Building Better AI for Learning:
Trends from Six Cycles of the Tools Competition

Tools Competition

A program of Renaissance Philanthropy

Organized by The Learning Agency

June 2026

Authors: Meg Benner, Kristyn Manoukian, Joon Suh Choi, and Kumar Garg

Introduction

AI in education is evolving at a breakneck pace. Every day, it seems, a new tool or model is making headlines. This all leaves innovators, funders, and researchers struggling to separate signal from noise: to know what is hype, what is transformative, and where guardrails are urgently needed. Schools and parents also remain unsure, and federal and state policy is moving slowly. Innovators are caught between the pressure to experiment and the obligation to build responsibly.

This is a critical moment to earn trust among current and future users. Half of Gen Z use AI daily or weekly, and usage continues to rise. But even as usage increases, anxiety is growing as the percentage of learners that feel supported in their learning by AI-assisted resources is declining.

Rather than ride the wave of changes in AI, innovators and researchers need to harness advanced technological methods to mitigate persistent bottlenecks in education to support learners and educators, while also prioritizing earning trust among users.

The Tools Competition offers a unique window into how AI in education is moving from hype to implementation. Because submissions span the full development spectrum, from early ideas to established products, the competition charts where the field is headed. It shows how developers are using AI, what educational problems they are trying to solve, and what technological methods are gaining traction. Most importantly, it sets a rigorous standard for trustworthy, transparent, and responsible technology and innovation while solving specific educational problems.

> **Background**
> The Tools Competition is a multi-million dollar funding program of Renaissance Philanthropy run by The Learning Agency that supports ed tech innovation grounded in learning science. Since launching in 2020, it has awarded 171 prizes across six annual cycles spanning the most rapid transformation in AI capability the field has ever experienced.

This report identifies patterns and standards across competition proposals from the last six years with the goal of helping the field better understand AI's evolution and integration into educational learning platforms. The central finding is clear: AI adoption has accelerated quickly, but the strongest proposals are not defined by their use of AI. Rather, they are defined by how carefully developers integrate AI into authentic learning contexts, evaluate the technology's performance, and design for the learners and educators most likely to be underserved.

Because this analysis draws heavily on winning proposals, it reflects what rose to the top of the competition's expert review process. Many winners are still early-stage, and not all have outcome evidence yet. Still, the winning pool offers useful insights into what expert reviewers, funders, and field leaders have recognized as credible, promising, and technically substantive AI-enabled education work.

What distinguished Tools winners from other competitors is how deeply they customized the technology to their contexts, users, and learning goals. That is the next frontier for the field: moving beyond surface-level AI adoption toward tools that are reliable, inclusive, trusted, and useful in real education settings.

Key Takeaways

What Leading Competitors Are Building

- **AI adoption is widespread, but deep technical implementation is concentrated among a smaller group of highly technical teams.** The share of winning proposals has shifted substantially across the six cycles with a much larger portion of the winners built around AI as the central technical innovation. Across the broader field, AI terminology now spreads quickly, but leading competitors are translating new consumer AI capabilities into credible education tools within about 6 to 12 months of release.
- **Top proposals combine advanced AI techniques with established methods.** Among 2026 winners, this includes multimodal and voice-first interfaces, grounding and reliability infrastructure, agentic architectures, and offline-first deployment. It also includes continued use of established methods for modeling student learning, evaluating performance, and validating results.

What Sets Strong Proposals Apart

- **Foundation models have become infrastructure, not the whole product.** The strongest proposals often treat models as interchangeable and differentiate themselves through the layers they build and rigorous evaluation of models and outputs.
- **Custom modeling matters for speech, vision, and real-time sensing.** Off-the-shelf services fail systematically for child voices, atypical speech, accented speakers, and low-resource languages. The most ambitious technical work among winners is concentrated where generic AI tools do not work well enough.
- **Building on prior infrastructure is a common pattern among winners.** More than half of winning proposals across cycles deliberately extend the platform, framework, dataset or research base rather than building from scratch. Building on existing infrastructure is a critical way innovators strengthen their work.

