



**AAB**

**AAB-STD-004**

---

## **AI Education Safety & Data Governance Guideline**

Version: Draft v0.1

Status: Public Draft for External Review

Classification: AAB Standards Series

Approval Authority: AAB Standards Council

Effective Date: 9/4/2025

Next Systematic Review: 12 months, or earlier when legal, safety, or platform conditions change



## Foreword

This document has been prepared as a draft AAB guideline for safety, privacy, human supervision, incident documentation, and public-release boundaries in AI education evidence records.

This document is not a final AAB standard, certification basis, endorsement, or public AAB position until approved through the applicable human review process.

### 1. Scope

This draft guideline defines minimum safety, privacy, supervision, and publication-boundary expectations for AI education activities documented or reviewed by AAB.

It applies to cases, pilots, resources, assessments, and policy records involving learners, educators, volunteers, community programs, schools, higher education, adult training, or workforce AI literacy.

### 2. Purpose

AI education records often involve minors, public learning spaces, third-party tools, generated content, student artifacts, and partner-submitted implementation notes. AAB needs a consistent safety and data governance vocabulary so that public records remain useful without exposing private, sensitive, or misleading information.

### 3. Core Requirements

#### 3.1 Human accountability

AI tools shall not be treated as responsible actors. Educational design, learner supervision, safety, interpretation, and local compliance remain human responsibilities.

#### 3.2 Data minimization

AAB records should preserve metadata, evidence summaries, and governance information rather than raw personally identifiable learner data.

#### 3.3 Age-appropriate supervision

Activities involving children or adolescents shall document the adult supervision model, facilitator role, and limits on learner autonomy.

#### 3.4 Tool role clarity

Records shall state whether AI is used as a co-creator, tutor, evaluator, simulator, automation tool, search tool, assistant, or infrastructure layer.



### 3.5 Output review

Activities shall define how AI outputs are reviewed, filtered, corrected, or discussed before being treated as educational material.

### 3.6 Privacy and consent boundary

Records shall describe whether personal data, images, voice, account logins, media, student work, analytics, or feedback are collected, retained, displayed, or published.

### 3.7 Incident and issue logging

Safeguard events, inappropriate outputs, technical issues affecting safety or integrity, major deviations, and unresolved concerns shall be documented.

### 3.8 Publication boundary

Public AAB pages shall display only approved metadata and summaries. Raw student data, private comments, unpublished artifacts, private partner notes, and internal review records shall remain outside public release unless expressly approved.

## 4. Minimum Safety Fields for AAB Records

Each relevant record should include:

- Learner age or grade band
- Setting
- AI tool or system type
- AI role
- Human supervision role
- Account/login requirement
- Personal data collection status
- Media/image/voice collection status
- Student artifact retention status
- Content moderation or output review approach
- Privacy safeguards
- Consent or notification notes, when applicable
- Known risks
- Observed incidents or issues
- Design adaptations
- Publication boundary
- Reviewer questions



## 5. Risk Tags

AAB should use controlled risk tags such as:

- Minor learner
- Personal data risk
- Media/image risk
- Voice input risk
- Open chat risk
- Hallucination risk
- Bias or representation risk
- Overreliance risk
- Academic integrity risk
- Emotional or psychological safety risk
- Inappropriate content risk
- Vendor or third-party platform risk
- Accessibility risk
- Language access risk
- Cybersecurity risk
- Unclear consent status
- Insufficient supervision

## 6. Safeguard Tags

AAB should use controlled safeguard tags such as:

- No individual logins
- No personal data collection
- No retained images
- No retained recordings
- Adult supervision
- Teacher or facilitator review
- Prompt scaffolds
- Pre-filtered prompts
- Content moderation
- Human editing phase
- Reflection discussion
- Local policy compliance
- Public-safe summary only
- De-identified record



- Reviewer-approved release

## 7. Special Rule for Minors

For records involving minors, AAB should default to:

- No unnecessary personally identifiable learner data
- No public names
- No public raw images or recordings
- No unreviewed public quotes
- No public student-level performance data
- No public partner files unless approved
- Clear adult supervision language
- Clear limits on AI autonomy

## 8. Alignment with Existing AAB Evidence

Visible AAB case pages already show important early patterns:

- Several activities avoid individual logins.
- Several activities state no personal data collection.
- Storytelling activities use prompt cards, sentence starters, facilitator framing, and human editing.
- Cases distinguish descriptive documentation from controlled assessment.
- Pilot records are intended to move through internal review, alignment, training, implementation, and final summary stages.

This guideline turns those recurring safeguards into a consistent review vocabulary.

## 9. Relationship to AAB-STD-003

The finalized AAB Pilot Framework already requires attention to privacy, consent, institutional review, safeguarding, neutrality, public release boundaries, conflict-of-interest disclosure, and incident documentation.

This guideline generalizes those safety expectations beyond pilots so they can also apply to cases, resources, assessments, frameworks, and policy records.

## 10. Open Review Questions

- Which risk tags should block public publication until resolved?
- Should AAB require a standard parent/guardian notification field for all minor-facing pilots?
- Should AAB publish tool names when risk review is incomplete?



## **AAB** AAB Standards Series

- How should AAB classify cases that use public generative AI tools but do not collect student accounts?
- What minimum safety fields are required for community-submitted cases?

### Revision History

Version	Date	Summary of Changes	Approved By
Draft v0.1	Draft	Initial working document generated from AAB evidence-layer drafts and ISO-style template.	AAB Standards Council