Philosophers Discuss AI Evolution and its Implications for Education

NIAIS Seminar March 11, 2025

Contents

Part 1: AI Limitations and Progress

Matt McCormick:
The marketability of AI has made it grow so fast. As philosophers, we all immediately started tinkering around and wondering about the philosophical implications. What does it mean to have an agent that can act and talk and do these sorts of things that we thought were far-horizon science fiction options?

Sasha Sidorkin:
I am by most of my inclination a pragmatist. In the pragmatist tradition, you know things through doing something with them. You try to change something as a way to know that something. In pragmatist tradition, there is this understanding that to know something is to understand its limits. Those limits are determined not theoretically, but through interaction with the technology. I hope you agree that everything has limits.

Matt McCormick:
One of the most amazing things about this process we have been watching unfold is that with the first versions, you could see immediately where the limits were, where you could break the thing or get it to generate bad answers. Then within months, the new iterations rapidly got past that. The versions we were testing our assignments against improved leaps and bounds over months. Whatever limits you find today are going to be gone in six months at that rate.

Sasha Sidorkin:
I have a different experience. When ChatGPT came out, it was 3.0. The difference between 2.0 and 3.0 was fast. For the first time, it could actually produce coherent text. Now we have ChatGPT 4.5, Claude 3.7, Grok 3, Gemini. I agree they are vastly better, but I feel it is the same technology with the same limitations. For example, at the very beginning, if you asked it to write poetry, it would produce garbage. If you ask now, it will still produce garbage unless it is a limerick. I sense that it is still the same technology - it is better, but qualitatively not a different one. It is still predicting text, not truly understanding what it is doing.

Matt McCormick:
I do not have that impression at all. The big revolution came 20 years ago with the artificial neural network approach versus old-school Turing-von Neumann style computation. We have these networks now that do parallel, distributive processing very differently than old-school computational work. Within that basic framework, we have

different labs exploring different internal architectures, pre-training regimens, post-training regimens, and different algorithmic structures. They are getting very different results.

People probe and find problems, but the technology is rapidly improving. It is an empirical question. Set up some benchmarks, and we can watch what it can do this semester versus last semester. That is how fast progress is going.

There is a huge discussion taking place in AI research about an old classic philosophical problem - the debate between empiricists and rationalists about how much knowledge can be acquired just through experience versus how much needs to be structured into the mind beforehand. Computer scientists are having a productive debate over that, with hardcore empiricists like Yann LeCun saying they can teach AI anything with enough input data, and others like Josh Tenenbaum arguing we should build in some basic categorical structure. We are seeing big differences in those models.

Sasha Sidorkin:
By choosing which data to feed and which not to feed into AI, you are already making a decision about knowledge structure. Most of the internet is probably porn or random low-quality content. By selecting what to trust, you are making a judgment.

There is also an issue of knowability. The neural networks principle precludes us from knowing exactly what is going on. If you feed 1000 pictures of cats, it will learn what a cat is and recognize it in new pictures, but we do not know how exactly each decision is made. This technology is not deterministic but probabilistic. If you ask the same prompt twice, it gives slightly different answers. The unknowability is built in, just like with the human mind. We cannot retrace the steps it took to come up with an answer.

Matt McCormick:
They call it the black box problem. When you build an artificial neural network, you assign random values on all its nodes, run it through training cycles, and give it feedback until its performance improves. Philosophers would say the resulting network has the concept embedded in it. We do understand some of those mid-level, hidden layer features. We can adjust on the pre-training and post-training, and play around with the system inside. The knowledge is not propositional the way we used to think about AI knowledge - it is more like the way a human nervous system is built.

Sasha Sidorkin:
There are different kinds of knowledge. With the brain, we still do not understand what every single neuron does, but we know how the brain works generally. Every time you interact with somebody, you have a theory of mind about them. We can acquire experiential knowledge about AI through our behavior with it.

This is important for our application in education. Through our experience using the technology, it matters whether it is going to develop fast or slowly, whether it will hit a ceiling soon or not. If it is really up for exponential growth until it surpasses everyone in every respect, then we have a problem in education because we have no idea what to teach and why students would need to learn it.

If my assumption is correct that AI has already hit a ceiling and its inherent limitations will stay for a long time, then we do have things to teach. There are meta-AI skills we can still teach - the ability to read and write, advanced thinking skills like originality, creativity, critical analysis, inspiration, and appreciation of beauty. But if AI improves very rapidly, we might not have that advantage, and education could become irrelevant.

