Principles for the Use of Artificial Intelligence in Education Research

Draft for Public Comment

FEBRUARY 2026

Contents

Exhibits

Principles for the Use of Artificial Intelligence in Education Research

This brief shares results from a collaborative Artificial Intelligence Principles for Education Research Design Lab facilitated by the American Institutes for Research® (AIR®) between December 2024 and December 2025. Recognizing the need for shared understanding and collaboration, this brief reflects contributions from many organizations and perspectives. All contributors agree that AI holds promise for education research and that the field has a responsibility to use it appropriately and ethically.

Contributors

Jill Burstein, Duolingo
Russell Cannon, Gates Foundation
William Corrin, MDRC
Kyle Fagan, American Institutes for Research
Rachel Garrett, American Institutes for Research
Laura Hamilton, Center for Assessment
Kristen Huff, Curriculum Associates
Orrin Murray
Julie Neisler, Digital Promise
Pati Ruiz, Digital Promise
Susan Therriault, American Institutes for Research
Amelia Vasquez, American Institutes for Research

In Attendance

Evan Heit, National Science Foundation
Matthew Soldner, Institute of Education Sciences

Acknowledgments

The contributors to this brief would like to extend their appreciation and gratitude to Tori Cirks (American Institutes for Research) for her time, expertise, and facilitation of the Design Lab.

The project team conceived, developed, and authored this brief. Microsoft Copilot was used for editorial refinement, such as improving clarity, structure, and tone. Responsibility for all substantive content rests entirely with the contributing authors.

The Need for Guidance

The development and use of artificial intelligence (AI) has risen rapidly in recent years. Today, AI is a central focus in both national and global contexts, and the education sector is no exception. AI is increasingly being integrated across the field of education, from federal policy directives to everyday classroom experiences, and even in education research, where AI-enabled approaches hold significant potential to advance and reshape the field.

Education research has long sought to understand education systems, policies, and practices with the goal of improving outcomes for teachers and students. As in other fields of research, trust is foundational in education research. Established research methods and agreed-upon practices have evolved in part to ensure the credibility, rigor, and beneficence of education research. Over the last several decades, the quality of research and methodological sophistication has improved substantially. At the same time, education researchers have made sustained, albeit imperfect and ever-improving, efforts to communicate findings in ways that are accessible and meaningful to policymakers and practitioners.

Education research is at an inflection point. Recent advances in AI have the potential to accelerate progress in education research and advance the field more rapidly. AI tools can help to address several long-standing challenges, including the speed at which analyses are conducted, particularly when working with large and complex data sets, as well as the cost and efficiency of research. In addition, AI offers new opportunities to improve the accessibility and usability of research findings for policymakers and practitioners. At the same time, the use of AI introduces potential risks to the established norms and standards that guide education research. For example, AI may challenge long-standing expectations around replicability, transparency, and responsible data use. Further, AI models may generate inaccurate or misleading outputs that could undermine the integrity and trustworthiness of research.

The education research community has a collective responsibility to proactively adapt to the use of AI while maintaining the integrity of the field. To advance this work, we convened a group of education researchers, industry experts, and funders to develop an initial set of core principles to guide the responsible use of AI in education research. This brief presents the outcome of that collaborative effort. We present these principles with the understanding that:

- **The evolution of these principles is guaranteed and accepted.** The development of the principles was an initial step. Even experts in AI cannot fully predict the pace, form, or impact of its evolution. As a result, these principles must be revisited regularly to ensure that they remain aligned with advances in AI and their implications, which cannot yet be anticipated.
- **These principles position AI as a tool or assistant, with humans retaining leadership and accountability for education research.** Although some envision a future in which AI replaces human work, these principles affirm that research intended to serve people must remain under human direction. This approach is explicitly human-centric and emphasizes thoughtful, moderated use of AI as researchers continue to learn about its capabilities, limitations, and appropriate applications in education research.

- **The AI principles are guideposts.** The proposed principles offer a foundation for organizing and guiding discussions and decision-making about whether and how to use AI in education research. They provide a shared framework that the education research community can use to surface, interpret, and discuss emerging developments related to AI use. At the same time, these principles are not intended to address every question or concern. Continued work is needed to understand and respond to the practical, on-the-ground implications of applying AI in real research contexts.
- **Education research community consensus on the responsible use of AI is essential.** As education researchers, we share a commitment to conducting ethical, responsible research that protects participants and offers clear potential benefits to those it serves. The AI principles for education research are grounded in, and extend from, these established norms.

