OFFICE OF
Educational Technology

Designing for Education with Artificial Intelligence:

An Essential Guide for Developers

July 2024

Designing for Education with Artificial Intelligence: An Essential Guide for Developers

Miguel A. Cardona, Ed.D.

Secretary, U.S. Department of Education

Roberto J. Rodriguez

Assistant Secretary, Office of Planning, Evaluation and Policy Development

July 2024

Examples Are Not Endorsements

This document contains examples and resource materials that are provided for the user's convenience. The inclusion of any material is not intended to reflect its importance, nor is it intended to endorse any views expressed or products or services offered. These materials may contain the views and recommendations of various subject matter experts as well as contact addresses, websites, and hypertext links to information created and maintained by other public and private organizations. The opinions expressed in any of these materials do not necessarily reflect the positions or policies of the U.S. Department of Education. The U.S. Department of Education does not control or guarantee the accuracy, relevance, timeliness, or completeness of any information from other sources that are included in these materials. Other than statutory and regulatory requirements included in the document, the contents of this guide do not have the force and effect of law and are not meant to bind the public.

Contracts and Procurement

This document is not intended to provide legal advice or approval of any potential federal contractor's business decision or strategy in relation to any current or future federal procurement and/or contract. Further, this document is not an invitation for bid, request for proposal, or other solicitation.

Licensing and Availability

This guide is in the public domain and available on a U.S. Department of Education website at https://tech.ed.gov.

Requests for alternate format documents such as Braille or large print should be submitted to the Alternate Format Center by calling 1-202-260-0852 or by contacting the 504 coordinator via email at om_eeos@ed.gov.

Notice to Persons with Limited English Proficiency

If you have difficulty understanding English, you may request language assistance services for U.S. Department of Education information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833); email us at Ed.Language.Assistance@ed.gov; or write to U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave. SW, Washington, DC 20202.

How to Cite

While permission to reprint this publication is not necessary, the suggested citation is as follows:

U.S. Department of Education, Office of Educational Technology, *Designing for Education with Artificial Intelligence: An Essential Guide for Developers*, Washington, D.C., 2024.

Acknowledgements

Project Team

Designing for Education with Artificial Intelligence: An Essential Guide for Developers was developed under the leadership and guidance of **Roberto J. Rodríguez**, Assistant Secretary for the Office of Planning, Evaluation and Policy Development, **Anil Hurkadli**, Interim Deputy Director for the Office of Educational Technology, **Bernadette Adams**, Senior Policy Advisor for the Office of Educational Technology, and **Kevin Johnstun**, Education Program Specialist for the Office of Educational Technology.

This work was developed with support from Digital Promise under contract (91990019C0076) and led by **Jeremy Roschelle** with **Anthony Baker**, **Pati Ruiz**, **Eric Nentrup, Gabrielle Lue** and **Sarah Martin**.

Contributing Members of the Developer Community

- Kristen DiCerbo, Khan Academy
- Teddy Hartman, GoGuardian
- Neil Heffernan, Worcester Polytechnic Institute
- Karl Rectanus, EDSAFE AI Alliance
- Steve Ritter, Carnegie Learning
- Sharad Sundararajan, Merlyn Mind
- Alyssa Van Camp, TeachFx
- Julia Winter, Alchemie Solutions, Inc.

We also thank the many developers, industry associations, and nonprofit organizations that attended our listening sessions and contributed their ideas for translating the Department's recommendations for Artificial Intelligence in education into practical guidelines.

Table of Contents

Designing for Education with Artificial Intelligence: An Essential Guide for Developers

Introduction

Today and in the future, a growing array of Artificial Intelligence (AI) models and capabilities will be incorporated into the products that specifically serve educational settings. The U.S. Department of Education (Department) is committed to encouraging innovative advances in educational technology (edtech) to improve teaching and learning across the nation's education systems and to supporting developers as they create products and services using AI for the educational market.

Building on the Department's prior report, *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations* (2023 AI Report), this guide seeks to inform product leads and their teams of innovators, designers, developers, customer-facing staff, and legal teams as they work toward safety, security, and trust while creating AI products and services for use in education. This landscape is broader than those building large language models (LLMs) or deploying chatbots; it includes all the ways existing and emerging AI capabilities can be used to further shared educational goals.

Our insights here are intended to support people who are managing teams in the design and development of products that leverage AI to improve teaching and learning. We have attempted to address topics that will be relevant across the continuum of edtech developers, which includes established firms and newcomers, as well as developers across research, nonprofit, and for-profit organizations. We address not only developers of products for formal education settings—including elementary and secondary schools, colleges, and universities—but also for educational uses at home, community, and other informal settings.

To this end, each section of this document is built around a core recommendation and includes a set of discussion questions that leaders in organizations can use to foster conversation, next steps to promote robust development processes, and resources that can provide additional support. Please note that the Department and other federal agencies are actively considering next steps to promote the safe and responsible use of AI. Thus, this document suggests "questions to ask" and "directions to pursue" to developers that are deliberately open-ended.

This guide provides non-regulatory, education-specific guidance that is aligned with federal guidelines and guardrails. This guide's coverage of existing federal guidelines and guardrails is not comprehensive or exhaustive. It is not intended to and does not enable a developer to establish its compliance with regulations. Also, it is not intended to and does not introduce any new requirements. Where examples are given, including links to non-U.S. Government websites, they are intended to be illustrative and not to restrict the application of this guide to additional forms of AI as they become available for use in education. We are providing these external links because they contain additional information relevant to the topic(s) discussed in this document or that

otherwise may be useful to the reader. We cannot attest to the accuracy of information provided on the cited third-party websites or any other linked third-party site. We are providing these links for reference only; linking to external resources does not constitute an endorsement by the Department. Developers can use this guide to increase their understanding of essential federal guidelines and guardrails to guide their work as they create AI applications for educational settings.

Responding to the October 2023 Executive Order

This guide is responsive to President Joe Biden's October 30, 2023, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Executive Order on AI), which states the following:

> *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.*

This guide is informed by an extensive series of public listening sessions with students, parents, and educators along with developers, industry associations, and nonprofit organizations. This included a cross section of developers representing a variety of company sizes, funding models, and organization types (for-profit/nonprofit). Session participants shared their current approaches to safety and security, the risks they and their users face, suggestions on supports and resources, and thoughts about opportunities to build trust in the future. Several additional listening sessions occurred with a smaller set of developers (listed in the contributing members section above) drawn from those who participated initially. Where this guide refers to listening sessions, it includes all these opportunities to hear from constituents.

This guide draws on a growing series of federal publications on AI, which includes these examples:

- **Office of Science and Technology Policy Blueprint for an AI Bill of Rights**: The 2022 White House white paper that broadly shapes a strategy informing and protecting citizens from AI and related technologies
- **National Institute of Standards and Technology (NIST) AI Risk Management Framework**: A seven-step framework to protect developer and end user interests alike, particularly relevant to edtech companies using AI components and forthcoming emerging technologies
- **FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House**: The White House's engagement with technology companies highlights shared responsibility among the federal government, developer organizations, and other constituents.

Defining "Artificial Intelligence" and "EdTech" Broadly

The Department takes a broad view of the terms "AI" and "edtech." This document's guidance applies broadly across the many types of AI that developers may integrate and the many ways their products may be used in educational settings.

Box A: Artificial Intelligence as defined in the Executive Order on AI.

The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to perform the following:

- Perceive real and virtual environments
- Abstract such perceptions into models through analysis in an automated manner
- Use model inference to formulate options for information or action

As a starting point for defining "AI," we note the statutory definition of AI (see Box A above), which also appears in the Executive Order on AI. As we noted in the Department's 2023 AI Report, AI is an umbrella term for many subfields of research and innovation. Following the 2023 AI Index Report, from the National Artificial Intelligence Research Resource Task Force, we continue to observe the extremely rapid evolution of AI capabilities in many domains, such as speech, vision, robotics, and text.

Box B: Defining Educational Technology (edtech)

As defined in the Department's **2023 AI Report**, edtech includes the following:

A. Technologies specifically designed for educational use

B. General technologies that are widely used within educational settings

Likewise, we define "edtech" broadly (Box B). In an independent market research firm's report, the global edtech market for both K-12 and higher education is valued at $123 billion as of 2022. This market includes startups, small businesses, nonprofit organizations, and larger corporations—companies with a mission specific to education and others with products that are used in education and other sectors. There is a broad spectrum of products ranging from infrastructure to student information systems to learning management systems to specific end-user applications and more.

Whereas a market is defined by buyers and sellers, we define the edtech "ecosystem" broadly to include the many different people and organizations working together to design and refine new products and services. Within this ecosystem, discussions of the Executive Order concepts of "safety, security, and trust" will be infused with concepts more specific to education (e.g., evidence, fairness, data privacy) as elaborated below. Ecosystem participants include

educational procurement departments, additional educational decision makers, educators, parents and guardians, students, nonprofits, postsecondary education institutions, and broader community members. The ecosystem includes people who directly create or use AI products and people involved in educational systems who are affected by AI products. Shared responsibility for building trust goes well beyond buying and selling; it thrives on open and engaged communication across the ecosystem.

Key Message: Shared Responsibility for Building Trust

Through listening sessions, the Department learned that one key recommendation in the 2023 AI Report, "Prioritize Strengthening Trust," resonated with developers as a call to action. Developers recognize the fundamental importance of "trust" in the edtech ecosystem in which they participate. Trust improves the co-design process between developers and educators so that together they can create and scale innovative products. Consequently, developers can benefit greatly from understanding how they can work with others in the ecosystem to strengthen trust. Specific key messages follow:

1. Trust is a shared responsibility.

The President's Executive Order on AI makes clear beginning with its title that the federal government has a responsibility to promote "safe, secure, and trustworthy development" and articulates a stance of shared responsibility (see quote below). Consequently, developers will find information within this document on where to locate federal laws and other federal resources that are directly applicable to their work in education. Because the technology is evolving rapidly, developers may find it valuable to go beyond attending to and complying with today's federal guidelines and guardrails to earn trust. One important example is the Software and Information Industry Association's Principles for the Future of AI in Education, which articulates seven principles (e.g., their evaluation principle calls for continually assessing the impact of AI). TeachAI has also developed Principles for AI in Education as part of its AI Guidance for Schools Toolkit to guide the effective development and application of AI in teaching and learning. As a third example, many developers have participated in the EDSAFE AI Alliance, which has produced a common framework, the SAFE Benchmarks.