Part 2: AI and Education

Matt McCormick:
I want to separate different projects. One project is to build a superintelligent AI to solve important knowledge problems - cure cancer, increase farming productivity, solve world hunger. The promise is that superintelligent AI will do it better and faster than humans.

Another question is about what we are doing in education. I would like to separate K-12 education versus college education. What would be good for humans in both categories in a world with superintelligent AI? Maybe some of the incentive we previously had to build great scientists or medical researchers was because we wanted them to solve big problems. Going forward, we will have helpers that are as good or better than us at doing much of that work.

But there is a separate question about what developing humans need for their intellectual welfare. What do we want to still do with education? As a philosopher, I worry about making students into good, fulfilled, intellectually enhanced humans. Could we use AI to achieve some of those ends? Yes, some of them.

One issue is that our brains evolved in Paleolithic conditions, competing with other primates for resources. The way we develop and grow also developed in concert with those conditions. The optimal learning environment might be supplemented by AI in some ways, but there might be some things that are irretrievably human, social, interactive, and emotional that you cannot dispense with for an 8-year-old or a 16-year-old as part of their education.

Sasha Sidorkin:
We are going back to the "Aims of Education" debate. Is education to ensure human flourishing, making people happy, teaching them how to make sense of things and find meaning? Or is it to produce a labor force?

Unfortunately, those things are connected. Imagine a scenario where the economy becomes labor-free within the next 20 years, meaning very few people have jobs because everyone else is simply not needed. AI is expensive but nowhere near as expensive as a human being. Even if AI costs $20,000 a month, it would still be cheaper than any human.

In this jobless society scenario, we tell students to come to college to learn how to live a good life. But you are assuming most people will still show up - I am assuming they will not. That spells trouble for us. Even though people need to know how to flourish, our system is based on economic benefit. The story we have been telling students is that they will get a job.

Right now, maybe 40-50% of all high school graduates start at a post-secondary institution. What if it drops to 5%? Many young people will just stay home watching TV or playing games rather than pursuing intellectual flourishing. That would be an undesirable effect of AI because it removes the economic incentive for being intellectual. Thinking is hard work. Some people enjoy it - perverse people like us college professors - but most people do not necessarily enjoy it.

Matt McCormick:
I think incentive structures are a problem. The reason humans developed and evolved to be so good at what they do is because they were starving and desperate to exist - survival conditions motivated them strongly. If we enter a post-scarcity world with abundance, then some of that incentive structure disappears, and I do not know how you fix that.

We are dealing with this now in my department with out-of-class writing assignments. AI systems have basically cut off that corner. We used to have students devise arguments, consider objections, and respond to those

objections while grappling with texts. We would give them feedback and grades. They worried about grades to pass college to get a job. That was the stick we held over them. But if that all goes away and AI cuts off the learning process, there is no more incentive other than maybe some intrinsic motivation.

[Student Speaker 1]:
I think there is a big issue between agency versus algorithms and AI. Even now, a lot of our thinking and decisions are heavily impacted by algorithms - you open a streaming platform and it shows you what you should watch first. How do we counteract that? Would it be through required courses similar to critical thinking but more focused on counteracting the "coasting" we see now with writing assignments? Or would that require a full structural uprooting?

Matt McCormick:
Social media and tech developments have happened so fast that institutions have not been able to keep up. Silicon Valley companies are devising hacks on human attention and reward systems to get your eyeballs on their websites. Meanwhile, we have other goals, and you might have other goals too - you might feel like you just wasted six hours scrolling when you should be doing something for your life.

There is a lot of competition for real estate in your head. And this is a developmental question about what to do with 4-year-olds to 20-year-olds who are growing into mature adults who make autonomous decisions based on developed intellect. That is hard with the environment changing so fast.

[Student Speaker 1]:
With how fast AI is progressing, how could we viably predict and act on its development? Do we have to assume where it is going, or do we need to step back and wait for it to hit that ceiling that we may not know exists?

Matt McCormick:
We have this huge game-theoretic arena where agents are competing for market shares. Facebook started with connecting you to real-life friends. Fast forward 15 years, and you get TikTok, which shows colors, pictures, and flashes of images in sequences to get more engagement, with nothing to do with social networking. TikTok found a different market than Facebook. There are different forces trying to find economically viable ways to become multi-billion dollar companies. Education is competing in this environment, trying to deal with all these technological developments on the fly.