We view this work as iterative and expect that it will be strengthened by the inclusion of additional perspectives and responsive to ongoing advances in AI. We invite others to participate as members of the field of education research. You can support these efforts by sharing these principles broadly, engaging in public discourse, offering feedback, and joining the effort to update these principles over time.

AI Principles for Education Research

Accepted guidance on the responsible use of AI is essential to support education researchers in making sound decisions about whether and how to use AI in their work. In developing and proposing the AI Principles for Education Research, our aim is to balance broad concepts with practical application, with the principles serving as a bridge between high-level vision and day-to-day practice. In doing so, the principles are designed to promote consistency, credibility, and consensus in the use of AI across education research.

We have developed three proposed principles for using AI in education research (see Exhibit 1): (a) Principle 1: Center human expertise; (b) Principle 2: Confirm model appropriateness; and (c) Principle 3: Operationalize transparency.

Exhibit 1. Overview of the AI Principles for Education Research

Principle 1. Center Human Expertise

The effective use of AI in education research depends on keeping human expertise at the core through intentional *co-design*, ongoing *human moderation*, and rigorous *human verification*. Together, these components will ensure that AI functions as a supportive tool guided by humans who apply professional judgement, ethical standards, and relevant domain knowledge.

Principle 2. Confirm Model Appropriateness

Model specification, *accuracy*, and *replicability* are fundamental to responsible and effective AI use in education research. Together, these practices will ensure that models are well aligned to educational contexts, evaluated for accuracy and

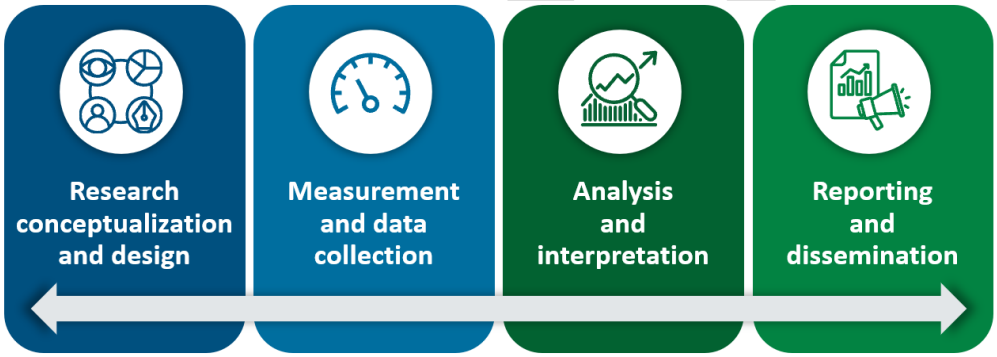
reliability, and able to produce consistent findings across data sets and settings, which, in turn, will support credible and actionable research.

Principle 3. Operationalize Transparency

Transparency is essential to ethical and rigorous AI-supported education research. Researchers should ensure *participant protection*, clearly *disclose how AI tools are used* and where human judgment applies, and appropriately *attribute contributions* from models, developers, and participants. Together, these practices support trust, accountability, and responsible reporting on AI in education research.

In the sections that follow, we examine each of the three principles in turn, illustrating their application across key activities of the education research process (see Exhibit 2).

Exhibit 2. Key Activities of the Education Research Process



Principle 1. Center Human Expertise

Co-Design | Human Moderation | Human Verification

The effective use of AI in education research begins with human reasoning, decision-making, and values. Amid all the automations and opportunities made possible by AI, human expertise remains paramount to realizing the potential of AI to extend traditional methods and advance improvements in education and education research. There are three core components to centering human expertise: co-design, human moderation, and human verification.

Co-Design. Responsible use of AI in education research begins with keeping people central to where and how AI is used across the research process. Keeping humans "in the loop" starts with the intentional co-design of AI use, including when and where AI is used, what decisions it informs, and what roles remain the responsibility of researchers. Although a growing number of AI-enabled methods and tools are currently available, relatively few are guided by clear norms for use in education research or informed by educational expertise. Education researchers are therefore uniquely positioned to shape appropriate uses of AI for the field by applying their domain knowledge to define use cases, establish guardrails, and integrate AI into research workflows in ways that align with existing standards and values. AI will not evolve in these directions on its own. By actively engaging the multiple constituencies that education

research seeks to serve, researchers can incorporate practitioner and community input to guide decisions about how AI is implemented, overseen, and verified in practice. This participatory approach supports responsible uptake of AI-supported processes and findings and helps to ensure that research remains relevant, transparent, and beneficial for its intended audiences.

