Renaissance Philanthropy

THE LEARNING AGENCY

Safe and Responsible AI in Ed Tech
Guidance for Funders

November 2026



AI is creating new opportunities to improve teaching and learning, from more individualized support for students to richer student data and time-saving tools for educators. At the same time, the use of AI in education raises real concerns. AI systems can produce inaccurate or inappropriate content, reinforce bias, encourage students to offload important thinking, and expose sensitive student information.

Funders can help shape how AI-powered ed tech tools are developed and used. By raising safeguarding questions early, and supporting grantees to address them, funders can help evidence-backed AI tools reach and support more learners while minimizing their potential for harm.

The *Guidelines for the Safe and Responsible Use of AI in Ed Tech*, developed by Renaissance Philanthropy and The Learning Agency for ed tech innovators, offer detailed recommendations across six principles:

1. Focus on educational outcomes.
2. Direct generative AI systems to support students' learning and wellbeing.
3. Ensure the privacy and security of user data.
4. Prioritize accessibility and fairness.
5. Promote transparency and explainability.
6. Give users control over their data.

This resource offers a practical way for funders to bring those principles into their grantmaking.



Encouraging the Safe and Responsible Use of AI in Ed Tech through Grantmaking

Not every AI project presents the same risks, and not every grantee will need the same approach. A student-facing AI tutor, for example, raises different questions than a tool that helps teachers with administrative work. A funder's expectations for grantees should reflect the product, its users, its stage of development, and the potential consequences if something goes wrong.

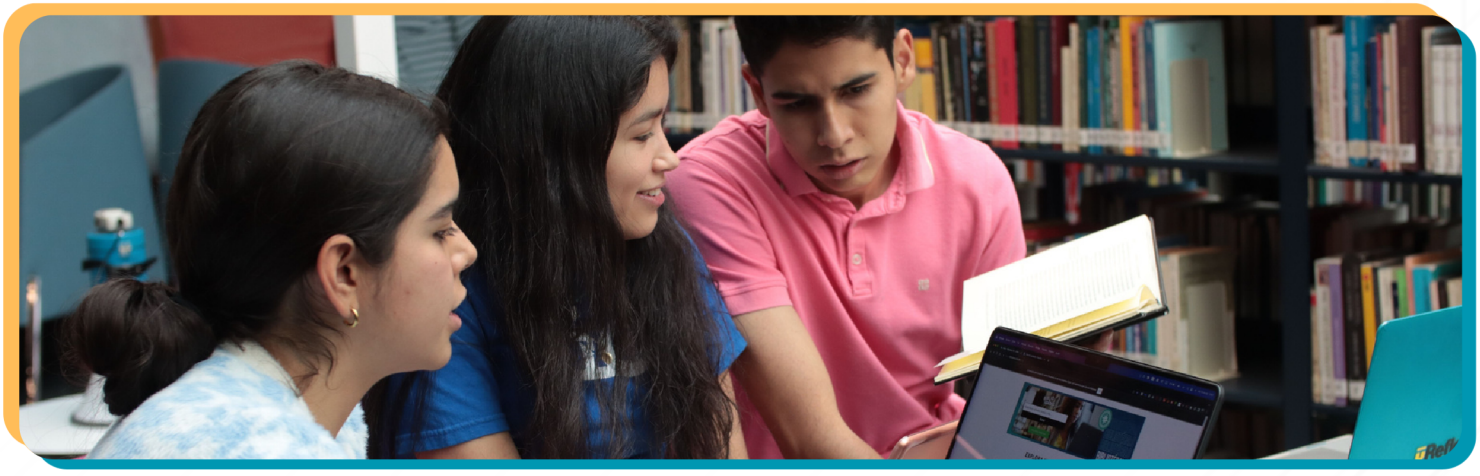
Depending on the project, a funder might ask safeguarding questions in Requests for Applications or during diligence; request additional review of the product within a particular area of risk; or include specific safeguarding activities or milestones in a grant.

For projects where AI plays a significant role, funders may also ask grantees to develop a **Safe and Responsible AI Strategy**. The strategy provides a way for grantees to identify the safeguarding elements most relevant to their work, explain how they are addressing them, and flag areas where additional work or support is needed.

Components of a Safe and Responsible AI Strategy

A Safe and Responsible AI Strategy does not need to address every recommendation in the *Guidelines for the Safe and Responsible Use of AI in Ed Tech*. It should focus on the components most relevant to the proposed project and be proportionate to the product's stage and risk.

Below are questions funders can ask grantees to learn about their Safe and Responsible AI Strategy.



Educational Impact

KEY QUESTION

How will AI support the educational outcomes you are trying to achieve, and how will you know whether it is working?

Grantees should be able to explain the role AI plays in the product, the evidence and user input informing the approach, and how they will assess whether it is having the intended effect. For student-facing tools, grantees should describe how they direct AI to support productive thinking and learning, and specific ways humans will be in the loop as monitors, editors, and validators of AI outputs throughout the workflow.

Read more: *Principle 1: Focus on educational outcomes.*

AI Safety

KEY QUESTION

In what ways could the AI negatively affect learners, and how will you test for and address those risks?

For generative AI, this may include inaccurate, off-topic, biased, or inappropriate responses as well as unexpected or concerning interactions with students. Relevant approaches might include ongoing model testing, red teaming, content moderation, pedagogical guardrails, and clear processes for users to flag problems.



The strategy should focus on the risks most relevant to the product and how the team will test whether its safeguards are working.