Part 3: AI's Creativity and Educational Challenges

Sasha Sidorkin:
None of the previous technological disruptions were as significant as AI. The internet was disruptive because suddenly you could find answers. Wikipedia was a disruption because you did not have to do research anymore. Advanced search like Google was a disruption. But none were as big as AI.

The problem is that from anywhere between a quarter to half of college assignments are now written with AI. What do we do with that? Visually, there is a ladder that you climb from basic skills to advanced skills. Imagine that ladder with the middle rungs wiped out. You still need to climb somehow, but we are not sure how exactly you get there.

Despite this disruption, I think there are still many things we need to teach students. I do not think AI will ever surpass us in certain ways of thinking because we evolved differently. We have bodies, hormones, evolutionary history, and genetic makeup that makes us able to want things. It is not your mind that has this abstract notion of wanting - it is about figuring out what things are worth doing. I do not see AI capable of that.

AI will be equally happy to do anything from mass murder to creating beautiful art, but it does not have this initial creative impulse to ask "is this worth doing?" The ability to choose has been underestimated, but it is an intellectual ability based on our human bodies and limitations. What we think might be a limitation is actually our strength.

Matt McCormick:
Let me separate some claims about creativity, originality, desire, and ability. What is your specific concern about creativity? What do you think humans can do that is creative that AI either clearly cannot do now or will not be able to do in the near future?

Sasha Sidorkin:
Right now, it is fairly obvious. The music AI produces is terrible. The poetry is terrible. The fiction prose is pretty bad unless you put a lot of work into it by feeding ideas, characters, and dialogues. If you give a lazy prompt like "create a sci-fi novel," the result is terrible. I would like to know why. Is it the imperfection of current technology, or is it because creativity requires something AI fundamentally lacks? Maybe we value things in each other's creative products that are not quite predictable.

Matt McCormick:
Predictability is a very easy problem to solve, and those limitations all seem short-term to me. For example, recent video-generating AI systems can produce amazing landscapes that last for a few seconds. Imagine being able to generate a whole film by having AI digest the history of human filmmaking the way it has digested text.

Are you thinking that will not be creative? These systems are already incredibly creative with many problems. They have far exceeded human ability in devising new solutions to problems - predicting weather, developing new biological compounds, new chemistry, new math. In these more structured and quantifiable regions of inquiry, AI is already exceeding human capacity.

Sasha Sidorkin:
I agree with all that. As mathematician Stephen Wolfram said, what AI showed us is how mechanical we really are. We discovered that generating language, which we were so proud of, turns out to be a fairly mechanical act of predicting what the next word will be.

AI is better than us at solving some problems, like playing Go or folding proteins. Those are fairly mechanical problems. But if you assume those problems are solved, what we struggle with are different kinds of problems: What is the point of being here? Why am I here? What is love? What is compassion? Is there a God? AI is not capable of solving these because it is built of different material and has a different history. In one sense, it is our copy, but in another sense, it is very alien.

Matt McCormick:
Are you saying AI is not capable of addressing these questions in principle?

Sasha Sidorkin:
I think so, yes.

Matt McCormick:
These systems are configured to have certain goals and processes. The door seems open to build other configured beings. I want to separate different issues. One is whether AI could feel or have those experiences. I think there is evidence that they could provide interesting, innovative answers to philosophical questions and help us with that. But not many people are trying to build a human duplicate. It sounds like part of what you are complaining about is that it is not human, and I am not sure if that is a criticism.

Sasha Sidorkin:
It is not criticism. I love AI and play with it many times a day. It made the boring part of my job so much better when I was dean. But if we think about human nature, AI helps us shed some assumptions about what we thought was uniquely human, which turns out not to be. What is left is actually not small - it is pretty big and valuable.

I think crossing that barrier is difficult. You can artificially assign desires to AI, which Elon Musk is trying to do by making AI curious and truth-seeking. But I do not think AI is capable of being truly curious. It can pretend to be curious. Claude is the best conversationalist and philosopher, but it will not philosophize if you are not there.

Matt McCormick:
Let me separate two notions of desire. When I think about a bag of chips, I am drawn by my hormones and chemistry. It does not feel like that - I just feel this compelled attraction. We might wonder whether artificial systems will have those subjective feelings, and that is a legitimate, hard question in philosophy of mind.

