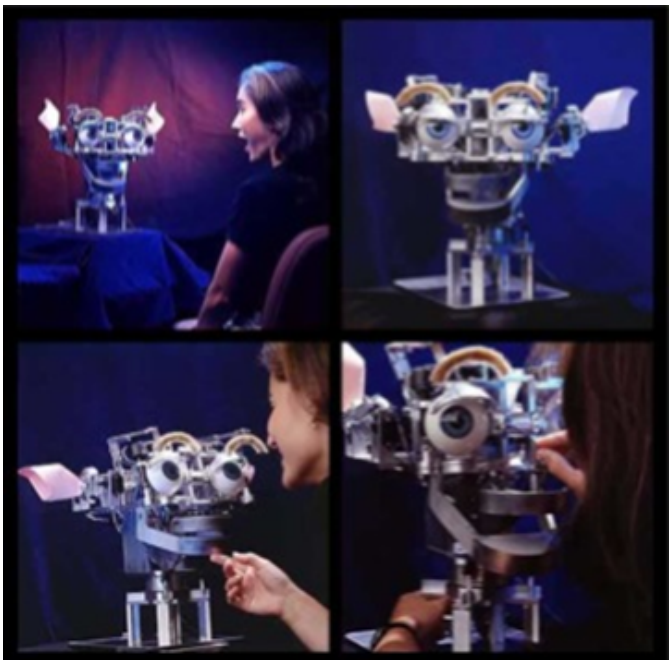

Figure 4: The First Social Robot Kismet (Kismet, 2025).

That tells us something: our brains are wired for social engagement, and under the right conditions, machines can trigger those same pathways. That opens up possibilities for human–robot interaction, especially in therapeutic or educational settings.

But as with everything in AI, it's still early. We don't really know what the long-term effects are. And we need to be thoughtful about where we draw the line between simulation and real connection.

Conclusion

This interview offers a window into how an experienced scientist is integrating generative AI into the everyday practices of research and graduate education. TM's reflections highlight both the possibilities and limitations of tools like ChatGPT: their capacity to accelerate routine tasks, stimulate conceptual exploration, and support mentoring, alongside persistent issues with accuracy, reasoning, and the risk of cognitive offloading. His experiences suggest that AI can be valuable when used by experts who can evaluate and challenge its outputs, and when it becomes part of a broader conversational process within labs and classrooms. At the same time, the discussion raises important questions about learning, assessment, and what counts as meaningful intellectual labor in an AI-mediated academic environment. Taken together, the conversation underscores the need for thoughtful, discipline-specific approaches to AI adoption that preserve human judgement, creativity, and connection at the core of research and doctoral training.

Practical Notes: Key Insights from the Interview

General Takeaways

- Use generative AI as a companion tool, not a replacement for expertise. GPT works best when paired with strong foundational knowledge that allows users to evaluate and challenge its outputs.
- Expect broad but shallow knowledge. AI can surface unexpected connections and generate helpful starting points, but deeper disciplinary reasoning still requires human judgment.
- Always verify citations and technical details. Because GPT may produce plausible but inaccurate references or conceptual errors, pairing it with reliable search tools remains essential.

Research Practices

- Leverage AI for routine tasks to enhance efficiency. Tasks such as summarizing papers, drafting code, and generating protocols can often be completed faster with GPT assistance.
- Use AI to support conceptual exploration. Quick, focused exchanges with GPT can stimulate hypothesis generation or help frame questions that bridge datasets or disciplinary boundaries.
- Recognize the limits of AI in writing. Writing remains an important part of scientific reasoning. AI-generated text may not capture desired precision or style.

Teaching and Supervision

- Be mindful of cognitive offloading. While AI can accelerate work, excessive reliance may reduce deeper learning, particularly for students still building foundational skills.
- Maintain clear expectations for independent thinking. Assessment should continue to emphasize students' ability to evaluate AI outputs and reason through ideas on their own.
- Experiment with triadic conversations. Three-way discussions among supervisors, students, and AI can generate dynamic, multi-perspective dialogue and push thinking into new directions.
- Consider AI as a tool for career reflection. GPT can help students brainstorm potential career directions, offering structured prompts for discussion and exploration.

Future Research Directions

This interview opens several pathways for empirical inquiry into the role of generative AI in research and graduate education. Future studies could adopt comparative designs across laboratories, departments, or institutions to map disciplinary differences in AI use. Longitudinal research could examine how faculty and students integrate AI tools over time and how their practices change with continued use. Qualitative studies that incorporate interviews, observations, and artifact analysis could also deepen understanding of how AI shapes cognition, mentoring relationships, and scientific workflows. Finally, collaborative studies involving both STEM and education researchers could examine how AI-mediated reasoning and supervision practices intersect across fields and offer a richer account of emerging academic work.

References

Ifargan, T., Hafner, L., Kern, M., Alcalay, O., & Kishony, R. (2025). Autonomous LLM-driven research—from data to human-verifiable research papers. *NEJM AI, 2*(1), AIoa2400555. https://doi.org/10.1056/AIoa2400555

Kismet. (2025). *Overview.* https://web.media.mit.edu/~cynthiab/research/robots/kismet/overview/overview.html

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education, 20*(1), 35. https://doi..org/10.1186/s41239-023-00404-7

VOA. (2016). *Google's alphago wins first game of go in matchup with human.* https://www.voanews.com/a/google-alphago-wins-first-game-in-matchup-with-human/3227479.html

Yang, R., Ye, F., Li, J., Yuan, S., Zhang, Y., Tu, Z., ... & Yang, D. (2025). The lighthouse of language: Enhancing LLM agents via critique-guided improvement. *arXiv preprint* arXiv:2503.16024. https://doi.org/10.48550/arXiv.2503.16024

Zou, Q., & Zhao, S. (2025). Exploring the impact of AI on research practices: A postdigital ethnographic study of doctoral students' engagement with ChatGPT. *Studies in Continuing Education,* 1-17. https://doi.org/10.1080/0158037X.2025.2535565

Copyright: © 2026. Timothy Mitchison and Yulu Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.