> *"Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society."*
>
> *— Executive Order on AI*

Other members of the edtech ecosystem also are accepting responsibility and developing conditions for trust. Educational leaders at all levels, including state, district, and building-level leaders, are writing their own guidance (Box C). Developers can look to these resources to understand the steps that educators are taking to build understanding and capacity; to strengthen procurement processes; to protect privacy, security, and fairness; and to manage other forms of risk. Further, in a forthcoming Toolkit for Educational Leaders developed

pursuant to the Executive Order on AI, the Department will be shaping the shared responsibility conversation among educators. Many nonprofits are also developing helpful toolkits and resources (see Box D). As key participants in the edtech ecosystem, developers are encouraged to interact responsibly with the ecosystem to develop trust.

Box C: State guidance resources about AI in education, as of June 2024

Arizona	Kentucky	Oregon
California	Mississippi	Utah
Connecticut	North Carolina	Virginia
Hawaii	Ohio	Washington State
Indiana	Oklahoma	West Virginia

Box D: Partial listing of nonprofits offering guidance resources about AI in education

- The Consortium for School Networking (CoSN) Gen AI Maturity Tool
- EDSAFE AI Alliance SAFE Benchmarks Framework
- International Society for Technology and Education 's Artificial Intelligence in Education Resource List
- Software & Information Industry Association's Principles for AI in Education
- TeachAI's AI Guidance for Schools Toolkit
- All4Ed's Future Ready Schools Emerging Practices Guide

2. Trust requires actively managing AI risks so that we can seize its benefits.

Through its conversations with developers, the Department observed an important shift in how developers are engaging with others around their work. Whereas it has been common to present "solutions"—how technologies can improve teaching, learning, and other educational processes—developers are now also openly discussing how they are managing risks. Some developers have publicly shared details on the process they went through to identify, prioritize, and manage risks. As developers openly discuss risk management, the Department suggests attending to two kinds of processes: (a) technical development processes that result in *trustworthy* systems and (b) engagement strategies that *build trusting relationships* among developers and other ecosystem members.

Thus, the Department understands the importance of discussing both opportunities and risks in a responsible manner. Box E lists examples of salient categories of risk.[1,2,3] Both risks and opportunities will be discussed in more depth later in this guide.

Box E: Types of risks of AI, ordered alphabetically, not by priority

- AI "Race-to-Release" Risks
- Bias and Fairness Risks
- Data Privacy and Security Risks
- Harmful Content Risks
- Ineffective System Risks
- Malicious Use Risks
- Misinformation Management Risks (including "hallucinations")
- Transparency and Explainability Risks
- Underprepared User Risks

In considering risks, it is important to note that AI is evolving rapidly. For example, just as educators are gaining familiarity with text-oriented chatbots, industry is advancing and releasing multimodal capabilities that add new layers to the potential risks. Hence, this guide outlines risks broadly and asks developers to adopt risk mitigation processes that address both the risks that are foreseeable today and those that will newly emerge.

Risks are not only intrinsic to the technology; risks also emerge at the interface of technology and human activity. As people use AI, both foreseeable and unforeseen risks will arise. In its 2023 AI Report, the Department recommended "humans in the loop" and yet, asking an educator to review every use of AI or every AI-based output is neither practical nor fair. Developers share responsibility to be "in the loop" to review uses and outputs of AI, both during the development process and as a product is used in the field. Building on Box E, we illustrate challenging scenarios where both developers and educators will likely need to attend to emergent risks, with a division of responsibility that is yet to be determined:

[1] *Potential Risks of Artificial Intelligence Integration into School Education: A Systematic Review.*

[2] *Getting To Know—and Manage—Your Biggest AI Risks.*

[3] *The Promises and Perils of Generative AI in Education: TFA's Evolving Perspective.*

- As teachers generate personalized lesson plans with AI services, who will review and revise the outputs to eliminate false information generated by large language models and confirm the content is accurate and aligned to educational objectives?
- As curriculum coordinators engage AI to support their work in curating instructional resources and formative assessments for use in their schools, who will weigh evidence for the efficacy of the resources and the validity of the assessments? Who will verify that resources address the needs of underserved and vulnerable populations?
- As guidance counselors use AI-assisted tools to recommend college and career pathways, who will detect and counter unfairness in the recommendations due to biases in historical data sets that were used to develop the AI model and which could harm vulnerable populations?
- As educators use AI to simplify their work of writing emails or other correspondence about their students' work, who is responsible for safeguards against disclosing a student's private information to unintended recipients, including the developer of the AI model?
- As administrators and school leaders procure early warning systems to identify students who may be "at risk," who has sufficient knowledge and time to evaluate whether the AI developer adhered to scientific, legal, and privacy standards necessary to safeguard students' civil rights?
- As educators deploy anti-plagiarism detectors to identify a student's inappropriate use of edtech, who has responsibility for recognizing weaknesses and biases in AI-based detectors that could lead to disciplining students unfairly or unequally? Who ensures that underserved and vulnerable populations are not unfairly targeted?

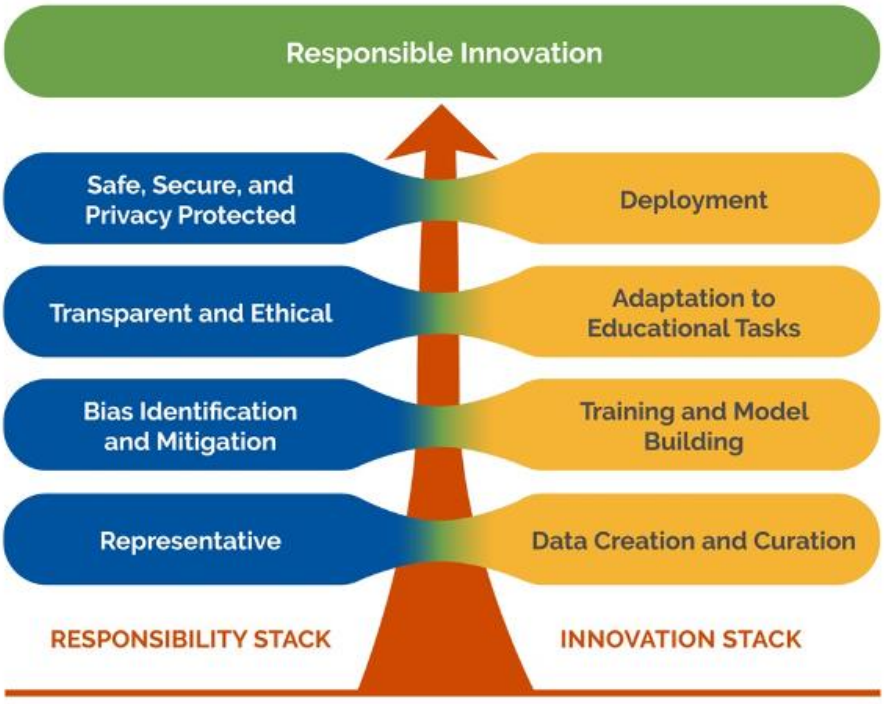
In each challenging case above, the Department respects the role of educators in overseeing educational decisions. However, AI developers should assume significant responsibility because it is unreasonable to ask educators to be primary reviewers of the data and methods used to develop AI models and related software. Building on an important emphasis of the Executive Order on AI, the Department calls on developers to pay special attention to identifying and mitigating potential harms to underserved and vulnerable populations from AI.

3. Coordinating Innovation and Responsibility Throughout Development: A "Dual Stack"

Developers use the visual metaphor of a stack to describe how products are built from layered or connected components. A development stack is a powerful way to coordinate the work of many innovators in a complex organization; it coordinates how the overall product or service will be produced and delivered to market. The Department strongly encourages development organizations to now define a parallel strength via a coordinated "responsibility" stack. This stack would establish how people in a complex edtech development organization work together to earn the trust of educational users of their products. See Figure 1.

Figure 1: Developers should integrate a Responsibility Stack with their Innovation Stack.

The specific elements of each stack are illustrative, not comprehensive. Each developer organization should elaborate the stacks to fit the specifics of their efforts.



Indeed, whereas developers may have previously emphasized one particular risk management office or role in their organization, such as a data privacy role, a single person who attends to safety will not be enough in an age of AI. While these existing risk management roles should continue, additional roles that address representation in training data, measure and mitigate algorithmic bias, rectify errors and misinformation in outputs, and tackle other concerns are also needed. Specific risks will emerge in particular components (such as a Large Language Model component vs. a Computer Vision component), in interactions between components, and through the different stages that take place while building and integrating AI systems (such as in collecting data sets for training, building AI models based on the data, and designing user interaction based on those models). Educational developers can examine how their internal roles and processes are organized to achieve consistent attention to responsible AI, and many are already doing this (Box F).

Box F: Examples of Developers' Public Documents about Responsibility

- DuoLingo released its Responsible AI Standards for its English Language Test, covering its approach to validity and reliability, fairness, privacy and security, and accountability and transparency. DuoLingo invited public comment on its document.
- In a document titled "What is Khan Academy's approach to responsible AI development?," the developer of the Khanmigo tutor discussed how it discloses risks, evaluates and mitigates risks, limits access to its software, and educates its developers about ethics.
- Instructure released an AI Governance Policy that discussed responsible AI use, transparency and accountability, bias and fairness, human-AI collaboration, training and education, and privacy.
- Grammarly has shared its Commitment to Responsible Innovation and Development of AI, which describes how it acts on commitments, e.g., "All models undergo bias and fairness evaluations. Our team of analytical linguists apply research and expertise to minimize bias and apply user feedback."
- Through work with many developer organizations, the Software and Information Industry Association produced seven Principles for AI in Education.

Many other development organizations have produced their own documents about their responsibility approach. In addition, as previously mentioned, eight major developers of AI models made voluntary commitments to manage risks (see White House Fact Sheet).

The Department observes that many educational and general AI developers are producing their own responsible AI principles (see Box F and the quote below). The public will benefit both from public commitment to responsibility and from accountability to such pledges. The Department therefore strongly urges developers to address how their "innovation stack" (components that work together to deliver new capabilities) can be integrated with a parallel "responsibility stack" (roles that work together to mitigate risks in every component and stage of development).

> *"Driven by a strong belief in purpose-built, beneficial AI, we prioritized domain specificity and safety from the outset. Through our Responsible AI initiative, we invite a diversity of voices to co-create solutions with us."*
>
> *— Sharad Sundararajan, Co-founder CDO/CIO, Merlyn Mind*

A Pathway from Designing for Education to Earning Trust

Through the listening sessions, the Department heard a wide range of specific concerns from developers and educators about AI in education.

Figure 2: A Development Process that Leads Toward Earning Trust

We categorized these concerns and organized this guide around five key overarching areas of shared responsibility, as follows (see also Figure 2):

1. **Designing for Education** recognizes that developers should begin by understanding values specific to education. Across many examples, the Department sees educators stepping up to articulate values such as centering humans in the loop and attending to priority educational challenges like reading, science, math, and computer science education. In addition, educator and student feedback should be incorporated into all aspects of product development, testing, and refinement to ensure student needs are fully addressed.

