Time for a Pause: Without Effective Public Oversight, AI in Schools Will Do More Harm Than Good.

Ben Williamson  
University of Edinburgh  

Alex Molnar and Faith Boninger  
University of Colorado Boulder  

March 2024  

National Education Policy Center  
School of Education  
University of Colorado Boulder  
nepc.colorado.edu
Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Patricia Hinchey
Academic Editor

Elaine Duggan
Production Design

Alex Molnar
NEPC Director

Kevin Welner
NEPC Director


Funding: This policy brief was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

Peer Review: Time for a Pause: Without Effective Public Oversight, AI in Schools Will Do More Harm Than Good was double-blind peer-reviewed.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The National Education Policy Center (NEPC), a university research center housed at the University of Colorado Boulder School of Education, sponsors research, produces policy briefs, and publishes expert third-party reviews of think tank reports. NEPC publications are written in accessible language and are intended for a broad audience that includes academic experts, policymakers, the media, and the general public. Our mission is to provide high-quality information in support of democratic deliberation about education policy. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence and support a multiracial society that is inclusive, kind, and just. Visit us at: http://nepc.colorado.edu

http://nepc.colorado.edu/publication/ai
I. Executive Summary

Ignoring their own well-publicized calls to regulate AI development and to pause implementation of its applications, major technology companies such as Google, Microsoft, and Meta are racing to fend off regulation and integrate artificial intelligence (AI) into their platforms. Concerns underlying calls for such a pause run the gamut; here, we focus on democracy and privacy. The weight of the available evidence suggests that the current wholesale adoption of unregulated AI applications in schools poses a grave danger to democratic civil society and to individual freedom and liberty.

Public education is a public and private good essential to democratic civic life. The public must, therefore, be able to provide meaningful direction over schools through transparent democratic governance structures. Yet important discussions about AI’s potentially negative impacts on education are being overwhelmed by relentless rhetoric promoting its alleged ability to positively transform teaching and learning. The result is that AI, with little public oversight, is on the verge of becoming a routine presence in schools.

These risks and harms follow years of warnings and precedents. The widespread use of pre-AI digital technologies in teaching and administration has already helped obscure educational decision-making and allowed student data to be exploited for non-school purposes. Without effective public oversight, the introduction of opaque and unproven AI systems and applications will likely intensify these problems and create many more.

The integration of AI into our computers, phones, cars, and homes is well on its way. Inevitably, this integration will impact schools, too—but we have choices about the nature and
extent of these changes. The core elements of learning and the types of interactions that take place in schools should be thoughtfully considered, with the resulting policies and practices following from deliberation decisions.

Advocates for AI claim that it will transform teaching and learning for the better. This will not happen, however, if integrating AI into the pedagogy of schools degrades the relationship between teachers and students. It will not happen if the AI imposes a rigid mechanistic model of instruction, corrupts curriculum with misinformation, and biases consequential decisions about student performance. It will not happen if integrating AI into schools’ administrative processes locks schools and districts into an expensive “stack” of corporate tech systems for many of their everyday operations, with the result that funds available for other uses—including the teachers who can develop deep connections with nation’s students—are increasingly shifted to corporate vendors. It will not happen if the AI exacerbates violations of student privacy, increases surveillance, and further reduces the transparency and accountability of educational decision-making.

All of these harms are likely if lawmakers and others do not step in with carefully considered regulations.

As existing school-focused products are updated to include newly created AI-based products, the immediate danger facing schools is not a future apocalypse as predicted by technology industry leaders. Rather, the danger is that—in the absence of responsible development, proper evaluation, or regulatory oversight—untested, opaque AI models and applications will become enmeshed in routine school processes.

The evidence set forth in this policy brief alarms us. The adoption of largely unregulated AI systems and applications would, we conclude, force students and teachers to become involuntary test subjects in a giant experiment in automated instruction and administration that is sure to be rife with unintended consequences and potentially negative effects. Allowing AI to become inextricably enmeshed in school processes and procedures invites both short- and long-term harms, because once it is enmeshed in those processes the only way to disentangle from it would be to completely dismantle those systems.

To forestall the far-reaching dangers posed by hasty implementation of AI in pedagogical and administrative systems, we recommend that school leaders pause adoption of AI applications until policymakers have had adequate time to fully educate themselves about AI and to formulate legislation and policy ensuring effective public oversight and control of its school applications. Any development of AI for schools should be conducted under “responsible AI” frameworks, in partnership with schools. In addition, we recommend that:

**Federal and state policymakers:**

- Stop promoting AI as a way to transform and modernize schools’ pedagogical and administrative practices.
- Prohibit schools from adopting AI-based educational applications until appropriate regulatory structures are established.
- Adopt regulations that prohibit schools from using any technology, including “black
AI models, whose workings are not transparent to state regulators, unless they have provided those regulators a well-developed rationale and justification for why the particular technology is the only way to achieve a clearly defined and valid school purpose, and how it offers an improvement over existing education practices.

- Reduce the pressure on schools to adopt AI in their administrative systems by reducing the data reporting requirements placed on schools and teachers.

**State policymakers:**

- Establish an independent government entity charged with ensuring the quality of digital educational products used in schools. Charge this entity with reviewing and approving the pedagogy and programming of any digital educational product a school proposes to use, both prior to implementation and periodically thereafter. Require that the programming of any digital educational product—explicitly including products that incorporate AI—used in schools be transparent and amenable to review.

- Create classroom contexts that allow teachers to spend more time with their students, such as enacting legislation to limit class size, so that teachers are not pressured to find ways to keep students quietly occupied with digital products.

**District policymakers:**

- Refrain from adopting AI-based educational applications until:
  - Strict transparency and accountability requirements are put in place as part of an overall technology accountability plan.
  - The public has been provided with compelling evidence or thoughtful and clear explanations as to how those applications are an improvement over other education practices that do not require digital technology.
Time for a Pause: Without Effective Public Oversight, AI in Schools Will Do More Harm Than Good.

Ben Williamson
University of Edinburgh
Alex Molnar and Faith Boninger
University of Colorado Boulder

March 2024

II. Introduction

Public education is a public and private good essential to democratic civic life. The public must, therefore, be able to provide meaningful direction over schools through transparent democratic governance structures. And yet artificial intelligence (AI), with little public oversight, is on the verge of becoming a routine presence in schools. The weight of the available evidence suggests that the current wholesale adoption of unregulated AI applications in schools poses a grave danger to democratic civil society and to individual freedom and liberty.

Background

In the spring of 2023, digital technology titans warned that artificial intelligence posed “profound risks to society and humanity,” and called for regulation of its development and a pause in implementation of related applications. But concurrently, major players such as Google, Microsoft, Meta, and Amazon raced both to integrate AI into their platforms and to fend off regulation. Tech industry marketing went into high gear and soon the popular press overflowed with industry hype and speculation about AI’s potential and pitfalls. The effect of all the attention-grabbing predictions, self-interested corporate behavior, flamboyant marketing claims, and uncritical reporting has been to obscure the immediate dangers posed by AI’s rapid implementation. Given this backdrop, it is not surprising that rhetoric promoting AI’s alleged ability to positively transform teaching and learning has dominated discussions about its impacts on education.
School administrators and teachers already use an array of digital educational technologies in teaching and management. Their use has increasingly obscured educational decision-making, made a mockery of student privacy rights, and allowed student data to be exploited for non-school purposes. In the absence of effective public oversight, the introduction of AI systems and applications in education will likely intensify these problems and create many more.

