

Integrating Digital Public Goods to Advance AI-Enabled Formative Assessment

Request for Proposals - *Applicant Instructions*



Grant Opportunity at a Glance

Detail	Information
Investment Name	Integrating Digital Public Goods to Advance AI-Enabled Formative Assessment
Investment Amount	Up to two awards, with a sum of \$1M available across the awards with a maximum individual award of \$600,000.
RFP Release Date	October 7, 2026
Proposal Due Date	November 20, 2026, 11:59 p.m., Anywhere on Earth (AoE)
Estimated Grant Period	18 months (3, 6-month phases)
Estimated Grant Start	February 1, 2027
Eligibility	This opportunity is open to any U.S.-based applicant, including edtech companies, for-profit organizations, non-profit organizations, and academic institutions. The outputs must clearly serve U.S.-based students. See Eligibility section below.
Grant Management	Both the proposal review and award monitoring will be managed directly by the K-12 AI Infrastructure Program including Digital Promise, Learning Data Insights, Massive Data Institute, DrivenData, and Catalyst @ Penn.
Application Link	Submit applications using this Form Form Template (view only)
Contact/Support	k12ai-infrastructure@digitalpromise.org

Purpose

The K-12 AI Infrastructure Program requests proposals to produce **a coherent package of open-source tools and workflows that demonstrate, document, and enable use of existing public goods to improve AI-enabled [formative assessment](#) in K-12 education contexts**. The educational problem that proposals should address is poor coordination of human teachers and AI-enabled technologies with regard to known failure points such as incoherent instruction, insufficient adaptation to learner variability, and limited teacher transparency into and control over AI outputs. Relevant technologies/tools include, but are not limited to, AI tutors, feedback and grading assistants, content and item generators, and classroom analytics dashboards. The learning sciences offer a theoretical and evidence-based language with which teachers and AI-enabled tools can coordinate on topics such as support for student reasoning and conceptual understanding; motivation, effort, and self-regulation; and ways in which pedagogy should adapt to learner variability. See [Required Phase 1 Focus](#) below for more details.

If the proposer's outputs are successful, developers will have access to a new, practical, data-informed approach to using public goods for an important purpose: enabling classroom teachers to experience greater agency with regards to AI-enabled technologies. This may include giving teachers:

- clearer insight into what AI-enabled tools are doing;
- more effective ways to modify and provide context to tools;
- greater ability to monitor these tools for known failure points; and
- stronger support for improving their own teaching practice.

In a future RFP, educators will be invited to submit proposals on a similar challenge but with different technical expectations. This RFP targets the third step in our logic model, summarized below, and specifically the bolded cell.

Table 1: K-12 AI Infrastructure logic model and activities, highlighting where this RFP fits.

K-12 AI Infrastructure Logic Model	K-12 AI Infrastructure Activities
(a) Funding new public goods (datasets, benchmarks, models)	The program's first two RFPs focused on this, resulting in more than a dozen projects developing new public goods .
(b) Increasing awareness of public goods	We have been increasing awareness through our repository and data competitions ; these activities will continue.
(c) Encouraging uptake and use of the public goods by developers	This RFP aims to ease the work and increase the benefits to developers of exploring and integrating education-specific public goods, with an overall aim of increasing the quality of teaching and learning using education-tested AI tools.

The RFP is inspired by the observation that "[agent harnesses](#)" are making it easier for developers to rapidly engineer productive uses of AI in domains like coding, conducting market research, and producing social media content. A harness orchestrates how LLMs and the various forms of context (system prompts, profiles, skills, stored knowledge or memories of past sessions, function calls, MCP servers, etc.) work together in reliable and effective workflows. Harnesses also can be self-improving; for example, they can write helper code (e.g., in Python) for deterministic parts of a complex workflow, which saves cost and increases reliability. Existing harnesses have not yet been tuned for education-specific use cases, especially regarding how to integrate the kinds of education-specific public goods we are funding. Inspired by how harnesses are helping in other domains, we seek a program of work to adapt or improve a harness to make education-specific public goods easier for developers to explore, test, and integrate. This focal challenge will be coordinating the work of teachers and AI-enabled tools, as described above and elaborated below in the [Required Phase 1 Focus](#) section.

Under the award to be made pursuant to this RFP, a technical team will produce an integrated package of software; tools; documentation; evidence of effectiveness (e.g., demonstrated through a technical white paper); and (optionally) supplementary assets like datasets, skills, knowledge, etc. Successful packages will build upon an enhanced open-source AI harness (alternative technical start points are acceptable if well-argued) designed to make education-specific public goods accessible, testable, and integrable for developers.

The package may also build upon existing open-source software, such as an existing open-source harness; however, a harness-based approach is not strictly required. The package and any dependencies must be licensed openly under MIT License, or Apache 2.0 for code and CC-BY-4.0 for datasets and other components (as well as any dependencies for the package to operate). These packages will be infrastructure in the sense that they intend to directly serve multiple developers of teacher-facing products. A demonstration application that engages teachers should be included to show the value of this work, yet the majority of the effort and the resulting outputs must serve multiple third-party developers and is not intended to only directly serve teachers.

