

Evaluating Artificial Intelligence Tools

PURPOSE AND SCOPE

The purpose of this evaluation is to provide a standardized method for evaluating Artificial Intelligence (AI) software and services. This rubric ensures that all AI tools adopted by the LEAs meet rigorous standards for Data Privacy, Pedagogical Alignment, and Ethical Safety.

COMPREHENSIVE EVALUATION RUBRIC

Evaluation Criteria	Level 4: Exemplary	Level 3: Proficient	Level 2: Developing	Level 1: Inadequate
Data Privacy & Security	Dedicated Education tier, Data never trains global models, clear user data ownership.	Verified FERPA/COPPA compliance; clear data encryption and "opt-out" for data training.	Policy exists but lacks clarity on 3rd-party data sharing or long-term storage.	No clear privacy policy; non-compliant with FERPA/COPPA/SOPPA.
Instructional & Cognitive Presence	Enhances higher-order thinking (analysis/synthesis) and metacognitive strategies. Seamlessly integrates with curriculum; provides real-time data for differentiated instruction.	Directly aligns with MSDE/District standards; supports cognitive tasks and learning efficiency.	Supports tasks but requires heavy teacher modification; focuses on lower-level skills.	No connection to state standards; lacks clear pedagogical utility; hinders cognitive engagement.
Bias, Equity & Access	Proactive bias mitigation; supports multiple languages and WCAG 2.1 accessibility.	Built-in filters for harmful content; vendor provides reports on algorithmic fairness.	Acknowledges bias but offers no active tools or filters to mitigate it.	No info on model bias or content filtering for harmful topics.

Evaluation Criteria	Level 4: Exemplary	Level 3: Proficient	Level 2: Developing	Level 1: Inadequate
Human-in-the-Loop	Full "kill-switch" control; teachers can manually approve AI-generated feedback/grades.	Teacher dashboard allows real-time monitoring of student prompts and AI responses.	Teachers can edit outputs but cannot monitor student-AI interactions.	AI acts autonomously; no mechanism for teacher oversight or approval.
Implementation & Pilot	Robust pilot program with clear success metrics and tiered scaling options.	Structured pilot phase available with dedicated technical support.	Offers a trial but lacks data exports to evaluate pilot effectiveness.	No support for a pilot phase; "all-or-nothing" licensing model.
Professional Development	Comprehensive PD plan included "train-the-trainer" resources and 24/7 support.	Live/recorded webinars available; intuitive UI with built-in tutorials.	Static PDF manuals provided, requires significant self-teaching.	No training provided; interface is unintuitive for staff.
LMS Interoperability	Full LTI integration with Canvas / Schoology; seamless data exchange and roster syncing.	Integrates via SSO (ClassLink / Clever); most features available within the LMS.	Mostly integrated, but some features require separate login or management.	Siloed application; operates independently of LMS; requires manual account creation.
Technical Performance	Scales perfectly without degradation; rapid and accurate responses; offline functionality available.	Occasional performance drops; works on most browsers and mobile devices.	High resource demands; limited mobile functionality or browser compatibility.	Consistent performance issues; poor scalability; requires specialized/latest equipment.
Environmental Impact	Designed for high energy efficiency; provider uses renewable energy and reports carbon footprint.	Reasonably energy-efficient; some sustainability info is available.	Provider has minimal sustainable practices; lacks optimization.	High operational power; no sustainability policy or efforts to mitigate carbon footprint.

IMPLEMENTATION CONSIDERATIONS CHECKLIST

The following items should be verified "YES" before a tool can be moved to the Pilot Phase.

- ☐ **Data Training:** Does the vendor contractually agree not to use District data (student or staff) to train their foundational models?
- ☐ **AI Transparency:** Does the tool clearly label AI-generated content to ensure users can distinguish it from human-generated work?
- ☐ **Error Reporting:** Is there a visible mechanism for users to flag "hallucinations," inaccuracies, or inappropriate content?
- ☐ **Right to Deletion:** Does the vendor provide a clear process for the LEA to request the permanent deletion of user data upon request?

SCORING AND DETERMINATION

Evaluators should total the points from Section 2 to determine the next steps:

24–28 Points (Exemplary): Approved for Pilot. High alignment with district goals.

18–23 Points (Proficient): Conditional Approval. Requires specific contract addendums or privacy tweaks.

Below 18 Points: Denied. Significant risks identified in privacy, equity, or instructional integrity.