As existing school-focused platforms and applications are updated to include AI, the immediate danger facing educators is not a future apocalypse. Instead, the danger is that AI models and applications will become enmeshed in school processes and procedures in ways that allow private entities to increasingly control the structure and content of public education, to reinforce surveillance practices, and to amplify existing biases and inequalities. For decades, academic researchers have worked on AI models for use in schools. Today, however, it is commercial enterprises that are aggressively pushing AI (and its attendant risks) into classrooms.

The campaign to promote AI in education follows the logic of a half century of commercial, political, and ideological efforts to privatize and commercialize education. Given this logic it is not surprising that, despite the known dangers, corporations, private researchers, and governments are aggressively promoting the use of AI before a statutory and regulatory framework has been put in place to ensure that AI programs are transparent and subject to effective public scrutiny and control. This puts schools under tremendous pressure to accept AI as an inevitable upgrade to existing processes.

Computer scientists and software developers focus primarily on technical engineering questions and corporate leaders and investors prioritize profit over the common good. Nevertheless, educators are being asked to trust that these people, who have no educational expertise and who stand to financially benefit when AI is used in schools, are best suited to imagine and lead educational transformation.

**III. Review of the Literature**

The term “artificial intelligence” (AI) was first introduced by computer scientists in the 1950s, though many of its underlying mathematical processes and mechanistic procedures can be traced back to early models of computing and manufacturing in the 1800s. Standard definitions of AI usually refer to computers performing tasks that only humans could normally do. Today, however, AI has become a slippery term without a widely agreed-upon meaning. This enables marketers to apply the label “AI” to almost any digital process or product they are selling.

**Development of Artificial Intelligence—From Rule-Based Systems to Generative AI**

AI has been developed over many decades by people and corporations with distinctive agendas and strategies. Their efforts are poised to shape AI implementation in schools, especially since AI development and marketing efforts are rapidly expanding.
Three principal phases of AI’s evolution began in the 20th century, when scientists and technicians focused on building systems that followed a sequence of pre-set rules, or algorithms, derived from the knowledge of experts. Eventually, over the past two decades, a machine learning approach emerged. This approach involves using sophisticated mathematical processes (known as learning algorithms) to analyze massive amounts of data (known as big data) to identify trends or commonalities within it. On the assumption that the vast amount of data analyzed confers validity on the trends identified, such programs generate predictions about behavior in complex events and phenomena—including human behavior. The success of such analyses, however, has been mixed. For example, one study asked six teams of data scientists to predict children’s life outcomes with huge quantities of big data and cutting-edge machine learning tools. None of them came close to a reasonable level of accuracy, casting serious doubt on the use of AI in social policy areas like education.

Since late 2022, a further evolution of machine learning, generative AI, has become an object of intense public, media, and political interest. Generative AI deploys even more complex learning algorithms to create original text, images, and audio from data collected from the web or other sources. This line of research has produced a series of program models often now called foundation models because they can be adapted for highly diverse purposes. For example, generative AI large language models predict which word is most likely to follow the words preceding it by breaking down words into a sequence of numbers and then calculating the most probable response.

Large language models have provoked widespread excitement because their promoters have promised productivity gains for businesses as AI programs relieve workers of everyday writing tasks. Similar support for teachers and students is also a promised benefit. It is, however, questionable that these promises will be realized in practice. For instance, while models often generate correct responses, computer programs cannot understand what words mean. Therefore, the accuracy of their responses or the value of their responses in any given circumstance cannot be assumed. Indeed, at the moment, AI is often “stupid.” It invents facts, mangles the results of analyses, destabilizes information sources and produces dangerously wrong assertions about matters of social, public, and cultural importance. When Google demonstrated its Bard language application as an educational tool by getting it to answer questions about the James Webb telescope in 2023, for example, the model gave a false factual response, prompting a market plunge in Google’s company value.

Concerns have also emerged about AI development processes. For technical and commercial reasons, AI models are not transparent—they are not publicly explained in any detail. For example, an analysis of 10 leading foundation models found limited information about where data for analysis came from, how much computing was necessary to create the models, and what specifics were embedded in analytical algorithms. Many machine learning models are black box models, meaning that their mechanisms are said to be too complicated to explain or not explainable at all. Others are hidden from public view by proprietary rights accorded to corporations.

Black box foundational AI models are key to large technology corporations’ plans to expand their proprietary models into all sectors, to grow global market share, and to generate maximum profit. In the absence of effective public oversight and regulation, running gen-
ervative AI programs is currently only possible by using “Big Tech” companies’ databases, high-powered computing capabilities, and financial resources. This makes it likely that the proprietary AI models of a few corporations will become the foundation for the vast majority of AI applications developed.\textsuperscript{53} Meanwhile, regulators and lawmakers will be left struggling to respond by creating a patchwork of after-the-fact regulatory protections.\textsuperscript{54}

\textbf{AI in Education}

Since the 1960s, scientists and technology companies have explored ways to apply AI in education. \textit{AI in Education} (AIED) is a major field of research and development.\textsuperscript{55} The AI applications being promoted to schools today were preceded in the 1960s and 1970s by “Intelligent Tutoring Systems” and “Computer Assisted Instruction” systems.\textsuperscript{56}

Since the early 2000s, researchers have gathered, stored, and analyzed massive quantities of educational data with the intention of informing institutional and instructional strategies.\textsuperscript{57} These approaches are now routinely considered synonymous with AIED, and have also been rapidly commercialized by the ed tech industry.\textsuperscript{58} Most AIED applications employ big data and machine learning to produce various predictions and automated actions—such as predicting that a student may fail an assessment or creating a “personalized” intervention intended to produce a desired learning outcome.\textsuperscript{59}

Research on AI in education has developed and tested various approaches and reported modest effectiveness on measurable learning achievement—performance on quizzes and tests, for example.\textsuperscript{60} Current excitement about its potential is motivating both public and private sources to generously fund researchers trying to find ways to improve learning outcomes using AI.\textsuperscript{61}

However, the assumption that AI in education can be understood primarily as a technical matter best addressed by scientists and companies is increasingly challenged by researchers who argue that a narrowly technical perspective may lead to both bad policy and bad pedagogy.\textsuperscript{62} They point out that AI exists in social, economic, and political contexts that shape its development and uses.\textsuperscript{63} How AI is adopted by different educational stakeholders (including AIED researchers, ed tech entrepreneurs, corporate leaders, and policymakers) will have significant implications for its use in schools.\textsuperscript{64}