Human Moderation. Human expertise remains essential throughout the entire research life cycle, not only during early conceptualization and design but also through analysis, interpretation, and the communication of findings. As AI systems become more capable and efficient, their value in education research will depend on sustained human judgment to guide appropriate use, assess outputs, and ensure alignment with research goals and ethical standards. Ongoing human oversight and moderation enable researchers to monitor processes, maintain and adjust guardrails, identify errors or unintended consequences, and course-correct when needed. This active supervision through the research process, including the ability to intervene when needed, helps ensure that AI models function as a supportive tool rather than a substitute for professional expertise, preserving rigor, contextual understanding, and responsibility in education research.

Human Verification. Human verification, validation, and confirmation are also essential safeguards when using AI for education research. This includes systematically reviewing AI-generated outputs to ensure that they are accurate, appropriate, and aligned with the original intent of the research. Rigorous quality control and quality assurance processes remain necessary because current AI models, including large language models, cannot reliably distinguish between true and false information or consistently differentiate facts from opinions. These systems are designed to generate responses and may "fill in the gaps" when information is incomplete or uncertain, at times producing plausible but incorrect or fictitious results. As AI is increasingly used to improve efficiency or automate components of the research process, the importance of human judgment will only continue to grow. The more responsibility is delegated to AI tools, the more critical it becomes to embed robust human checks and verification points. Researchers must therefore remain active in assessing outputs, results, and findings to ensure that AI-generated information is valid, reliable, and ultimately trustworthy.

Across all three components of centering human expertise, the involvement of the right human informants is essential. Neither humans nor AI can fully understand complex educational phenomena on their own; it is the combination of human judgment and AI capabilities that creates the potential for greater advances. Education researchers are uniquely positioned for this role. In practice, this means extending well-established research procedures, such as methodological review, validation, and interpretation, into AI-enabled contexts rather than treating AI as a fundamentally separate paradigm. When thoughtfully integrated, AI can help researchers access information that has historically been difficult to capture or analyze systematically, supporting deeper, more nuanced understanding on the part of the researchers. Ultimately, trust in AI-enabled research depends on approaches that are developed with appropriate expertise and on researchers' continued responsibility to assess outputs, uphold quality, and ensure that the research produced is both contextually valid and meaningful.

From research conceptualization and design through reporting and dissemination, researchers must clearly articulate the processes to ensure that humans remain in the loop as AI is leveraged in both planned and emergent ways. Exhibit 3 provides short examples of ways to do this.

Exhibit 3. Examples of *Principle 1. Center Human Expertise* in Practice

	**Run co-design sprints with practitioners** to refine research questions, define intended AI uses, and surface contextual constraints; document how teacher, student, and family input directly shapes the study design and tool selection. **Develop a "human-in-the-loop" plan** that names roles (e.g., principal investigator, subject matter experts, community advisors), decision points requiring human judgment, escalation paths for pausing or overriding AI use, and the ways that participant voices will inform any new AI uses that arise during the study. **Prespecify human verification protocols** (e.g., double-coding samples, expert review, acceptance criteria) so that any AI-assisted outputs meet documented quality thresholds before their inclusion in the research workflow.
	**Engage teachers, students, and relevant community members** in co-designing or refining measures, ensuring they validate what is being collected, how it will be collected, and why. **Confirm representativeness of data sources and samples** by involving human experts who can assess whether the artifacts, populations, and contexts used for AI-assisted measurement are appropriately generalizable and contextually relevant. **Use human reviewers to verify AI-generated or AI-coded measurement outputs,** such as constructs or classifications, ensuring accuracy and appropriateness before integrating results into the research data set. **Create human-driven quality assurance checkpoints** that require researchers to audit AI-enabled data collection tools (e.g., automated transcription, scoring) for errors, unintended outputs, and misalignments with the intended construct.
	**Use human experts to verify AI-assisted analyses,** such as AI-generated codes, themes, or classifications, ensuring outputs are accurate, appropriate, and aligned with the original analytic intent. **Conduct human-led validity and robustness checks**, for instance, reviewing AI-produced summaries, patterns, or statistical relationships, to identify errors, detect unintended consequences, and ensure alignment with research goals. **Maintain human oversight in interpreting AI-generated insights,** ensuring that contextual understanding and domain expertise guide conclusions rather than relying on AI-produced outputs at face value.
	**Require human verification of all AI-generated summaries, visualizations, and translations** to ensure they accurately represent findings, reflect appropriate nuance, and avoid misinterpretation before dissemination. Verify that the AI has not introduced fabricated citations, misstated findings, or omitted essential limitations. **Use human reviewers, such as practitioners or intended end users, to assess the clarity, relevance, and usability** of AI-assisted products (e.g., condensed briefs, podcasts, or practice-focused summaries), ensuring that outputs meet the needs of real audiences. **Establish human-led review protocols for sensitive or prepublication content,** avoiding the use of non-enterprise AI tools for drafts or findings that must remain confidential.

