

Child Impact Assessment

Student Dropout Risk Predictor | Detailed child rights, safety, fairness, and governance assessment

International Children's Education Nonprofit | Simulated AI Governance Consulting Engagement

Portfolio note

All organizational details, AI systems, roles, risks, scores, and data in this portfolio artifact are fictional or synthetic and are provided solely to demonstrate applied AI governance consulting methodology.

Version 1.0 | August 2026

1. Executive Summary

This Child Impact Assessment evaluates the simulated BrightBridge Student Dropout Risk Predictor, a predictive model intended to flag students who may benefit from additional retention support. The assessment focuses on whether the system can support earlier assistance without creating stigmatization, unfair labeling, privacy harm, discriminatory intervention, inappropriate automation, or cross-country inequity.

Proposed governance decision

CONDITIONAL GO - Limited pilot may proceed only after country-level validation, documented human review, feature minimization, subgroup fairness testing, parent/student transparency appropriate to age and context, and quarterly monitoring. No automated denial of services, discipline, or punitive action may use the score.

2. System Profile

Attribute	Assessment Value
System ID	BB-AI-001
System	Student Dropout Risk Predictor
Purpose	Identify students who may benefit from additional retention support.
Users	Program staff and educators.
Affected Stakeholders	Students, families, educators, regional program teams.
Data	Attendance, grades, school engagement, household support indicators.
Model Type	Predictive classification.
Decision Role	Advisory only. Program professional must validate before intervention.
Deployment	Simulated multi-country pilot.
Risk Tier	High.

3. Intended Use and Prohibited Use

Intended Use

- Identify students for supportive outreach or resource review.
- Help program staff prioritize follow-up where resources are constrained.
- Support, not replace, educator judgment and local knowledge.

Prohibited Use

- Automatic denial or reduction of educational support.
- Disciplinary action, expulsion, or punitive labeling.
- Automated referral to child protection or law enforcement.
- Ranking children for worthiness, donor storytelling, or generalized social scoring.
- Use of the risk score outside the approved retention-support purpose.

4. Affected Child Groups

Child Group / Context	Why It Matters	Potential Impact
Age	Approx. school-age children and adolescents in BrightBridge programs.	Developmental stage affects comprehension, agency, and ability to challenge decisions.
Socioeconomic status	Students from low-income households may be overrepresented in risk indicators.	Proxy discrimination and stigmatization risk.
Disability	Students with disabilities may have irregular	Risk of false positives and exclusion.

Child Group / Context	Why It Matters	Potential Impact
	attendance or learning patterns.	
Language	Students operate across multiple languages.	Data quality, translation, and interpretability risk.
Gender	Different dropout patterns may affect model performance.	Subgroup error-rate differences.
Rural / remote communities	Attendance and access may be shaped by transport or infrastructure.	Contextual variables may be misread as disengagement.
Migration / displacement	Interrupted schooling may create non-standard patterns.	Risk of inappropriate inference from instability.
Country / region	School calendars, program design, data quality, and social context differ.	Pooled model may not generalize fairly.

5. Rights, Safety, and Well-Being Assessment

Assessment Domain	Key Question	Inherent Concern	Required Safeguard
Safety & Safeguarding	Could a false negative mean a child needing support is missed? Could a false positive trigger intrusive or stigmatizing intervention?	High	Require human validation; no safeguarding action solely from score; escalation through existing safeguarding procedures.
Privacy	Does the model use more household or behavioral data than necessary?	High	Minimize features; restrict access; define retention; prevent secondary use.
Fairness & Non-Discrimination	Do error rates or intervention rates differ materially across protected or vulnerable groups?	High	Pre-deployment subgroup testing; country validation; quarterly fairness monitoring.
Dignity & Stigma	Could being labeled high-risk change how adults treat a child?	High	Neutral language; restrict score visibility; emphasize support need rather than deficit label.
Participation & Voice	Can children, families, educators, and local teams influence implementation and raise concerns?	Moderate	Age-appropriate notice; feedback channel; local advisory review.
Accessibility & Inclusion	Does the system work for children with disabilities, low connectivity, or non-standard attendance?	High	Alternative pathways; accessibility review; context codes; no single-channel dependence.
Explainability	Can an educator understand why a student was flagged and challenge it?	High	Reason codes; feature explanations; override process.
Human Oversight	Does the reviewer have authority, time, context, and training to disagree with AI?	High	Named reviewer; documented decision; override rationale; periodic quality sampling.

6. Data Assessment

Data Element	Usefulness	Risk	Mitigation
Attendance	Potentially useful but context-sensitive	School closures, illness, disability, transport, caregiving, displacement may drive absence.	Validate local calendar/context; do not treat absence as intent.
Grades	May indicate academic difficulty	Different grading systems and teacher practices across countries.	Normalize locally; avoid cross-country direct comparison.
Engagement indicators	Potentially subjective	Teacher-recorded engagement may encode bias.	Limit subjective features; audit inter-rater consistency.
Household support indicators	High proxy risk	Can encode poverty, family structure, migration, or disability.	Strict necessity test; fairness analysis; consider exclusion of features.
Free-text notes	High sensitivity / bias risk	May include unverified opinions or safeguarding information.	Exclude from model unless separately approved; restrict sensitive narratives.