The fact that entrepreneurs and corporations funded by venture capital and private equity are rushing to promote AI in education will inevitably narrow possible applications to those preferred by stakeholders with financial interests.\textsuperscript{65} Small-scale ed tech start-ups and Big Tech corporations alike see AI as an opportunity,\textsuperscript{66} leveraging popular hype to market such education products as personalized learning programs, automated lesson plan generators, and AI tutoring chatbots, called “tutorbots,” to schools.\textsuperscript{67} Compelling evidence for the effectiveness of tutorbots in education remains scarce,\textsuperscript{68} though this does not prevent entrepreneurs and researchers from proclaiming their usefulness.\textsuperscript{69}

Policymakers routinely invoke AI rhetorically, calling on schools to embark on “digital transformation,”\textsuperscript{70} often with little attention to social, economic, legal, or ethical implications.\textsuperscript{71} These calls dovetail with existing political priorities on performance monitoring, account-
ability, efficiency, and effectiveness—all of which require extensive collection of data about students. Although systems of test-based accountability have existed in schools since the 1990s, they will expand and intensify as AI is used to continuously monitor and assess student learning. As a result, commercial AI systems will increasingly serve as private actors in public education as schools, districts, and governments relinquish key tasks, functions, and responsibilities to third-party technology vendors.

Existing and potential uses of AI in education are not merely innovative technical add-ons to teaching and learning practices or engineering solutions to schools’ existing pedagogic and administrative problems. Rather, AI in education has been spurred by multiple forces: longstanding efforts by scientists to measure, predict, and support learning processes and outcomes; commercial aspirations to profit from selling products to schools; and the political objective of being perceived as having improved school efficiency and accountability while cutting costs. As things currently stand, these ambitions have begun to coalesce into a vision of AI-driven schooling in which commercial products assess student learning, automate teaching, and make decisions about student progress.

**Inadequate Research Base**

Despite the extensive research in the field of AI in Education (AIED) and the burgeoning research on machine learning, there is remarkably little evidence to support claims of AI’s ability to “transform” schools. While AIED researchers have produced many research findings, their studies tend to focus primarily on measures of individual student engagement and performance (assessed by standardized achievements tests), or on “engineering” problems such as designing increasingly sophisticated algorithms and enhancing machine learning effectiveness.

Overall, AIED studies tend to find ambiguous results, lack independence and scale, and fail to address more fundamental questions about educational goals. AIED research therefore often promotes a view of education transformation as improving measurable individual outcomes despite very limited evidence that AI “works.” In effect, such studies reduce well-researched and nuanced theories of how humans learn to whatever can be made into a mathematical model (however complex), and they ignore the contested terrain of exactly which goals and curriculum public schools should embrace. Moreover, claims that AI can solve major educational problems—such as lack of qualified teachers, student underachievement, and educational inequalities—rely to a considerable extent on conjecture rather than evidence.

Even more problematic are the serious methodological flaws in machine learning research that call into question the validity of hundreds of studies. The nature of the flaws, in general, leads toward “over optimism” with respect to the usefulness and value of machine learning applications in a variety of fields. These findings are particularly concerning because they call into question not only commercial marketing claims, but also the scientific evidence base supporting the widespread implementation of AI systems in all sectors, including education.
Finally, because of the very high computing costs associated with running machine learning models, most researchers have to rely on systems from the dominant AI companies themselves in order to conduct research—the same corporations that often fund AI studies. This makes research dependent on corporate resources, funds, and business practices, giving AI firms considerable influence over not only AI development, but also the academic research that depends on their systems. It also compromises an important part of the research process, which is reproducing findings to verify their validity. When a company changes or stops supporting a particular model, researchers cannot reproduce studies conducted earlier. This renders the research base unstable and unverifiable—and thus unusable as a basis for assessing subsequent models.

IV. Recent Developments

In November 2022, OpenAI released ChatGPT, accelerating the race to develop and market generative AI platforms and applications. Companies that have developed artificial intelligence foundation models, including OpenAI, Google, Meta, and Amazon, intend to expand rapidly and “scale up” in every sector they enter—including education. While media attention has focused on students using ChatGPT, these companies are swiftly expanding AI in education in several ways:

1. Selling access to digital systems to schools. For example, Amazon sells schools and districts access to cloud computing facilities, enabling them to use its AI systems to analyze institutional and student data.

2. Adding AI features to products that schools already use. Google, for instance, has begun introducing AI into its Workspace suite for schools, which includes the Classroom platform used by schools worldwide.

3. Integrating AI applications into new or upgraded products. OpenAI, for example, partners with ed tech companies to integrate its language models into services it promotes as “AI teaching assistants,” and is also exploring its own educational applications of ChatGPT.

4. Building AI into services such as search engines and other everyday applications commonly used in classrooms. For example, Microsoft introduced new AI applications in its Office software, based on OpenAI technologies, which are promoted for educational use.

Current promotion of AI in education focuses largely on the pedagogical uses of AI applications that provide automated language and image producing capabilities. For students, reporting has, for example, focused on their use of automated tools, such as ChatGPT, to write assignments, and of personalized learning “tutorbots” to mimic a one-on-one tutoring experience. For teachers, it has focused on their use of AI “assistants” to create lesson plans, develop grading systems, or review student progress, among other tasks. Proponents assert that such AI applications offer students personal assistance in learning and offer teachers time-saving support. Entrepreneurial educators have produced guidebooks and training materials for teachers, and OpenAI has launched a Teaching with AI guide to train teachers to use its applications.
Accepting proposed benefits as real, international and governmental organizations have supported the use of AI products in schools. For example, the Organisation for Economic Co-operation and Development (OECD), worried that AI will soon outperform humans on many cognitive tasks, has called for the urgent modification of formal education systems so students can learn skills to complement AI rather than skills for tasks that could soon be automated. The US Department of Education’s Office of Educational Technology encourages teachers to involve themselves in developing and evaluating AI applications for education. This all results in teachers and schools being pushed to accommodate untested and opaque commercial AI applications or risk being “left behind.”

Recent regulatory proposals show little sign of slowing the rapid advance of AI into schools. An executive order released by President Biden in October 2023 outlined directives to, among other things, increase federal oversight of foundation model testing, protect data privacy, and “promote innovation and competition.” It mandates the Department of Education to create an “AI toolkit” incorporating “appropriate human review of AI decisions, designing AI systems to enhance trust and safety and align with privacy-related laws and regulations in the educational context, and developing education-specific guardrails.”

However, the fact sheet accompanying the executive order described the Department of Education’s mandate differently as: “shap[ing] AI’s potential to transform education by creating resources to support educators deploying AI-enabled educational tools, such as personalized tutoring in schools.” Although the Department of Education will base its toolkit and guidance on its goal to “keep humans in the loop,” two assumptions guide its approach: that “learning” is defined as and limited to those things that digital programs can measure and that AI programs can and should “optimize” learning as so defined.

V. Discussion and Analysis

As noted earlier, despite the risks of rushing untested tech into classrooms and the lack of implementation criteria or regulatory controls, schools face unrelenting pressure to “modernize” by adopting artificial intelligence. While AI applications are marketed as ways to address teaching and learning problems and to streamline school administrative processes, they carry with them all the limitations, problems, and risks inherent in the AI models used to run them. Rushing AI into schools increases the likelihood that these technologies will reproduce or intensify many problems, making any potential benefit less significant than the potential harms.