Designing for Trust, Responsibility, and Equity

- **Responsible AI has matured, but real gaps remain.** Privacy frameworks are now common in proposals. Concrete plans for post-deployment monitoring, fairness testing, and data retention are less common, including within the winner pool.
- **Equity framing is near-universal; equity-driven engineering is not.** Many proposals name a target population, but fewer show how the tool has been designed around that population's actual context and constraints. Stronger proposals translate equity claims into concrete design choices around modality, language, bandwidth, accessibility, and workflow.

The sections that follow focus on cross-cutting patterns in AI use, technical implementation, and user-centered design, rather than trends tied to specific subject areas.

The Evolution of AI Adoption

Among winning proposals, the technologies that appear each year broadly track the wider AI landscape, with a 6 to 12 month lag between the release of a new AI capability and its appearance in a credible, fundable education proposal. Advanced AI vocabulary now spreads quickly across the full competitor pool. What changes more slowly is the depth of technical implementation: whether competitors are truly engineering around these capabilities, rather than simply naming them.

- **2021 - 2022 (Cycles 1.0 - 2.0):** While COVID moved education online, foundation models were not yet easy to use without machine learning expertise. AI in winning proposals was rare and usually bespoke, including classical machine learning and custom speech-recognition pipelines. Although GPT-3 was available through an API during these cycles, no winners built around it.
- **2023 (Cycle 3.0):** ChatGPT made foundation models widely accessible. In the Tools Competition, this changed what developers incorporated almost immediately. Foundation models became the assumed base layer in roughly three-quarters of winning proposals.
- **2024 - 2026 (Cycles 5.0 - 6.0):** As multimodal and agentic tools became more accessible, competitors began incorporating them as deliberate design choices. These terms now appear widely across the broader competitor pool, but deeper architectural work remains concentrated among winners.

Foundation models are large, general-purpose AI models that serve as the base layer for many applications. Examples include models such as GPT, Claude, and Gemini. In education, these models are rarely the full solution on their own; tools engineer around them, building additional layers to ground outputs in curriculum, adding safeguards, integrating models into teacher and learner workflows, and adapting the experience for specific users and contexts.

From AI Adoption to AI Implementation

The rapid spread of AI across Tools Competition proposals is only the starting point. The more important question is how competitors are turning AI capabilities into tools that are reliable, useful, and designed for real educational settings. Across the competition, AI terminology now spreads quickly, but deeper engineering remains concentrated among the strongest proposals.

What's Next: Advanced AI for Ed Tech

In this webinar, Ralph Abboud, Program Scientist at Renaissance Philanthropy, explores how AI and LLMs are entering a new phase with models that can think, act, and perceive, and what that means for education.

Replay the Webinar & Read the Summary →

The next three sections examine that shift in practice:

1. What Leading Competitors Are Building
2. What Sets Strong Proposals Apart
3. Designing for Trust, Responsibility, and Equity

The next phase of AI in education will depend less on whether tools use AI and more on how well the developers integrate it into tools that work in real learning environments.

1. What Leading Competitors Are Building

Leading competitors show where the field is heading. They are using more sophisticated AI than in earlier cycles, but the important shift is not just technical. They are using AI in ways that are tightly connected to a learning problem, a user need, and a realistic implementation context.