The successful awardee will conduct an 18-month program of work that enables many downstream developers to more easily, conveniently, and effectively make use of public goods to improve their instructional resource. The 18 months will have three phases, each about six months in duration, starting February 1, 2027:

1. **Phase 1:** Address the specific challenge of coordinating teacher and AI-enabled technologies with regard to [formative assessment](#), as specified below, using public goods that are already published. Milestone: Feasibility is demonstrated through a working prototype; emergent issues and risks are clarified and a plan is made to address them.
2. **Phase 2:** Extend the output of Phase 1 to incorporate a [subset of newly available public goods](#), to be released by the K-12 AI Infrastructure Program in summer 2027. Milestone: An upgraded version incorporates the subset of additional public goods and demonstrates promise to a downstream developer audience. Further, an initial benchmark or set of eval results demonstrate improvement over status quo approaches.
3. **Phase 3:** Refine the output of Phase 2 to better address the needs of developers uncovered during the work, and publish a final package of resources. Milestone: Use cases from

downstream developers are identified; a releasable version and benchmarks/evals address this set of use cases and is released publicly after a safety review with the K-12 AI Infrastructure Program team.

Background on the K-12 AI Infrastructure Program

The K-12 AI Infrastructure Program invests in public goods that enable a wide variety of products and services to provide high-quality, AI-enabled formative assessment to every student, with particular attention to the needs of U.S. students furthest from opportunity. Public goods supported by the Program are openly licensed resources that can be broadly used and incorporated, directly or indirectly, by developers of AI-enabled tools and infrastructure for K-12 education. In this RFP, we seek to reduce barriers to this use while addressing a known educational challenge: the need for better pedagogical and learning sciences coordination between teachers and AI-enabled tools.

The Program focuses on [formative assessment](#) grounded in the learning sciences; many developers of AI-enabled technologies could benefit from being able to coordinate the work of teachers and AI-enabled tools in learning sciences terms. Formative assessment is about understanding what students know and can do and using that evidence to inform adaptive educational practices. Teachers and AI-enabled tools could usefully coordinate on learning sciences-grounded issues, including: eliciting students to improve their conceptual understanding or strong reasoning skills (and not just right answers); managing cognitive demand, cognitive load, and potential cognitive offloading; sharing information about misconceptions as well as strengths in student reasoning; using consistent, correct, and clear language and representations to guide student reasoning; and contextualizing pedagogy by the same curriculum resources, standards, frameworks, and other guiding principles. We encourage applicants to review our write up on [Formative Assessment and Public Goods for AI in Education](#) for more insight into Program expectations.

We are particularly interested in public goods and technical solutions that operationalize evidence-based insights about formative assessment in ways that can be incorporated into practical applications. This would be demonstrated in changes that the AI-enabled tool or teacher could make that adapts to student evidence; for example, in wake of identifying a common misconception among students, subsequent lesson plans and tutoring session plans are modified to address that misconception in coherent ways.

The Program also adopts a [Targeted Universalism](#) approach: supporting advances that can benefit all learners while intentionally attending to populations of students whose strengths and needs may not be adequately represented or served by existing AI-enabled technologies. Applicants should consider learner variability and how their proposed work can support developers in creating high-quality formative assessment experiences that appropriately adapt to learner variability.

For the purposes of this Program, [public goods](#) are modular, foundational building blocks that technical users can adopt to develop or improve educational applications. Public goods funded through the Program include datasets, benchmarks, and models, as well as associated tools, code, documentation, and other resources that facilitate responsible use.

Through its first two RFPs, the Program is funding more than a [dozen projects developing public goods](#) supporting AI-enabled formative assessment, including development of resources in mathematics, writing, science, and reading, across modalities such as text, speech, student work, and classroom and tutoring interactions. See [Appendix 1](#) for details.

Our program emphasizes infrastructure. We are not funding the creation of proprietary products that directly serve teachers and students. Rather, we are funding openly licensed infrastructure that can help many developers improve their products and services for students and teachers. As long as this emphasis is clearly respected in the proposal budget and workplan, a teacher-facing demonstration can be a useful aspect of the work.

Technical Goals

The successful applicant will address the problem that the *availability* of education-specific public goods may not be enough to enable *uptake* by educational developers, even though these education-specific assets have the potential to improve many developers' products. This could be because it takes time and effort to explore, test, and integrate public goods. If uptake is too costly, too difficult, too inconvenient, or too arcane, developers may not make it a priority. We ask applicants to propose solutions that would help many developers overcome these barriers, as well as benchmarks or evaluations that would demonstrate the advantages.