Two heavily promoted applications illustrate the perils of uncritical adoption of AI in schools: tutoring chatbots, based on large language models, which promise to personalize and automate teaching; and adaptive learning platforms, which use large quantities of student data to make predictions, customize classroom resources and activities, and automatically intervene in pedagogical processes. The dangers that such products pose include costing teachers more time than they save, artificially restricting the definition of “learning,” sideling teacher expertise and relationships with students, introducing curricular misinformation, and increasing bias and discrimination in classrooms and schools. In light of such dangers, before expanding or allowing AI applications into schools, policymakers and

http://nepc.colorado.edu/publication/ai
education leaders should consider what it would mean to use AI responsibly, whether its potential benefits outweigh the costs, and whether its adoption in education is truly inevitable.

**Dangers in Teaching and Learning**

**Restricted Teaching and Learning**

Khan Academy has developed several offerings for technology-based schooling, expanding them beginning in March 2023 with heavy marketing of its Khanmigo “tutorbot” to parents, teachers, and school districts. Khanmigo’s marketing claims that the tutorbot can engage in personalized and “conversational” pedagogic interaction with students, thus enhancing “personalized learning.”

Khanmigo incorporates OpenAI’s GPT-4 language model to generate customized educational content, lesson plans, and assessment, and to perform such other tasks as monitoring student progress. Its marketing materials promise one-on-one “virtual tutoring” for students and a “personal teaching assistant” for teachers. These materials also claim that Khanmigo can understand and respond individually to students’ academic and career goals and save teachers time by providing a content library for lesson planning, reports of student progress (along with “recommendations for what to do with that information”), and a “personal concierge service” that can help teachers “just do more.”

Such marketing suggests Khanmigo can make decisions that mimic a real teacher's thinking and responses, and so satisfy the needs both of students who require a teacher’s attention and teachers who cannot offer that attention because of other demands on their time. It also reveals assumptions about the nature of teaching and learning embedded in the programming that are unavoidably imposed on real teachers, students, and communities that work with the product.

Algorithms in personalized learning applications implement a narrow understanding of learning and a highly constrained pedagogic model. “Learning” in these applications typically consists of students engaging in computer-based activities and producing “correct” or acceptable answers within a limited range of predetermined responses. The applications present teachers with reports of student “progress” on the measurable tasks and suggest strategies to improve students’ “performance.” Any school adopting these applications tacitly accepts that learning can largely be reduced to a narrow range of behavioral responses and tacitly rejects more complex child-centered understandings of learning.

The theory of learning built into AI chatbot applications like Khanmigo prioritizes data-based numerical profiles of students—considered “objective” measures of performance—while undermining teachers’ ability to make professional judgments about their students and devaluing teachers’ subjective experience, subject matter expertise, classroom interactions, and contextual knowledge of a given child, class, or social setting.

A supposed advantage in such AI-based programs is that they reduce administrative burdens on teachers and allow them more time to teach. Khanmigo’s advertising promises, for example, “your evenings are yours again without compromising quality.” This is persuasive
marketing for overburdened teachers, who might welcome outsourcing their overwhelming administrative duties to a computer program, or might be relieved to occupy students with technology that theoretically supports their learning while they wait for teacher time in an overcrowded classroom.

Delegating tasks to a tutorbot, however, inserts a digital intermediary between students and teachers who are pedagogical experts in their fields and who know their students and understand their context. Moreover, despite generative AI’s humanlike communication, it is limited to responding to prompts and queries that fall within a product’s established parameters. As a result, teachers must devote time to understanding and mastering such limitations and then teach students how to ask questions and verify responses. Even as AI automates some administrative tasks, then, it is likely to introduce other time-consuming pressures and burdens on teachers.

Rather than reducing demands, AI of the kind currently promoted by private commercial enterprises can add further complexity to teachers’ workloads by ultimately positioning them as servants of the technology tasked with ensuring its smooth classroom operation. Reconfiguring the core systems of education—instruction, curriculum, and assessment—to accommodate AI will demand laborious efforts by educators to adjust their professional practices. At the same time, it will deny them a voice in determining whether proposed changes have real value and should or should not be implemented. Academic AIED research and development, informed by partnership with teachers and students, will struggle to be heard above the current promotional and speculative rhetoric around commercial AI applications for schools. AI therefore presents a critical challenge to the institutional autonomy of schools’ decision-making processes, and particularly to teachers’ autonomy to make professional decisions about their pedagogic practices.

Curricular Misinformation

A key challenge of AI for teachers is that generative AI technologies are trained to produce text that seems convincing even though it might contain false information. Applications that use generative AI to support teachers’ lesson planning and resource generation, for example, could flood the classroom with misleading inaccuracies or false information. This is not a problem likely to be resolved, because the data such programs depend on may degrade further as automated content spreads across the web. The danger is that the information environment will be overrun by AI-generated text, making it impossible to ascertain the authority or authenticity of any online source, and therefore rendering online sources useless or misleading for educational purposes.

As has been documented in both language- and image-generating AI, instabilities and errors in underlying data make AI chaotic and unreliable. No doubt this is why Khanmigo recommends that users not rely on its responses but rather verify its information using such resources as “textbooks, articles, or other trusted sources.” And why it also restricts the time students can spend interacting with the tutorbot, noting that “extended interactions are more likely to lead to poor AI behavior.” These recognized limitations beg the question: Why spend time using this application or any AI application when it seems obvious
their use will result in more work and very likely undermine effective teaching?\(^\text{139}\)

Outstanding questions remain, then, about ed tech and Big Tech market leaders becoming powerful gateways to online learning content. Do their AI applications improve the quality of information taught in the classroom? How much should the content produced by AI, having been trained on material scraped from the web by companies like OpenAI, be trusted by teachers and students? What are the implications of students encountering “poor AI behavior” from a tutorbot? It is also not clear whose responsibility it must be to check the quality of the content produced by AI. Khanmigo implicitly makes already overworked teachers responsible for doing the hidden labor of checking for errors and providing feedback that the product’s programmers can then fix one-by-one.\(^\text{140}\) This scenario makes it very likely that such errors will be overlooked and that the accuracy of information that students encounter via AI will degrade over time.\(^\text{141}\)

AI poses a real threat to the accuracy of school knowledge and thus to the validity of curriculum materials. In a context where students and many teachers already source content online (for example, through sites such as Teachers Pay Teachers\(^\text{142}\)) it may become increasingly difficult to tell an authoritative and accurate source from a plausible but fallacious one produced by AI. This is likely to become an even bigger problem with OpenAI’s release of “GPTs”—versions of ChatGPT that can be customized by users—as individuals are able to create their own educational AI applications and share them freely online, with few quality checks and controls.\(^\text{143}\)

