

Children's Privacy at Stake? Assessing Data Breach Risks In the Pratham - Anthropic AI collaboration under India's DPDP Act

The use of Artificial Intelligence (AI) in school education has generated debates and discussion that goes beyond the question of efficiency to the question of safety and rights of children. The risk of misuse of children's data is being acknowledged worldwide as an uncomfortable reality. A group of multiple UN bodies including UNCRC, UNESCO, UNICRI, Office of the United Nations High Commissioner for Human Rights, issued a joint statement highlighting the need for protection and privacy of data of children collected by the AI systems, tools and platforms¹. In 2025, a lawsuit has been filed in Texas, USA against edtech giant PowerSchool, a student information system, for a breach that allegedly exposed the sensitive information of more than 60 million students and 10 million teachers, sometimes including their Social Security Numbers². A pilot study conducted in India revealed that Indian educational institutions suffered more than two lakh cyberattacks and nearly four lakh data breaches in just nine months.³

This global and national context is relevant to the recently declared partnership between Pratham and Anthropic. Pratham is a non-government organization working in the field of education and Anthropic is a US-headquartered AI company that develops AI systems. The Pratham has partnered with Anthropic where Claude (a large language model) accesses children's handwritten responses, academic performance data, and personalised feedback for children. The Anytime Testing Machine (ATM) is the first product of the Pratham–Anthropic partnership announced in February 2026. Powered by Anthropic's Claude, the ATM generates curriculum-aligned tests; digitises handwritten student answers; applies rubric-based grading using Claude; and delivers bilingual (Hindi-English), personalised feedback.⁴ In light of India's DPDP Act the assessment model may raise certain risks.

The *first* risk relates to verifiable parental consent. Section 9 (1) of the DPDP Act mandates that a Data Fiduciary must obtain verifiable consent of the parent or lawful guardian before processing any personal data of a child (defined as any individual under 18 years). The Draft DPDP Rules, 2025 (Rule 10) provide further guidance on mechanisms for obtaining such consent, including OTP-based parental consent and integration with government-issued IDs. However, in the case of ATM using data of children, parents may not fully comprehend that their child's handwritten work is being photographed, uploaded to a cloud-based AI system, processed by a US company's servers, and analysed by a large language model. For the

¹ International Telecommunication Union (2025). *Joint Statement on Artificial Intelligence and the Rights of the Child*.
https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-CYB_JOINT-2025-PDF-E.pdf

² https://publicinterestprivacy.org/ai-changes-everything-data-minimization/?utm_source=chatgpt.com

³ <https://ciso.economictimes.indiatimes.com/news/data-breaches/over-2-lakh-cyberattacks-4-lakh-data-breaches-hit-indian-educational-institutions-in-9-months/123294482>

⁴ <https://claude.com/customers/pratham-international>

consent to be valid, it needs to be 'informed' that therefore must include clear, understandable information about how data will be used. The parents and guardians in India may not have adequate knowledge or practical ability to verify US-based Anthropic's data retention policies, procedure of withdrawal of consent, or exercise rights of the data principal given under the DPDP Act.

Moreover, taking the above point forward, when children's data is processed by an international organization, the enforcement of Indian children's rights becomes complex. The parents in India may not have access to the laws, authorities and courts of another jurisdiction. This is an important reason to avoid routing children's data through foreign private platforms.

Secondly, Section 9 (3) states that a data fiduciary shall not undertake tracking or behavioural monitoring of children or targeted advertising directed at children. The aim of developing the ATM is to give personalised feedback to children based on their performance. To achieve this, the system would rely on repeated assessments and examining the responses of children over a period of time.

Further, the DPDP Act under section 8 (1) entrusts the Data Fiduciary to be responsible for complying with the provisions of this Act and the rules made thereunder in respect of any processing undertaken by it or on its behalf by a Data Processor. This is irrespective of any agreement to the contrary or failure of a Data Principal to carry out the duties provided under this Act. The Digital Personal Data Protection Rules, 2025, provide an 18-month, phased window for full compliance, culminating in May 2027. During this period, all Data Fiduciaries processing children's data are expected to put in place robust mechanism for gaining consent, security, breach-notification and grievance-redress mechanisms, failing which they may face penalties. In this case since Pratham collects data during the pilot/survey, any personal data breach of any kind would attract liability for Pratham under Section 8, including potential penalties for inadequate safeguards. This is applicable to all other data collection exercises conducted by Pratham with or without collaboration with government departments.

The above provisions align with the guiding principles of the wider child protection system in the country. The Juvenile Justice (Care and Protection of Children) Act, 2015 lays down general principles that must guide all authorities and agencies dealing with children, including the *principle of best interest* which states that *all decisions regarding the child shall be based on the primary consideration that they are in the best interest of the child and to help the child to develop full potential*. It also re-enforces *principle of right to privacy and confidentiality* by all means and throughout the judicial process.