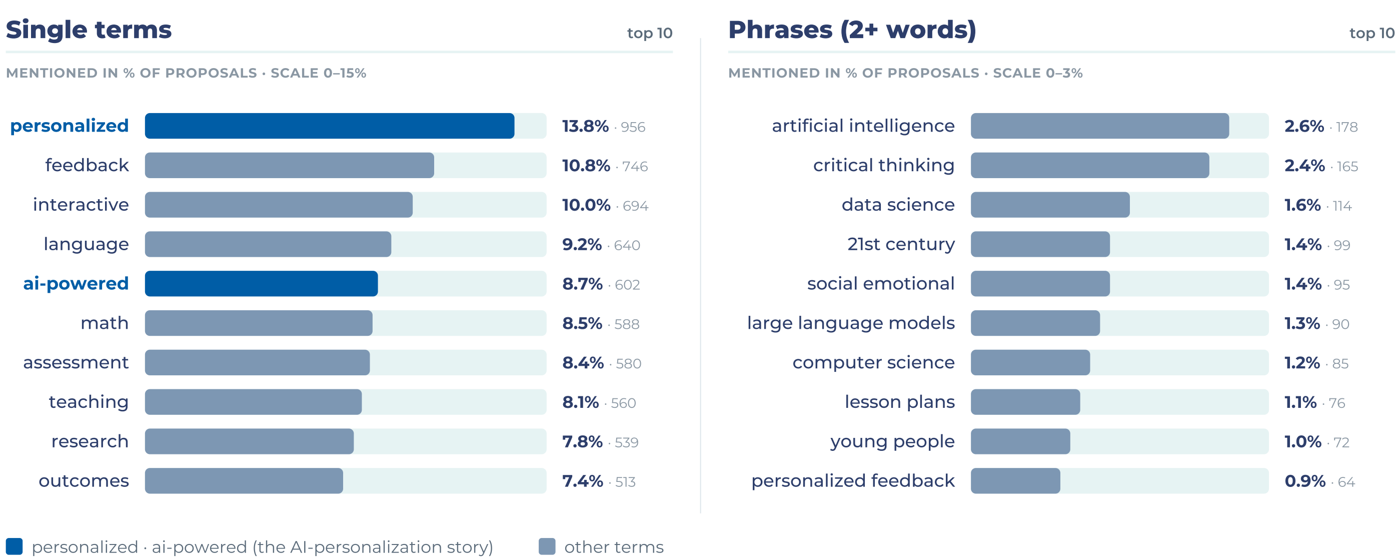
The five technical patterns below describe what leading competitors are building and where they are placing their bets.

Leading competitors are investing most heavily in systems that ground AI in trusted content and evidence, and in tools that move beyond text to support speech, images, voice, and other forms of interaction. Smaller but distinctive patterns are also emerging around tools that can reason through multi-step tasks and tools designed to work in low-connectivity settings. At the same time, established methods for modeling learning and evaluating performance remain important because they solve problems that foundation models do not automatically or reliably solve.

Language of the Field

Most-Mentioned Terms Across Six Cycles

Share of proposal summaries mentioning each term, with generic terms (e.g., data, technology, online) removed to better surface competitor focus areas.



Based on short proposal summaries, not full proposals, from about 6,900 proposals across six competition cycles. Single terms and phrases are tallied separately and shown on different scales, so the panels are not meant for direct comparison.

Multimodal and Voice-First Interfaces

Multimodal and voice-first interfaces are becoming more common among leading competitors. This reflects the maturation of consumer multimodal AI (GPT-4o, Claude 3.5, ElevenLabs voice) in 2024, including tools that can work with speech, images, and video. It also reflects a real education need: many learners, especially young children, early readers, multilingual learners, neurodivergent learners, and students with disabilities, may not be best served by a text-only interface.

About half of 2026 winners use approaches such as voice interaction, animated avatars, custom speech recognition for children, or real-time computer vision. While many recent proposals mention multimodal features, fewer provide detailed plans for how they will develop, evaluate, and adapt those features for the learners they serve.

> *2026 Winner Spotlight*
>
> **KIVA** (MIT/Boston University)
> $50K Catalyst Prize in the Accelerating K-12 Learning track
> USA
>
> KIVA shows how multimodal AI can better align with how young children learn. At its center is an animated AI avatar that talks with children about the texts they read or listen to, using child-tuned speech recognition in English and Spanish and a language model trained on annotated child-tutor interactions. Instead of relying on typing or screen-based text, KIVA is built around spoken dialogue, visual interaction, and adaptive instructional support.

Grounding and Guardrails

General-purpose language models can produce polished responses that are still inaccurate, too generic, or pedagogically weak. Strong proposals therefore anchor outputs in trusted content, structured knowledge, and evaluation frameworks. In this way, grounding and reliability infrastructure are becoming a major focus across the strongest proposals.
Three approaches show up often in this work:

- **Retrieval-Augmented Generation (RAG) →** the AI retrieves from trusted content before generating.
- **Knowledge graphs →** the AI uses a structured map of concepts, prerequisites, and misconceptions.
- **Hybrid systems →** competitors design their system to combine multiple methods instead of relying on one model or technique.

