



SuperDataScience

SDS PODCAST

EPISODE 1036:

IS AI HOLDING KIDS

BACK FROM

LEARNING?



Jon Krohn: 00:00 This is episode number 1036 on whether AI is holding kids back from learning. Welcome back to the Super Data Science Podcast. I'm your host, Jon Krohn. Today's episode tackles a question that parents, teachers, and anyone who cares about the future workforce are asking. That's does AI stop children from learning? Let's start with how widespread student AI use has become. A survey last year by the ed tech firm, Chegg, found that 80% of undergraduates in wealthy countries were using AI in their studies, and more recent polls in Britain and Germany put that figure at 93 to 94% of students. Teachers, unsurprisingly, report grading stacks of suspiciously similar essays that look like they came out of ChatGPT. Until recently though, rigorous evidence on what all this AI use does to learning has been thin. That's changing. And the data released over the past few months tell a story that's more nuanced and more useful for even us as adults than either AI is rotting kids' brains or AI is the best thing to happen to education since the printing press.

01:13 Those are two extremes. All right, let's tease them apart. The biggest new study comes from researchers at Stockholm University and the University of Hong Kong. Published in June, it uses two and a half years of data on nearly 30,000 Chinese students in grades seven to 12, so roughly ages 12 to 18. China makes for a fascinating natural laboratory because AI adoption among students there has been fast. About 80% of students reported using chatbots like DeepSeek or ByteDance's Dubao, while the remaining 20%, just a fifth, did not. Because students started using AI at different times, the researchers could exploit that staggered adoption in what's called a difference in differences design, comparing how each student's outcomes changed after adopting AI against the trajectories of students who hadn't adopted yet. That's a big step up from the typical one-off correlational snapshot that we get in big studies of human behavior.



- 02:11 The data cover monthly closed book exams, high school and college entrance exams, and homework scores and completion times across nine subjects. Here's what they found. After adopting AI, students' homework scores rose by 18% and the time they spend on each assignment fell by about 30% from an average of 64 minutes down to 45 minutes. That sounds like a productivity miracle, but monthly exam scores of course fell by 20% within six months and the damage compounds. On high stakes entrance exams, scores fell by 18% and 24% with the full penalty emerging only after about two years. So as you might expect, it sounds like AI is not so much a productivity miracle as being used to cut corners on homework. That leads me to quite an interesting finding. Homework scores used to be a solid predictor of exam performance. With AI in the mix, however, that relationship has flipped so that the students posting the highest homework scores are now more likely to do worse on exams, presumably because AI is on average being abused to race through homework as opposed to being used as a tool to learn.
- 03:21 If you're a teacher using homework grades as a signal of who understands the material, that signal is now broken. The losses were largest in social sciences followed by STEM and then languages, and the losses were especially large for junior students, for high achievers and for boys. So should schools ban AI like they have recently done in New York City where I live? Well, not so fast, I'd say. The learning losses were concentrated among roughly 80% of AI users whose behavior looked like homework outsourcing, exceptionally short completion times paired with high homework scores. Meanwhile, among AI users who spent between 50 and 65 minutes on homework about the same as non-AI users, exam scores were effectively identical to those of non-users, even though their homework scores were significantly higher. In other words, those 20% of students were using AI as a learning



tool, presumably to do things like explain concepts or get unstuck rather than to produce answers wholesale.

- 04:23 This kind of pattern, it isn't unique to China. A field experiment with nearly a thousand high school math students in Turkey led by a researcher at the Wharton School in Pennsylvania gave some students a standard GPT-4 chat interface and others a version called GPT Tutor with safeguards in that GPT tutor designed to protect learning. Each GPT tutor prompt ran to more than 500 words compared with about 50 words for the standard version. During practice sessions, performance improved by almost 50% with the standard GPT-4 and by 127% with GPT tutor. But on the exam that followed, without AI, students who had used the standard version scored 17% worse than students who had had no AI access at all, while the GPT tutor group's results were statistically indistinguishable from the control groups, the control group of course not having access to AI at all. So having the standard GPT-4 improved performance in practice sessions, but having this program GPT tutor designed for learning, students performed much better.
- 05:31 And then on the exam that followed, people who had access to normal GPT-4, they underperformed relative to the control group, but the folks with the GPT tutor did not. So yeah, it seems like there's something to potentially work with here, or at least it's not all bad news with AI. One interesting side effect from this Turkish study is that students who used AI were overly optimistic about how much they'd learned, even the high achievers were overly optimistic. So weaker skills paired with greater confidence is a dangerous combination, but it's unsurprising that the AI users would have inflated egos given how flattering LLMs tend to be to us. I'm sure you're used to that experience. So far, all I've discussed is students with AI access at best performing comparably on exams to students with no AI access, but there are some



new studies suggesting AI can be leveraged in promising ways.