We adapt the classic Rogers (1962) "Diffusion of Innovation" to express the factors that could reduce barriers. The open-source package produced by a successful awardee should address the following goals (however, it is not necessary to address these one-by-one; they can be grouped):

- **Observability:** Provide runnable examples (e.g., demonstration instance) of using a public good to improve an instructional activity, so a developer can observe a successful case of using a public good. Answer the question: How can a public good asset become an active ingredient in an AI-enabled application?
- **Trialability:** Make it easier and faster for a developer to explore how they could use a public good and tune it to their specific needs. Answer the question: How can developers experiment with public goods and determine whether and how they fit the improvements they want to make to their products?
- **Complexity:** Reduce the perceived complexity of using a public good to improve one's own instructional product. Answer the question: How can a standardized package of resources simplify what a developer needs to know in order to explore, try, and integrate public goods?
- **Quality Assurance:** Enable the developer to use their own as well as the included evals and benchmarks to establish that the integration of a public good meets their own requirements for improvements to quality. Here, we creatively reinterpret Rogers' concept of "compatibility" to emphasize quality assurance in educational contexts. Answer the question: How can a developer ascertain that integrating public goods retains and improves upon desired educational qualities of AI output or AI-mediated interactions?
- **Relative Advantage:** Make it easier to show the educational (e.g., teaching or learning) value of using an education-specific public good. Answer the questions: If a developer integrates a

public good, what advantages are teachers likely to experience? How can the developer evaluate and demonstrate these advantages to teachers?

We are open to the proposer's conception of the best technical approach to achieve these and other benefits, and thereby to make developers more willing to explore the use of public goods in their product and more willing to integrate public goods into their work.

1. As indicated above, one possibility is to modify an open-source AI harness so that it is tuned to work with education-specific public goods and customizable to a developer's desired exploration, testing, and integration. Harnesses ideally will be available without cost, give the user choice of which models and API keys they will use to access models, support open-weight models as an option, have active maintainers, and include a permissive license to enable downstream developers to freely productize what they produce with the harness.
2. A second, more traditional approach is to create a well-documented public collection of sample code and tools, for example, in a GitHub repository.
3. A third approach, which could be taken in combination with the above approaches, is to develop a "reference architecture" that is a minimal viable product that shows the use of public goods for instructional purposes and provides source code that can be reused to ease the work of integrating public goods into a developer's own educational product.

It is possible to propose an MVP starter kit that could enable a next generation of builders to start from public goods.

Note: Harnesses manage context and therefore the case should be made that the chosen approach will minimize irrelevant context (e.g., harness capabilities that exist only to solve an out-of-scope problem, such as helping a YouTube creator build content) and will seek to right-size the context that is incorporated.

Required Phase 1 Focus: Teacher-AI Coordination

To streamline proposal review and Phase 1 startup, we specify the educational target for Phase 1 work. Proposers must be clear about how their work will address this target. Plans for Phases 2 and 3 will necessarily be more emergent (based on public goods that become available, interest from developers, and results from applied research) and should build upon Phase 1 work.

The target is **improving educator and AI coordination with regards to formative assessment**. For example, a recent [Stanford SCALE study](#) argued for the benefits of teacher-AI integration: "The effectiveness of AI-led tutoring depends as much on integration, such as by teachers in classrooms or parents at home, as on the quality of the software itself." Yet, most teachers do not feel well-informed by edtech products nor do they feel they know how to provide input. Conversely, edtech and AI products are often underinformed about what an educator knows that is relevant to formative assessment. Lack of coordination between teachers and AI-enabled tutors is and will continue to be a major barrier to adoption, transparency, and trust unless it is better addressed.

Proposals **must describe at least four tasks** that their harness will do that would be helpful to a developer and would result in better teacher-AI coordination. We provide examples below. A

proposer can choose among these, modify these, or add their own. The goal is to do tasks that would be helpful to downstream developers and that would give teachers a meaningful experience of more ability to learn from, monitor, and control what happens in sessions.

Phase 1 Task Examples

Please note that in the "use of public goods" in this table, we suggest that proposers use public goods in two distinct ways:

1. As placeholder data, that a developer or teacher would replace with their own data when this task is implemented in a downstream applied setting.
2. As substantive additional input, such as a misconceptions database, that informs the context for doing the task.

Sample Task	What It Does	Use of Public Goods
Adapt teaching plans based on tutoring session data	Show how evidence from AI-based tutoring could be made relevant to a teacher's upcoming lesson: for example, evidence of misconceptions, what kinds of student assignments (e.g., math problems) revealed them, what strengths students showed. Align on the basis of curricular objectives and teacher plans, as well as clarifying conversations with the teacher about their focus.	Doing this well should include multiple data sources: • Tutoring session data • Misconceptions data • Annotated tutoring moves (to pull out examples of student interactions that illustrate to the teacher what is going on) • Knowledge Graph to specific content in tutoring and teaching • Teacher lesson plans • Specific input from the teacher on what they want to focus on
Differentiate the plan for specific students' upcoming tutoring sessions based on teacher's input and classroom data	Show how recent classroom data could be analyzed with a teacher to gain a sense of where low-, medium-, and high-performing students are and variation in what students are struggling with. Elicit the teacher's input on a tutoring plan that is differentiated for different ability-based groups of students or other specific needs of the student. Establish a monitoring plan for the future tutor session so teachers get a report back on what happened in tutoring and how it related to their recommendation to the tutor.	Doing this well should interpolate among different types of data: • Classroom transcripts • Classroom student performance data; how students answered quiz problems, for instance • Lesson plans, teaching objectives, specific teacher input on what they want to focus on • Data on problem or assignment difficulty, to give the tutor specific examples of appropriate student tasks or problems • A description of ways in which the tutor can adjust based on teacher input (this