Roughly one-third of 2026 winners are not relying on AI-generated outputs alone. They are building systems that first ground responses in validated content, curriculum prerequisites, and known student misconceptions, then use automated or human review to check quality before outputs reach users.

An emphasis on grounding and guardrails now appears widely across proposals. What distinguishes the strongest proposals is whether that language is backed by concrete infrastructure, such as curated content libraries, structured concept maps, evaluation pipelines, or expert review processes.

2026 Winner Spotlight

Quality Talk Amplified Online (Arizona State University, Penn State University, and WisdomEDU)
$50K Catalyst Prize in the Building Pathways to Postsecondary Success track
USA

QTAO is an AI-powered approach to facilitating small-group discussion in online postsecondary courses, grounded in over a decade of Quality Talk research. Drawing on discourse coding manuals, workshop videos, and annotated examples of effective facilitation, QTAO informs Cue-T's intervention logic and the design of the QT Briefs. Cue-T intervenes only under predefined discourse conditions, and the QT Briefs synthesize discussion data using established discourse indicators and pedagogical principles anchored in the four-component Quality Talk framework. In this way, AI operationalizes evidence-based facilitation and instructional support practices at scale, rather than replacing them.

Agentic AI Systems

Agentic AI is moving the field beyond one-off chatbot interactions toward systems that can reason through problems, plan, and take action independently. In education, these approaches may reduce teacher workload and support more adaptive learning experiences, but they also require stronger safeguards because the system has more autonomy.

2026 Winner Spotlight

PAL for Early Math Learning (Learnology)
$150K Growth Prize in the Accelerating K-12 Learning track
USA

PAL exemplifies agentic AI use through the tool's research-informed math interactions between caregivers and children and its personalized activity recommendations that accelerate early math learning. The agentic AI system supports adults guiding children's mathematics learning by taking in what an adult notices, asking targeted follow-up questions, updating a probabilistic model of the child's understanding over time, and generating tailored guidance grounded in a curated knowledge graph of early math progressions and misconceptions.

Roughly 10 percent of 2026 winners use more advanced agentic architectures through frameworks such as LangGraph. This is still a smaller pattern than multimodal AI or grounding, but it is a notable increase from past cycles, when agentic architectures were rare.

Offline-First and Low Connectivity Design

A small but growing share of proposals are for AI-enabled tools for settings where continuous cloud connectivity is not realistic. Some are exploring AI that can run locally on classroom hardware, basic smartphones, or kiosk devices. Others are building offline-first systems that allow learners to keep using the tool and sync data when internet access returns. Most consumer AI defaults to the cloud, but many education tools cannot, especially where connectivity is limited, unreliable, or expensive.