2. **Providing Evidence of Rationale and Impact** is important to making decisions about which edtech products to adopt or procure, especially where the goal of the product is to improve student outcomes. Both the Elementary and Secondary Education Act of 1965 (ESEA) and educational decision makers call for developers to provide evidence that products or services improve student outcomes. In procurement, for example, educational institutions are making clear demands for the evidence they require.

3. **Advancing Equity and Protecting Civil Rights** is an essential commitment of the Department and the Administration and a centrally important concern of constituents in the public listening sessions, both among developers and educators. Developers should be vigilant, for example, about issues of representation and bias in data sets, algorithmic discrimination in systems, and ensuring accessibility for individuals with disabilities.

4. **Ensuring Safety and Security** are emphasized in the Executive Order on AI and related Administration guidance. Educational decision makers are articulating their data privacy and security requirements with clarity and elaborating additional requirements—such as civil liberties—in an age of AI. To participate responsibly in the ecosystem, developers will need to detail the actions they will take to ensure the safety and security of users of AI.

5. **Promoting Transparency and Earning Trust** is an important overarching goal. Earning trust requires attending to all the values above and, in addition, has an important communication dimension that goes beyond output. For example, trust requires transparency and other public commitments to building mutual confidence among technology suppliers and users. Mutual engagement in defining and acting collaboratively among developers, educators, and other constituents builds trust.

Recommendation 1. Designing for Teaching and Learning

Developers of AI-enabled products and services should start with strong attention to education-specific values and visions, which are articulated in a mix of federal, state, and local resources, along with resources to support the use of AI in education produced by nonprofits and industry associations. Attending to ethics is an essential area of shared responsibility.

Key Ideas

- Resources developed by governments (federal, state, and local), nonprofits, and industry associations can provide developers with a good starting point for anchoring their work in educational values and visions and avoiding negative outcomes.
- Human-centered and humans-in-the-loop approaches that proactively include educators are emerging as key values that educational decision makers are demanding, and educator and student feedback should be incorporated throughout the development and testing process.
- Developers should attend to key ethical concepts such as transparency, justice and fairness, non-discrimination, non-maleficence/beneficence, privacy, pedagogical appropriateness, students' and teachers' rights, and well-being.
- Human factors, and human-centered design, and longstanding software development practices can provide a starting point for developers.

What to Know

The 2024 National Educational Technology Plan (NETP) presents a forward-thinking approach to reframing and realizing the potential of edtech to enhance the instructional core, reduce achievement gaps, and improve student learning in our schools. These are three critical national priorities. During an event to release the plan, U.S. Secretary of Education Miguel Cardona stated, "As we work to Raise the Bar in education, it's essential we focus on empowering teachers to become designers of active learning, using technology in effective ways to engage and inspire students." Inside the 2024 NETP, developers will find a wealth of information regarding educational visions, goals, and values related to the use of technology.

Notably, the 2024 NETP is not directly about AI. That is because a valid educational purpose and important unmet need should be the starting point for development, not excitement about what a particular technology can do.

At the federal level, the Department's 2023 AI Report provides additional guidance on aligning generative AI to what educators care about and need. The report has sections about student learning, supporting teachers, and improvement assessments, with many examples regarding

how AI could lead to advances in these areas. The Department has observed that the following two recommendations in the Report resonate consistently with educators in state, local, and international forums: (a) emphasize humans in the loop, and (b) align AI models to a shared vision for education. These provide important shared touchstones with educators as developers begin their design process with AI-enabled technologies.

State (Box C) and local resources provide additional guidance about AI in education. These resources partially anchor visions for AI in education on changing expectations regarding career readiness preparation for students' future employment. They highlight that developing teacher and student AI literacy is essential to responsible use of AI. Many of the resources provide advice on improving equity and inclusion. Data privacy, security, and safety are key risk areas that are important for developers to address and school leaders to manage. Overall, the resources provide developers with a better understanding of how educators are conceptualizing opportunities. In Box G, we list some opportunities that are frequently featured across these resources.

Box G: Opportunities for using AI in education

The Department sees many opportunities for improving academic outcomes, as well as accessibility and inclusion of students in academic programs. Here are some examples:

- Improving academic outcomes, accessibility, and inclusion for children and students with disabilities
- Providing students with more and better feedback and guidance as they learn curricular subject matter
- Addressing learner variability (including both access and inclusion) in all its aspects by better matching learning resources to each individual student's strengths and needs, as well as addressing the needs of historically underserved student populations
- Saving administrative time for teachers and enabling teachers to focus on their students
- Enabling teachers to incorporate research-based pedagogical principles, like those found in the What Works Clearinghouse practice guides, into their instructional plans
- Improving teacher professional learning by including opportunities to practice specific pedagogical strategies with simulated classrooms and students
- Reducing the cost to customize learning resources to build on strengths and interests of students in the locale of various educational communities
- Achieving efficiencies in school operations, such as school bus schedules

Of course, developers may surface equally important opportunities through their own engagements with educators. In the case of generative AI, the capabilities are still too new to be certain about where the best opportunities lie. Developers can strengthen alignment and validation of purpose by establishing strong feedback loops with educational communities at

every stage of product design, development, deployment, and refinement. When seeking feedback, it is important not only to include those in the most influential roles (e.g., superintendent, technology director, and/or other decision-maker), but also those who will be most affected by the educational product or service (e.g., classroom teacher, special education teacher, student, and family). Hearing diverse input is one way for developers to keep humans in the loop. Co-design with educators, a recommendation in the Department's 2023 AI Report, is a strong way to enact shared responsibility. However, as previously noted, involving educators in both designing and monitoring the use of AI is not a panacea and can impose responsibilities on educators that developers should rightfully own.

Developers frequently raised ethics as an important area of focus during the Department's listening sessions. Researchers and educators have been working together to develop ethical guidelines, and the Department expects this work to continue. (See the "Resources" section of this recommendation, for example.) A review of major ethics frameworks for AI in education found that the following general ethics concepts were applicable to education: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom, and autonomy.

In addition to adapting these principles to more closely fit education settings, the review identified four additional principles specific to education: pedagogical appropriateness, children's rights, AI literacy, and educator well-being. These ethical principles are core components of designing AI systems that interact with children. For example, is it ethical to present an animated coach in a product as a humanoid persona, blurring the line between people and algorithms? Or, if AI is included in a system to support an aspect of teachers' well-being, what is the standard of care, and when should human caregivers be involved? Additional examples of ethical considerations in the design and implementation of AI-enabled tools will become clear as the use of these tools in education expands. (Figure 3). Regarding AI Literacy, the review states, "AI literacy underlines the educational importance of children and youth learning about AI so that they may become critically informed, as well as the need to build teachers' professional knowledge and parent awareness of AI."

Figure 3: A Synthesis of Ethics Themes

SAMPLE OF VALUE-CENTERED DESIGN PRINCIPLES FOR ETHICS IN EDUCATION	
General Ethics Themes	**Education Ethics Themes**
• Transparency • Justice and Fairness • Non-maleficence • Responsibility • Privacy • Beneficence • Freedom and Autonomy	• Pedagogical Appropriateness • Children's Rights • AI Literacy • Teacher Well-Being • Responsiveness to Student Needs

In February 2024, researchers at NIST suggested that building on long-standing concepts in the 1979 Belmont Report (which established principles for the protection of human subjects in research)—beneficence, respect for persons, and justice—can organize how developers address ethics in the age of AI. For example, beneficence can be established by collecting and sharing evidence that a product delivers expected benefits and by mitigating any situations in which individual users might experience harm (even though effects on average are positive). This applies across the life cycle of a solution, from prototyping to productizing. Additional ethics concepts can likewise be translated into practical steps that developers could take (and build on with measures many educational developers already routinely practice).

Likewise, developers should attend to a foundation of rights and respect for human dignity as they create new AI-enabled applications. This is consistent with language found in the Administration's Executive Order on AI:

> *"The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected. . . . [which requires] appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI."*
>
> *— Executive Order on AI*

International resources such as The European Commission's Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence put human dignity at the foreground of ethical considerations.

Product leads and their teams should not only be aware of ethical concerns but should also find ways in which ethics can be interwoven throughout the life cycle of product development. This applies from the initial ideation and prototyping stages and product deployment and continues in perpetuity as the solution improves—both autonomously and with human intervention—across the product's life cycle. Value-centered design is one approach that weaves people, purpose, and ethics together (see Resources). Also, the Association of Computing Machinery has useful guidance on ethics for developers (also in Resources).

Many organizations are already paying attention to human factors in their development approaches. AI may be relatively new to education, but simple to sophisticated examples have been prevalent for decades in other fields such as manufacturing, aviation, and retail. Further, the concept of human-centered design has a long history in computer science (see Resources section below) and is something educational developers already incorporate into their processes to varying degrees. In listening sessions, educators strongly advocated for their involvement not only in initial design but also in the development process, both to improve the system and to participate in explaining the use of AI in the system to other educators. Similarly, educational researchers have been cultivating space for youth voices throughout the development process. Most generally, it is important to include not only the users who are powerful but also those who will be most affected by an application of AI in education.

"As a LatinX and neurodiverse student, I think it's especially important to mitigate the systemic biases that are entrenched within AI models. Assembling a diverse team of people who specifically work to tackle this issue is essential to building safe AI."

—Nicholas Gertler, student and chair of AI Issue Advisory Council and AI & Education Advisor, Encode Justice

However, in some cases, human-centered design may only receive attention in the user experience (UX) layer of development or in the work of one specific team or department in an edtech company. For example, giving feedback to a student might be considered to be a UX feature. However, giving appropriate feedback to a student may depend on the quality of the product's records of a student's past learning, earlier pedagogical interventions, and what kinds of feedback work well for that student. Thus, the database aspects of giving high quality feedback are equally important to the UX aspects of giving high quality feedback. For an AI product or service to be human-centered, developers will likely need to incorporate human-centered methods throughout the layers of system design and deployment. Human-centered development of AI for education should occur throughout a developer's responsibility stack.

Questions to Ask

1. What can we learn from how written educational strategies (such as the NETP) and our educational customers describe the most important and equitable purposes for using AI?
2. How can our work align to the "humans-in-the-loop" recommendation in the Department's 2023 AI Report?
3. Through what feedback loops are we continually learning more from our users about how to align to educational purposes to meet the needs of diverse students and to respect the role of educators?
4. How does our team understand ethical considerations for AI in education, and what steps can we take to integrate ethics into our work?
5. How have children and youth, families, and educators from underserved settings been consulted and involved in design decisions?
6. How can we continually strengthen development disciplines (such as human factors and human-centered design) to address emerging AI-related features and risks?