**Potentially Amplified Bias and Discrimination**

Because AI models are trained on either internet data or historical data, they incorporate biases that can transfer to their educational applications.\(^\text{144}\) For example, when ChatGPT was released in November 2022, educators expressed anxiety about students cheating on written assignments, prompting ed tech companies to develop automated AI detectors. The makers of Turnitin, a product already used internationally to detect student plagiarism, added AI detection functionality in early 2023, arguing it would be able to detect distinctive markers of AI-generated text.\(^\text{145}\) However, independent studies have found that such AI detectors are prone to error, leading to a surge in false accusations of cheating.\(^\text{146}\) Such accusations are disproportionately biased against non-native English speakers, who tend to write in simpler sentences that AI flags as suspicious.\(^\text{147}\)

Similarly, the automatic AI essay grading programs used in many states are prone to bias against certain demographic and ethnic groups, falsely awarding high grades to work that features sophisticated language and structure, regardless of the meaning or quality of the writing.\(^\text{148}\) This is because AI programs do not “know,” in any real sense, what makes for a good or bad essay, but can only search for patterns that correlate with higher or lower human-assigned grades. Likewise, GoGuardian, which uses algorithms and AI to monitor students’ social media and web browsing activity for “suspicious” content or behavior, routinely flags multiple categories of non-explicit material as harmful or dangerous—including material on general educational sites.\(^\text{149}\)

AI plagiarism detectors, automated essay graders, and student monitoring platforms are
built on the false assumption that automated analyses of data provide objective accounts of students’ practices and behaviors. They replace human discretion with automated suspicion scores generated by machine learning algorithms, with subsequent judgments on assignments, grades, and behaviors that are difficult, if not impossible, for students to challenge. Thus, if AI is integrated into the existing structures of schooling, it is likely to worsen rather than reduce many of the inequalities that characterize contemporary education. Opaque machine learning systems will make hidden decisions affecting students’ courses of learning, outcomes, and prospects based on unknowable mathematical calculations.

Dangers in Administration

Increased Costs

Learning management systems already used in many schools, such as Google Classroom, Blackboard, and Canvas, are beginning to integrate AI into their platforms. Google Classroom, with its suite of nominally “free” software and low-cost Chromebook hardware, dominates the market. It has already announced the launch of AI-based adaptive learning add-ons to Classroom, with associated additional costs for schools, as well as plans to upgrade Classroom further with generative language AI. “Practice Sets” is Google’s AI-based adaptive learning system for education, and “Duet AI” is its “collaboration partner” for teachers. In addition to any pedagogical implications associated with using Google Classroom, its integration of further AI and automation into many aspects of school functioning also carries potentially significant administrative implications.

The most significant of these is to obscure the rationale for administrative decisions about critical institutional issues when decision-making is ceded to opaque machine learning systems controlled by tech firms. Google Classroom, for example, integrates with hundreds of other ed tech products and can synchronize with a school’s student information systems. It offers Google cloud services such as single sign-on, identity management, and device management, as well as plagiarism detection, automated grading, teaching templates, student grouping, and administrative analytics to facilitate “data-driven decisions.” Such management systems facilitate the transfer of control of schools from the public to private corporations by acting as central conduits through which all of a school’s digital activities must pass—making it hard for educators or administrators to see how any decisions based on the data have been made.

Because running AI is costly, the use of AI programs in schools will necessarily require schools to pay for operating costs for an increasing number of pedagogic and administrative AI applications. The promise that AI can save schools money by reducing staffing costs is likely illusory, as schools will probably be required to pay costly fees for accessing AI facilities. In other words, rather than saving money, administrative applications are more likely to shift existing funds to monopolistic technology providers.

Khanmigo and Google Classroom already illustrate how this works. Khan Academy, when it provides Khanmigo to districts, currently charges those districts $60 per student for annual use, citing high computing costs associated with OpenAI’s GPT-4 as the justification for the
Likewise, districts must also pay for Google Classroom’s AI upgrades. To access its latest adaptive learning application, Practice Sets, they must switch from the free basic offering to a for-fee premium package. In other words, tech firms are extracting value from school budgets to defray the high computing costs associated with AI (and grow company value).

**Increased Threats to Student Privacy**

AI applications collect and aggregate data in order to function. In so doing, they normalize digital surveillance and privacy invasions in school. In practice, education technology companies use applications like Google Classroom to routinely collect as much data as possible, well beyond that required to perform their assigned tasks.

Although proponents of using AI in education tend to emphasize the efficiency of data-driven administrative systems, privacy-related threats to equity are inherent in it. This is because AI models are built using massive data sets that can be used to profile, compare, and assess individuals who are then subject to potentially discriminatory decisions based on “statistical dossiers” of their personal lives. Thus, a significant danger of digital technology in general, and of the privacy-invasive model of AI in particular, is that they can reproduce and amplify existing forms of inequality in education by using datasets containing examples of historic bias and discrimination. For example, if a big data set indicates that certain marginalized groups have underperformed historically, then a software application may be biased against individuals from such groups in the future, singling out and targeting them as “at-risk” and closing down or limiting their opportunities to access information and resources.

Moreover, school data systems are vulnerable to breaches, hacks, ransomware, and denial-of-service attacks. A data breach at the student-tracking ed tech company Illuminate, for example, compromised the educational data of at least a million public school students and prompted New York City’s Department of Education to ask schools to stop using Illuminate’s products. School data systems feature highly detailed and intimate student information, including personal and demographic data, grades, attendance, behavioral information, and other confidential information. Increasing AI capacity in ed tech products may exacerbate these vulnerabilities, as student data are collected at even greater scale by a wide range of companies—including AI companies—that offer only vague data privacy protections.

**Reduced Transparency and Accountability**

Finally, enabling AI to play a role in school administration will reduce the transparency and accountability of decision-making. Many digital products already used in schools are neither transparent nor accountable because current law and regulation allows companies to shield the inner working of their products behind proprietary protections.

AI is even more opaque than other digital programs. Black box machine learning and AI models are so complicated that their outputs are often impossible to explain or interpret. Although in many cases simpler and more accessible statistical models can produce equally
accurate results, companies benefit from selling access to proprietary models that require customers to trust the systems and simply accept being unable to verify results.\textsuperscript{174} If the system makes a mistake, it might never be identified or redressed and the public suffers the consequences. For example, the facial identification systems used for remote testing often fail to accurately identify individuals or mistakenly flag student behaviors as suspicious, but they are very hard for students to challenge.\textsuperscript{175}

In high-stakes decision-making in a sector like education, allowing such impenetrable models to assume responsibility for key administrative procedures necessarily means the creation of schools in which school leaders and teachers will be unable to exercise judgment, provide a rationale, or take responsibility for classroom and institutional decisions.\textsuperscript{176}

**Considerations for the Future**

**Is AI Development Responsible?**

The rapid creation of AI applications for schools raises the urgency of prioritizing ethics, student rights, and social responsibility in their development.\textsuperscript{177} *Responsible AI* development would ensure that products are safe and trustworthy, designed to benefit people, communities, and society, and mitigate harms.\textsuperscript{178} As yet, there is little indication that such values are adequately addressed in education applications.\textsuperscript{179} Unfortunately, academic AIED researchers have tended to ignore them or delegate addressing them to the educational tech industry and policy centers.\textsuperscript{180} This complacency—along with the money and power held by commercial actors—enables commercial rather than educational imperatives to guide the development of AI and furthers political interests promoting relentless testing and school surveillance.\textsuperscript{181}

Responsible governance would require the companies developing AI to commit to transparent and responsible product design, and also to monitoring, understanding, and mitigating the continuous impacts of AI in various contexts. Of particular concern is the automation of decisions with “irreversible and severe consequences.”\textsuperscript{182} For example, technologies to identify emotions are currently being developed to assess if a person is lying and cheating.\textsuperscript{183} These technologies are inherently inaccurate, however, and an inaccurate judgment that a student has cheated or that a witness is lying could have dire consequences for their lives. Responsible AI governance might lead to delaying or indefinitely pausing development of such technologies.