		will eventually vary per downstream developer; proposer should plan to make a proxy.) • Evals that help the developer track whether they are adapting as desired.
Coordinate on pedagogical strategies for challenging tutoring situations	Identify examples from both tutoring and teaching where students are struggling to make progress, possibly even with support. Prepare materials for an interactive discussion of what productive and non-productive struggle looks like. Discuss high-level strategy for forthcoming plans (see tasks above) so that teacher and tutor can better coordinate in an approach to adjust these struggles.	Data for this can be similar to what is above. Please note that this task is included because it likely works well for both mathematics and writing and thus could be an opportunity to show how coordination on how to support productive struggle could occur across curricular domains. More generally, self-regulated learning could be an area for cross-curricular coordination.
Customize improvement goals and tracking for mutual teacher and tutor improvement	Discuss a range of possible improvement areas with a teacher that could occur simultaneously in tutoring and teaching (for example, "I want my students to show more agency and reasoning," "I want to develop better strategies for getting students to meaningfully reflect on their writing before rewriting," or "I want students to have feedback that is better aligned to research principles.") and ways that these could be evaluated both in teacher talk and tutor moves. Customize a set of evals and a monitoring plan for coordinated improvement.	This example would leverage evals and benchmarks that already exist as public goods, or ones that could be adapted from public goods to meet the specified goal. Further, classroom and tutoring transcripts could be used as "dry runs" to give the teacher feedback on what the evals would capture and show, enabling tuning the evals to ensure the teacher gets a concise, meaningful, and actionable report, and not a wall of text. This could tie into the tasks above, to guide teacher and tutor planning to incorporate these goals.

Using Learning Sciences in Proposed Tasks

Proposed approaches should be grounded in existing educational theory and help to extend it. As discussed above, we see the learning sciences as offering a *lingua franca* for discussing coordination between AI-enabled technologies and teachers by providing a theory-motivated vocabulary for what kinds of coordination are meaningful. A non-exhaustive list of examples of coordination issues and useful learning sciences concepts and approaches are provided below. Incorporate these or other concepts where they strengthen your proposal. There is no need to address every listed concept; the proposer should have the expertise to choose concepts that best fit their approach.

Coordination Issue Teachers should have agency to align AI-enabled technologies on:	Learning Sciences Concepts and Approaches Teachers can leverage AI-enabled tools to coordinate:
Strengths and weaknesses of student reasoning and conceptual understanding (and not just correctness)	Misconceptions, claims-evidence-reasoning structure, multiple representations, sense-making, concept mapping
Effort, motivation, persistence	Cognitive load, cognitive demand (what students are asked and supported to do), cognitive offloading, self-regulation, productive struggle, productive failure
Learning variability	Multiple means of expression, action, and representation; funds of knowledge and assets that support student engagement and reasoning; zone of proximal development
Pedagogical strategies	Teaching moves, understanding student reasoning, principles of providing high-quality feedback, (for mathematics) mathematical quality of instruction

Available Public Goods for Building Coordination Capabilities

In Phase 1, proposers should plan to use available public goods from any source (not limited to the K-12 AI Infrastructure Program public goods, which may not yet be available). Proposers may also add their own data, benchmarks, evals, or annotations to enable rapid progress, but there must be a plan whereby either the proposer or a downstream developer can substitute for them later so that it is possible to use the harness without reliance on the proposer's private IP.

Considerations that may be helpful for Phase 1 and beyond include:

- In Phase 2**, public goods should specifically emphasize deliverables from K-12 AI Infrastructure Program projects and may continue to include existing public goods. [Appendix 1](#) provides a starting point for identifying public goods that applicants should explore and integrate into their proposed work, with the goal of increasing the quality of teaching and learning using education-tested AI tools. Before Phase 2 (e.g., during **Phase 1**), the proposer should plan to get to know these in depth. Our team will support access to pre-public release of public goods from our actively funded cohorts.
- In Phase 3**, the proposer will work with downstream developers to better serve their needs and to demonstrate uptake of the harness and K-12 AI Infrastructure Program public goods. This may again involve working with some private information (e.g., the developer's IP). This is fine, as the expected use of the harness should be as a tool that downstream developers can use to interweave public goods with their own product-specific IP.

Community Goals

The proposer will work with K-12 AI Infrastructure Program awardees and developers. The K-12 AI Infrastructure Program will have separate funds available to provide incentives to awardees and outside developers to participate, and thus proposals do not need to budget for this.

Working with K-12 AI Infrastructure Program Awardees. Proposers must plan to work with multiple education-specific public goods that are becoming available via the K-12 AI Infrastructure Program and can plan to work with other available education-specific public goods that would (a) allow them to start development sooner and/or (b) complement what is available in our repository. They must plan to work across multiple public goods, not only one. However, we are not requiring a proposal to address every K-12 AI Infrastructure Program public good.

We will facilitate interactions between the awardee under this RFP and other creators of public goods that have already been funded. We will reserve funds that can be allocated to the creators of public goods to compensate them for working with the awardee under this RFP.