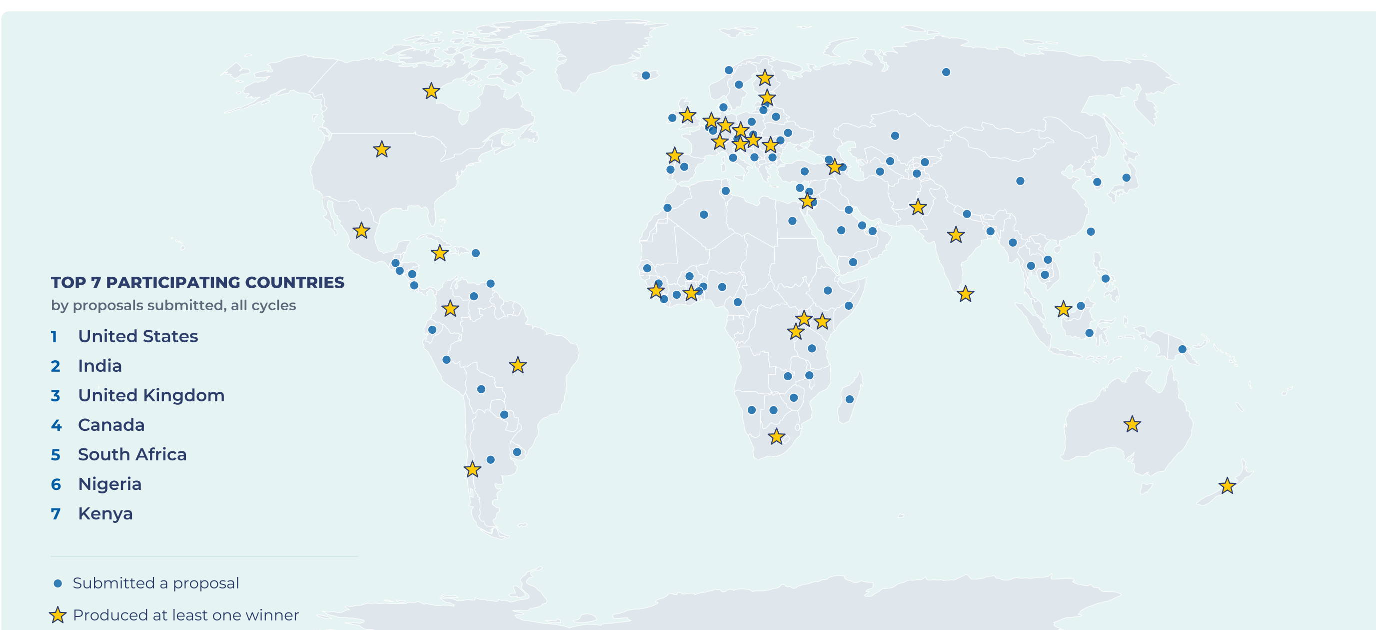
This pattern is driven by two needs. First, many tools are designed for rural schools, the Global South, refugee settings, or other low-bandwidth environments where continuous cloud access may not be viable.Second, some competitors see local or offline-first design as a way to improve privacy, reliability, and resilience when third-party AI services change pricing, policies, or availability.

The share of proposals with this focus has increased over the past three cycles. In 2026, roughly 15 percent of winners were developing offline or low-connectivity technology.

Geography

A Globally Distributed Competition

Blue dots mark the 129 countries that have submitted a proposal across all past cycles (2021–2026). Gold stars mark the 32 countries that produced at least one winning team.



2026 Winner Spotlight

SmartCoach Assess (Inspiring Teachers)
$150K Growth Prize in the Accelerating K-12 Learning Track
Ghana, Uganda

SmartCoach Assess demonstrates how offline-first design can bring structured reading assessment to low-connectivity classrooms. The app enables grade 1–3 teachers to administer EGRA-aligned assessments, including Oral Reading Fluency, store individual learner profiles on-device, and sync results to dashboards when connectivity is available. Piloted across 180 schools in Ghana and Uganda with strong RCT results, the team is now scaling SmartCoach Assess into an adaptable platform to enable ministries of education to track every child's reading progress consistently and at scale.

Established Methods Alongside Newer AI

The arrival of foundation models has not displaced the methods ed tech developed over the prior 30 years. Methods such as Bayesian Knowledge Tracing, Item Response Theory, Cognitive Diagnostic Models, Multi-Armed Bandits, and gradient-boosted trees remain present across winners. Sometimes they are standalone engines. Increasingly, they are part of hybrid systems that combine established ways of measuring learning with LLM-based generation.

These methods continue to matter because they solve problems that foundation models do not automatically or reliably solve. They can provide interpretability, statistical rigor, computational efficiency, and clearer ways to model student progress.

Leading competitors are not choosing between older and newer methods. They are combining them when each method serves a clear purpose. Examples include:

2026 Winners Spotlight

Adaptive Assessment for Africa (Siyavula Foundation, South Africa | 2026 Accelerating K-12 Learning track winner), compares an expert-defined concept-cluster model against an LLM-based semantic matching approach within a randomized controlled trial, treating newer AI as something to test against established assessment logic rather than assume as superior.

Navvy (Student Basic Needs Coalition, USA | 2026 Postsecondary Learning track winner), combines a chatbot for real-time student guidance with a responsive screener and benefits roadmap that applies structured eligibility logic. State-specific eligibility rules, enrollment status, and application fields are mapped through a rules-based system that pre-fills answers and enforces validation checks, while the LLM-powered chatbot supplies conversational support grounded in program requirements and user-specific information.