Since, Anthropic is a US-based company, it has to comply with the US laws and has to provide data whenever is required by the US authorities. Giving access of important data related to

children to another country does not align with the fundamental principle of protecting sovereignty and integrity of India. Therefore, here the question is not about the 80% accuracy of the ATM system, but whether it is acceptable for children's educational progress be continuously tracked with the help of a system developed by an international private organization over which India's institutions and parents have limited control.

It is important to highlight that recently, Anthropic has been in news indicating issues between the US Department of Defense and Anthropic⁵. This point towards the concern that Pratham has tied up with a company who is in conflict with the interests of its own government. In another incident, as reported, Anthropic agreed to pay \$1.5bn (£1.11bn) to settle a class action lawsuit filed by authors who said the company stole their work to train its AI models. The company holds more than seven million pirated books in a central library, according to Judge Alsup's June decision, and faced up to \$150,000 in damages per copyrighted work⁶. One of Anthropic's latest AI models, Claude Opus 4, drew attention not just for its coding skills, but also for its ability to scheme, deceive and attempt to blackmail humans when faced with shutdown⁷. Reportedly, the representatives of Anthropic are summoned by a district Court after a case has been filed by Anthropic Software Private Limited, a local IT company that has accused the American company of using the same name.⁸

Another key point is that constructing learner profiles would be unavoidable to achieve Pratham's three-year goal of evolving the ATM into a "learning and credentialing engine that recognizes competencies gained through non-linear pathways". It needs to be ensured that neither Pratham nor Anthropic exploits children's data collected through ATM for commercial purposes in any manner. It has been argued that AI systems in education undermine multiple rights enshrined in the UNCRC and that children's data exploited by commercial AI developers effectively amounts to children "working as unpaid labour".⁹

Most importantly, Pratham is one of India's largest education NGOs. Over the years, Pratham has conducted numerous programmes such as Annual Status of Education Report (ASER) on learning outcome of children. The report influences how governments understand learning outcomes and design school-education policies at national and state levels. Pratham's collaboration or dependence on a foreign AI company may influence India's future policy decisions around children. However, its collaboration with Anthropic has larger implications. This collaboration may make way for very large volume of data, collected over years, being

⁵ <https://timesofindia.indiatimes.com/technology/tech-news/pentagon-may-designate-anthropic-as-supply-chain-risk-what-this-means-for-the-company-its-customers-and-partners/articleshow/128465866.cms>

⁶ <https://www.bbc.com/news/articles/c5y4jpg922qo>

⁷ <https://www.axios.com/2025/05/23/anthropic-ai-deception-risk>

⁸ <https://www.thehindu.com/news/national/karnataka/anthropic-usa-representatives-fail-to-appear-in-belagavi-court-in-brand-name-case/article70638378.ece>

⁹ Holmes W (2025) AI, education, and children's rights. *Front. Educ.* 10:1656736. doi: 10.3389/feduc.2025.1656736.
<https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1656736/full>

shared with the foreign private company. The huge dataset may then be used indifferent ways, potentially for behaviour analysis and profiling of children, to shape the government policies or to draw commercial benefits out of it. If any data is to be collected or analysed, it has to be done after following all the due procedures as per the DPDP Act.

Similarly, Anthropic is collaborating with Central Square Foundation for education of children from underserved communities. As part of this collaboration, Anthropic will provide technical expertise, mentorship, and API credits to organizations developing AI-enabled tools, including personalized tutors, teacher coaching solutions, and assessment-driven instruction, with the goal of reaching more primary school students across India.

To sum-up, these concerns suggest that given the frequent cybercrimes and instances data breaches worldwide, we must be careful in giving access of children's data. It is also important to consider whether such collaborations align with the Hon'ble Prime Minister's vision to build an AI future that integrates human value such that technology and human trust move together. Further, whether the use of AI in assessment of children, primarily the 'underserved' could be in violation with the RTE Act, 2009 and vision of National Education Policy (NEP), 2020. NEP emphasises holistic, formative, and competency-based assessment, grounded in child-centric pedagogy and teacher-led evaluation processes. Contrary to the provision made under the Section 29 (1) of the Act, the evaluation procedure would then no longer solely be determined by the 'academic authority' notified by the government. The evaluation procedure would be partly determined by which assessment AI model are the schools using. Also, assessment, feedback and instruction, are integral part of teachers pedagogic duties. So, it may be inferred that the fundamental benefit of using AI lies in reducing teachers' workload. From a rights-based perspective, the questions become whether it is necessary to expose children's personal data to potential risk of violating their fundamental rights, and who are the ultimate beneficiaries of such collaboration? Apparently not the children.