In Phase 1, the proposer should plan to gather information about what public goods will become available in summer 2027. In Phase 2, the proposer should plan to extend their work to incorporate multiple (but not necessarily all) public goods that have become available. In Phase 3, the proposer may interact with producers of public goods as needed to refine their work.

Working with K-12 AI Developers. Proposers must plan to work with multiple AI developers who are potential users of public goods. Proposals will benefit by obtaining letters of support from one or more developers who would be willing to participate in the work. Once funded, awardees should plan to engage at least three developers (the Program can help facilitate connections, as needed).

In Phase 1, interactions with developers should focus on identifying the specific factors that enhance usability, feasibility, convenience, and perceived value to support successful adoption (e.g., What makes diffusion of innovation work in this case?). In Phase 2, awardees should collect developer feedback on early-stage demonstrations to guide and refine ongoing work. In Phase 3, the K-12 AI Infrastructure Program will reserve funds to incentivize developers to spend time to test the resulting resources and undertake integration of public goods into their own work.

Across these phases, awardees should treat public-goods creators and downstream developers as partners in the iterative development process. Feedback from both groups should inform the design, refinement, documentation, usability, and feasibility of the resulting open-source package.

Eligibility and Expectations

Eligibility

This opportunity is open to all U.S.-based entities (including U.S. territories), such as edtech companies, for-profit or non-profit organizations, school districts, and higher education institutions. Non-profits under fiscal sponsorship are also eligible to apply.

Submission Limits:

- **Standard Organizations:** Limit one application per lead organization.
- **Large Organizations (>2,000 staff):** Limit one application per department. Must be distinct team composition per proposal.
- **Universities (>2,000 staff):** Limit one application per major academic division or school (e.g., College of Natural Sciences, Graduate School of Education). Must be distinct team composition per proposal.

Partnerships combining technical, learning sciences, and community-engagement expertise are strongly encouraged.

Allowable costs include:

- Staff time, which can include buy-out time for educators who work on the project
- Consulting fees or stipends to experts and advisors (who may be located internationally)
- Equipment and computational resources, (including devices, internet access, privacy-compliant analytical applications, computing power, and storage)
- Travel where necessary to producing or disseminating the public goods
- Costs of Institutional Review Board (IRB) review, privacy oversight, legal support for data sharing agreements, etc.
- Software or tools that are essential and allocable to the public good being developed
- Subawards or subcontractors performing project-specific tasks
- Translation or accessibility services to support equity and access
- Institutional overhead *up to a maximum of 15%*

Unallowable costs include:

- Entertainment, social, or networking events
- Lobbying or advocacy-related expenses
- Pre-award costs
- International travel not directly tied to project deliverables
- Expenses that are included in institution overhead such as:
 - General office supplies not directly allocable to the project
 - Administrative or clerical salaries not directly tied to the project
 - Capital expenditures (e.g., buildings, renovations)
 - Equipment purchases over \$5,000 unless pre-approved and justified

Evaluation Criteria

Proposals will be reviewed by both peer reviewers and the program team on four equally important criteria:

1. **Significance:** WHY the completed project will improve the coordination between teachers and AI-enabled tools in a meaningful way, while advancing many developers' interest in and ability to use public goods to increase use of the learning sciences to create educational value in their products. The potential for downstream developers to use the outputs of the completed project and thereby to increase use of public goods overall is important. We recognize that

since the program's funded projects are still in development, applicants may need to make assumptions about their affordances.

2. **Assets and Capabilities:** WHAT and WHO will enable project success. Two kinds of expertise must be included: AI/technical and educational/learning sciences. Capabilities can include what the team has previously developed and any existing data or assets that enable a rapid start.
3. **Project Workplan:** HOW the project will be completed. The reviewers will look for compelling detail on how Phase 1 will be executed. As Phases 2 and 3 will evolve, reviewers will look for a basic plan that fits, as well as a clear description of how the team will gather information to refine the plan and to develop metrics for success that address known and emergent requirements.
4. **Release and Dissemination:** OUTPUTS of the project and how they will support adoption and use. Reviewers will look for a harness-like output that could be used by multiple developers under the required licenses without dependencies that block use. Additional documentation and examples that support use are desired. The quality of your dissemination plan will be evaluated.

An FAQ is available on the [K-12 AI Infrastructure Program website](#). We will also offer office hours to respond to questions about the RFP.

Proposal Guidance & Structure

Proposals should use the criteria below to describe the proposed work. Applicants should respond directly to the guidance provided, while avoiding unnecessary repetition across sections. The total proposal narrative is limited to 6,000 words.

Significance

Significance of Proposal

- Explain the significance of the proposal, including how the project will coordinate teachers and AI-enabled tools around known failure points as well as reduce barriers to uptake of public goods as described above. Note how you will track and measure success.
- Relating to the Program's emphasis on formative assessment, provide a clear, concise, and well-grounded overview of how you understand formative assessment in the context of this specific work (no more than 1-2 paragraphs).
- Describe **at least four tasks** that your harness will do that would be helpful to a developer and would result in better teacher-AI tool coordination.
- Provide an overview and high-level rationale for your technical approach.
- Describe clearly what openly licensed resources will result from the project and which developers or types of developers you expect will use them.