Principle 2. Confirm Model Appropriateness

Model Specification | Accuracy | Replicability

Selecting and using appropriate AI models is critical in education research because models are increasingly being used to analyze learning outcomes, generate predictive forecasts, identify trends across qualitative data and artifacts, and inform policy decisions. Model appropriateness ensures that researchers use models consistently and reliably across different types of data sources, student and adult populations, and educational contexts. Key components of appropriateness include sound model specification, accuracy, and replicability across settings and data sets.

Model Specification. In education research, appropriate specification requires ensuring alignment of AI models closely with the educational context and with the research questions they are intended to address. Although AI tools are often used as general purpose supports across the research life cycle, no single model is well suited for all tasks. Applying a generic model trained on data from a different educational context (for example, using a model trained by urban student population data to analyze rural school populations) can produce misleading results due to contextual misalignment. Moreover, efforts to fine-tune or retrain models may introduce unintended and difficult-to-detect failures. Clearly checking and knowing a model's scope, assumptions, limitations, and intended use is therefore essential to ensuring that AI-generated outputs are valid, interpretable, and actionable.

Additionally, the use of benchmarks (e.g., robustness techniques, inter-rater reliability) and evaluation frameworks are essential for assessing the appropriateness of AI models because these tools provide standardized criteria for evaluating accuracy, robustness, and generalizability. These tools also enable researchers to systematically identify the strengths and limitations of AI models across varied tasks and data sets, and to evaluate whether models behave as intended and reliably in practice. Benchmarks also support meaningful comparisons across models, promoting transparency, rigor, and accountability when using AI for research.

Accuracy. Limitations in AI models can undermine the generalizability and validity of research findings when models systematically reflect incomplete or unrepresentative data. To ensure that results are accurate and broadly applicable, these limitations must be proactively identified and addressed throughout model use. For example, an AI system designed to recommend students for gifted programs that is trained primarily on data from affluent schools may fail to accurately identify talent in under-resourced communities. Similarly, accuracy challenges can arise in large language models used for automated essay scoring, where differences in dialect or writing style may lead to systematic misclassification of student performance. Addressing these risks requires the use of representative training data, validation across relevant populations, and ongoing monitoring and evaluation to ensure that results remain robust, appropriate, and trustworthy.

Replicability. Replication in education research is critical for demonstrating that AI-driven findings are not artifacts of data sets, assumptions, or modeling choices. For instance, a model used to identify effective teaching strategies should produce consistent results when applied across different schools, districts, and contexts. Replication efforts may include rerunning analyses with independent or publicly

available educational data sets (e.g., from the National Center for Education Statistics [NCES] or the Programme for International Student Assessment [PISA]) to test the stability of findings. Transparent documentation, shared code, and the use of open-source tools help enable reproducibility, strengthen credibility, and build trust in the use of AI within education research.

Taken together, these components of model appropriateness are further reinforced through rigorous vetting, validation, and quality assurance of both the models and the technology used to develop and deploy them. There is a clear opportunity for the research community to collaboratively develop and strengthen benchmarks and evaluation frameworks across specific areas of study and the field as a whole. In addition, there will likely be a need for ongoing adjustment to these tools as the field's use of AI expands. Exhibit 4 provides examples of this principle across the key research activities.