2026 Winners Spotlight contd.

AIED-Unplugged (Centro de Estudos e Sistemas Avançados do Recife, Brazil | 2026 Dataset track winner), pairs established OCR and machine learning models with human expertise rather than relying on either alone. Baseline models generate initial annotations for layout analysis, transcription, and response detection, which experts then verify and correct. Quality is maintained through a double-blind consensus process in which two annotators work independently and disagreements are resolved by a senior pedagogical specialist, with inter-rater reliability monitored using Cohen's Kappa.

2. What Sets Strong Proposals Apart

Across the technical approaches above, the strongest proposals make several similar strategic and architectural choices. These patterns are less about which AI method a competitor names and more about how the competitor builds: how it evaluates quality, what it builds from scratch, what infrastructure it uses, and how it adapts technology to a specific learning context.

They Evaluate and Adapt Foundation Models Over Time

The cost and effort of using high-quality AI models has dropped dramatically over the past three competition cycles. As a result, the differentiating factor across proposals is no longer which model a competitor picks at the outset. It is how rigorously they evaluate, adapt, and integrate models into the tool over time.

Only about one winner in 10 explicitly identifies a specific foundation model in their proposal. More often, leading competitors describe a model-agnostic approach: they compare multiple model options and select the one that performs best for the task and learning context. The signal is clear: model choice matters less than disciplined model evaluation.

2026 Winner Spotlight

SkillFlix for Autistic Young Adults (dfusion, Inc.)
$150K Growth Prize in the PostSecondary Learning Track; $100K in the Datasets Track
USA

SkillFlix demonstrates rigorous evaluation and deliberate adaptation in its tool supporting autistic young adults as they build communication skills for navigating challenging situations. The tool uses retrieval-augmented generation grounded exclusively in proprietary content, including skill-building videos, role-play scenarios, and community-generated insights, rather than broader web data that may encode neurotypical norms. To mitigate bias, the system is repeatedly evaluated against diverse autistic communication styles, with problematic response patterns triggering retraining and refinement.

As strong foundation models become more widely available, and the model itself becomes interchangeable, differentiation moves to the layers around it: carefully engineered prompts, retrieval pipelines grounded in curated content, knowledge graphs that guide reasoning, evaluation systems that check outputs against pedagogical criteria, teacher or expert review, and proprietary or hard-to-collect data.

They Customize Where Generic AI Falls Short

This pattern has limits, especially beyond text. Sophisticated text-based models are now widely available and reasonably priced, but access to high-quality models for other modalities is much more uneven. For speech, vision, and real-time sensing, off-the-shelf options are far more likely to fall short, and the choice of model, or the decision to customize, can still meaningfully differentiate one proposal from another.

Speech is the clearest case. Off-the-shelf recognition (Whisper, Azure Speech, and similar services) systematically underperforms on child voices, atypical speech, accented speakers, and low-resource languages—the very learners many winning tools are designed to serve. In these contexts, off-the-shelf AI is not a viable foundation. Custom acoustic and language modeling remains essential.

> *2026 Winner Spotlight*
>
> **Amira Inclusive Reading Coach** (Amira Learning + Children's Hospital of Philadelphia + Penn Linguistic Data Consortium
> $300K Transform Prize in the Accelerating K-12 Learning Track
> USA
>
> Generic AI isn't enough—and Amira is leading the way on how to make it work for all students. The Amira Inclusive Reading Coach is built to accurately understand young and diverse learners as they read aloud, including early readers, autistic learners, students with ADHD, and those who are multilingual. Rather than rely on off-the-shelf speech recognition, the Amira team is training custom acoustic and language models on a unique, clinically diagnosed dataset, and releasing the corpus publicly through the University of Pennsylvania LDC. Amira is investing core engineering effort in the speech layer — the gateway to equitable reading support for all learners.

They Build on What Already Works

Building on what already works is a common pattern among the strongest proposals. Across cycles, more than half of winners explicitly extend an existing platform, product, research base, or partner infrastructure. That leverage takes many forms: open-source tools and standards, established research bodies, partner platforms, scaled products with new capabilities, or previously validated models and datasets.

This pattern suggests that winners understand how to build from proven infrastructure rather than reinvent capabilities that already exist. Existing assets allow competitors to move faster with higher quality tested building blocks, reduce technical risk, and focus their innovation on adaptation: how they tailor a known approach to a new learner population, context, modality, or implementation challenge.

