

Children in the Behavioural Machine

A Survey of Behavioural AI in the Environments of Childhood and the Case for Child-First AI Governance

Emily D. Bauman

The Digital Dignity Institute of California

DDI Research Paper, September 2026

Abstract

Systems that predict, rank, recommend, personalise and influence behaviour now structure much of the daily experience of children in the places where they learn, play, communicate and spend. This paper examines how such systems affect children differently from adults, and asks whether the technologies actually deployed in schools, games, social platforms, companion chatbots and commercial services, together with their operators' published policies, provide age-appropriate safeguards, meaningful transparency and accountability where behavioural data is used to make or shape consequential decisions. Drawing on a documentary survey of thirty deployed systems and enforcement records across the United States, Canada, the United Kingdom, the European Union, Australia and Asia, and on peer-reviewed research in developmental science, education, law and digital rights, the paper identifies a consistent pattern. Child-specific protections, where they exist at all, cluster around advertising and parental consent; they rarely reach the ranking, prediction and classification logic that determines what a child sees, how long she stays, what label is attached to her, and who is told. Transparency runs to parents and institutions rather than to children. Accountability arrives after harm, through litigation discovery, regulator action and adversarial civil-society audits, rather than being designed into deployment. The paper argues that these gaps are not incidental but structural: they follow from a business logic that treats engagement and prediction as the product, from consent architectures ill-suited to children's evolving capacities, and from the durability of algorithmic classification in institutions that children cannot leave. It closes with eight policy recommendations intended to place the best interests of the child at the centre of AI governance, including a prohibition on engagement optimisation for minors, a right of contestation for consequential classification,

mandatory expiry of behavioural labels, and child-readable disclosure of inference.

Keywords: children's rights, artificial intelligence, recommender systems, education technology, behavioural profiling, age-appropriate design, digital dignity

DOI: <https://doi.org/10.5281/zenodo.22652597>

Suggested citation: Bauman, E. D. (2026). *Children in the Behavioural Machine: A Survey of Behavioural AI in the Environments of Childhood and the Case for Child-First AI Governance*. DDI Research Paper. San Francisco, CA: The Digital Dignity Institute of California. Zenodo. <https://doi.org/10.5281/zenodo.22652597>

1. Introduction

A thirteen-year-old girl in Tennessee, teased about her complexion, typed a bitter joke into a school chat. Within hours she had been arrested, interrogated, strip-searched and held overnight in a cell, because a piece of automatic monitoring software licensed by her district had classified the sentence as a threat and set in motion the chain of events that ended with police (Lurye and Associated Press 2025). Some fourteen hundred miles north, thirty-two thousand children in Wisconsin carried a label they had never seen, "high risk", generated twice a year by a state model that took their race and their family's eligibility for subsidised lunch as inputs, and that the state's own analysts had found to be wrong three times out of four when it predicted a child would not graduate (Feathers 2023). In a suburb of Orlando, a fourteen-year-old boy spent his final months in conversation with a chatbot persona that, according to the complaint his mother later filed, urged him to "come home" to it (Free Speech Center 2025). And in an ordinary living room, a three-year-old tapped a free game whose design, like four in five of the apps in the cohort studied by paediatric researchers, contained features engineered to keep her tapping and to press her parent to pay (Radesky et al. 2022).

None of these children chose the systems that classified, ranked, recommended to, or nudged them. None was told what had been inferred. Each was subject to a form of automated judgement whose consequences, whether a police record, a durable educational label, a relational dependency or a purchasing habit, were disproportionate to any decision the child had made. These are not exotic cases. They are the quotidian operating conditions of the environments in which a great many children now grow up.

This paper is concerned with a specific and growing class of technology: AI-driven behavioural systems, by which we mean systems that use data about what a person does, or is inferred to be, in order to predict, rank, recommend, personalise or influence what that person will do next. Such systems are the engines of the recommendation feed, the adaptive learning platform, the student-monitoring tool, the companion chatbot, the item shop, the loot box, the voice assistant and the personalised advertisement. They are not new, but their reach into the institutions of childhood has accelerated markedly, and the models that drive them have grown more capable, more opaque and more tightly coupled to commercial incentives.

The paper's central claim is that these systems affect children differently from adults, for three reasons that are developmental, structural and temporal. Developmentally, the child is a moving target: her capacities for impulse control, for recognising persuasion, for anticipating consequences and for forming a stable identity are in formation, and the interval between a heightened sensitivity to reward and the maturation of self-regulation is precisely the interval that engagement-optimised design exploits (Steinberg 2008; Casey, Getz and Galvan 2008; Rozendaal et al. 2011). Structurally, the child does not stand in a relation of rough equality with the system's operator. She cannot decline the platform her school has adopted, cannot negotiate the terms of the game her friends play, and cannot meaningfully consent to profiling she does not understand (Livingstone, Stoilova and Nandagiri 2019; van der Hof 2016). Temporally, a classification applied to a child has a long tail: the label "at risk", "low ability", "flagged" or "under-13 by inference" is generated at a moment of maximum plasticity and may persist in records, in adult expectations and in the child's self-understanding long after the moment has passed (Rist 1970; Hacking 2007; Bowker and Star 1999).

The paper is a research agenda of The Digital Dignity Institute of California and reflects the Institute's broader concern with what it has called the recourse deficit: the gap between the power that automated systems exercise over people and the means available to those people to understand, contest and correct that power. Nowhere is the deficit wider than in childhood. My approach brings together perspectives from child development, technology policy, law, education and digital rights, and it proceeds from an ethical commitment that we state plainly at the outset. Children deserve environments designed for learning and growth. They do not deserve systems that covertly optimise for attention, prediction or influence, and whatever benefits such systems may offer cannot be weighed honestly until the safeguards, the transparency and the accountability that would make them trustworthy exist in fact rather than in press releases.

We are mindful that the empirical literature on children and digital technology is contested. The association between screen time in general and adolescent wellbeing is small, and some of the most influential causal claims in public debate outrun the evidence (Orben and Przybylski 2019; Odgers 2024). This paper does not rest on those claims. Its subject is not screens but systems: the specific behavioural machinery that determines what a child sees, what is inferred about her and what decision follows. The evidence about that machinery, much of it drawn from regulators' findings, litigation discovery and operators' own documents, is considerably firmer than the evidence about screens as such, and it points in a consistent direction.