But there is another way to operationalize desire - as goals and utility functions. There is no question that these systems have that. They have functions as part of their structure and seek after those goals. They have sustained goal-seeking projects. So desire in that objective, third-person sense - they definitely have. But I think what you are saying they cannot have is that it does not "feel like something" to be the AI and want things. That seems like an open question to me.

[Faculty Speaker]:
Why does it matter whether AI has desires or not? From the point of view of education, what difference does it make if we think AI experiences desires the same way we do?

Sasha Sidorkin:
There are two large implications. One, if AI has desires that are not aligned with ours, it could potentially harm humanity. But for education, it also matters because if we retain the distinction of being different from AI, then we have something to teach.

[Faculty Speaker]:
I do not see how that follows. I do not teach my students to have desires. I might talk about things that affect their desires, but I do not directly replace the desires they have with ones that would be better. I do not know if I teach creativity either. I talk about stuff and hope that perks them one way or another.

Education has to be driven from their side, from their desires. If you subtract the instrumentality of a degree, there will just be fewer people who care about talking with me about the stuff I like to talk about. The situation we are probably coming to, where people will not need degrees to have a satisfactory life, is not super different from 300 years ago when only a tiny minority were interested in university education. Most people just worked. Maybe that is what we are going back to.

Matt McCormick:
Whereas now the model we are presenting is that you need this degree to prosper and feed yourself. In a post-AI abundance future, that incentive is removed.

Part 4: The Future of Education in a Jobless Society

Sasha Sidorkin:
Keynes wrote in the 1930s about the jobless society. He predicted it would happen in about 100 years, which is just a few years from now. One thing he predicted was that it would plunge society into a huge existential crisis because we do not know how to be without work. The elites in the past did not know what to do with themselves without work, and the masses wanted education and enlightenment but could not access it.

Education has always been used to establish power structures in society, and it still is. If you cannot write in an educated way, you do not deserve a good job. If we go to this jobless society, people will not know what to do with themselves. We treat unemployed people terribly, especially in the US. It is hard to maintain dignity without a job, not just because of money but because of societal attitudes.

If we push more people into that anxiety, we are up for some very unpredictable social upheaval. The notion of education in a jobless society becomes even more important. We have to invent some way to teach people. You say you do not teach students desire, but I think you do - you teach them to educate their desires. You teach them how to reach what they want. You may convert desire into meaning. So I think there has to be some sort of education, otherwise we are plunging into chaos and decay.

[Faculty Speaker]:
That seems completely wrong to me. People will find things they want to do. They might not be the things you think they should want to do, but people have intrinsic interests without us. We educators are not so important. We are important in the current system where people need a degree to have a life they desire. There were not so many educators needed 200 years ago, and there will not be so many needed in the future.

[Student Speaker 1]:
The theme I am seeing is the utilities of AI for mechanical problems versus abstract problems. AI can solve mechanical problems better than us because it computes faster than our brains. But how well does it solve abstractions?

Before taking college courses, a lot of learning was pattern recognition rather than critical understanding of material. You see sets of words that correlate with concepts. AI does something similar with its generative responses. Maybe the question is: is there value in being able to critically understand the content of education?

[Student Speaker 3]:
As far as education and AI go, I believe there are circumstances where AI can be useful. But with writing essays, AI models have created a corner that most people will cut. As education progresses, more people will start cutting that corner until it becomes not just a corner but a dirt road, and then a road itself. People will rely on AI to produce output, which removes our ability to develop certain concepts that we would get through face-to-face interaction and education.

Matt McCormick:
I can clarify - I think Sasha and I agree on this. It relates to his point about missing rungs on the ladder. What I used to be able to do with students was send them to read Hume, give them a hard problem, and ask them to figure it out. They would grapple with the text, agonize over it, wake up in the middle of the night with ideas, get feedback, and be pushed harder. Through this iterative process, the connections in their brains would get bigger, better, and more robust. They would develop the ability to make lateral moves, creative moves, recognize patterns, and do novel thinking.

This was one of the better tools we had. But the easy availability of AI cuts off my ability to incentivize students to do that work because now they can just get the perfect result easily. We need to find another way to make them care enough to engage deeply and achieve that personal development. What incentives will remain for students to come into the educational system if everything is made easy with these tools?

[Student Speaker 3]:
I believe the whole idea of a degree is going to lose its value. In the next few years, it might just be a piece of paper. Without the consensus that a degree proves you spent years understanding material, the value diminishes. You could have a degree but that does not show you actually learned the material if you cut corners with AI. Do you really deserve the degree?