7. Fairness and Bias Evaluation Plan

- Compare false-positive and false-negative rates across gender, disability status where available/appropriate, language group, country/region, and socioeconomic proxy groups.
- Compare intervention selection rates, not only model accuracy.
- Test intersectional slices when sample sizes permit and avoid drawing conclusions from unstable small samples.
- Investigate feature contribution and proxy effects for household, location, attendance, and engagement variables.
- Do not deploy in a country where validation is insufficient or materially worse than approved thresholds.

Illustrative Fairness Thresholds for Pilot Governance

Metric	Illustrative Threshold	Action
Country-level sensitivity	>= 0.75 and within 10 percentage points of pooled baseline	Below threshold: local redesign or no deployment.
False-positive rate difference	<= 10 percentage points between material subgroups	Breach: investigate features and threshold.
Intervention selection-rate ratio	0.80-1.25 unless justified by evidence	Outside range: fairness review before continued use.
Unknown / missing subgroup rate	< 15% for metrics used in fairness conclusions	Higher missingness: do not claim subgroup fairness.

These thresholds are simulated portfolio assumptions, not universal standards. A real deployment would require organization-specific validation and expert review.

8. Human Oversight Design

1. Model flags a student as potentially needing support.
2. Authorized program professional reviews the reason codes and local context.
3. Reviewer checks for known contextual explanations, recent events, disability/access factors, and data quality issues.
4. Reviewer may accept, reject, or defer the recommendation and records rationale.
5. Any safeguarding concern follows the established safeguarding process; the dropout model does not make a safeguarding decision.
6. Intervention is supportive and proportionate; the student is not labeled or penalized based on the score.
7. Quarterly quality review samples decisions, overrides, and outcomes.

9. Child Impact Risk Table

Risk ID	Child Impact Risk	Inherent	Key Mitigation	Residual
CIA-01	Proxy discrimination from household variables	High	Feature minimization; fairness tests	Moderate
CIA-02	Country-level underperformance	High	Local validation; suspend where insufficient	Moderate
CIA-03	Stigmatization of flagged children	High	Neutral language; restricted access; training	Moderate
CIA-04	Overreliance by educators	High	Mandatory review/override; training	Moderate
CIA-05	False negative misses support need	High	Sensitivity threshold; monitoring; non-AI referral channels	Moderate
CIA-06	False positive creates intrusive follow-up	High	Human validation; proportional intervention	Moderate
CIA-07	Sensitive data secondary use	High	Purpose limitation; access control; retention	Low-Moderate
CIA-08	Accessibility-related false positives	High	Accessibility/context review; alternative indicators	Moderate
CIA-09	Poor explainability	Moderate	Reason codes; documented	Low

Risk ID	Child Impact Risk	Inherent	Key Mitigation	Residual
			rationale	
CIA-10	Data quality errors	High	Quality checks; correction workflow	Moderate
CIA-11	Student/family lack of transparency	Moderate	Age/context-appropriate notice; feedback channel	Low-Moderate
CIA-12	Model drift	High	Monthly performance; quarterly subgroup review	Moderate
CIA-13	Use for punitive action	Critical	Explicit prohibition; access control; audit	Low
CIA-14	Unauthorized cross-country data transfer	High	Data map; approved transfer mechanisms; minimization	Moderate
CIA-15	Vendor/tool change undermines safeguards	High	Change notification; reassessment trigger	Moderate

10. Consultation and Participation Plan

- Regional educators review whether model indicators reflect local school and community realities.
- Safeguarding specialists review any variable or workflow that may overlap with child-protection concerns.
- Inclusive education specialists review disability and accessibility implications.
- Age-appropriate student and family feedback is collected for pilot communication and intervention design where safe and appropriate.
- Community representatives are consulted on culturally sensitive features and potential stigma.

11. Decision and Conditions for Pilot

Condition	Required Before / During Pilot
1	Complete country-level validation before local use.
2	Remove or justify high-proxy-risk features through documented necessity and fairness review.
3	No automated intervention; human confirmation required for every flagged case.
4	No punitive, disciplinary, eligibility, or safeguarding decision may use the score as determinative input.
5	Use neutral support-oriented language; limit score visibility to authorized staff.
6	Provide appropriate transparency and a mechanism to correct underlying data.
7	Monitor performance monthly and subgroup fairness quarterly.
8	Reassess after material model/data changes or any serious child-safety incident.

Residual Risk Conclusion

Residual risk remains Moderate if all conditions are implemented and evidenced. Without documented country validation, human review, feature controls, and fairness monitoring, residual risk would remain High and the pilot should not proceed.

Portfolio deliverable result

The CIA demonstrates how BrightBridge would assess a child-affecting AI system through safety, rights, privacy, fairness, accessibility, human oversight, local context, mitigation, and a documented governance decision.