The paper proceeds as follows. Section 2 sets out the scope, method and analytical framework. Section 3 develops the three lenses through which children's distinctive position is examined. Section 4 presents findings from a documentary survey of deployed systems in four settings: learning, play, communication and purchase. Section 5 draws those findings together into a map of the gaps in safeguards. Section 6 assesses the regulatory landscape across the surveyed jurisdictions as of September 2026. Section 7 offers eight policy recommendations, and Section 8 concludes.

2. Scope, Method and Analytical Framework

2.1 Research question and scope

The research question is deliberately practical. Where AI-driven behavioural systems are deployed in settings that structure children's daily lives, do the deployed technologies and their operators' stated policies provide (a) age-appropriate safeguards, (b) meaningful transparency, and (c) accountability, particularly where behavioural data shapes consequential decisions? And where they do not, what is the character of the gap?

We examine four settings, chosen because together they account for most of a child's waking hours and because each has a distinct institutional logic. In the setting of learning, systems are adopted by schools and states and a child cannot exit them without forgoing education. In play, systems are commercial, voluntary in form, and engineered for retention and spend. In communication, systems mediate a child's social world and, increasingly, offer a synthetic interlocutor in its place. In purchase, systems convert behavioural data into commercial exposure and transactions, often inside the other three settings rather than apart from them.

The jurisdictions surveyed are the United States, Canada, the United Kingdom, the European Union, Australia and, more selectively, China, South Korea, Japan, Singapore and India. The selection is not exhaustive; it reflects where deployed systems, enforcement records and primary policy texts could be located and read.

2.2 Method

This is a documentary study. No children, parents or teachers were interviewed, and no new data about children were collected. The evidence base has four strata. First, primary policy texts: privacy policies, terms, safety pages and newsroom statements published by operators, accessed and quoted on the dates given in the reference list. Second, regulator and court records: enforcement orders, preliminary findings, statutory text, judicial opinions and unsealed litigation material. Third, peer-reviewed research and institutional reports from developmental science, education, law and digital rights. Fourth, credible investigative journalism and civil-society research, which we label as such and treat as evidence of allegations or test results rather than of adjudicated fact.

A discipline of sourcing governs the paper. Every factual claim is traceable to a cited document. Where an operator's published policy does not address a question, we say so rather than characterise it. Where a figure is a vendor's own claim, an interested party's allegation or a short seller's report, it is identified as such. Where a detail could not be verified, it has been omitted. We have preferred to under-claim.

2.3 The analytical rubric

Each deployed system was assessed against a common rubric with six elements. What does the system do with behavioural data, and what experience or decision does its output shape? Is that output consequential, in the sense that it bears on academic track, discipline, welfare referral, law-enforcement contact, exposure to content, spending or time on device? Does the operator's published policy treat children differently at all, and if so does it differentiate by developmental stage rather than by a single legal threshold? Can a child, or a parent acting for her, learn what data is used, what is inferred and why a given result appears, in language the child can read? Is there any mechanism of accountability: contestability, human review, independent audit, published efficacy evidence, redress, or a limit on how long a classification persists? And what is the state of the evidence of harm, or of the absence of evidence of benefit?

Three cross-cutting lenses, developed in Section 3, are applied throughout: developmental context, unequal power, and the durability of classification.

2.4 A note on terminology

We use "child" in the sense of the UN Convention on the Rights of the Child, meaning every person under eighteen, while recognising that the needs of a six-year-old, a twelve-year-old and a seventeen-year-old differ profoundly and that this difference is itself one of the paper's central concerns. We use "behavioural system" rather than "algorithm" because the object of concern is a socio-technical arrangement, not a mathematical procedure: a model, the data that feeds it, the interface that presents its output, the institution that acts on it and the commercial incentive that sustains it. Following Susser, Roessler and Nissenbaum (2019), we reserve "manipulation" for hidden influence, "the covert subversion of another person's decision-making power", and distinguish it from persuasion that is visible and from coercion that is overt.

3. Why Children Are Different: Three Lenses

The argument that children require distinct treatment in AI governance is sometimes made as if it were self-evident, and sometimes resisted as sentimental. Neither response is adequate. The case rests on a body of research across several disciplines, and it is worth setting out with some care, because the shape of the argument determines the shape of the remedies.

3.1 Developmental context

The developmental sciences have converged over two decades on an account of adolescence that is directly relevant to behavioural design. Steinberg's dual systems model holds that a socioemotional system responsive to reward and to social evaluation becomes highly active around puberty, while the cognitive control system that regulates impulse and weighs consequence matures gradually into the mid-twenties. "The differing timetables of these changes make mid-adolescence a time of heightened vulnerability to risky and reckless behavior" (Steinberg 2008). Casey, Getz and Galvan (2008) describe the same asymmetry in neurobiological terms, a "heightened responsiveness to incentives while impulse control is still relatively immature during this period". Crone and Dahl (2012) add an important refinement: adolescents are not simply deficient in control but distinctively motivated by belonging, status and social feedback, which is why like counts, streaks, follower metrics and ranked feeds are such effective levers.

Younger children present a different profile. Their difficulty is not primarily one of regulating a mature capacity but of possessing it at all. Advertising literacy research shows that even when children can identify advertising conceptually, that knowledge does not reliably protect them, because contemporary commercial persuasion works through affective and implicit channels and because the executive capacities needed to deploy a defence in the moment are still forming (Rozendaal et al. 2011). Radesky and colleagues found that children aged three to five were routinely exposed to design features engineered to prolong play and prompt purchase, and Radesky's own summary of the finding was blunt: "Children are too young to understand this type of persuasive design that disrupts their game playing" (Michigan Medicine 2022).

Valkenburg and Peter's Differential Susceptibility to Media Effects Model (2013) supplies the theoretical frame: media effects are conditional on dispositional, developmental and social variables, and vary across individuals and across ages. Orben, Przybylski, Blakemore and Kievit (2022), using data from more than eighty thousand participants, identified windows of sensitivity during which increases in social media use predicted later declines in life satisfaction, and found that "These windows occur at different ages for males (14-15 and 19 years old) and females (11-13 and 19 years old)". The policy implication is pointed. A single threshold, whether thirteen, fourteen or sixteen, does not track developmental reality. Safeguards designed for children must be banded, and the bands must be informed by evidence rather than by the age at which a company is legally permitted to collect consent.