Directions to Pursue

- Developers should familiarize themselves with relevant resources that express educational visions and strategies, including resources that are available at the national, state, and local levels as well as internationally.

- Developers should deepen their understanding of historical discrepancies in opportunity to learn and in learning outcomes for diverse student populations and how their products could contribute to equity for all students.
- Developers should be intentional about strengthening their feedback loops with educational user communities throughout the product life cycle, from defining the product's purpose to refining how it operates.
- Developers should engage ethics experts to guide their work and build their team's understanding of ethical issues in the day-to-day work of developing, deploying, and continually improving a product.
- Developers should involve educators and youth throughout the product development process, seeking to include not only those with power but also those who will be most affected by design choices in a product or service.
- Developers should explore obstacles to human judgment (e.g., a tendency to defer to the suggestions coming from a technology), to misunderstand limitations of AI-based inferences, or to underappreciate the potential for more risks to emerge once AI is deployed in educational settings and expand their understanding of human factors to encompass all the ways an AI-enabled system may enable or impede sound instructional and educational decisions.

Resources

- Batya Friedman & David G. Hendry's Value Sensitive Design: Shaping Technology with Moral Imagination
- Center for Humane Technology's Potential Policy Reforms Toolkit
- Center for Democracy & Technology's AI Policy Tracker
- NIST's Human-Centered Design Principles
- Organisation for Economic Co-operation and Development (OECD) & Education International's Opportunities, guidelines and guardrails for effective and equitable use of AI in education
- United Nations Educational, Scientific and Cultural Organization's (UNESCO's) Artificial Intelligence and the Futures of Learning project
- U.S. General Services Administration's Human-Centered Design Guide Series

Recommendation 2. Providing Evidence for Rationale and Impact

By clearly articulating how they have incorporated evidence-based practices into their products and how they intend to build new evidence about a product's usability and efficacy, developers and deployers of AI systems in education can work together toward responsible use. Developers of AI-enabled products and services share responsibility to explain how research informs the rationale (or logic) of their offering, to document and analyze data to make improvements and address risks, and to evaluate the impact on educators and students, especially those in historically underserved groups or settings.

Key Ideas

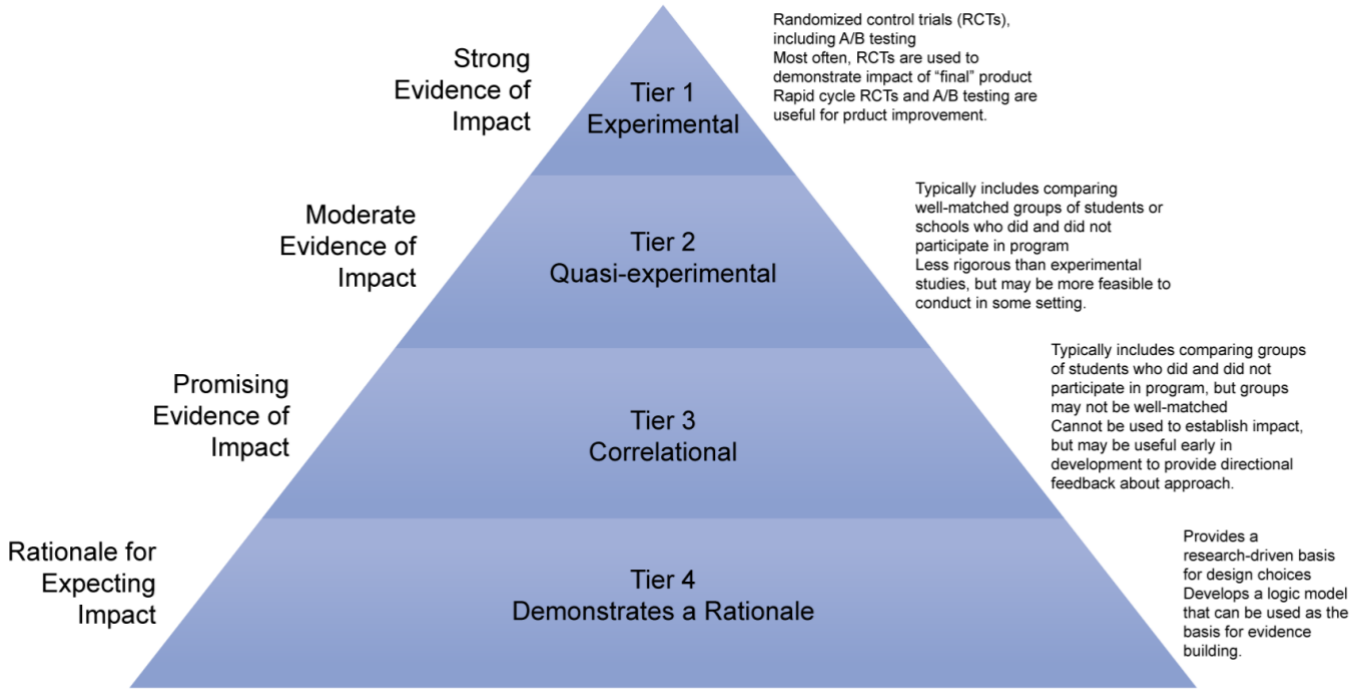
- The ESEA specifies four tiers of evidence (see ESEA section 8101(21)) that characterize the quality of research evidence that establishes whether an educational product or service has been shown to improve student outcomes. (See also the Department's September 2023 non-regulatory guidance.) Educational leaders increasingly demand such evidence when making adoption and procurement decisions.
- Developers should clearly articulate how existing evidence-based practices inform the design of their product or service. When specific evidence is not yet available for potentially innovative features, developers should be clear about the more general scientific rationale that justifies including the features in a product's design.
- Developers should be clear about the student outcomes their product or service is meant to improve. When the product or service provider seeks to measure those outcomes, the measure used should be high-quality (i.e., demonstrate acceptable levels of validity and reliability for the students and settings in which the product or service will be used).
- Developers should seek to build evidence about the risks and outcomes associated with, and impacts of, the use of their product and services. The evaluation process should be framed to measure to type of outcome the product targets and should take into account the importance of identifying and mitigating potential risks, especially risks that might differentially impact vulnerable and underserved populations. Undertaking the most rigorous study (e.g., a randomized controlled trial) may at first seem daunting. If this is the case, developers may first use less rigorous forms of evidence building (e.g., correlational studies) to describe how student demographic information and product use is associated with observed student outcomes. Over time, developers may proceed to more rigorous evaluations that support statements about a product's causal impact on student outcomes, such as by using randomized controlled trials or other rigorous approaches to evidence building.
- When making adoption decisions, states, districts, and schools typically consider the extent to which the evidence supporting a given product or service meaningfully includes the students and settings they seek to serve. As such, when building evidence about a program or service's relationship to, or impact on, student outcomes, developers should disaggregate their findings to be clear about "what works, for whom, and under what conditions." Considerations of "for whom" should address vulnerable and historically underserved student populations; for example, an AI-enabled product for mathematics teaching and learning should address the types of disparities in mathematics achievement documented via the National Assessment for Educational Progress via policy-relevant demographic variables.

What to Know

In 2015, the Every Student Succeeds Act reauthorized the ESEA. The ESEA encourages school and district decision makers to choose educational products, services, and interventions that have been shown to improve student outcomes through high-quality research and evaluation or that at least demonstrate a rationale that they will be effective. The Department also provides a document titled Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments, which includes the evidence framework and definitions in the Education Department General Administrative Regulations. In addition, the Office of Educational Technology offers an Edtech Evidence Toolkit with resources including one-pagers, case studies, and examples. Developers of AI-enabled tools and platforms should know that required characteristics of educational evidence are specified in law and used in practice when educational leaders make adoption and procurement decisions and, with this knowledge, specifically define the outcomes their solutions can provide and how to measure them as evidence.

The ESEA defines four tiers of evidence, which are summarized in the Department's What Works Clearinghouse as shown below in Figure 4.

Figure 4: Tiers of Evidence



Early-stage developers may start by using evidence, including scientific theories, to develop the rationale for how their product or service is intended to strengthen student learning. The National Academies of Sciences, Engineering, and Medicine 2018 report "How People Learn II: Learners, Contexts, and Cultures" recommends assessing the degree to which learning environments are learner-centered, knowledge-centered, assessment-centered, and community-centered. Here are some examples of how to do this:

- Learner-Centered: How can an AI-enabled product support learners to be more active and collaborative as they strive to make sense of new information?
- Knowledge-Centered: How can an AI-enabled product activate a learner's prior knowledge and engage them in processes that actively strengthen understanding like elaborating, explaining, or critiquing?
- Assessment-Centered: How can an AI-enabled product provide students with more timely, relevant, and useful guidance and feedback when they encounter a difficulty?
- Community-Centered: How can an AI-enabled product support social interactions in which a student's peers, teachers, and other community members actively support an individual's strengths and needs?

"How People Learn II" also provides a list of five effective strategies to support learning:

- Retrieval practice
- Spaced practice
- Interleaved and varied practice
- Summarizing and drawing
- Explanations: elaborative interrogation, self-explanation, and teaching

There are many additional principles specific to age levels or subject domains, such as the science of reading, learning mathematics, or pursuing science and engineering. The Department's Institute of Education Sciences (IES), through its What Works Clearinghouse, provides general as well as domain-specific recommendations for evidence-based practices through its series of Practice Guides. Working with a teacher professional association dedicated to target school content is also often a good way to uncover learning principles specific to a domain. Another way is to contact scholarly societies or associations seeking a connection to a researcher active in the area the developer will pursue.

To hone their own rationale, which might be based on the above or additional modern learning principles, developers may seek feedback from pilot group users to inform design and development adjustments to the solution to improve efficacy. Evidence should inform both how a narrow product component has been designed (e.g., the ways in which mathematics is represented in symbols and graphics) and provide a scientific rationale for the foundational logic of their approach to improving learning. Foundational logic is sometimes called the

"theory of action" and is understood to be the integrated set of mechanisms that together lead to improvements to student outcomes (see Resources for more information).

Researchers and developers often follow the best practice of documenting their theory of action through a logic model. A logic model traces the connections from inputs (e.g., that a product or service provides) to teaching and learning processes (e.g., that teachers and students enact with support from the product) to outcomes (e.g., increased student achievement in a particular subject). Within a logic model, a major point of focus for developers, as they begin a journey toward strong evidence, should be defining a clear problem that is informed by robust root cause analysis, identifying an appropriate set of target outcomes, designing products or features that are likely to achieve those outcomes, and designing processes to measure those outcomes.