Exhibit 4. Examples of *Principle 2. Confirm Model Robustness* in Practice

	**Assess whether the AI model aligns with the research questions and educational context,** ensuring that the model's assumptions, training data, and intended use match the population, setting, and constructs under study. **Identify potential risks to accuracy early** by examining whether the model was trained on data that over- or under-represent certain student groups or contexts. **Review existing benchmarks or establish new evaluation frameworks** to understand how the selected model performs on tasks similar to how humans would perform. **Define a plan for replicability from the outset,** documenting how model specifications, prompts, parameters, and evaluation steps will be captured to support reproducibility across settings or with new data sets.
	**Assess whether the AI tool used for measurement was trained on data similar to the study context** (e.g., comparable student populations, instructional settings, artifacts) to ensure the outputs are valid and not distorted by contextual misalignment. **Require performance and validity checks for AI-enabled measurement tools,** such as automated scoring or transcription, to verify they function as intended. **Prespecify quality assurance and robustness criteria** for AI-generated measurement outputs, including error rate thresholds, human comparison checks, and follow-up review if the model behaves inconsistently across data sets or settings.
	**Rerun AI-assisted analyses using independent or publicly available data sets** (e.g., from NCES or PISA) to test whether findings hold across contexts and are not artifacts of a particular data set or modeling choice. **Review the AI model's assumptions, training data characteristics, and known limitations** to identify where outputs may be misaligned with the educational construct being analyzed and adjust your interpretation accordingly. **Replicate AI-generated findings with transparent documentation,** including model specifications, prompts, parameter settings, and evaluation procedures, to enable verification by peers and future researchers. **Cross-validate AI-generated insights with alternative analytical approaches** (e.g., human coding, statistical checks, or model comparisons) to ensure that results are not driven by idiosyncrasies in a single AI model.

Report how model performance was validated (e.g., accuracy, reliability, robustness checks) and whether those results varied across subgroups or contexts, ensuring that other researchers and beneficiaries can assess the credibility of AI-supported conclusions.

Clarify whether models or data sets are open source or proprietary, because this distinction affects replicability and the ability of others to verify or extend the findings.

Share supplementary materials when possible, such as model descriptions, code, evaluation rubrics, or anonymized data sets, to enable peer review, replication, and responsible reuse of AI-enabled research outputs.

Principle 3. Operationalize Transparency

Participant Protections | Disclose AI Use in Reporting | Attribute Others' Contributions

Emerging technologies pose new challenges to research reporting and dissemination compared with traditional tools. The complexity and opacity of many AI systems can obscure how decisions are made, how data are interpreted, and how conclusions are reached. As AI becomes increasingly integrated into education research, transparency will remain essential for upholding ethical standards, methodological rigor, and public trust. In education, where research findings directly shape teaching practice, student experiences, and policy decisions, researchers must commit to existing transparency norms and operationalize them in new technological contexts throughout the research process. Important dimensions of transparency that education researchers should consider when using AI in their work include ensuring participant protection, disclosure of AI use in reporting, and attribution to others' contributions.

Participant Protection. Protecting participants is a critical component of operationalizing transparency, with direct implications for consent, privacy, and future data use. Transparency requires clearly informing students, families, educators, and administrators of when and how AI is used in research processes, particularly when findings may influence instruction, assessment, or support services. Informed consent should explicitly address the role of AI in data collection, analysis, and decision-making; the types of models used (e.g., commercial or researcher developed); whether data will be analyzed in closed or external systems; and whether participant data may be reused or contribute to training AI models beyond the scope of the study at hand.

Privacy protections are essential. Researchers should also communicate that data may be used in unforeseen ways over time and that de-identified data carries a risk of re-identification as technologies evolve. By foregrounding participant protection through transparent disclosures, Institutional Review Board (IRB) review, and robust privacy safeguards, researchers can build trust and ensure that AI-enabled research respects the rights and expectations of those it seeks to serve.