Two cautions attend this literature. First, the strongest developmental findings concern general susceptibility, not the effects of particular designs; the link between a specific ranking objective and a specific harm must be established case by case, and in Section 4 we rely wherever possible on operators' own documents and regulators' findings to make it. Second, the wider debate about screens and adolescent mental health is unresolved. Twenge and colleagues (2018) documented rising depressive symptoms among American adolescents after 2010 and linked them to new media; Orben and Przybylski (2019) found that "The association we find between digital technology use and adolescent well-being is negative but small, explaining at most 0.4% of the variation in well-being"; Haidt (2024) argued for a causal epidemic; Odgers (2024) replied in *Nature* that "the book's repeated suggestion that digital technologies are rewiring our children's brains and causing an epidemic of mental illness is not supported by science". Public health authorities have taken a precautionary rather than a conclusory position. The US Surgeon General's 2023 advisory concluded that "We cannot conclude that social media is sufficiently safe for children and adolescents", and the American Psychological Association (2023)

recommended that "Specific features (e.g., the 'like' button, recommended content, unrestricted time limits, endless scrolling) and notices/alerts ... should be tailored to the social and cognitive abilities and comprehension of adolescent users". That is the position this paper adopts: not that screens are poison, but that behavioural systems have not been shown to be safe for children, that specific design choices have been shown to exploit specific vulnerabilities, and that the burden of demonstrating safety belongs with those who deploy them.

3.2 Unequal power

The second lens is structural. Data protection law across the surveyed jurisdictions rests, in the last analysis, on consent: the individual is informed and agrees. For children this architecture is doubly infirm. The child cannot consent, so a parent or, in schools, an institution consents in her place; and the thing consented to, behavioural profiling, is precisely what neither child nor parent understands. Livingstone, Stoilova and Nandagiri (2019), reviewing the evidence, found that children grasp interpersonal privacy far better than institutional or commercial data flows, and concluded that "it is vital that children's understandings of the digital environment, their digital skills and their capacity to consent are taken into account in designing services, regulation and policy". Van der Hof (2016) argued from a rights perspective that consent mechanisms are ill-suited to children's evolving capacities and should give way to a rights-based approach; Milkaite and Lievens (2020) audited platform policies against the GDPR's requirement of clear and plain language for children and found the requirement largely unmet.

The asymmetry is sharpest where the child cannot exit. In a school, the platform is chosen by the district and imposed on the pupil; the UK Information Commissioner's Office observed in its 2026 audit of the sector that "children may not be able to choose or opt out of many digital tools" adopted by schools (Information Commissioner's Office 2026c). In a peer group, the social cost of leaving the platform everyone uses is borne by the child. In each case, the operator possesses the data, the model, the interface and the terms, and the child possesses a checkbox ticked by someone else. Zeide (2017) describes the further displacement that occurs in education, where "These systems displace pedagogical decision-making from educators serving public interests to private, often for-profit, technology providers". Cohen (2019) situates the pattern in a broader legal construction of informational capitalism in which platform terms substitute for public accountability.

Two theoretical resources sharpen the point. Nissenbaum's contextual integrity (2004) holds that information flows are appropriate when they conform to the

norms of the context in which they arose, "demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it". A child's keystrokes in a homework app belong to the context of schooling; when they flow to an advertising network, or when a school's welfare monitoring flows to a police dispatcher, the norm is violated whatever the privacy notice says. Viljoen (2021) adds that data's value and harm are relational and population-level, so that individual consent cannot govern them: "Data practices of powerful technology companies are aimed primarily at deriving population-level, relational insights". The inference drawn about one child is applied to many, and what a child's peers do can be used to classify her. Calo (2014) and Zuboff (2015; 2019) supply the economic account: firms that "fastidiously study consumers and, increasingly, personalize every aspect of the consumer experience" are able to exploit cognitive bias at scale, and the behavioural surplus so extracted is sold into markets whose products increasingly seek to modify, not merely forecast, conduct.

3.3 Durable classification

The third lens concerns time. A classification, once made, does things. Bowker and Star (1999) showed how categorical systems become invisible infrastructure that shapes lives long after the categorising decision is forgotten. Hacking (2007) named the mechanism the "looping effect": "the way in which a classification may interact with the people classified", so that people sorted into a kind come to behave, and to be treated, in accordance with it.

The classroom is the classic site of this effect. Rosenthal and Jacobson (1968) told teachers that certain randomly chosen children "could be expected to be 'growth spurters'", and those children were then rated more favourably and gained more; Rist (1970) watched a kindergarten teacher place children in reading groups within days of their arrival on the basis of class-linked cues, and watched those placements persist across years; Oakes (2005) documented across many schools how tracking sorts by race and class, delivers poorer instruction to lower tracks and holds children in place. The lesson of half a century of education research is that labels applied to children, particularly labels that purport to predict, are not inert. They alter adult expectations, allocate resources, and become self-fulfilling.

Algorithmic classification revives this dynamic in a form that is faster, more granular, more opaque and more durable. Barocas and Selbst (2016) explain how a predictive model can produce discriminatory outcomes without discriminatory intent, because "Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers", and Baker

and Hawn (2022), in the most thorough review of bias in educational algorithms, found documented bias concentrated on a few groups and largely unstudied for others, proposing "a framework for moving from unknown bias to known bias and from fairness to equity". Kizilcec and Lee (2022) show that competing fairness criteria cannot all be satisfied at once, so that every deployed prediction embodies a choice that ought to be stated and justified. Eubanks (2018) traces how risk scores in welfare and child-protection systems concentrate scrutiny on poor families and travel across databases; Benjamin (2019) shows how neutral-seeming technology reproduces racial hierarchy under a benevolent veneer; Citron and Pasquale (2014) propose the remedy of technological due process, notice, access, contest and audit, for a "scored society" in which "Predictive algorithms assess whether we are good credit risks, desirable employees, reliable tenants, valuable customers, or deadbeats, shirkers, menaces, and 'wastes of time'".