Although several responsible AI initiatives have produced principles, frameworks or checklists for safe and trustworthy AI development and accountability,\textsuperscript{184} these agendas can be manipulated through various forms of industry lobbying and efforts to water down their scope or possibilities of enforcement.\textsuperscript{185} Expanding responsibility for product safety to include the wide range of people or organizations that build and use AI—rather than leaving it to technicians and business alone—would mitigate such dangers.\textsuperscript{186}

Among the many obstacles to the implementation of responsible policies governing AI is their cost. The goal of profit-seeking business is to shift to the public as many costs as pos-
sible while garnering the highest possible private rate of return on investments. Public oversight of AI necessarily entails either public ownership or a comprehensive regulatory regime adequately financed to achieve its mission. The question is, where will the money come from?

Moreover, the required regulation flies in the face of 50 years of policy devoted to deregulation and privatization. It would demand a fundamental rethinking of the government’s relationship to commercial interests. Such rethinking would, without a doubt, be attacked by self-interested parties as not only too costly but also as stifling innovation and promoting inefficiency. While these arguments may be relevant in individual circumstances, they are neither generally nor self-evidently true.

From the perspective of education, responsible governance of AI therefore entails significantly more commitment than the simple principles of responsible development issued by industry. It also requires costly and ongoing monitoring of the effects of AI in classroom contexts. It may also require delays and indefinite pauses in development where warranted—such as, for example, in cases where commercial AI providers seek to introduce products into schools with insufficient evidence that they produce beneficial outcomes, or when those products automate professional judgement with potentially negative consequences, or when they inadequately address questions of AI ethics directly relevant to education.

**Is AI Inevitable?**

AI products are moving into schools at dizzying speed. As we have noted, this is in part the result of the pressure on schools to “modernize” by adopting the latest products that the technology industry offers. There is already a consensus of sorts that the move to AI is inevitable. The director of educational technology at Newark Public Schools made the case to the *New York Times* when he explained why his district adopted Khanmigo: “It’s important to introduce our students to it, because it’s not going away.”

The de facto requirement that students serve as a technology company’s experimental subjects might be explained by the initially low entry cost for school districts. Struggling districts, especially, might be willing to gamble that a technological innovation might turn things around for their students. However, before placing that bet it would be valuable to first ask some fundamental questions. Computer scientist Joseph Weizenbaum posed such concerns 50 years ago, essentially arguing that no technology—including AI—should be implemented unless we know that it is both necessary and good.

**Is AI Necessary?**

The technology industry has tried for a century to profit from schools. In the 1920s, creators of mechanical “teaching machines” claimed they could help students work at their own individual pace and ease administrative demands on teachers. The same claims are now attached to digital personalized learning platforms designed to “meet students where they are,” whenever and wherever they are ready to learn. And, as we have noted, technology companies are eagerly incorporating AI into personalized learning and other platforms to
conclude virtual “tutoring,” generate lesson plans and tests, and collect and organize data for administrative decision-making.

AI programs may perhaps, in some well-defined circumstances, efficiently perform useful tasks. However, the question of whether these tasks can be done as well as or better without using AI programs and accepting their negative consequences is unclear—and rarely discussed. We cannot now know which, if any, AI programs offered to schools can legitimately be said to be necessary.

Is AI Good?

AI programs are unreliable, tend to hallucinate and break, and will almost certainly produce manifold unintended consequences. Instability is inherent in AI programs, in part because they quickly and unpredictably change as they “learn” from how they are used. Very few studies in machine learning research address values-based issues such as societal need or negative potential, instead prioritizing commercial and engineering values like system performance, efficiency, and novelty. There is a danger that AI programs will propagate these built-in commercial and engineering values without the public ever having been considered or having agreed to adopt them.

AI algorithms are neither objective nor neutral. To further complicate matters, even if it’s clear which assumptions govern AI training data and how a program weights the data it processes, when programs actually run they produce results based on a vast number of mathematical probabilities that are unknowable to both their creators and their users. As a result, even a person with knowledge of a black box AI program’s initial coding cannot explain how it produced its results. Not surprisingly, machine learning inevitably produces outcomes that may be incomprehensible, untrue, or incorrect in a variety of unknowable ways. If such systems are used, neither teachers nor administrators will be able to understand, explain, or justify the conclusions the programs reach, much less audit or document their validity.

We are now on the verge of creating a world in which students set on a particular academic path or denied advancement by an AI-based learning application will have absolutely no way to challenge the judgments made about them. Nor will their teachers and administrators have a meaningful way to respond to students and parents who challenge AI decisions. As AI captures the pedagogy and administrative processes of schools, as it becomes embedded in the full range of school routines, it therefore poses a direct and powerful threat to the democratic governance of schools. In this sense, implementing AI in schools cannot be considered “good.”

We cannot conclude with any confidence that AI is, in general, either necessary or good for students, teachers, or schools. On the contrary, the weight of the available evidence suggests that the wholesale adoption of artificial intelligence in schools poses a grave danger to democratic civil society and to individual freedom and liberty. The adoption of opaque, untested, and largely unregulated AI systems and applications would force students and teachers to become involuntary test subjects in a giant experiment in automated instruction and ad-
administration that is sure to be rife with unintended consequences and potentially negative effects. Allowing AI to become inextricably enmeshed in school processes and procedure invites disaster, because once it is enmeshed in those processes the only way to disentangle from it would be to completely dismantle those systems.

VI. Recommendations

To forestall the far-reaching dangers posed by hasty implementation of AI in pedagogical and administrative systems, we recommend that school leaders pause adoption of AI applications until policymakers have had adequate time to fully educate themselves about AI and to formulate legislation and policy ensuring effective public oversight and control of its school applications. Any development of AI for schools should be conducted under “responsible AI” frameworks, in partnership with schools. In addition, we recommend that:

Federal and state policymakers:

- Stop promoting AI as a way to transform and modernize schools’ pedagogical and administrative practices.
- Prohibit schools from adopting AI-based educational applications until appropriate regulatory structures are established.
- Adopt regulations that prohibit schools from using any technology, including “black box” AI models, whose workings are not transparent to state regulators, unless they have provided those regulators a well-developed rationale and justification for why the particular technology is the only way to achieve a clearly defined and valid school purpose, and how it offers an improvement over existing education practices.
- Reduce the pressure on schools to adopt AI in their administrative systems by reducing the data reporting requirements placed on schools and teachers.