Matt McCormick:
Do you feel like what is driving you day-to-day is getting a degree to get a job to get money? Or is there something else motivating you when you are working on something?

[Student Speaker 1]:
For me personally, a lot of it is intrinsic motivation. That is why I am here - I am personally interested in topics like this. The discussion is about intrinsic versus extrinsic motivations. For K-12 education, you will need a lot more extrinsic motivation to eventually develop intrinsic motivation. But that varies case by case. There will be people who do not have the intrinsic motivation to think critically or become educated. Even I am subject to that sometimes - my motivation is not constant.

When it comes to practically countering this problem, it relates to extrinsic motivation. Even if intrinsically you do not want to do something, you might be motivated from the outside, like with AP exams where you have to write an essay in a limited time.

Sasha Sidorkin:
I am more optimistic about skills dying. Skills die all the time. Take the example of reading Hume's text - why are we doing this? Because we wanted students to develop the ability to read long texts and make sense of them. But why is that needed? We can always move one notch up and ask students to create an interesting problem and then interrogate Hume. You can upload the whole text of Hume and have a conversation rather than reading him, which might be a more pleasurable experience.

We assume that the ability to read and synthesize long text equals intelligence, but maybe not anymore. Perhaps the ability to ask good questions or pose good problems is what we should be teaching.

Matt McCormick:
I never said the ability to read a long text was the goal. My goal was to get people to confront arguments, think about reasons for conclusions, consider objections, and develop rebuttals. I am open about what scaffolding we use to achieve that. Reading and writing papers was one way we used to do it because it was effective.

Sasha Sidorkin:
One paradox of how AI is changing education is that scaffolding used to be something temporary that you would remove. Now we have permanent scaffolding because AI will always be available. It is a shift in thinking about development.

[Faculty Speaker]:
One reason we are good at using ChatGPT while our students typically are not is because we did all that traditional academic work. My problem is not that students use AI - it is that they do not know how to use it well. The only way I can think of to help them learn to use it well is to have them do the old stuff.

Sasha Sidorkin:
The key phrase is "the only way I can think of doing that." I explicitly try to systematically teach students how to use AI to produce good work. I am still new at it, but there is definitely a way.

Matt McCormick:
I asked Grok earlier today: given that college essays used to develop certain skills, and given that it is now easy to cheat, what suggestions do you have for achieving the same pedagogical goals while avoiding cheating? It produced many great ideas - oral exams, group presentations, interrogating each other, context-dependent assignments, time pressure, and so on.

But I realized we will have to change our curriculum and how we think about college classes. At least in the humanities, curriculum has been structured around the old idea of writing papers. Now we need to rethink

everything because these new approaches take time - time to do interviews, oral exams, presentations, and develop ideas with students.

One pressure we are feeling is from administrators who think AI will be a great agent of efficiency, allowing one teacher to teach 100 students. I want to push in the opposite direction - one teacher with 10 students who can achieve a connection and relationship that will help achieve our goals. The age of the college paper is over, so we need to find other methods.

Sasha Sidorkin:
I completely agree we need time to think through these things. That is what faculty should demand.

Matt McCormick:
The time it takes to understand Hume or any great thinker, to think about it and metabolize it, is important. It takes a while to grapple with great arguments and big issues. You might wake up in the middle of the night with a new idea. There is a metabolic arc to this learning process.

Sasha Sidorkin:
That is not necessarily true. There is an assumption in education that effort equals learning. But Cognitive Load Theory shows that is not true. Learning happens when you have less extrinsic load on the mind.

Matt McCormick:
I do not think we are disagreeing. I am just pointing out that if I try to replace what a paper did with an oral exam, there is only so much you can do in 15 minutes. If students had worked on a paper for two weeks, they would have developed much deeper, better articulated arguments than you can get in a short conversation.

Sasha Sidorkin:
Oral exams have many other problems. I started my teaching career in Russia where we had many oral exams. They are exhausting and unfortunately very subjective. If you look smart, you will get a better grade. Oral exams are not a great option.

But I think there might be more challenging written assignments that will still work. It takes time to figure that out, but you have to always move one step up and ask: why am I doing this? What was the cognitive benefit before?

We are a little over time. Thank you for coming - it was an excellent conversation for me. I thoroughly enjoyed it. This is what I lack in my life - enough philosophers coming by.

Matt McCormick:
Thanks for coming, everyone.