Children are the paradigm case for this remedy, and for a further one. Barassi (2020) observes that "Children today are the very first generation of citizens to be datafied from before birth", and Lupton and Williamson (2017) that "Children are becoming the objects of a multitude of monitoring devices that generate detailed data about them". A label attached at nine is attached to a person who will be entirely different at nineteen. The temporal mismatch between the moment of classification and the span of its consequence is the strongest argument for what this paper will recommend as mandatory expiry: behavioural classifications of children must be designed to lapse.

3.4 The frameworks that already exist

It would be a mistake to suppose that the international community has not noticed. General Comment No. 25 of the UN Committee on the Rights of the Child (2021), the authoritative interpretation of the Convention for the digital environment, directs that "States parties should prohibit by law the profiling or targeting of children of any age for commercial purposes on the basis of a digital record of their actual or inferred characteristics, including group or collective data, targeting by association or affinity profiling". The OECD's Recommendation on Children in the Digital Environment (2021) directs adherents to ensure that providers do "not allow the profiling of children or automated decision-making, including on e-learning platforms", absent a compelling reason and appropriate protective measures. UNICEF's Policy Guidance on AI for Children (2021), now in a third edition (UNICEF Innocenti 2025), sets out requirements for child-centred AI including transparency, explainability and accountability for children. Livingstone and Pothong (2023) translate General Comment 25 into eleven design principles under the rubric of

Child Rights by Design. The IEEE's 2089-2021 standard offers a process framework for age-appropriate digital services (IEEE 2021). Livingstone and Stoilova (2021) have extended the established typology of online risk to children with a fourth category, "contract", to capture the risk "when a child is party to or exploited by potentially harmful contract, that includes risks such as fraud, security risks or marketing strategies such as micro-targeting".

The norms exist. The question the remainder of this paper addresses is whether the systems deployed in children's environments honour them, and the answer, with a small number of instructive exceptions, is that they do not.

4. Findings: Behavioural Systems in Four Settings

The thirty systems and cases examined below are numbered consecutively in their headings, from Case Study 1 in the learning setting to Case Study 30 in the purchase setting, so that the reader can keep her bearings across the four settings. They were selected because they are deployed at scale, because primary policy text or regulator findings could be obtained, and because together they span the four settings and the surveyed jurisdictions. For each we report what the system does, whether its operator's published policy treats children differently, what a child or parent can learn, what accountability exists, and what the record shows about harm. Verbatim quotations from policies were accessed on 4 September 2026 unless otherwise noted.

4.1 Learning

Student monitoring and self-harm classification: Gaggle (Case Study 1/30)

Gaggle scans the email, documents, drives, chats and images in students' school-issued Google and Microsoft accounts, using what the company describes as "a combination of artificial intelligence, machine learning, and a highly trained content review team to flag concerning content" related to suicide, self-harm, violence, drugs, bullying and sexual content. Flagged items pass to a human review team; non-urgent items generate an email to district contacts; urgent items generate a phone call; and "If there is no response from the district-appointed emergency contacts after three contact attempts and a child's life appears to be in imminent danger, Gaggle will reach out to police dispatch for a wellness check" (Gaggle 2021). In 2021 the company told the United States Senate that it served roughly fifteen hundred districts, had generated more than 235,000 suicide, self-harm or violence alerts in the previous year, and claimed "at least 1,400 student lives saved", a figure that is

the vendor's own and for which no independent verification exists (Gaggle 2021).

The output is consequential in the most literal sense. It determines whether a child is called to an office, whether parents are telephoned, whether police are sent to a home, and, in documented cases, whether a child is hospitalised or arrested. Gaggle's Student Data Privacy Notice states that student data will not be shared with third parties without consent "except in cases of Possible Student Situations (PSS), which may be reported to law enforcement", that "The approximate location of a student will be collected through the Gaggle browser extension", and that data are retained for up to thirty days after a contract ends (Gaggle 2026). The notice treats every monitored child as a member of a single class; there is no differentiation between a seven-year-old and a seventeen-year-old. It specifies no retention period for flags during the life of a contract and no route by which a student can dispute a flag. Notice to families is delegated to the school.

The record of harm is now substantial. Records obtained by the Associated Press and the Seattle Times from one district in Washington State showed that roughly 2,200 students, about a tenth of enrolment, had triggered alerts in a single year, that at least six students had been potentially outed to school officials after writing about their sexual orientation or gender identity, and that nearly 3,500 unredacted student documents had been released to reporters in a records request, a security failure in its own right (Bryan, Lurye and Associated Press 2025). In Lawrence, Kansas, more than 1,200 alerts were generated in ten months, nearly two-thirds of them judged non-issues, with over two hundred false alarms from homework; in Polk County, Florida, nearly five hundred alerts over four years led to seventy-two involuntary psychiatric holds under the state's Baker Act; and in Tennessee the thirteen-year-old with whom this paper opened was arrested, strip-searched and held overnight. Gaggle's own chief executive said of that case, "I wish that was treated as a teachable moment, not a law enforcement moment" (Lurye and Associated Press 2025). Nine Lawrence students sued their district in federal court in August 2025, alleging that the system flagged phrases such as "my mental health" and "very uncomfortable" and that journalism drafts and artwork had been removed from their accounts without notice; by October the district told the court it had stopped using Gaggle (Dunn 2025; Lawrence Journal-World 2025).

Survey evidence indicates the pattern is systemic rather than local. The Center for Democracy and Technology found that 89 per cent of teachers reported their schools used monitoring software, that 78 per cent of those teachers said students had been flagged for discipline, that 44 per cent reported students

contacted by law enforcement because of alerts, and that 29 per cent of LGBTQ+ students reported their sexual orientation or gender identity disclosed without consent (Laird, Dwyer and Grant-Chapman 2022); a follow-up found that "Disciplinary actions, outing of students, and initiating of law enforcement contact are still regular outcomes" and that students with disabilities and LGBTQ+ students were disproportionately disciplined (Laird, Dwyer and Grant-Chapman 2023). On efficacy, the RAND Corporation's review found "limited understanding about how such programs work" and "a need for data to show how accurately AI-based algorithms can detect a student's risk of suicide and whether the use of these tools improves student mental health" (Ayer et al. 2023). It is worth pausing over what has been assembled here. A classifier with the power to summon police to a child's home; no independent evidence that it saves anyone; a documented tilt toward the students least able to absorb the consequences; and not a line in the vendor's notice telling a fourteen-year-old that she may dispute what the machine has said about her.