State policymakers:

- Establish an independent government entity charged with ensuring the quality of digital educational products used in schools. Charge this entity with reviewing and approving the pedagogy and programming of any digital educational product a school proposes to use, both prior to implementation and periodically thereafter. Require that the programming of any digital educational product—explicitly including products that incorporate AI—used in schools be transparent and amenable to review.
- Create classroom contexts that allow teachers to spend more time with their students, such as enacting legislation to limit class size, so that teachers are not pressured to find ways to keep students quietly occupied with digital products.

District policymakers:

- Refrain from adopting AI-based educational applications until:
  - Strict transparency and accountability requirements are put in place as part of an
overall technology accountability plan.

- The public has been provided with compelling evidence or thoughtful and clear explanations as to how those applications are an improvement over other education practices that do not require digital technology.
Throughout this brief, we recognize that various forms of AI exist, each with different histories, technical foundations, and functions. Our focus is on two particular forms of AI that are of contemporary significance in education: data systems that use machine learning algorithms to measure and make predictions about behaviours or outcomes, and large language models (one version of so-called “generative AI”) that can automatically generate original text.


Perhaps the most extreme and exaggerated example of industry spin was produced by the tech investor Marc Andreessen in a blog post titled “Why AI Will Save the World.”


The ‘true threat of AI’, argues technology critic Evgeny Morozov, is an expansion of corporate power over the public domain.
Likewise, Timnit Gebru and her colleagues note that, “The harms from so-called AI are real and present and follow from the acts of people and corporations deploying automated systems. Regulatory efforts should focus on transparency, accountability and preventing exploitative labor practices.”


In one example, a computer scientist—with no background in the field of education—predicted that AI-driven personalized tutoring systems could make traditional schools and teachers redundant.


In July 2023, the assistant director of UNESCO’s education unit warned:

The speed at which generative AI technologies are being integrated into education systems in the absence of checks, rules or regulations, is astonishing . . . [T]oday, in most national contexts, the time, steps and authorizations needed to validate a new textbook far surpass those required to move generative AI utilities into schools and classrooms. In fact, AI utilities often required no validation at all. They have been ‘dropped’ into the public sphere without discussion or review. I can think of few other technologies that are rolled out to children and young people around the world just weeks after their development. In many cases, governments and schools are embracing a radically unfamiliar technology that even leading technologists do not claim to understand . . . Education, given its function to protect as well as facilitate development and learning, has a special obligation to be finely attuned to the risks of AI—both the known risks and those only just coming into view. But too often we are ignoring the risks.


http://nepc.colorado.edu/publication/ai
In his Executive Order of October 30, 2023, President Biden charged the Department of Education as follows:

To help ensure the responsible development and deployment of AI in the education sector, the Secretary of Education shall, within 365 days of the date of this order, develop resources, policies, and guidance regarding AI. These resources shall address safe, responsible, and nondiscriminatory uses of AI in education, including the impact AI systems have on vulnerable and underserved communities, and shall be developed in consultation with stakeholders as appropriate. They shall also include the development of an “AI toolkit” for education leaders implementing recommendations from the Department of Education’s AI and the Future of Teaching and Learning report, including appropriate human review of AI decisions, designing AI systems to enhance trust and safety and align with privacy-related laws and regulations in the educational context, and developing education-specific guardrails.


See also:


33 The AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC) repository has collected over 1,200 examples of real-world problems caused by the introduction of AI, algorithms and automation across different industries and sectors.


Bommasani and colleagues noted, “While the societal impact of foundation models is growing, transparency is on the decline …” Their analysis assessed the 10 leading foundation models on 100 indicators of transparency, and found the average score to be 37 out of 100 (Meta’s Llama 2 scored highest, at 57). They found especially low transparency with respect to how the models are built (e.g., related to the training data) and on downstream impacts of use (e.g., how users might seek redress for harm).


Indicative lists of the range of research and applications of AI in education can be found in the two main research journals in the field. The aims and scope of the *International Journal of Artificial Intelligence in Education* (originally launched in 1988) feature 50 relevant topics, including adaptive systems, statistical methods, cognitive models, robotics, intelligent agents, tutoring systems, neural models, student modelling and cognitive diagnosis, and theories of learning.


A new journal, *Computers and Education: Artificial Intelligence*, launched in 2020, refers to using AI “to facilitate teaching, learning, or decision-making. With the help of AI technologies, which simulate human intelligence to make inferences, judgments, or predictions, computer systems can provide personalized guidance, supports, or feedback to students as well as assisting teachers or policymakers in making decisions.”


See also:


A recent statistical meta-analysis of AIED research results suggested some effectiveness of AI on “learning achievement,” as measured by standardized tests. Although only 24 articles met the study’s criteria for inclusion, most of these were very short time-scale quasi-experimental studies, and the socio-economic or geopolitical contexts of the reviewed studies were not made clear.


A few examples can provide some sense of the range of applications and businesses promoting AI in education. The established education business Duolingo integrates with OpenAI’s large language model technology to run a conversational language learning application. Microsoft promotes the application of AI in education, and has also incorporated OpenAI language technology into its products. The learning management systems Blackboard and Canvas have both announced plans to incorporate generative AI technologies. Blackboard has launched an “AI Design Assistant” powered by OpenAI to support teachers’ lesson planning. Canvas is also actively exploring and promoting the application of AI in schools.

Anthology, Inc. (Blackboard) (2022). AI design assistant [webpage]. Retrieved November 27, 2023, from https://help.blackboard.com/Learn/Instructor/Ultra/Course_Content/Create_Content/AI_Design_Assistant


A recent meta analysis of research on AI chatbots in education, based on a sample of 24 studies, found some evidence of effectiveness on learning performance on selected skills and outcomes in some contexts, though mostly in short time-scale conditions with some degree of “novelty effect,” and primarily in higher education settings. The authors acknowledged that some studies found no significant difference in learning performance between the use of chatbots and control groups, the risk of information overload from chatbot use, a lack of cultural specificity in studies, and its conclusions remained highly speculative.


See, for example:


http://nepc.colorado.edu/publication/ai


Similarly, in February 2024, PowerSchool announced that it would be incorporating its AI assistant, PowerBuddy, into its suite of products ( Schoology Learning, Performance Matters, Naviance CCLR, ContentNav, Analytics and Insights, PowerSchool SIS, and My PowerSchool).


http://nepc.colorado.edu/publication/ai


See also:


In some contexts, AI tutoring is also marketed to parents as a cheaper alternative to in-person private tutoring services.


For example:


The OECD’s Artificial Intelligence and the Future of Skills project is developing a program to assess the capabilities of AI and robotics, particularly in relation to literacy and mathematics, and their impact on education and work.


http://nepc.colorado.edu/publication/ai


In late summer 2023, the United Nations Educational, Scientific and Cultural Organization (UNESCO) published *Guidance for Generative AI in Education and Research*, a report that proposed “key steps for governmental agencies to regulate the use of generative AI” and “enable ethical and effective uses of this technology in education.”