Developer Use Cases

- Describe how you intend to support downstream developers: What use cases will you support and what will be the benefits and advantages to them?

- List any developers who have provided letters of support and how the proposed use cases fit their needs.
- Describe your willingness to collaborate with a K-12 AI Infrastructure Program officer and with others in the Program to map and refine shared targets for Phases 2 and 3 during the period of performance.
- State how you would work with a set of developers in Phase 3 to support actual uptake. (Note that the Program plans to keep financial resources on reserve to incentivize greater developer involvement particularly in Phase 3.)

Educator Use Cases

- Describe how you intend to support educators: What use cases will you support, and how will they improve coordination between educators and AI-enabled technologies?
- Describe how the proposed use cases will increase educator agency, including educators' ability to understand what AI-enabled technologies are doing, provide relevant context and guidance, monitor student experiences and outcomes, and influence subsequent interactions.
- Explain how educator input and classroom evidence will inform AI-enabled technologies, and how information from these interactions will be made useful and actionable for educators in their own instructional planning and decision-making.
- Describe how you will engage educators in testing and refining the proposed use cases, including how educator feedback will inform the design, usability, feasibility, and educational value of the resulting tools and workflows.

Assets and Capabilities

Team Qualifications

Describe the expertise and roles of key personnel and explain how the team collectively has the technical, learning sciences, and community-engagement expertise necessary to complete the proposed work. Identify who will be responsible for the major technical and project responsibilities, and describe the relevant experiences of key personnel in producing and/or using open-source software and/or public goods. Include:

- Technical expertise (key personnel that establish eligibility and their output or publications)
- Educational and learning sciences expertise (key personnel that establish eligibility and their output or publications)
- Prior success in producing and/or utilizing public goods
- Prior track record in working with a community of both suppliers and users to achieve benefits to the ecosystem.

Project Plan

Project Management/Timeline Contingencies

Describe how the project will be managed across the 18-month period of performance, including team roles and responsibilities, decision-making, and coordination across partner organizations, as well as processes for monitoring progress against milestones and deliverables.

- Provide a project timeline covering all three phases, including major milestones, deliverables, and responsible team members.

- Describe the most important technical or implementation risks and how the team will respond if they arise.
- Explain how the team will achieve a rapid start following the anticipated February 1, 2027, project start date.

Community Interaction Plan

- Describe how you will interact with the K-12 AI Infrastructure Program community across the three phases (see phase-by-phase expectations in the [Purpose](#) and [Phase 1 Task Examples](#) section).
- State how you plan to work with a set of creators of public goods during each phase (see guidance above), including how these interactions will inform the development and refinement of the proposed resources.

Public Goods To Be Used

- List K-12 AI Infrastructure Program public goods that you intend to use (particularly for Phase 1; the list will evolve in Phases 2 and 3).
- List other education-specific public goods that you intend to use.

Technical Approach

- Address Rogers (1962) "Diffusion of Innovation" as noted in the [Technical Goals](#) section: observability, trialability, complexity, quality assurance, and relative advantage.
- Be specific and detailed about Phase 1:
 - Explain and argue for your overall technical approach.
 - Clearly articulate what the milestones and deliverables will be.
- Knowing that Phase 2 and 3 will depend on insights from Phase 1, explain what you will do to integrate emerging public goods from the Program in Phase 2 to better support developers in Phase 3. Provide additional details that support your argument that you will produce a compelling public good with demonstration of developer use by the end of phase 3.

Release and Dissemination

Responsible AI

- Ethics: Explain how you will align your work with established data privacy, security, and ethical standards to ensure safety, transparency, and accuracy. Explain how you will mitigate bias, ensure fairness, and preserve human-in-the-loop oversight and user agency.
- Targeted Universalism: Define universal goals that support high-quality AI-enabled formative assessment benefits for all K-12 learners. Describe approach for addressing student population needs and structural and/or contextual barriers for specific groups (i.e., multilingual or neurodiverse learners) that also support learning for all students. Describe implementation strategies (i.e., accessible/inclusive user interfaces) that ensure public goods effectively support all learners.

Public Goods, Licensing Terms, and Dissemination

- List the major public goods that will be newly produced and released by the end of the project, along with licensing terms.

- The K-12 AI Infrastructure Program will publicize these goods. The proposer should list specific ways in which they can support dissemination and early uptake, such as giving workshops or webinars, creating tutorials or user success stories, continuing to provide technical support for a period of time, etc.
- Describe how public goods will respond to FAIR Principles (Findable, Accessible, Interoperable, Reusable).
- Describe how the project team anticipates working with the Program team to review and approve public goods prior to release to ensure quality and protections for privacy, data security, and other safety concerns.
- Describe the documentation and supports that will accompany the released package, such as technical documentation, examples, tutorials, sample code, demonstrations, or other resources that make adoption and integration more likely.
- Your plan should include release of your public goods on the [K-12 AI Infrastructure Program Platform](#) (hosted by DrivenData).

When and Where to Submit

Submissions will be accepted until November 20, 2026, at 11:59 p.m., Anywhere on Earth (AoE), [on this website](#). You will be asked to register in order to access the application.