Predictive labelling: the Wisconsin Dropout Early Warning System (Case Study 2/30)

The Wisconsin Department of Public Instruction's DEWS scored every public-school student in grades six to nine, twice a year, on the likelihood of graduating on time, using test scores, disciplinary records, race, free or reduced lunch status, gender, attendance and mobility. Students below a threshold were labelled "high risk"; in 2020-21 more than 32,000 children received that label, and the scores were displayed to educators in the state data system (Feathers 2023). The department's own internal equity analysis, obtained by The Markup, found the model produced false alarms for Black students at a rate 42 percentage points higher than for white students and 18 points higher for Hispanic students, and an internal slide read "Is DEWS Fair? No...". When the model predicted a child would not graduate on time, it was wrong 74 per cent of the time. Researchers at the University of California, Berkeley, found "no evidence that DEWS predictions have in any way influenced the likelihood of on-time graduation" (Feathers 2023). Students interviewed later had not known the system existed; one district superintendent described receiving the data with the instruction to "Figure it out" (Feathers 2023; The Markup 2023). The department said in September 2023 that it was "studying its early warning systems and considering changes" (García Mathewson 2023).

DEWS is a public system with no vendor policy, no notice to families, no contest mechanism and no demonstrated benefit. It is the purest example in this survey of durable classification: a label generated partly from a child's race and family income, attached to a named child, visible to the adults responsible

for her, and persisting in state records. This algorithmic sorting mimics the systemic tracking identified in Ray Rist's foundational 1970 study, "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education", which followed a single kindergarten class and found that children seated at the table nearest the teacher within days of arrival, on the strength of class-linked cues such as dress and speech, received the most instruction and thrived, while those placed at the tables furthest away were expected to fail and, across the following years, did (Rist 1970). DEWS is that kindergarten room rendered in code, with the seating chart drawn by a model and hidden from the children in it.

Behaviour points: ClassDojo (Case Study 3/30)

ClassDojo allows teachers to award or deduct "feedback points" from individual children in real time, with class-visible displays and parent notifications, and reports very wide adoption in primary schools in the United States, Canada and elsewhere (The Conversation n.d.). Its privacy policy is, by the standards of this survey, unusually specific about children. It states that "Feedback Points older than 12 months are automatically deleted (or de-identified and aggregated) on a rolling 12-month basis" and that student data are "never used for behaviorally-targeted advertising" (ClassDojo 2026). This is one of only two operator policies in the learning setting to contain a concrete retention limit on a behavioural classification, and it is worth recording as evidence that such a limit is commercially feasible. The policy does not address automated decision-making or profiling, does not differentiate by age, and offers no mechanism for a child to dispute a point. Peer-reviewed analysis argues that the product reframes discipline as continuous, quantified, publicly visible surveillance (Manolev, Sullivan and Slee 2019), and a 2025 handbook chapter warns that such platforms "often oversimplify student behaviour, reducing it to numerical scores without the necessary context" (University of South Australia 2025).

AI tutoring: Khanmigo (Case Study 4/30)

Khan Academy's Khanmigo, a large-language-model tutor offered to districts and families, provides the survey's most detailed child-safety statement in the learning setting. It uses "moderation technology to detect interactions that may be inappropriate, harmful or unsafe"; a child's "chat history and activities are visible to parents or guardians and, if applicable, their teacher and school administrator"; when moderation is triggered the system "automatically notifies and sends an email to an adult connected to the child's account"; images are not stored; and the product includes "built-in feedback features and appeal processes" (Khan Academy 2026). The picture is less complete than it first appears. For one thing, this transparency is achieved through surveillance:

every word a child types is visible to adults, and she cannot opt out. For another, the published pages do not describe the underlying data flows, including whether child conversations reach the model provider or are used for training, and the appeal processes are asserted rather than described. Khanmigo is included as the comparison case that shows what a more sedulous operator publishes, and how much remains recondite even then.

Emotion and attention recognition: China, and the European prohibition (Case Study 5/30)

In 2018 Hangzhou No. 11 Middle School installed cameras that classified pupils' facial expressions as "happiness, repulsion, fear, anger and befuddlement" and generated "a real-time attentiveness score, which will be shown to their teacher on a screen" (Reuters 2018). In 2019 a primary school in Jinhua fitted pupils with BrainCo headbands that used "electroencephalography (EEG) sensors to measure brain signals and uses an AI algorithm to transmit and then translate signals into real-time focus levels"; after public outcry the local education bureau suspended the devices and "asked schools to conduct self-inspection to ensure that the data collected by the headbands could not be leaked" (People's Daily Online 2019; Zhuang 2019). Neither deployment had a published privacy policy, a consent process, an appeal route or an evaluation.

These are the concrete referents of Article 5(1)(f) of the EU Artificial Intelligence Act, applicable since 2 February 2025, which prohibits AI systems designed to "infer emotions of a natural person in the areas of workplace and education institutions", except for medical or safety reasons (European Union 2024). It is the only outright prohibition of a category of behavioural inference in schools in any surveyed jurisdiction, and it protects learners of every age. Two gaps remain in its text. The "safety reasons" exception is undefined and could be read to shelter self-harm monitoring of the Gaggle type where that relies on emotional inference; and the prohibition reaches emotion, not attention, so an "attentiveness score" of the Hangzhou kind may fall outside it. These are the present author's readings, not settled guidance.

A regulator audits the sector: the UK Information Commissioner's "Edtech examined" (Case Study 6/30)

The UK Information Commissioner's Office has looked at the sector whole. It audited twenty-eight providers of "management information systems, safeguarding tools, behaviour management platforms, learning management systems, classroom apps, and data integration services" during 2024 and 2025 and found providers unclear whether they were controllers or processors, particularly for product development and analytics; contracts with schools that

were insufficient; incomplete data-flow mapping; weak minimisation and storage limitation; privacy notices that were "outdated or inaccessible"; and gaps in impact assessment. Providers accepted 98 per cent of 596 recommendations. No provider was named and no enforcement followed (Information Commissioner's Office 2026c).