After establishing that their product or service "demonstrates a rationale" for use, developers may wish to consider how they can build evidence that it is effective. The strength of evidence supporting claims about a product or service's effectiveness can vary from relatively lower (Tier 3 or Promising Evidence) to relatively higher (Tier 1 or Strong Evidence). Lower-quality evidence (e.g., Tier 3) offers less confidence about effectiveness and speaks only to an *association* or *relationship* between a product or service and student outcomes. Higher-quality evidence (e.g., Tier 1) offers more confidence about effectiveness, leading to conclusions that a product or service *caused* a student outcome.

As shown in Figure 4, the primary factor that distinguishes lower-quality evidence from higher-quality evidence is the type of research that is used to evaluate a product or service's effectiveness. Not all developers will have experience in designing and conducting high-quality research and, as such, will seek out the assistance of outside experts. Notably, even developers with research capacity may elect to partner with an independent evaluator to lend additional credence to their effectiveness findings.

Successful evidence-building at any tier depends on both robust forethought and adequate resources. It is almost always easier to design a program or service from the beginning with the goal of high-quality evidence building in mind than to add that goal when a program or service is nearing completion or final. As such, developers should consult experts in evaluation early in the design (or improvement) process. Although evidence-building can occur with lower costs when thoughtfully incorporated into a new or redesigned program, additional financial resources supporting evaluation—particularly an external evaluator—may be needed. Grant funding for research is available in multiple federal programs (see Resources) and through some philanthropic foundations.

Measurement of educational outcomes is always imperfect and always a focus of improvement for the field. The Department's 2023 AI Report includes a section on formative assessment which emphasizes measuring what matters; specifically, that educators call for broader ways of capturing what students know and can do (compared to traditional multiple-choice assessments, for example). Evidence from standardized end-of-course measures (culminating or summative assessment) will continue to play an important role in assessing the efficacy of policies, systems,

and practices that support student learning. However, formative and diagnostic evidence collected using additional relevant measures (continuous or formative assessment) also provides valuable information for educators, parents, and students about learning and academic performance. This sort of evidence also provides faster feedback cycles for all engaged parties. Developers may wish to engage with outside experts to learn more about how to connect their vision to measurable outcomes and how to identify measures with strong psychometric qualities (e.g., high levels of validity and reliability).

The tiers of evidence describe rationale and impact but leave an additional area of evidence open: how developers document risks and safeguards in their products by collecting data, analyzing what is working and what needs change, and guiding improvements. It is a common practice for many educational developers to provide case studies of how they worked with schools or districts to field pilot and test their product. They may use A/B testing to identify the features of their product that support needs across a variety of educational settings.

In terms of shared responsibility, educational decision-makers also value transparency for the means employed by the solution developer to field test and improve their product, as well as what product certifications have been secured through evaluations by trusted third-party learning organizations working in edtech. With regard to continuous improvement, some developers are exploring how to streamline the process of engaging with external researchers who could test improvements to their platform (see the Institute of Education Sciences' funded SEERNet, for example), while safeguarding privacy and other considerations. For example, developers are working with researchers to validate in-platform measures that can be used to study learning outcomes while reducing data collection costs. Developers can also incorporate interfaces into their platform that enable researchers to specify how to deliver variable resources to students while automatically collecting data—without requiring researchers to know the details of the developer's platform code. Developers can also standardize mechanisms for protecting the privacy of student data to enable researchers to perform analyses without enabling researchers to identify specific students.

Finally, evaluations are typically more valuable to decision makers when they address "for whom and under what conditions" because schools and districts want to know if the results are likely to generalize to their population and setting. With such a dynamic range of diversity in our nation's schools and districts, an approach that is only tested with students in a large metropolitan school district in one state may not work optimally for staff and students in a small rural school or a district in another state. Likewise, other questions arise, such as whether children and students with disabilities were included in the study population, and were there any noteworthy differential impacts for these students? Was the product effective for both students with low and high prior performance or only for students with average performance? Does the product work well for multilingual learners? There are many other relevant questions about "for whom and under what conditions." Developers can purposely vary the populations and settings in which they evaluate their offering to accumulate evidence that respects variability among learners and across settings.

This approach to addressing variability between context and learners has significant relevance for products that use AI to adapt to students' strengths and needs, as products may seek to strongly serve a greater diversity of students or function in a wider range of conditions. That AI technologies avoid harm and produce "the greatest good for the most people" cannot be taken for granted and must be specifically investigated. (An important and related concept appears in the equity-centered section; developers should seek to demonstrate their process for developing AI algorithms and models to minimize unfair bias.) The Department acknowledges it is expensive and time-consuming to study all possible variables. Developers should weigh the potential for harm and the level of confidence that a risk has been mitigated. Schools should protect students from harm, for example, regarding infringing students' civil rights—and thus algorithms that unfairly impact students' broad opportunities to learn and advance could be high risk and high priority for strong evidence. An example of a lower risk application might be a productivity tool that helps teachers do routine aspects of their work, such as configuring classroom edtech for a particular activity that the teacher has planned, and which can easily be supervised and corrected by the teacher. High-quality evaluation studies, similarly, can detect not only whether an educational resource benefits students on average but also whether it is ineffective for groups of students.

Questions To Ask

1. How could our team fulfill the evidence-based rationale or Tier 4 level and continue to expand the evidence base for diverse populations?

2. How could our product rationale build on or align with existing bodies of evidence or theory about how people learn or teach effectively, especially including students in historically underserved populations?

3. How (including with what partners) can we begin collecting evidence to demonstrate the potential for positive impacts for a diversity of students related to our use of AI in education?

4. How can we ensure that educational decision makers have access to the evidence and other information to make responsible choices about using AI applications?

5. What is our long-term plan to generate rigorous evidence, including for whom and under what conditions our application works and for expanding the populations for whom the application is effective?

Directions to Pursue

- Developers should ask potential customers how they use specific kinds of evidence in their decision-making process and "what success would look like," which would go beyond asking potential customers to list relevant local, state, or other procurement requirements.

- Developers should seek to form partnerships with researchers early in their design work so that modern learning principles can be employed to maximum advantage.
- Developers should seek to form partnerships with educators and users to conduct field tests throughout the life cycle of their product.
- Developers should involve those who will be most affected by a product in collecting and interpreting evidence.
- Developers should collect evidence not only on efficacy, but also related to safety, security, trust, and other issues.
- Developers should collate the workstreams from above into cogent and living documentation that is regularly updated and housed publicly online for the sake of transparency.

Resources

- Relevant funding sources
    - IES Small Business Innovation Research Program
    - IES Education Research and Special Education Research Grant Programs
    - The Department's Education and Innovation Research Program
    - The National Science Foundation's Division of Research of Learning
    - Many philanthropies also provide funds for educators, researchers, and developers to work together, with shared responsibility for evidence. The Bill and Melinda Gates Foundation's AIMS Collaboratory is one example.
- Resources on methods for building evidence
    - The Institute of Education Science Resources
        - The Logic Model Workshop Toolkit
        - IES Standards for Excellence in Education Research
        - IES/National Science Foundation Common Guidelines
        - Companion Guidelines on Replication & Reproducibility in Education Research
        - What Works Clearinghouse Find What Works Guides and Practice Guides
    - The Office of Educational Technology's EdTech Evidence Toolkit

    - The U.S. Department of Health and Human Services' Rapid Cycle Evaluations
- Organizations that provide information about evidence building
    - LearnPlatform
    - LeanLab
    - Common Sense
    - International Society for Technology in Education (ISTE)
    - Digital Promise
    - EdTech Evidence Exchange
    - International Certification of Evidence of Impact in Education

Recommendation 3. Advancing Equity and Protecting Civil Rights

Developers and educators share responsibility for advancing equity and protecting students' civil rights. Developers who are equity-centered will better address the potential for algorithmic discrimination, guard against civil rights violations, advance accessibility for all users, especially children and students with disabilities, and move to close overall gaps in design, use, and access of edtech.

Key Ideas

- Algorithmic discrimination could result in unfair distribution of opportunities to learn, resources and supports for learning, or outcomes of learning. NIST has identified three major categories of AI bias to be considered and managed: systemic, computational, and human, all of which can occur in the absence of prejudice, partiality, or discriminatory intent.
- Civil rights in educational settings are established in law and enforced by the Department's Office for Civil Rights. Developers should be well-informed about existing civil rights laws that apply to educational settings and design to comply with these laws. Existing civil rights laws apply no matter to what degree AI is implicated in the violation.
- Pre-existing and forthcoming AI training data sets should seek to reduce bias and represent educational user diversity. Educators are already expressing high awareness of these potential issues.
- Inclusion and accessibility are areas in which the capabilities of AI to support multiple forms of human interaction and augment human strengths and needs may be particularly beneficial.
- Digital equity encompasses attention to gaps in design, use, and access.

What to Know

Advancing equity is a broad concept but also points to a set of more specific considerations that developers should center in their work. This guide names civil rights[4], algorithmic discrimination, and accessibility as specific equity-related topics and uses the term "digital equity[5]" to point to additional issues (i.e., fairness in the affordability of technologies) that are important yet may not rise to the level of violations of laws. Existing laws (including civil rights laws) are paramount and apply to situations where any variation of AI leads to any discrimination across the continuum of a child's learning experience. Civil rights laws protect students against discrimination based on protected characteristics[6]— and apply to learning experiences inside and outside the classroom during the school day.

The Executive Order on AI directs many federal agencies with advancing the connection of AI equity and civil rights, both separately and in coordination across agencies. Developers should monitor agency websites to seek to stay abreast of policy guidance and other resources advanced by the Department and by other agencies, such as NIST and the U.S. Department of Justice, related to advancing equity and protecting civil rights.

Further, the Executive Order on AI directs the Assistant Attorney General in charge of the Civil Rights Division to meet with all federal civil rights offices "to discuss comprehensive use of their respective authorities and offices to prevent and address discrimination in the use of automated systems." This meeting occurred on January 10, 2024 (see Readout). The Department's Office for Civil Rights may evaluate and/or investigate allegations of civil rights violations stemming from the use of AI-enabled systems in educational settings.

Algorithmic Discrimination

Likewise, the Biden-Harris Administration made its position on the potential for algorithmic discrimination abundantly clear in the Executive Order on AI:

> *"My Administration cannot—and will not—tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice."*

The Office of Science and Technology's Blueprint for an AI Bill of Rights defines "algorithmic discrimination" and outlines what developers should do as follows:

[4] As appropriate, developers should consider the civil rights of educators, organizational employees, or other end users. Developers may find this overview of the Department of Justice's Civil Rights Division's work on AI and civil rights of interest in making such determinations. Importantly, this guide does not attempt to discuss all of the civil rights provisions that could potentially apply to AI use in education settings.