Read more: *Principle 2: Direct generative AI systems to support students' learning and wellbeing.*

Data Privacy and Security

KEY QUESTION

What user data will be collected and how will it be protected?

The strategy should describe a detailed data management plan aligned to the implementation context in which the grantee is operating. The plan should articulate the user data the product collects, why it is needed (e.g., for a tool's functionality, personalization, continuous improvement), and how it will be stored and protected (e.g., multi-factor authentication, data encryption), used, and shared, as well as how long it will be retained before deletion. Particular attention should be paid to data shared with authorized third parties or research collaborators, and any data sharing agreements that ensure personally identifiable information (PII) remains anonymized and will not be reidentified. The strategy should specify what the team does to anonymize PII, with semantic anonymization (i.e., replacing personal details with contextually appropriate surrogates) as the ideal. The strategy may include the frequency and results of data management audits.

Grantees should also have a clear plan for responding if a privacy or security problem occurs.

When possible, collected data should be anonymized and structured in such a way that allows it to become a public good.

Read more: *Principle 3: Ensure the privacy and security of user data.*

Accessibility and Fairness

KEY QUESTION

How will you ensure the product meets the needs of the diverse learners it is intended to serve?

Meeting the needs of diverse learners means ensuring the fairness and accessibility of an AI-enabled tool. The tool should be fair, performing effectively across all types of users. This involves designing for the diversity of the product's intended users, considering factors such as age, gender, race/ethnicity, language, learning abilities, and socioeconomic background. The product must also be designed for accessibility, enabling all of its intended users to use it with relative ease, whether that includes students without high-speed internet, students with disabilities, or other groups that may face limitations.

The strategy should identify how the team will routinely test for biases, disparities, and accessibility barriers and what it will do when issues surface. It may describe the representative data on which the AI is trained if a foundation model is not being used, or any accessibility audits that have been conducted on the product.

Read more:

- *Principle 4: Prioritize accessibility and fairness.*
- *Accessibility Standards for Inclusive Ed Tech.*

Transparency

KEY QUESTION

What will users know about the use and limitations of AI?

Students, educators, families, and schools should have clear, easy-to-read information about when and how AI is being used, what it is intended to do, its known limitations, and relevant evidence about how well it works.

Read more: *Principle 5: Promote transparency and explainability.*

User Choice

KEY QUESTION

What choices will users have about their data?

Users should understand and have meaningful choices about what personal data is being collected and how their data is being used, and have the opportunity to consent to or opt out of its collection and sharing. Different approaches should be considered for K-12 and adult learners. Additionally, users should be offered an easy way to request the removal and deletion of their data.

Read more: *Principle 6: Give users control over their data.*



Putting the Strategy Into Practice

A Safe and Responsible AI Strategy can be requested as part of a proposal, developed during diligence, or included as an early grant deliverable. For some projects, a few paragraphs may be enough. Higher-risk or early-stage products may warrant a more developed strategy.

Importantly, the strategy does not need to describe a complete, comprehensive plan for the product's safe and responsible use of AI. It can identify gaps and lay out what the team plans to address during the term of the grant. A grantee might commit to conducting red teaming before a student-facing pilot, completing an accessibility or security review, establishing an incident-response process, or generating more robust evidence about the accuracy of its AI-powered features.

Those commitments can become grant activities or milestones. Funders should also consider whether the grant includes enough funding for the safeguarding work. Safety testing, evaluation, accessibility audits, privacy and security reviews, and outside expertise all have costs.

The strategy can then be revisited during the grant as the product develops. What has the team learned? Have new risks emerged? Has testing or user feedback surfaced problems? What needs to change?

Funders should invite and make it easy for grantees to raise these issues. Identifying a problem through testing is not the same as failing to address it. What matters is whether a team is looking for problems, responding when they arise, and continuously improving ts approach.



SAMPLE PROPOSAL LANGUAGE

Safe and Responsible AI Strategy

For projects that involve the development or use of AI, please describe your approach to safe and responsible AI as part of the proposed work. Focus on the components most relevant to your project:

- **Educational impact:** How will AI support the educational outcomes you are trying to achieve, and how will you know whether it is working?
- **AI safety:** In what ways could the AI negatively affect learners, and how will you test for and address those risks?
- **Data privacy and security:** What user data will be collected and how will it be protected?
- **Accessibility and fairness:** How will you ensure the product meets the needs of the diverse learners you intend to serve?
- **Transparency:** What will users know about the use and limitations of AI?
- **User choice:** What choices will users have about their data?

Your strategy may include practices already in place as well as work you plan to complete during the grant period. Where relevant, identify significant gaps and any additional resources or expertise needed to address them.

Please refer to the *Guidelines for the Safe and Responsible Use of AI in Ed Tech* for more detailed guidance and recommended practices.

Supporting Safe and Responsible AI

The *Guidelines for the Safe and Responsible Use of AI in Ed Tech* provide the detailed practices developers can use to strengthen their approach. Funders do not need to reproduce that guidance or become technical experts themselves.

Instead, funders can make safeguarding a critical part of how AI projects are funded: encourage grantees to raise the right issues early, set expectations that make sense for the project, and provide grantees with the resources to address them.

By asking grantees to articulate and refine their Safe and Responsible AI Strategy, funders can increase awareness of and commitment to safeguarding best practices among their grantees. With a few short questions, the safe and responsible use of AI in education can spread field-wide.