The Department of Education appears unlikely to provide meaningful oversight of technology companies’ activities in schools. It has, for example, never enforced a FERPA violation, despite that no school can satisfy the Department’s requirement that it directly control technology vendors’ use and maintenance of student education records in their role as “school officials.”

For definition and requirements of “school official,” see:


CRPE reports that only two U.S. states—California and Oregon—have provided any official guidance to school districts on using AI in schools for the 2023-2024 school year. Only 11 states are in the process of developing guidance.


Chatbot tutors can be defined as AI technologies that can hold conversations and produce all kinds of text in response to prompts, and are being developed extensively by ed tech industry companies.


Google promotes adaptive learning in its core AI offerings for education, as: “a type of learning where students are given customized resources and activities to address their unique learning needs . . . What’s new is applying recent AI advances to this concept, which opens up a whole new set of possibilities to transform the future of school into a personal learning experience.”


Salman Khan, the founder of Khan Academy, claimed, “We’re at the cusp of using AI for probably the biggest positive transformation that education has ever seen. And the way we’re going to do that is by giving every student on the planet an artificially intelligent but amazing personal tutor. And we’re going to give every teacher on the planet an amazing, artificially intelligent teaching assistant.”


See also:

The analytics functionality in Khanmigo is called “Class Snapshot” and provides a breakdown of student progress from their interactions with the application. See the Khanmigo Teacher Tools page. Retrieved November 15, 2023, from https://support.khanacademy.org/hc/en-us/articles/14799047733045-What...
teacher-tools-are-available-on-Khanmigo-


122 Statistical performance gains from students using AI—based on studies in controlled experimental classroom settings, highly prescriptive tasks, and short-term interventions, and measured through standardized assessments—can also be generalized as speculative claims about the potential of foundation models like GPT-4 as "personal tutors." See, for example:


126 Khan Academy (n.d.). Khanmigo is a ground-breaking AI tool designed to improve learning and teaching. [webpage]. Retrieved October 19, 2023, from https://blog.khanacademy.org/teacher-khanmigo/


129 Highlighting the speculative nature of promotional rhetoric of AI in education, in a recent interview Salman Khan imagined that “one day, you’re going to have AI, and you’re going to be in an immersive world, and you’re gonna have a link straight into your brain . . . That’s likely to happen in the next 15 years, and it’s likely to be very accessible.”


132 Smith, M.S. (2023, June 23). The Internet isn’t completely weird yet; AI can fix that. *IEEE Spectrum*.
Technology critic and educator Matthew Kirschenbaum calls this a “textpocalypse.”


The ed tech company Pearson suggests it will mitigate the tendency of generative AI to produce false information by only using internally generated material to create and train its AI applications. It remains to be seen whether Pearson’s efforts will succeed—or perhaps more accurately, it remains to be seen whether any third-party independent entity will be able to verify Pearson’s claims.


When Newark, NJ, schools piloted Khanmigo, for example, teachers found an error in how the program answered a math question. They alerted Khan Academy, who fixed that particularly error. What Khan Academy could not do, however, was prevent errors more generally—and AI is prone to errors.


In an apparent effort to protect the GPTs marketplace from criticism, amidst public concern over children’s access to harmful online content, OpenAI partnered with Common Sense Media to create guidelines on “safe,
responsible use of ChatGPT for educators and young people.


163 The Federal Trade Commission has recently promised to crack down on indiscriminate data collection by the ed tech industry. A child’s education, it stated, must not be “conditional” on surveillance.


Rudin, C. & Radin, J. (2019). *Why are we using black box models in AI when we don’t need to? A lesson from an explainable AI competition*. *Harvard Data Science Review*, 1(2). Retrieved November 15, 2023, from https://doi.org/10.1162/99608f92.5a8a3a3d

Computer scientist Cynthia Rudin explains and advocates for the superiority of inherently interpretable AI models to black box models. Interpretable models are not the same as open source models.


According to the Edinburgh Declaration on Responsibility for Responsible AI:

All who exercise and unleash the material and economic power of AI/AS, even with the best of intentions, must start attending — and be required to attend — responsibly to the vulnerabilities that AI/AS create, perpetuate, amplify or relieve. ... Some applications of AI/AS, ... such as the automation of moral and political human judgments with irreversible and severe consequences, arguably require delay or indefinite pause.


Putting the lid on Pandora’s Box: How community power shapes AI [blog post]. Promising Trouble. Retrieved January 10, 2024, from https://www.promisingtrouble.net/blog/2023-12-pandoras-box


Similarly, Arizona State University President Michael Crow explained the University’s partnership with OpenAI by noting: “ASU recognizes that augmented and artificial intelligence systems are here to stay, and we are optimistic about their ability to become incredible tools that help students to learn, learn more quickly and understand subjects more thoroughly . . . “


In 1966, Joseph Weizenbaum developed the ELIZA computer program that was capable of chatting with humans like an empathic psychologist (i.e., an early “chatbot”). He later published papers that called on scientists and engineers, in particular, to critically assess when and how computer technologies should be applied to human affairs. His 1978 paper asked the following:

- Who is the beneficiary of our much-advertised technological progress and who are its victims?
- What limits ought we, the people generally and scientists and engineers particularly, to impose on the application of computation to human affairs?
- What is the impact of the computer, not only on the economies of the world or on the war potential of nations, etc., but on the self-image of human beings and on human dignity?
- What irreversible forces is our worship of high technology, symbolized most starkly by the computer, bringing into play?
- Will our children be able to live with the world we are here and now constructing?


After the release of ChatGPT and the subsequent rush to implement AI in all sectors, Jack Stilgoe, in August

http://nepc.colorado.edu/publication/ai
2023, returned to Weizenbaum’s questions, summarizing them as we have here: Is it good? Do we need it?


See also:


Algorithms built into educational software are presented as “neutral” and “scientific,” and as embodying “truth” or fact. They cannot be neutral, however, because they are created by people—and people are not neutral.


In their critique of using black box AI models to make significant decisions, Cynthia Rudin and Joanna Radin explain that “Even if one has a list of the input variables, black box predictive models can be such complicated
functions of the variables that no human can understand how the variables are jointly related to each other to reach a final prediction.” (p. 3)

Rudin, C. & Radin, J. (2019). Why are we using black box models in AI when we don’t need to? A lesson from an explainable AI competition. Harvard Data Science Review, 1(2). Retrieved November 15, 2023, from https://doi.org/10.1162/99608f92.5a8a3a3d

200 Currently, their programming could feasibly be audited if regulatory mechanisms required it and independent specialists were available. Such auditing is recommended in, for example:


203 McQuillan, D., Jarke, J., & Pargman, T.C. (2023). We are at an extreme point where we have to go all in on what we really believe education should be about. Postdigital Science and Education. Retrieved January 16, 2024, from https://doi.org/10.1007/s42438-023-00433-5