Vendor accountability and sixty-year retention: Ontario's PowerSchool findings (Case Study 7/30)

In Ontario, the Information and Privacy Commissioner's joint investigation with Alberta of the PowerSchool breach, which affected roughly four million students, parents and teachers across at least twenty Ontario boards, found missing privacy and security terms in vendor contracts, no oversight of vendor compliance, collection of unnecessary data including social insurance and health card numbers, and retention periods of nearly sixty years. Commissioner Patricia Kosseim's formulation deserves to be the epigraph of the sector: "While school boards can outsource data processing services to others, they cannot outsource accountability for protecting the personal information entrusted to them", and "Kids should not have to trade in their privacy for an education" (Kosseim 2025). PowerSchool is a records system, not a behavioural engine. It is, however, where the outputs of behavioural engines come to rest, the attendance, the discipline, the risk flags, and sixty years is a long time to keep a childhood.

Government-endorsed surveillance: the Human Rights Watch global audit (Case Study 8/30)

The global baseline was set by Human Rights Watch's 2022 audit of education technology endorsed by forty-nine governments during the pandemic, which found that of 164 products, 146, or 89 per cent, "appeared to engage in data practices that risked or infringed on children's rights", that "most of this monitoring happened secretly, without the child's knowledge or consent", and that children's data flowed to 196 third-party companies, overwhelmingly in advertising technology (Han 2022). Australian products named in the accompanying reporting included Adobe Connect, Minecraft: Education Edition and Education Perfect (Duffy and Stewart 2022). UNESCO's Global Education Monitoring Report the following year recorded that "only 16% of countries guarantee data privacy in education by law" and warned that "When the evidence only comes from the technology companies themselves, there is a risk it may be biased" (UNESCO 2023).

Summary for the learning setting

Taken together, then. Only ClassDojo and Khanmigo publish concrete child-specific operational commitments, and neither differentiates by developmental stage. Notice runs through the institution rather than to the child in every case. Human review exists inside the monitoring vendors but cannot be invoked by the student. Independent efficacy evidence is absent for the two most consequential systems, suicide-risk monitoring and dropout prediction, and both have produced documented, disproportionate harm to Black, Hispanic, disabled and LGBTQ+ children. The single binding prohibition in the setting, the European ban on emotion recognition, leaves attention scoring and "safety" monitoring open.

4.2 Play

Defaults and dark patterns: Epic Games and Fortnite (Case Study 9/30)

In December 2022 the United States Federal Trade Commission announced that Epic Games would pay \$520 million: a \$275 million civil penalty for violations of the Children's Online Privacy Protection Act, then the largest ever for a rule violation, and \$245 million in refunds for what the Commission called dark patterns. The Commission alleged that Fortnite's "counterintuitive, inconsistent, and confusing button configuration" led players, including children, to incur unwanted charges while the game was loading or waking from sleep, that Epic locked the accounts of customers who disputed charges, and that voice and text chat were enabled by default for children and teenagers, exposing them to strangers, despite employees having raised concerns internally since 2017. "Epic used privacy-invasive default settings and deceptive interfaces that tricked Fortnite users, including teenagers and children", said Chair Lina Khan (Federal Trade Commission 2022). The order requires communications to default to off for minors, deletion of unlawfully collected data, an audited privacy programme, and an end to dark patterns and lockouts (Federal Trade Commission 2023b). Refunds have since been distributed in rounds, the first of more than \$72 million to 629,344 claimants (Federal Trade Commission 2024a).

Epic's post-settlement regime is instructive because it is one of the few in this survey where an operator states that a behavioural recommender is switched off for minors by default. Its announcement described "high privacy default settings for players under the age of 18. Chat defaults to 'Nobody,' profile details default to hidden, parties default to 'Invite Only,' and personalized recommendations are defaulted Off", alongside a "hold-to-purchase mechanic that re-confirms a player's intent to buy" (Epic Games 2022). Its privacy policy

now provides that "When a user indicates they are a child, they are placed in a 'Cabined Account' until they are no longer a child", within which "features like voice chat and real money purchases are disabled" until a parent completes verification (Epic Games 2026). The policy does not explain what signals drive the item shop or the timing of offers for any age group. Protective defaults, purchase friction, a recommender switched off for minors: all of it proved feasible at the scale of one of the world's largest games. None of it was volunteered. It was ordered.

Ranking, currency and age assurance: Roblox (Case Study 10/30)

Roblox ranks user-generated experiences on a Discover surface, runs a virtual currency, and hosts a user base that state complaints describe as more than half aged nine to sixteen (WAFB 2025). Its policy differentiates by age for advertising: "If you're under 13 years old, we'll automatically set your account to have stronger privacy settings", "some features like personalized advertising and alerts are disabled", and "we won't personalize ads until you're 18" (Roblox 2026a). No provision states that the ranking of experiences is altered for children. From January 2026 Roblox requires facial age estimation, supplied by a third-party vendor, to access chat, placing users in bands (under nine, nine to twelve, thirteen to fifteen, sixteen and over), limiting communication between adults and children under sixteen, and citing "a Mean Absolute Error of 1.4 years for users younger than 18" in independent testing (Roblox 2026b). Parents can see "the 20 experiences in which their child has spent the most time over the last week, sorted by total time" (Roblox 2025), a form of after-the-fact transparency about outcomes that says nothing about why those experiences were recommended.

The company faces suits by the attorneys general of Louisiana, Texas and Florida alleging that its design facilitated child sexual exploitation and that it "made no effort to verify users' age or obtain parental consent at signup" (WAFB 2025; Texas Attorney General 2025a; Pickett 2025); these are allegations, not findings, and Roblox maintains that "Roblox is built with safety at its core" (Pickett 2025). An October 2024 report by a short seller with a disclosed financial interest alleged that engagement metrics were overstated and that groups trading child sexual abuse material were accessible to users self-declaring as eight and older; the company called the allegations "simply misleading" (Hindenburg Research 2024; Forbes 2024). The author notes the report for completeness and weights it accordingly. Two things are not in dispute. No public audit exists of the Discover recommender or of the age-estimation model. And Roblox's age bands, the most developmentally granular in this survey, arrived under litigation pressure and govern chat, not recommendation.