[5] Section 60302(10) of the Infrastructure Investment and Jobs Act defines "digital equity" as "the condition in which individuals and communities have the information technology capacity that is needed for full participation in the society and economy of the United States."

[6] Title VI of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000d, 34 C.F.R. Part 100; Title IX of the Education Amendments of 1972, as amended, 20 U.S.C. § 1681, et seq., 34 C.F.R. Part 106; Section 504 of the Rehabilitation Act of 1973, as amended, 29 U.S.C. § 794, 34 C.F.R. Part 104; The Americans with Disabilities Act of 1990, as amended, 42 U.S.C. §§ 12131, et seq., 28 C.F.R. Parts 35 and 36; The Age Discrimination Act of 1975, as amended, 42 U.S.C. § 6101, et seq., 34 C.F.R. Part 110.

> *Algorithmic discrimination occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections. Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way.*

Developers should proactively and continuously test AI products or services in education to mitigate the risk of algorithmic discrimination. In addition, training educators on both proper and inappropriate use of their solutions may mitigate issues of algorithmic discrimination in specific educational applications of AI (or prior uses of machine learning) according to researchers. Outcomes of algorithmic use of such data in education could deny students in a protected class from equitable opportunities in learning and achievement.

What is important is that the overall impact should not be inequitable. This applies to the use of AI enabled solutions in curriculum, as well as any technology or solution that can be used for monitoring behavior, classroom management, or discipline. Educators have many specific bias concerns related to computer vision algorithms and yet recognize that other forms of input such as speech recognition could also prove problematic.

The Department notes that the potential for algorithmic discrimination will not be limited to an application that makes obviously big decisions such as guiding student course or career selection but could also occur in a series of smaller decisions (for example, in the pacing or content of technology-based lessons), which in aggregate effect of the smaller decisions leads to an inequitable learning opportunity for students. Developers can address such concerns early in the process of building and training an AI, including by collecting representative data, exercising care in how data are curated and how algorithms are selected, testing for bias, and more.

Regarding shared responsibility, throughout the Department's listening sessions, educators consistently expressed strong concerns about bias in the data sets that are used to train AI models, both in foundation models and in tuning models to educational applications. Bias in the model's performance may occur because existing data sets inherently include historical biases and omissions—data sets may not be representative of the educational population who will participate as users. Developers should likewise build in opportunities for human review and strengthen features in products that increase transparency about the outputs that AI generates or the reasoning behind AI-based recommendations. They should also proactively design their products for learner variability, as will be discussed below.

"To simultaneously keep students safe and secure and train AI for equitable solutions with broad and diverse data is a defining challenge for this space. Simply saying 'no data use' will not keep all students safe or equip the equitable access we know all students deserve. As this field is rapidly evolving, thoughtful solutions must be identified to ensure safety, accountability, fairness, and effectiveness."

—Karl Rectanus, edtech entrepreneur

While risks exist, early research suggests[7],[8] that AI could be useful in helping better support children and students with disabilities, multilingual learners, and other populations that have long encountered barriers to learning with resources designed with less accommodation to their needs. The Department's 2023 AI Report outlines additional key areas of equity in the design, development, and deployment of AI-enabled systems in education with these examples:

- The report observes that adaptive algorithms in past edtech products often were more focused on deficits, weaknesses, mistakes, or gaps. Although addressing mistakes and errors is important in the development of AI-enabled tools, developers should consider balancing this approach in the development of such tools with a more asset-based design that deploys additional AI capabilities focused on assets to build on students' strengths and interests, also aligned with available community resources and assets.

- The capabilities of AI to support a wider range of inputs (e.g., speech, gesture, drawing) and outputs (e.g., translations among languages, annotations of images with language, automated production of American Sign Language) can provide additional supports to children and students with disabilities and can do so more evenly across varied learning activities and instructional resources as well as assessments.

- Similarly, the newer capabilities of AI can be aligned to achieve advances in equitable support for multilingual learners via translation support and identification of culturally responsive resources to accompany instruction.

  "We are excited about the potential of AI to allow us to adapt content more easily to students' linguistic needs and personal interests. For example, we have been field testing an AI-based system that prompts students for their interests and writes a word problem that fits their interests as well as the mathematics topic they are learning."

  —Dr. Steve Ritter, Founder and Chief Scientist at Carnegie Learning

More generally, the 2023 AI Report highlights how AI could be used to adapt instructional resources to all aspects of learner variability; whereas prior edtech services may have been most effective for students most similar to a developer's intended target population, developers could seek to use AI to serve a "long tail" (wider distribution) of student strengths and needs.

[7] https://journals.sagepub.com/doi/abs/10.1177/00400599241231237

[8] https://library.iated.org/view/MEHIGAN2024CON

Accessibility

As they seek to support learner variability, developers should be aware of and follow requirements of the Individuals with Disabilities Education Act (IDEA, as amended in 2004). Also developers should review Section 504 of the Rehabilitation Act of 1973 (Section 504). IDEA highlights the importance of educational resources that *leverage students' strengths* and not only resources that address their challenges or needs. Section 504 prohibits discrimination on the basis of disability by recipients of federal financial assistance. Digital accessibility is a component of accessibility for students with disabilities. Developers should look both broadly and narrowly for support toward incorporating digital accessibility into their solutions.

The Web Content Accessibility Guidelines (WCAG), developed and maintained by the World Wide Web Consortium (W3C), is an approved ISO standard and is recognized by developers as the benchmark for creating content that is accessible without limitations to all users. AI can enable edtech developers to support interactions with people using a broader range of modalities (e.g., speech, gesture, American Sign Language, etc.). Incorporation of such capabilities could help not only specific learner populations but all learners as well. More focused on education solution design, Universal Design for Learning (UDL) is one well-established framework for guiding design of tools that "improve and optimize teaching and learning for all people based on scientific insights into how humans learn." Developers may find UDL's three broad guidelines to be a good starting point for conceptualizing how AI could improve learning via their product or service:

1. **Engagement**. Presenting educational content in new interactive formats can increase student engagement. As an example, with appropriate guardrails, AI can support U.S. or World History courses by enabling students to interact with historical interviews of real people.

2. **Representation**. Presenting educational content options via multiple entry points of representations (e.g., text, audio, graphics, animation) benefits learners; in a simple example, AI may support generating more useful "alt text" to accompany images accessed by screen readers, even if humans remain in the loop to verify AI generated alt text.

3. **Action and Expression.** Giving students enhanced opportunities to act and express themselves improves learning; in one research-based example, students can learn science by engaging in an animated, interactive narrative where AI makes the plot, characters, and dialogue adaptive.

Digital Equity

More broadly, the Office of Educational Technology's 2022 report, Advancing Digital Equity for All, describes ongoing gaps in access to devices and broadband connectivity that impact educational opportunity but goes beyond access to also include affordability and unequal adoption as concerns, as well. We provide examples of potential pitfalls that developers should be aware of in developing and deploying AI systems:

- Access: AI-enabled applications may first appear in districts of more wealthy communities, or conversely, under-resourced schools may rely upon more affordable AI resources instead of human resources.

- Affordability: In more affluent households, guardians may be able to afford premium subscriptions for the most powerful versions—especially if students are expected to use AI-enabled tools at home.

- Adoption: Lack of community buy-in for valuable AI-enabled educational products may occur when communities are less informed or less involved in the design and marketing of those products or when other necessary resources (such as teacher professional learning and development and accessibility of tools) are differentially available.

Developers serving edtech interests have equity-related considerations other sectors may not; they should be aware of the potential pitfalls and the steps they can take toward equitable access, affordability, and adoption. The 2024 NETP has additional recommendations regarding equity that may provide useful guidance to developers. Specifically, it discusses three types of gaps:

1. Digital Use Divide: addressing opportunities to improve how students use technology to enhance their learning, including dynamic applications of technology to explore, create, and engage in critical analysis of academic content and knowledge; and ensuring that students have equitable opportunities to use technology for learning.

2. Digital Design Divide: addressing opportunities for educators to expand their professional learning and build the capacities necessary to design learning experiences enabled by technology that serve the diversity of their students.

3. Digital Access Divide: addressing opportunities for students and educators to gain equitable access to educational technology, including connectivity, devices, and digital content. This also includes accessibility and digital health, safety, and citizenship as key elements of digital access.

Questions To Ask

1. How can we connect civil rights and digital equity to specific ideas that inform our work as developers and ensure we are following applicable federal laws?

2. How does the role of the education leaders in protecting civil rights relate to our product or service?

3. What steps can we take to audit and remove the potential bias or algorithmic discrimination in our product, with special attention to mitigating any impacts for vulnerable or underserved populations?

4. How could we leverage AI in our product specifically to enhance accessibility and inclusion?

5. What is our long-term strategy to be a positive force for digital equity in all its dimensions, from design all the way through distribution and use?

Directions to Pursue

- Developers can infuse their organizational culture with equity and civil rights priorities from underlying data sets used for training algorithms to UI/UX choices.
- Developers may establish or improve a review process/checklist for new platforms, enhancements, and/or extensions to ensure broad representation in solution performance.
- Developers may build feedback loop mechanisms with organizations and relationships with expert practitioners for equitable learning experience design.
- Developers can work to stay current with both mainstream and edtech standards bodies and their pending work to address racism and other forms of algorithmic discrimination in AI-enabled products and services.
- Developers may participate in regular third-party review processes for eliminating bias from underlying databases, algorithms, and even UI/UX design elements that exclude certain cohorts from equitable user experiences.

Resources

- NIST is working on a standard for identifying bias in AI, with three main ways in which bias occurs (systemic, statistical/computational, and human), which intersect with different foci for mitigating bias (in datasets and models, during development of applications, and as applications of AI are used in the field).
- CAST's Universal Design for Learning is a research-based framework that can be used to guide applications of AI to build on students' strengths.
- The Center for Democracy and Technology provides extensive resources and guidance on equity, civil rights, and AI.
- Culturally Responsive Teaching and other design guidelines of processes are available to do equity-relevant design.
- The Leadership Conference works to ensure that new technologies further civil rights protections, and also works on related Digital Equity issues.
- The Learner Variability Navigator is a tool to find research-based strategies to support whole-child learning.
- Several nonprofits have design frameworks and services aimed at addressing equity. Two examples are the National Equity Project and Center for Inclusive Innovation.

Recommendation 4. Ensuring Safety and Security

Educational leaders and decision makers are looking to developers for partnership in strong plans to both address well-known risks and manage the broader range of potential risks so that AI technologies that improve teaching and learning can be implemented safely for students, staff, and communities. Developers who are safety- and security-centered will prioritize protecting students' and teachers' data security and privacy; developers will also acknowledge that risks now go beyond these broadly known, longstanding edtech risks and therefore conduct risk identification, prioritization, and management throughout their development and deployment processes.