Disclose AI Use in Reporting. Transparency in AI-enabled education research requires clear disclosure of what tool or model was used, how it was used, and where researcher judgment shaped the process. Researchers should disclose AI use and findings in plain language, beginning with informed consent and extending through the dissemination of results to educators, school leaders, policymakers, students, and families. At a minimum, researchers should disclose the type of AI model employed; its intended

purpose, key assumptions, and the prompts or instructions used; the known characteristics of the training data; and any modifications made to the base model. To address the opacity of many AI tools, researchers should adopt explicit strategies for "showing their work," with fuller documentation, such as detailed model descriptions, code, and sensitivity analyses, made available upon request. These practices are essential for enabling rigorous peer review and accountability in a context where AI tools are dynamic rather than static.

At the same time, researchers must clarify whether AI tools and data sets are publicly available or privately held, as this distinction has implications for replicability. Open-source tools and data facilitate replication and peer validation, while proprietary systems require careful disclosure of functionality, validation, and limitations. Given that AI systems evolve over time, the field may also need shared routines and protocols that redefine acceptable standards for replication, including agreement on parameter ranges. Responsible dissemination demands clear documentation of how AI was used, including model descriptions, input data characteristics, and validation procedures, alongside established routines for reporting in publications and responding to requests for deeper review. Establishing these norms are important for sustaining rigor, transparency, and credibility.

Attribute Others' Contributions. The third component of transparency is attribution. Researchers should clearly credit the AI models used, their developers, and the human contributors who shaped model development, training, validation, and interpretation. They should also acknowledge that models may be trained on participant-generated data. Transparency in attribution also requires naming the limits of what can be known about proprietary or closed-source systems. As with the disclosure of AI use in reporting, responsible dissemination should include attribution to the contributions of others. Together, these practices clarify who contributed what to the research process and how AI influenced design, analysis, and findings, without compromising participant privacy. When educators, students, or community members contribute their expertise or lived experience to shaping research questions, tools, or AI-enabled interventions, their intellectual contributions should be explicitly recognized in ways that are consistent with consent agreements.

As AI tools grow more complex, ensuring transparency in education research requires participant protection, clear disclosure, and appropriate attribution to uphold trust and rigor. Exhibit 5 provides examples of this principle across the key research activities.

Exhibit 5. Examples of *Principle 3. Operationalize Transparency* in Practice

Clearly communicate during study design how AI will be used, including the types of tools to be used, their intended purposes, and decision points, so that participants, partners, and reviewers understand the role of AI from the outset.

Integrate explicit transparency expectations into IRB materials and consent plans, specifying whether AI tools will support data collection, coding, or preliminary synthesis and what data (if any) may flow through external or proprietary systems.

Establish routines for "showing your work" early in the design phase, such as drafting model descriptions, prompt templates, and anticipated validation steps, ensuring that transparency is built into downstream reporting.

	**Provide clear, plain language explanations to participants,** including students, families, and educators, about how AI will be used in data collection, what types of data will be captured, and how those data may be processed or reused. **Include transparent disclosures in consent materials** specifying whether data will flow through commercial AI systems, whether models are closed or external to the research team, and any risks of future reuse or re-identification as technologies evolve. **Make privacy protections explicit** by explaining what safeguards are in place, how data will be stored, what oversight bodies (e.g., IRBs) have reviewed the approach, and how participants can ask questions or opt out when appropriate. **Clarify attribution for data-generating processes,** acknowledging when AI tools, developers, or participants contribute to data creation or labeling, and noting any constraints on transparency for proprietary or closed-source systems.
	**Provide clear descriptions of data flows and model behavior,** including whether analyses used closed or external systems, what training data characteristics might influence outputs, and how results were validated. **Disclose how AI contributed to analytical decisions,** including which models were used, what prompts or parameters shaped outputs, and where human judgment intervened in interpreting results. **Maintain transparent records of verification steps,** such as human review of AI-generated codes, cross-checks against alternative methods, or sensitivity analyses that reveal how interpretations change under different conditions. **Clarify attribution for analytical contributions,** identifying where AI models, developers, or human reviewers influenced interpretations and ensuring that credit is appropriately distributed without overstating model capabilities.
	**Clearly disclose how AI tools contributed to reporting,** including which models supported summarization, visualization, or drafting; what inputs the models received; and where human judgment shaped or corrected the outputs. **Explicitly acknowledge all contributors,** including AI tools, developers, human reviewers, and community collaborators, so that responsibility, intellectual contributions, and tool lineage are clear to audiences. **Explain any privacy-related constraints on transparency,** noting when proprietary models, closed data systems, or sensitive participant data limit what can be shared, and clarifying how these constraints were reviewed and addressed.