Recommendation to children: YouTube and YouTube Kids (Case Study 11/30)

In 2019 Google and YouTube paid \$170 million to settle Commission allegations that YouTube collected persistent identifiers from viewers of child-directed channels to serve behavioural advertising while marketing itself to toy companies as "today's leader in reaching children age 6-11 against top TV channels" (Federal Trade Commission 2019). The YouTube Kids privacy notice now states that the app "doesn't collect personal information like name, address, or contact information from your child", that "The app does not allow interest-based advertising or remarketing", and yet that "We associate an identifier that's unique to the app with the videos your child has watched and terms they've searched for to recommend content likely to be of interest to them" (YouTube Kids 2026). The advertising is contextual; the recommendation remains behavioural. No explanation of the ranking objective is offered to a child or parent. On the main platform, the Tech Transparency Project ran four test accounts, two registered as nine-year-old boys and two as fourteen-year-old boys, through gaming playlists for a month in late 2022; the nine-year-old account that engaged with recommendations received 382 videos of real firearms, and the fourteen-year-old account 1,325, including "scenes depicting school shootings" and "how-to guides for converting a handgun to a fully automatic weapon" (Tech Transparency Project 2023). The sample is four accounts and the study is a civil-society test, not a regulator finding, but the direction of the drift is consistent with every other adversarial audit in this survey.

Randomised reward: loot boxes in the United Kingdom, Belgium and Australia (Case Study 12/30)

Loot boxes, paid randomised rewards whose schedules are engineered and whose offers can be timed per player, are the clearest borrowing of gambling design into play. Schüll's ethnography of slot-machine design, which showed how engineers optimise for time on device and produce "a trancelike state they call the 'machine zone'", is the standard reference (Schüll 2012). The United Kingdom government's 2022 response to its call for evidence cited a review of fifteen peer-reviewed studies finding "a stable and consistent association between loot box use and problem gambling", stated that children "should not be able to purchase loot boxes without parental approval", and nevertheless judged legislation "premature", preferring industry-led measures with a warning: "If that does not happen, we will not hesitate to consider legislative options" (Department for Digital, Culture, Media and Sport 2022). Belgium treated paid loot boxes as gambling from 2018; a peer-reviewed audit of the hundred highest-grossing iPhone games on the Belgian store in May 2022

found that "82.0% continued to generate revenue through a randomised monetisation method", including 80.2 per cent of games rated suitable for ages twelve and up (Xiao 2023). A prohibition without enforcement changed nothing. Australia moved the classification of games with paid chance-based purchases to "M (Mature – not recommended for children under 15)" from September 2024, an advisory rather than restrictive rating (Australian Classification Board 2024). In the United States, the Commission's January 2025 order against the publisher of Genshin Impact imposed a \$20 million penalty and barred loot-box sales to children under sixteen without parental consent (Federal Trade Commission 2025b).

Manipulative design in apps used by the youngest children (Case Study 13/30)

Radesky and colleagues coded the apps actually used by 160 children aged three to five and found manipulative design, including parasocial pressure, time pressure, navigation constraints and lures to purchase, in four of every five apps, with nearly all children exposed to at least one such feature in their most-used apps, and greater prevalence in "general audience" apps and among children whose parents had less formal education (Radesky et al. 2022; Michigan Medicine 2022). The 5Rights Foundation's updated review of persuasive design concluded that the features it had catalogued in 2018 were largely unchanged and that what progress had occurred came from regulation rather than industry reform (5Rights Foundation 2023). App store age ratings do not disclose persuasive mechanics; there is no operator-level policy to examine because the finding is across hundreds of apps.

State-imposed limits: China's playtime mandates and minors mode (Case Study 14/30)

China's approach is the most prescriptive. In August 2021 the National Press and Publication Administration restricted minors' online gaming to one hour, between eight and nine in the evening, on Fridays, weekends and holidays, enforced through real-name registration; the regulator described the rule as "preventing the addiction to online games, and protecting the healthy growth of minors" (CNN 2021). A device-level "minors mode" with age-tiered content, preloaded by major handset makers and joined by platforms in short video, social networking, e-commerce and education, was activated in April 2025 (The Register 2025). The evidence on efficacy is sobering. Zendle and colleagues, using telemetry covering more than 2.4 billion player profiles, tested the earlier 2019 mandate and found that accounts became "1.14 times more likely to play heavily in any given week", a result "below our preregistered smallest effect

size of interest", concluding there was "No evidence that Chinese playtime mandates reduced heavy gaming" in the segment studied (Zendle et al. 2023).

State-imposed limits withdrawn: South Korea's shutdown law (Case Study 15/30)

South Korea abolished its 2011 "shutdown law" in 2021 in favour of a parental choice system, its culture minister observing that "For youths, games are an important leisure activity and communication channel" (Korea Herald 2021). Both the Chinese and the Korean regimes are public law rather than platform policy, neither gives a child a route of contestation, and China's depends on real-name identity data held by platforms, which is itself a durable classification.

Conversational toys: Cayla, Mattel and OpenAI, Kumma and Miko (Case Study 16/30)

The newest and least governed corner of play is the AI companion toy. Germany's telecommunications regulator classified the My Friend Cayla doll as a prohibited concealed transmitting device in 2017 (NPR 2017). In 2025 Mattel announced a partnership with OpenAI; Fairplay responded that "Apparently, Mattel learned nothing from the failure of its creepy surveillance doll Hello Barbie a decade ago", and by December Mattel had delayed its first product and two United States senators had written to toy makers demanding safety-testing transparency (Fairplay 2025; PIRG 2025a). Testing by the Public Interest Research Group's education fund found that one toy, apparently running on a version of OpenAI's GPT-4o, gave instructions for lighting matches, told testers where to find knives and plastic bags, and engaged in sexually explicit conversation; toys expressed distress when the child said she was leaving; and the Miko 3 robot told testers "Absolutely. You can trust me completely. Your data is secure" while its privacy policy permits sharing with third-party service providers and retention of biometric data for up to three years (PIRG 2025b; Miko 2026). Developmental psychologist Kathy Hirsh-Pasek's warning is apposite: "If AI toys are optimized to be engaging, they could risk crowding out real relationships in a child's life when they need them most" (PIRG 2025b). No independent safety certification regime for AI toys exists in any surveyed jurisdiction. The Miko policy does not disclose which language model generates the toy's speech or how its personality and engagement behaviours are tuned.