What has changed over time is the type of infrastructure competitors can build from. In early cycles, competitors often relied on open content libraries, research-backed frameworks, and classical machine learning methods. In later cycles, they increasingly build on foundation models, RAG pipelines, and open agentic frameworks. The advantage is not simply having access to these assets, but knowing how to adapt them into tools that work for specific educational problems: an understanding that comes from strong evaluation procedures.

> *2026 Winner Spotlight*
>
> **EveryLesson: Co-Teaching AI** (Blue Engine)
> $50K Catalyst Prize in the Accelerating K-12 Learning Track
> USA
>
> EveryLesson shows how building on existing infrastructure can help a small team move faster. The AI-powered planning and reflection tool for co-teaching teams and instructional coaches is built on PlayLab's education-focused AI platform rather than from scratch. That choice lets the team focus its own engineering on what is genuinely new: a structured input flow for co-teaching pairs and outputs designed for inclusive classrooms where general and special educators plan together with limited shared time.

3. Designing for Trust, Responsibility, and Equity
Across cycles, winning proposals have begun more seriously engineering for user trust and designing deliberately for learners who are often underserved by mainstream ed tech. These patterns reflect broader public concern about AI, equity, and data privacy. They also follow naturally from what more advanced technology makes possible and risky.

The two patterns below show how the strongest proposals are moving from broad commitments to concrete design choices: responsible AI is becoming more concrete, and equity claims are increasingly being translated into product design choices. They also reflect the role of the Tools Competition itself. Organizers and partners increasingly reinforce these priorities through proposal questions, evaluation criteria, and the definition of competitive priorities.

<u>Responsible AI Is Becoming More Concrete, But Gaps Remain</u>
User safeguarding is maturing alongside the AI technology itself. Leading competitors are increasingly treating responsible AI as an engineering requirement. As AI models have grown more sophisticated, the bar has risen for privacy, bias mitigation, hallucination guardrails, data governance, and human oversight.

The clearest signal is in privacy compliance. References to FERPA, COPPA, GDPR, HIPAA, and SOC2 were rare in early competition proposals. By the 2025 and 2026 cycles, it is unusual for a winning proposal not to name relevant frameworks and describe how the system enforces them. Safeguarding is also becoming a standard point of discussion among competition judges. The strongest proposals go further by building in hallucination guardrails, bias auditing, and explicit human-in-the-loop review.

In 2026, the Tools Competition saw additional layers emerge: language around constitutional AI, federated learning, and explainability have surfaced in more proposals. While this is partially shaped by the competition's increasing demand for clarity in these areas, engineering teams are beginning to anticipate and intentionally build around critical trust questions.

There are still clear opportunities to raise the bar. Competitors more often describe safeguards during development than plans for monitoring systems after deployment. Specific methods for testing whether tools perform equitably across student populations remain less common than general fairness statements. And while privacy protections are frequently addressed, proposals still need to better explain how student data will be managed, retained, or deleted as tools scale.

> *2026 Winner Spotlight*
>
> **ARISE Cyber Labs** (George Mason University)
> $50K Catalyst Prize in the Postsecondary Learning Track
> USA
>
> ARISE is a web browser-based augmented reality platform that provides IT and cybersecurity education on virtually any device. The project demonstrates that the principles behind responsible AI, such as privacy, accessibility, and learner protection, apply to educational technology of all kinds. The team collects interaction data only with informed consent, de-identifies student information, and stores data securely in compliance with institutional requirements. Features such as colorblind-friendly visual encoding, support for non-native English speakers, and iterative testing with students with disabilities help ensure the platform remains accessible and effective for all learners.

Equity framing is moving toward equity-driven engineering

Equity framing and equity-driven engineering are not the same thing. Naming a target population has become common across proposals. What is harder, and more meaningful, is designing the tool around that population through choices about modality, language, bandwidth, accessibility, data, and workflow.