Moving Forward with Intention

In this brief, we propose an initial set of three principles for using AI in education research. Centering human expertise, confirming model appropriateness, and operationalizing transparency are central to the effective and ethical use of AI in education research. Although we were thorough in our development approach, we recognize that this is just the beginning of gaining consensus around a shared set of principles and engaging in AI-enabled research responsibly.

Realizing the potential of these principles requires the field to actively examine how the principles take shape across the full education research ecosystem. We encourage researchers, funders, institutional leaders, IRBs, journal editors, professional associations, and practitioners to investigate where and how these principles influence their routines, standards, and decision-making. This includes rethinking

proposal reviews, data governance practices, consent procedures, methodological norms, replication expectations, peer-review processes, and the design of tools, trainings, and policies that guide daily work. By scrutinizing the touchpoints where AI intersects with research, from conceptualization to dissemination, we can surface gaps, adapt existing infrastructures, and build new ones where needed. We offer these principles not as a static framework, but as a shared basis from which the field can engage, refine, test, challenge, and ultimately embed in ways that strengthen the rigor, ethics, and impact of education research.

Training and professional learning are essential for helping these principles take hold. As AI-driven tools become increasingly prevalent in education research, there is a pressing need for robust training and professional learning to help researchers harness their potential while adhering to safe and responsible practices. Navigating the nuances of AI in education research is not merely a technical task; it is a foundational component of responsible innovation. Building this capacity is essential to ensure that the potential of AI is realized in ways that protect participants and ultimately benefit all learners, now and into the future.

The safe and ethical use of AI in education research depends on a clear understanding of the technology's capabilities, limitations, and broader societal implications. Robust training and professional learning are therefore foundational for the following reasons:

- **Demystifying AI.** Many researchers encounter AI as a "black box," with limited visibility into how models function or how outputs are generated. Training helps break down these barriers, enabling individuals to critically engage with AI tools, understand underlying assumptions, and make informed decisions about the use of AI.
- **Understanding the potential to improve practice and outcomes.** AI holds significant promise to advance education research by enabling new methods of inquiry and expanding what can be considered "data." AI-enabled analytical approaches allow researchers to process and analyze large volumes of previously underutilized information, supporting deeper investigation and, in some cases, new research questions.
- **Recognizing and mitigating risks related to accuracy, privacy, and safety.** AI systems can produce inaccurate results when trained on low-quality or unrepresentative data. Well-designed professional learning builds researchers' capacity to identify these risks and to apply strategies to improve model accuracy and validity across contexts and populations. Given that education research often involves sensitive student information, training should also address ethical data stewardship, compliance with privacy regulations (such as the Family Educational Rights and Privacy Act [FERPA]), and best practices for secure data management to ensure the responsible and trustworthy use of AI.
- **Promoting transparency, accountability, and ethical decision-making.** Training is critical for reinforcing the importance of transparency in AI-enabled research processes, including documenting methods and clearly communicating AI-driven findings to multiple audiences. Professional learning also supports ethical reflection on issues such as consent and the appropriate role of automation in education research and practice.

Implementing effective training and professional learning is not without challenges. Common barriers include resource constraints, varying levels of baseline knowledge, and resistance to change. The rapid pace of AI innovation further complicates efforts to keep training current with emerging technical and

safety considerations. Addressing these challenges will require sustained commitment and investment in professional learning across both the public and private sectors.

As a field, we share a collective responsibility to steward the use of AI in ways that strengthen, rather than compromise, the integrity of education research. This responsibility extends beyond individual researchers to the norms we establish, the expectations we set, and the infrastructure we build together to support ethical, high-quality work. AI is a powerful tool that offers opportunities to transform how we understand teaching and learning, particularly by helping us make sense of the vast and often underutilized information generated through students' daily educational experiences. When used responsibly, AI-enabled methods can help synthesize data from classroom activities, student work, and interactions into usable evidence and enabling insights into learning and development that are deeper, timelier, and more cost-efficient. Realizing this potential, however, depends on collective action to align technical innovation with shared principles, safeguards, and professional judgment. By working together to advance AI use that is grounded in integrity, transparency, and care for those we serve, we can further improve the relevance and validity of education research and build more nuanced understanding that ultimately can better support all students.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit www.air.org.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

29616_01/26