- 06:30 In a randomized experiment, researchers at Middlebury College in Vermont had undergraduates in proctored in-person sessions learn about an unfamiliar topic and write an analytic essay with or without off-the-shelf generative AI and then take unaided assessments immediately and one week after. AI access raised immediate test scores by 0.27 standard deviations, which is quite a bit, and those gains persisted a week later. Interestingly, essay quality barely changed while students had AI, but it improved a week later when they wrote unaided, and that delayed gain was larger among augmentation users who had AI explained concepts, whereas automation users who had to generate text saw their short run gains vanish once the AI was removed. The mechanism for all this, students with AI shifted time away from drafting text and toward reading and searching for information, and they reported enjoying learning more. There's additional good news. At Harvard, a randomized study of nearly 200 introductory physics students published in June of last year found that students learned more than twice as much from a custom AI tutor as from an active learning classroom and in less time.
- 07:43 That tutor was no generic chat bot. It was fed the lecturer's own solutions and the researchers instructed it to keep responses to a few sentences to avoid cognitive overload. And in another separate study in Nigeria, a World Bank randomized trial of a six-week afterschool program using Microsoft Copilot powered by GPT-4 improved scores by 0.31 standard deviations, again, quite a bit, equivalent to one and a half to two years of business as usual schooling. Wow. And this outperformed this GPT-4 randomized trial solution outperformed 80% of educational interventions that were evaluated through randomized trials in the developing world. The World



Bank team behind that study is careful to note that the results reflect the full package of structured prompts, curriculum alignment, teacher support, and peer learning plus the AI itself. So that intervention was way more than just plopping kids down in front of a ChatGPT session.

08:38 So back to my question at the top of this episode, does AI stop children from learning? The data suggests it depends on how AI is used. Learning requires effortful thinking. Cognitive scientists call this desirable difficulty. When AI removes that effort by producing the answer before a student has wrestled with a problem, the homework looks great and the learning doesn't happen. When AI is used to explain concepts, offer hints and give feedback, it can be an outstanding tutor. Given all that, I've got three practical takeaways from all this for you. First, if you're a parent or educator, time on task looks like a useful warning signal. A student whose homework quality jumps while homework time collapses is probably outsourcing to AI. Second, tool design matters. The Turkish, Nigerian and Harvard results that I talked about in this episode all point toward AI configured to tutor rather than to just answer.

09:37 So it's worth turning on the study mode or the learning mode that many AI platforms now offer. Third, assessments completed without AI such as in-class closed book exams become more important now, not less, because homework grades no longer tell you what a student knows. This also makes me think of new testing methods that might make sense. For example, Professor Jon Renne with whom I co-instruct an introductory course on AI at Tulane University in New Orleans. Professor Jon Renne is the co-creator of an AI tool called oralexam.ai. You can check that out online and it has students engage in a bespoke automated conversation about their written work after they submit it to the platform. Jon will be my guest on this show in a few weeks, so look out for more on oralexam.ai and keep an



eye out for more clever AI-powered educational solutions to AI-created problems in the near future.

- 10:31 Finally, as might have been obvious to you as you listen to this episode, these lessons from the studies I covered in this episode aren't only for kids. Every one of us using AI at work faces the same trade-off. Finish the task faster today or build the skill that lets you tackle harder tasks tomorrow. My guidance would be to try the problem yourself first, ask the model to explain its reasoning and use AI to augment your thinking rather than replace it. Use that way. AI may be the most powerful learning tool humanity has ever had. Go put it to work. All right, and haven't done this in a while, but it is time to check on our Apple Podcast reviews and see if anyone has made any new ones, and we do have some. So we have one here titled Great Podcast from the very clever and mysterious username user 21309.
- 11:26 And user 21309 says, "Love the work you do on this podcast, returning to this community after years and super happy to see how much it has grown." We're super happy to have you super listening. Thanks for all the recent ratings and feedback on Apple Podcasts on Spotify and all the other podcasting platforms out there, as well as for likes and comments on our YouTube videos. Bonus points if you leave written feedback, which you can do on Apple Podcasts. If you do that, I'll be sure to read your feedback on air like I did today. At this time, I really only see when people do that in the US, I think, but someday I will look around the world and see where else I can find reviews. All right, that's the end of today's episode. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science podcast with you very soon.