Your submission should include a completed form including an abstract, the required PDF attachment, and a separate spreadsheet for the budget. The abstract (350-word limit) will be used to select appropriate peer reviewers. Specifically, the narrative and budget should contain:

Narrative PDF:

- Please include your abstract in this PDF as well as the form field.
- Include a description of your project and narrative addressing the topics above (6,000 word limit). Proposers should note that while this RFP provides detailed guidance on how proposals will be evaluated, many of the guidelines can be addressed concisely. A separate paragraph for each guideline is not required.
- Include references and citations, not included in the word count
- Please aim for legibility and follow standard proposal formatting, such as 12-point font size for text, at least 1.15 line spacing, a minimum 10-point font for images/tables, and 1-inch margins.
- Use the file name convention
PILASTNAME_K12AI_RFPIntegratingDigitalPublicGoods_Narrative.PDF

Supplemental PDF:

- Supplemental materials may include images, figures, tables, visual database schemas, mock-ups/demos, or other visual materials that are essential to understanding the proposed work. The supplemental PDF should **NOT** be used to provide additional narrative beyond the proposal page limit.
- Include resumes or CVs for the project director and any other key staff
- If additional organizations or consultants are named in the project description, include short letters stating that they have reviewed their role and are willing to serve (e.g., similar to NSF

Letters of Commitment). These letters are required for documentation purposes only and will not be reviewed by peer reviewers.

- Use the file name convention
PILASTNAME_K12AI_RFPIntegratingDigitalPublicGoods_Supplement.PDF

Budget:

- Submit as an .XLS or .XLSX file.
- Include the project's budget in the template provided, along with justification.
- [Link to forced copy of Google Sheet](#)
- [Link to download .XLSX file](#)
- Download a copy of the template and save your version with the file name convention
PILASTNAME_K12AI_RFPIntegratingDigitalPublicGoods_Budget.XLS

Appendix 1: K-12 AI Infrastructure Program Public Goods & Additional Resources

Appendix 1 provides a starting point for identifying public goods that applicants should explore and integrate into their proposed work. It includes public goods currently being developed through the K-12 AI Infrastructure Program, as well as other relevant education-specific resources. Applicants may use resources beyond those listed here. **Note: This is not a comprehensive list.**

K-12 AI Infrastructure Program Public Goods (currently funded, under development)

The K-12 AI Infrastructure Program grantees are developing a wide range of public goods, primarily focusing on datasets, benchmarks, models, and open-source tools to advance AI in education. Below is a summary of the public goods being developed by each grantee organization and project. More will be added on our [website](#).

Stanford University (Project: KB-TutorBench)

- Datasets: A multimodal knowledge-building dataset containing ~3,000 semi-synthetic dialogues across five subjects and ~3,000 authentic 1:1 tutoring transcripts.
- Benchmarks & Models: A benchmark evaluation suite with standard scenario seeds and rubrics, alongside fine-tuned 7B pedagogical model weights (Mistral-7B/QLoRA).
- Tools: An open-source GitHub repository housing data collection tools, annotation code, and evaluation scripts, supplemented by a technical report.

Cornell University/NTO (Project: Benchmarking ASR in Educational Contexts)

- Benchmarks & Models: An open educational Automated Speech Recognition (ASR) leaderboard on Hugging Face and fine-tuned children's speech ASR baseline models (Whisper/wav2vec variants) optimized for minoritized speech.
- Datasets: An expanded tutoring audio dataset with annotations for diarization, accents, and disfluency.
- Tools: Reproducible training recipes, evaluation harnesses, and reporting standards.

Learning Equality (Project: Identifying Science Misconceptions)

- Datasets: A structured assessment package dataset covering grades 6–12 science with free-response items, rubrics, and misconception maps.
- Benchmarks & Models: A free-form science misconception benchmark and locally deployable reference classifiers (fine-tuned, quantized sub-3B small language models).
- Tools: Technical reports, failure mode analysis, and datasheets.

Princeton University (Project: Simulated Student Models)

- Models & Pipelines: An open neural network simulating student problem-solving strategies and errors in fraction arithmetic, supported by a synthetic student reasoning generation method and a two-stage neural network training pipeline.
- Datasets: Preprocessed human reasoning datasets containing structured fraction think-aloud answers and verbal explanations.

Maine Mathematics & Science Alliance (Project: AMR Pilot)

- Datasets: A multimodal interview dataset of 2,400 student reasoning episodes (grades 2–5) with visual artifacts and transcripts, plus a dataset of high-resolution iPad stroke-log traces.
- Benchmarks: A proof-of-concept reasoning benchmark task to evaluate AI's capability in scoring student additive-to-multiplicative reasoning.
- Tools: A DDI-compliant codebook and annotation manual.

University of Maryland College Park (Project: Formative Assessment Datasets)

- Datasets: The EDSI Small-Group Discourse Dataset (~1,000 math lessons with multi-channel audio) and an enhanced NCTE Classroom Dataset of 1,660 elementary math transcripts.
- Models: An automated group-work detection classification model trained on audio and transcript signals.
- Tools: A comprehensive relational data schema and linking documentation.