Key Ideas

- Developers must be aware of the federal laws and related guidance regarding privacy and data security, such as these:
    - Family Educational Rights and Privacy Act (FERPA)
    - Protection of Pupil Rights Amendment (PPRA)
    - Children's Online Privacy Protection Act (COPPA)
    - Children's Internet Protection Act (CIPA)
- Protecting privacy and enhancing cybersecurity are key issues that school technology leaders manage as they procure, implement, and monitor technology in their educational institutions.
- Developers of AI applications for education must identify and mitigate risks that go beyond privacy and cybersecurity (see Box E).
- NIST has produced an AI Risk Management Framework that can guide a comprehensive and continuous process for identifying, prioritizing, and addressing risks.

What to Know

Privacy and data security are the aspects of edtech where the strongest guidelines and guardrails already exist. Most participants in the edtech marketplace have been actively addressing privacy and cybersecurity for many years before generative AI became widely available and will continue to require strong safeguards. The Center for Democracy and Technology reports privacy as a chief concern among parents and students using edtech while learning at school—especially for students with individualized education plans or 504 plans.

Educational leaders are committed to procuring tools that protect their students, staff, and communities. The Department provides support around federal requirements and best practices for protecting student privacy. For example, the Department's Student Privacy Policy Office (SPPO) offers extensive information on protecting student privacy in accordance with applicable federal laws and regulations. SPPO resources include information on the Family Educational Rights and Privacy Act (FERPA) and the Protection of Pupil Rights Amendment (PPRA), which apply to educational agencies and institutions receiving Department funds. The Department's Office of Educational Technology and the Privacy Technical Assistance Center (PTAC) host content and events to support developers in safeguarding student data. In addition, the Children's Online Privacy Protection Act (COPPA) adds protections for students younger than 13 years old, and the Children's Internet Protection Act (CIPA) addresses risks related to harmful content in schools and libraries. Developers must know these laws.

Developers should be aware that both states and associations provide leadership and direction to the work of leaders in local educational agencies (LEAs) as they tackle these issues. As one example, the Utah State Board of Education has a Student Data Privacy Team focused on supporting LEA efforts to protect student data. Most states have similar initiatives. The Consortium for School Networking (CoSN), an organization that supports school information technology professionals, has produced a parallel NIST cybersecurity framework for education to assist with classifying and understanding resources. Their membership and resources reflect strength in both cybersecurity and its extension to AI-enabled products and services, including a readiness checklist developed in conjunction with other learning organizations. The Data Quality Campaign provides extensive coverage of student data privacy and security policies, including information on how state and local jurisdictions go beyond federal protections. Additional resources are offered by the Student Data Privacy Consortium, which brings together constituent groups in the edtech marketplace to set expectations about student data privacy.

> *"AI isn't a monolith. When developers think about responsible design and use of AI tools in education, it's important to consider the range of AI applications and align their risk mitigation strategies to the relative risk of each tool. One strategy is to keep humans in the loop—for example, involving educators at optimal moments that help minimize the risks and maximize the value of these new technologies."*
>
> *—Teddy Hartman, Head of Privacy and Trust, Pear Deck Learning*

Risks of AI in educational settings extend beyond the well-known challenges of data security and privacy. The Department urges developers to rapidly begin to identify, prioritize, and manage additional risks. Consider the following examples, which use the categories in Box E and are based on publicly reported incidents:

- In an example of a race to release risk, generative AI chatbots became widely available to schools before any guidance was available to teachers and students on acceptable use, distracting educators from managing the core functions of teaching and learning.

- In an example of a Bias and Fairness risk, test proctoring systems that rely on facial recognition algorithms were suspected of unfairly and disproportionately disciplining non-white students.
- In an example of Harmful Content risks, students often use tools to create images for use in their school projects, and generative AI has been found to offer negative and hurtful stereotypes when given queries to construct an image for "black girls."
- In an example of a Malicious Use Risk, students have been found to use generative AI in cyberbullying, for example, constructing false and negative narratives about fellow students or false images about fellow students.
- In an example of Hallucination Risk and Wrong Information Risk, generative AI has been found to produce output that describes historical figures who never existed and to give wrong answers to math problems.
- In an example of Transparency Risks, educational procurement officers face pressure to allow various AI-enabled technology into schools, well before adequate information is available that would allow them to apply good judgment and ask hard questions about data sources, algorithms, risk mitigation, and other requirements.
- In an example of Underprepared User Risks, human decision makers sometimes defer to algorithms, which weakens the important role of educator judgment in how to support students and undermines the human in the loop recommendation from the Department's 2023 AI Report. Likewise, human decision makers sometimes absorb biases from the output of AI, interfering with their own better judgement.

An important starting point for mitigating AI-related risks is NIST's AI risk management framework. Although this framework is applicable to sectors broader than education, it can offer valuable guidance as developers identify and implement specific risk mitigation approaches that are more uniquely tailored to the education sector and encompass all identifiable risks. See Figure 6 and Figure 7.

Regarding **risk identification**, risks can occur at multiple scales, from harming individual people to harming an ecosystem. In addition, per the earlier discussion, AI-enabled systems for education will be developed with components from organizations outside the primary developer, and thus risks may emerge along a supply chain of components that are utilized in an overall product or service. Supply chain vulnerabilities can occur when a component is judged to be "safe" at an earlier time but then changes its behavior without notice to the component user; for example, foundational models have been observed to degrade at times for unknown reasons or to suddenly begin making errors as developers tweak their algorithms. Regarding **risk prioritization**, the European Union AI Act defines levels of risk. Currently, there is no similar official definition of levels of educational risks in the United States; however, developers are beginning to index specific risks in their applications and develop targeted

mitigation strategies. The Department urges developers, working in their own organizations and in shared responsibility with educators, to consider how to identify and prioritize risks.

> *"Our team developed a detailed list of potential risks as we designed our AI-enabled math tutoring product. Then we prioritized the risks and developed our responses. We'd love for educators to ask us about these specific risks, and we'd welcome feedback on how we've addressed them."*
>
> *—Kristen DiCerbo, Chief Learning Officer, Khan Academy*

Figure 6: Categories of Harm Described in NIST's AI Risk Management Framework

The Department, following the lead of NIST, suggests developers take a lifecycle approach to managing these and other risks that must be identified (Figure 7). The four aspects of a lifecycle risk management framework are:

1. **Govern**: Cultivating a culture of risk management within every edtech developer organization, for example, ensuring that developers in every phase or component of product development are aware of potential risks, their responsibilities to identify product-specific risks, and how to engage with product managers to mitigate risks.

2. **Map**: Recognizing the special features and challenges of school contexts and identifying specific risks that arise in those contexts. Developers can look throughout this document for examples of the specific kinds of risks that arise in school contexts as well as their special responsibilities and obligations under law when working with minors and students.

3. **Measure**: Documenting, analyzing, and tracking risks not only as the product is developed, but also as it is field tested in classrooms or other educational settings and more widely distributed and used. For example, developers can apply research-based methods to measure bias; the Department's 2023 AI Report included an extensive

discussion on how techniques from educational psychometrics could be helpful in ensuring fairness of AI-enabled educational resources or assessments.

4. **Manage**: Prioritizing risks and acting to protect students, teachers, and educational communities. Throughout listening sessions, the Department learned about development organizations that are prioritizing risks early in their development process, during the process of building each product feature or component, and after the product is launched.

Figure 7: Overview of the NIST AI Risk Management Framework

Educators look to developers to share responsibility for managing risks. On the one hand, state guidance resources (such as those mentioned above in Box C) instruct school districts to govern, map, measure, and manage risks. On the other hand, risk management capacity for AI in schools is likely limited and, thus, assumptions that risks known to a developer or known to the public can be managed by those closest to teaching and learning are unlikely to be borne out. Educational institutions will need partnerships with developers to help them manage risks. Shared risk management measures, implemented in customer service agreements between solution providers and end users in schools, can provide an added value that is not only ethical but also fosters trust through deliberate attention and thorough, regularly updated documentation to mitigate any potential risks, known or unknown. For example, PTAC offers a resource on developing a model terms of service.

Questions to Ask

1. What are we hearing as top safety and security concerns from educators regarding our product or similar products?
2. How do applicable federal, state, and local laws and guidance, such as the White House Blueprint for an AI Bill of Rights, apply to our product or service?
3. What specific privacy, security, or other safety risks are most relevant to our product and to vulnerable groups, and how can we mitigate them?
4. How can our organization systematically govern, map, measure, and manage AI risks beyond privacy and security?
5. What overall risk management strategy throughout our organization would lead to growing our positive reputation for delivering AI-enabled edtech with the strongest protections for privacy, security, and safety?

Directions to Pursue

- Developers should draft clear and plain language disclosures about how an organization protects student data security and privacy.
- Developers should bolster accountability efforts via audits or other procedures for inspecting and testing protections, as well as obtaining feedback on risks from end users, with special attention to vulnerable and underserved populations.
- Developers should collaborate across product lines or with other companies to articulate shared standards or approaches for addressing risks in educational products and services that incorporate AI, creating approaches to address issues of risk not only within an organization but also with upstream suppliers and downstream consumers.
- Developers should cultivate awareness of how the public, users, and regulators perceive levels of risks related to AI educational products and services and should respond if noteworthy risks emerge in an offering.
- Developers should prioritize staying abreast of rapidly evolving legislation and other governance activities related to AI in the federal government, in states, and locally.
- Developers should be thoughtful about addressing the interplay between state and federal policies. At the time of publication, more states are developing and releasing AI policies.

Resources

- The National Science and Technology Council's National Strategy to Advance Privacy-Preserving Data Sharing and Analytics.
- Cyber and Information Security Agency's Secure by Design Framework
- Software and Information Industry Association principles on developing AI-enabled edtech products
- Data Quality Campaign resources on protecting student privacy
- CoSN's NIST Cybersecurity Alignment for K-12.
- Additional resources at the Center for Democracy and Technology.

Developers who are working internationally may also wish to consider the European Union AI Act's levels of consequentiality of educational risks; more generally, staying abreast of developments worldwide may be important. One useful example is the Canton of Zurich's Legal Best Practices regarding AI in education.

Recommendation 5. Promoting Transparency and Earning Trust

As a sound business approach to the education market, developers engage in two-way communications with educators, students, and others in the ecosystem about AI, including promoting transparency for how AI has been implemented in an educational application, addressing concerns, and working together to expand the strength of shared responsibility.