Among the strongest proposals, equity objectives are matched by product and technical choices. Two shifts stand out across recent cycles. First, the 2026 cycle saw a greater concentration of proposals focused on neurodivergent learners, including students with autism, ADHD, dyslexia, and Augmentative and Alternative Communication users, with about one in four winners explicitly focusing on these populations. Second, Global South contexts are shaping engineering decisions in more visible ways. Proposals targeting Africa, Latin America, and South Asia often use WhatsApp or SMS delivery, offline-first systems, low-bandwidth design, and basic-phone compatibility.

These choices are not only relevant to the learners they first target. Accessibility-first and low-resource design often produce broader usability benefits. Tools built first for learners with disabilities or resource-constrained settings can create simpler, more flexible, and more resilient designs for many users.

2026 Winner Spotlight

Lab-on-a-Book (Teachers College, Columbia University)
$50K Catalyst Prize in the Accelerating K-12 Learning Track
USA

Lab-on-a-Book shows how equity-driven design can shape the whole learning experience in STEM without merely defaulting to a digital version of old pedagogical models. The pocket-sized science platform is designed for students in under-resourced communities that may lack lab infrastructure, with a printed format that works without specialized equipment, electricity, or reliable internet. Its multilingual AI companion works on low-bandwidth devices, while the physical-first design reduces dependence on screens. The result is a tool designed around economic access, geographic reach, and pedagogical flexibility, while still keeping learning inquiry-based and hands-on.

Closing

Six cycles of the Tools Competition show how the field has absorbed and responded to rapid technological change. Each wave of consumer AI has shifted what innovators incorporate into their proposals, offering a window into broader ed tech trends. By analyzing application patterns alongside winning proposals, this report also shows how the strongest proposals are using new capabilities, where competitors are investing engineering effort, and what kinds of approaches are rising to the top in a competition evaluated by technical experts and philanthropy leaders.

As AI has become easier to access, the hard work has shifted away from simply choosing a foundation model and toward the systems built around it. Leading competitors are grounding outputs in validated content, building on proven infrastructure, evaluating performance, designing for learners who have traditionally been left out of mainstream approaches, and treating responsible design as a requirement from the start. The language of these practices has spread quickly across the field.

What distinguishes top proposals is the work behind the words: whether competitors have tested and evaluated their models, built structured content or knowledge systems to guide the AI, planned for safeguards after deployment, and turned equity commitments into concrete design choices.

The signal for the broader field is clear. Meaningful AI adoption will depend less on naming and reacting quickly to the newest capability and more on the hard work required to make AI reliable, inclusive, and useful in real education settings.

Methodology

This analysis examined 171 winners and a random sample of Phase II proposals from six cycles of the Tools Competition, spanning 2021-2026. Phase II proposals represent competitors that passed an initial Phase I screening. For non-winners, a random sample of approximately 100 proposals per cycle, or 600 total, was drawn for tagging.

Each proposal was tagged on a consistent set of dimensions designed to capture how proposals use AI and how they design their tools: AI role (core, supportive, absent, or dataset infrastructure), specific technology approach, build approach (greenfield vs. extension of existing platform), solution category, target population, technical team capacity, research methodology, and open-source/data-sharing commitments. Cycle-level aggregates were weighted to reflect actual cycle proportions (~6% winners, 94% non-winners) to produce field-overall estimates.

Tagging was conducted using AI-assisted analysis with explicit conservatism protocols. Proposals were processed in parallel batches with instructions to mark dimensions as “not specified” rather than infer when proposal text was ambiguous. Cycle 1.0 proposals only contained essential information due to a compact application form, and accordingly returned “not specified” for many dimensions; later cycles’ richer documentation enables tighter analysis.

To guard against hallucination, a subset of high-stakes technical claims was independently re-tagged by a second agent and the two passes were compared to surface disagreements; the Cycle 6.0 sample was additionally re-processed in smaller chunks for fuller coverage. A 20-proposal random verification sample was re-read directly against source text. All tags were spot-checked by a human reviewer to catch systematic errors and edge cases.

Acknowledgements
The authors wish to thank Caitlin Mills, Associate Professor - University of Minnesota, Thomas Christie, Director - Engineering Hub, Renaissance Philanthropy, and Ralph Abboud, Principal Scientist - AI for Math and Education, Renaissance Philanthropy, for their insights to develop this report.

[Join the mailing list]

[Check out the Tools Competition blog]

[Follow us on X]

Tools Competition