Harvard University (Project: OpenLiteracy)

- Benchmarks & Models: A phoneme-level reading error benchmark of 20,000 authentic K-3 child speech utterances and fine-tuned phonetic error detection speech models.
- Datasets: A pipeline that generated 50,000 synthetic speech utterances embedded with decoding errors.
- Tools: Standardized data governance infrastructure, including speech de-identification and voice-masking pipelines.

National Trail Local School District (Project: Discipline-Specific AI Writing Feedback)

- Datasets & Benchmarks: A dataset of ~14,832 de-identified writing samples across ELA, Science, and Social Studies, paired with a feedback quality benchmark featuring subject-specific rubrics.
- Models & Tools: Enhanced on-premise Moodle AI plugins integrating local Ollama LLMs, a discipline-specific prompting framework, and a privacy-preserving deployment guide for rural IT staff.

University of Florida (Project: Storiza Corpus of Early Reader Recordings)

- Datasets: ~80 hours of K-3 dialect-inclusive audio recordings covering all five major U.S. dialect regions.
- Models & Tools: Reading error and ASR baseline models, alongside automated annotation and assessment algorithms for oral reading fluency and comprehension.

University of Pittsburgh (Project: Multimodal Writing and Feedback Dataset)

- Datasets: A multimodal first-year writing corpus of ~2,000 multi-draft student essays paired with 300–450 hours of student-instructor conference audio.
- Benchmarks & Models: A held-out evaluation test set, an open-source public leaderboard, and multimodal writing support AI baseline models.
- Tools: Data processing, layout extraction, and anonymization pipelines.

Quill.org (Project: The "Productive Coaching" Benchmark)

- Datasets: A training dataset of 10,000+ authentic teacher-authored feedback examples, plus a multi-turn AI coaching evaluation dataset of 1,000+ coaching sessions.
- Benchmarks & Tools: A public benchmark leaderboard for evaluating AI coaching quality and a "Productive Coaching" framework/adoption package grounded in learning sciences.

Playlab (Project: AI-Enabled Formative Writing Feedback in Science)

- Datasets: A proof-of-concept annotated dataset of multi-turn claims, evidence, and reasoning (CER) feedback conversations (10,000–15,000 interactions) for grades 8–12 science.
- Tools: An annotation protocol and rubric for coding CER quality, an open-source de-identification pipeline, and extensive data preparation documentation.

Building and evaluating AI for K-12 education requires open, well-documented resources. View our new [K-12 AI Infrastructure Platform](#) (hosted by DrivenData) to access a trusted hub for data competitions, datasets, models, and benchmarks.

This [Public Goods for AI in K-12 Education Scanner](#) is our working inventory of those resources. It tracks openly available datasets, benchmarks, models, and competitions relevant to K-12 education, recording what each one covers, which grade levels and subjects it addresses, how it is licensed, and where it came from. It refreshes automatically each month as new resources are published. ***It is a starting point, not a finished map.***

Learning Commons

[Learning Commons](#) provides open-source products:

- [Knowledge Graph](#) provides machine-readable educational datasets that give AI products trusted educational context.
- [Evaluators](#) score AI-generated content against expert-built educational rubrics and return dimension-level results with reasoning.
- [Agent Skills](#) provide expert-informed instructions that help AI assistants complete teaching and learning tasks.

DrawEduMath

Authored by: Worcester Polytechnic Institute and University of California, Berkeley with ASSISTments

[DrawEduMath](#) is an dataset of math problem images concatenated with students' hand-drawn responses. Each image is annotated by a human expert math teacher describing the student's written answer. This dataset is useful for training models to interpret elementary students'

hand-written math responses.

FoundationalASSIST: An Educational Dataset for Foundational Knowledge Tracing and Pedagogical Grounding of LLMs

[FoundationalASSIST](#) is introduced as the first comprehensive English educational dataset for LLM research in education, enabling studies on student modeling and pedagogical understanding through detailed interaction data and alignment with educational standards.

National Tutoring Observatory

The National Tutoring Observatory (NTO) provides [researchers with access to a rich, large-scale repository of tutoring data to advance the science of learning](#). It offers access to diverse datasets—including session transcripts, audio/video recordings, and student performance metrics—through secure platforms. By providing this data, along with open-source tools for analysis and de-identification, it aims to lower the barrier to entry for impactful educational research. NTO fosters a collaborative community through workshops, meetings, and data competitions, enabling researchers to investigate pressing questions, validate new AI models, and discover effective tutoring practices that can be applied at scale.

Relatedly, [Sandpiper](#) orchestrates multiple LLMs to produce reliable, highly accurate annotations at scale, starting with tutoring transcripts, built for any field.

References

powell, j. a., Menendian, S., & Ake, W. (2019). Targeted universalism: Policy & practice [Primer]. Othering & Belonging Institute, University of California, Berkeley.

https://belonging.berkeley.edu/sites/default/files/targeted_universalism_primer.pdf

Rogers, E.M. (1962) Diffusion of Innovations. Free Press, New York. - References—Scientific Research Publishing. (n.d.). Retrieved October 6, 2026, from

<https://www.scirp.org/reference/referencespapers?referenceid=2008938>