Key Ideas

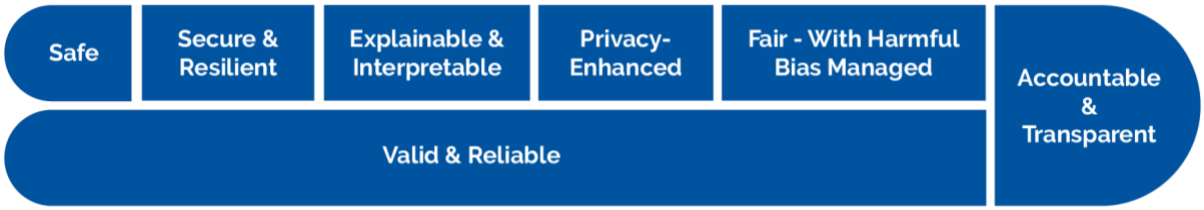
- Transparency contributes to trust.
- "Trustworthy systems" is an important technical area of research and development (e.g., the NIST AI Risk Management Framework), and trust is a relationship of mutual confidence among those who create AI-enabled educational systems and those who use them.
- Developers can grow trust by contributing to AI literacy among educators, parents and caregivers, students, and other constituents; conversely, without strong AI literacy, assurances by developers may fail to earn trust.
- To sustain trust, developers should tend to the level of ethics training among their teams.

What to Know

Developers who attended the Department's listening sessions viewed all four recommendations previously addressed in this guide (i.e., designing for education, evidence, equity and civil rights, and safety and security) as essential to developing trust in their education markets and communities they serve. Another way of representing this multifactor approach to earning trust is NIST's analysis of the characteristics of trustworthy AI systems. See Figure 8.

> *"Schools deserve AI solutions that are purpose-built for learning. That means solutions must adhere to the highest levels of safety, privacy, and security with models that are hallucination resistant and trained on vetted education-specific content. That's why we're building AI models specifically for education."*
>
> *—Latha Ramanan, Head of Responsible AI, Merlyn Mind*

Figure 8: Characteristics of Trustworthy AI Systems



The Department encourages developers to attend both to *trust* as the mutual confidence of two parties (e.g., relationships among developers and adopters of edtech) and *trustworthiness* as ascertainable properties of a technical system (as indicated in the image above).

An example based on the potential for AI to support "stealth assessment" may clarify the importance of transparency and relationships to trust. In its original meaning, "stealth" was a synonym for "unobtrusive" and was intended to provide authentic and supportive feedback to students while maintaining high student engagement in learning[9]. However, "stealth" can also imply surveillance—that students may be measured invisibly, without their knowledge, and with no direct and obvious benefit to their learning. With regard to the original meaning, students and teachers want to reduce the amount of time taken away from learning for testing and thus may appreciate unobtrusive assessment, especially when it's clear how the outputs help teachers and students immediately and directly. Yet, with regard to the second meaning, teachers, students, and parents may be rightly concerned if sensitive learning data is shared beyond the classroom without their knowledge, and that could have unforeseeable consequences for the students' performance, well-being, and future opportunities. Clearly communicating the purposes and limits of unobtrusive data collection—as well as involving teachers and students in meaningful data use—can make the difference in trust.

Due to the relational nature of trust, earning trust is also nurtured by how developers engage in communication with other ecosystem participants. Examples that developers raised in listening sessions included these:

- Trust is buoyed by transparency and disclosure.
- Trust requires effective listening and sharing.
- Trust is enhanced by demystifying why, how, and what AI is employed.
- Trust is increased when developers, educators, and researchers (among others) work together in feedback loops to identify problems and address them.
- Trust is strengthened when developers are active in forums that advance public interests.

[9] Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), Computer games and instruction (pp. 503-524). Charlotte, NC: Information Age Publishers.

Some edtech developers have already released voluntary commitments or voluntary disclosures about how they mitigate risks of AI as they develop and improve their products (see Box F).

The Department also notes that developers are already participating in public forums and initiatives where work on safe, secure, and trustworthy AI is occurring through mutual contributions of developers, educators, researchers, policymakers, funders, and more. Box D includes a list of nonprofits that are active in creating such forums. These forums are places where developers can contribute to the public interests by participating in shaping further guidelines and guardrails.

Indeed, the Department views voluntary commitments, voluntary disclosures, and participation in forums as a first step toward articulating a "dual stack" approach where developers have equally strong coordinated systems to ensure responsibility and to achieve innovation. Transparency starts with commitment but should go beyond commitment to also discussing how edtech organizations are orchestrating their development processes to be responsible from product conception to delivery and from foundational models to educational applications.

Transparency includes participating in forums but should go beyond sharing information to working together over time to build trust. The Department encourages developers to act on a belief that a healthy ecosystem requires them to step up not only with competitive innovations but also with contributions that advance the broad public interest in safe applications of AI in education.

Further, developing trustworthy edtech will likely require that developers' teams have training in AI ethics, equity, and related concerns. The Association of Computing Machinery is active in developing ethical principles for AI and other emerging technologies, and their code of ethics is designed specifically to shape the work of software developers. Trust is fortified when developers are aware of ethical principles and publicly discuss how they apply the principles in their work.

Regarding transparency, for example, LLMs are black boxes to developers and users. Further, competitive factors can overpower desires and the best intentions to communicate transparently, for example, about sources for training data. And yet, listening session attendees encouraged the Department to call for clarity of expectations: developers and educators mutually require an articulate sense of what improvements in teaching and learning can be expected. Here are some examples:

- Developers can provide opportunities to interact with their product or service and explore it deeply before an educator purchases or extensively uses it.
- Developers can provide strong training and professional development around the roles and responsibilities necessary for safe and effective use of their product or service.
- Developers can provide service guarantees around how they are available to quickly respond to and mitigate any issues that arise, including opportunities to override aspects of the product or service that are not meeting expectations.

Clarity of expectations, coupled with strong feedback loops among developers and educators, can lead to not only achieving those expectations but also exceeding them. For example, in one high-quality practice that already exists in the edtech industry, customer success managers support educators' usage with fidelity in the field, trust is fortified, and the solution provider builds a robust feedback loop for informing enhancements and updates to the solution—hopefully with fewer iterations than if designers and developers tried to innovate in a vacuum. And thus, conversations with developers in listening sessions led to an understanding of how progress through transparency is possible, highlighted in Box H.

Box H: A cycle of shared responsibility and transparency that can earn trust.

Conversations with developers led to articulating this repeatable cycle of steps that can increase trust:

- Shared responsibility begins with aligning expectations about what AI can deliver in an educational setting to be described responsibly to achieve shared educational visions, without hype or neglect of potential risks.
- Further transparency can emerge among developers and educators as they collaborate closely to address ethics, evaluate evidence, protect civil rights, address equity, and manage specific risks related to their innovations.
- Shared responsibility also increases when developers engage in forums that focus on the public interest in safe AI and where the developers can share norms, standards, and values that, if widely adopted, could make AI safer for all ecosystem participants.
- Exceeding expectations both in the AI-driven product experiences and in public contributions around responsible AI develops strong relationships with customers.
- With strong relationships around meaningful educational value and shared risk management comes trust.

Pursuing this virtuous cycle makes sense, and yet capacity may be limited by the AI literacy levels in non-developer populations. Presently, levels of AI literacy vary widely among people in the education sector, and developers can also play an important and positive role by strengthening AI literacy. Key areas for helping educators include explaining basic data governance concepts and skills, as well as how they relate to AI-specific terminology and functionality. Developers who demystify their usage of AI and provide users with an accessible concept of how AI works in their educational system show respect for their users and contribute to the growth of human agency as people adopt AI tools in education, which in turn, develops trust in the solution provider. Many learning organizations have entered the space to focus on preparing the field, and developers can learn from as well as support these initiatives for mutual benefit.

Questions To Ask

1. How can we engage with customers with balanced attention to our innovations and our responsibilities?
2. How could the NIST AI Risk Management Framework help our team to develop trustworthy AI systems?
3. What steps toward transparency could we take regarding our use of AI in products and services?
4. How can our organization contribute to AI literacy in the broader edtech ecosystem?
5. What is our long-term plan to achieve respect for our responsible use of AI in support of the overall value of our products and services to students, teachers, and others in education?

Directions to Pursue

- Developers should work to promote transparency. Developers can demonstrate a commitment can be addressed in marketing by highlighting the dual stacks of responsibility and innovation.
- Developers can share more openly written commitments and disclosures but should also emphasize two-way communication and collaborations with educators during product development and improvement.
- Developers should support efforts to build AI literacy in the ecosystem.
- Developers may consider how to publicly describe the characteristics of trustworthy system architecture achieved in their products and services. Some characteristics such as explainability may be hard to achieve in the short term, but related concepts like interpretability may be possible now while research and development continues.

Resources

- Trustworthy artificial intelligence (AI) in education: Promises and challenges | OECD Education Working Papers | OECD iLibrary
- Common Sense Media's AI Initiative
- Trust The Process: How To Choose and Use EdTech That Actually Works - EdTech Evidence Exchange
- Building trust in EdTech: Lessons from FinTech
- The Association for Computing Machinery's Code of Ethics

Conclusion

Educational decision makers express cautious optimism for new products and services that leverage new capabilities of AI. As indicated throughout this guide, educators see a wealth of opportunities to use AI to achieve the vision of their educational institutions—and yet they must be well informed of risks that must be addressed. Educational decision makers thus stress the duality of focusing on important opportunities and taking strong, clear steps to address risks. This duality shapes the market opportunity for today's educational developers.

> *"Would I buy a generative AI product? Yes! But there's none I am ready to adopt today because of unresolved issues of equity of access, data privacy, bias in the models, security, safety, and a lack of a clear research base and evidence of efficacy."*
>
> *—Patrick Gittisriboongul, ED.D, Asst. Superintendent of Lynwood Unified School District, California*

By organizing the many areas of opportunity and concern into five topics, the Department aims to sharpen developers' attention to topics of enduring importance:

1. Designing for Education
2. Providing Evidence for Rationale as well as Impacts
3. Advancing Equity and Protecting Civil Rights
4. Ensuring Safety and Security
5. Promoting Transparency and Earning Trust

"Earning the public trust" is of paramount importance as developers launch new applications of AI in education. The Department envisions a healthy edtech ecosystem highlighting mutual trust amongst those who offer, those who evaluate or recommend, and those who procure and use technology in educational settings. The Department has found an e-bike analogy to be a good starting point for discussions across the ecosystem, offered in the Department's 2023 AI Report: that teachers and students should be in control as they use the capabilities of AI to strengthen teaching and learning. Just as a cyclist controls direction and pace but preserves energy with the assistance of an e-bike's drivetrain, so should participants in education remain in control and able to apply saved energy and time for the most impactful interactions and activities when technology amplifies their choices and actions. Now, continuing the analogy, developers should take precautions to design AI-enabled educational systems for safety, security, and to earn the public's trust, just as riders would expect e-bike developers to ensure their rider's safety and security, and to earn the public's trust.