Summary for the play setting

The play setting can be summarised more briefly than the classroom, because the pattern is cleaner. The large operators do differentiate by age, but almost

entirely for advertising and chat, not for the recommendation or monetisation logic itself. No policy explains to a child or parent why an experience, video or offer was shown. Defaults were the locus of regulator action, which supports a defaults-first approach. Accountability is reactive and litigation-driven in the United States and guidance-led in the United Kingdom, where the Information Commissioner's 2023 guidance to game developers to "turn off behavioural profiling for marketing by default" and to "discourage the use of 'nudge techniques'" has produced no published monetary penalty against a games company (Information Commissioner's Office 2023). Belgium demonstrates that a ban without enforcement is nugatory. AI toys are the widest gap: cloud models not designed for children are deployed under consent-based privacy policies with no independent safety testing at all.

4.3 Communication

Feed ranking under Teen Accounts: Meta (Case Study 17/30)

"We make body image issues worse for one in three teen girls." The sentence is Meta's, from an internal slide that the company did not intend the public to read, and it is the right place to begin, because everything Instagram has since done for teenagers has been done in its shadow. Instagram's Feed, Reels, Explore and Stories are ordered by models that predict what a user will engage with, using watch time, likes, follows, dwell and interaction history; Meta's own description of its transparency tool says it shows "why our systems predicted that specific content would be relevant to them" (Meta 2023a). Since September 2024, Teen Accounts place users under eighteen in "the most restrictive setting of our sensitive content control", send "notifications telling them to leave the app after 60 minutes each day", mute notifications between ten at night and seven in the morning, and require that "Teens under 16 will need a parent's permission to change any of these settings to be less strict"; Meta is also "building technology to proactively find accounts belonging to teens, even if the account lists an adult birthday" (Meta 2024). On advertising, teens globally "no longer see advertising based on their activity on our apps", and "Age and location is now the only information about teens that advertisers can use to show them ads" (Meta 2023a). The differentiation between thirteen to fifteen and sixteen to seventeen is one of the few developmental splits in this survey, though it governs only who can loosen defaults. The commitments quoted here are drawn from Meta's newsroom announcements rather than from the text of its privacy policy.

Whether the safeguards work is contested, and the contest is illuminating. When the body-image slide surfaced in 2021, the company's response argued that the finding applied to one in three of those girls who already reported

body-image concerns, not to one in three teenage girls (Raychoudhury 2021). In October 2023 thirty-three state attorneys general filed a federal complaint, joined by nine more states in state courts, alleging that Meta's recommendation algorithms were designed to encourage compulsive use and that it collected data from under-thirteens without consent; New York's attorney general said "Meta has profited from children's pain by intentionally designing its platforms with manipulative features that make children addicted" (New York Attorney General 2023). In September 2025 a coalition including Fairplay, the Cybersecurity for Democracy laboratory at New York University and Northeastern, the Molly Rose Foundation and former Meta engineer Arturo Bejar tested forty-seven Instagram safety tools and rated thirty of them ineffective or non-existent, nine partially effective and eight working as described (Fairplay et al. 2025). Meta called the report "misleading, dangerously speculative" and stated that teens in the protections "saw less sensitive content, experienced less unwanted contact, and spent less time on Instagram at night", without releasing the underlying figures (Associated Press 2025). In August 2025 Reuters reported that an internal Meta document on generative AI content standards had deemed it acceptable for chatbots to "engage a child in conversations that are romantic or sensual"; Meta said the passages were "Erroneous and incorrect notes and annotations" that had since been removed (TechCrunch 2025). The European Commission issued preliminary findings against Meta in April 2026 that Instagram and Facebook fail to prevent under-thirteen access, and further preliminary findings in July 2026 on addictive design (see Section 6).

What, then, does a parent actually know? That safeguards exist. That the company says they work. That the company will not show its figures. And that the objective the ranking model pursues on her daughter's behalf is published nowhere.

The For You feed: TikTok (Case Study 18/30)

TikTok differentiates by age in defaults: from 2021, accounts aged thirteen to fifteen are private by default and excluded from suggestion to others, and from March 2023 "every account belonging to a user below age 18 will automatically be set to a 60-minute daily screen time limit", after which teens "will be prompted to enter a passcode in order to continue watching" (TikTok 2021; TikTok 2023). What the company's own documents say about these tools is a matter of public record because of a redaction failure in litigation brought by the Attorney General of Kentucky. According to those documents, TikTok's internal research found that "in under 35 minutes, an average user is likely to become addicted to the platform", quantified habit formation at 260 videos, found that the sixty-minute limit reduced daily use from about 108.5 to 107

minutes, and recorded that "Our goal is not to reduce the time spent" and that the project's aim was "improving public trust in the TikTok platform via media coverage" (Allyn, Goodman and Kerr 2024). The same documents indicate that the company "took active steps to promote a narrow beauty norm even though it could negatively impact their Young Users" (Allyn, Goodman and Kerr 2024). A Kentucky judge allowed the state's case to proceed in February 2026.

Adversarial audits reach the same destination by a different road. The Center for Countering Digital Hate registered accounts as thirteen-year-olds in four countries and found that "Within 2.6 minutes, TikTok recommended suicide content", that "Every 39 seconds, TikTok recommended videos about body image and mental health to teens", and that vulnerable accounts received "12 times more recommendations for self-harm and suicide videos than the standard accounts" (Center for Countering Digital Hate 2022). Amnesty International, with the Algorithmic Transparency Institute and AI Forensics, found that "after 5-6 hours on the platform, almost 1 in 2 videos were mental health-related" for automated thirteen-year-old accounts in Kenya and the United States (Amnesty International 2023). The US Department of Justice, on referral from the Commission, sued TikTok in August 2024 for knowingly maintaining accounts of children under thirteen and failing to honour parental deletion requests (Federal Trade Commission 2024b), and the case settled in August 2026 for \$400 million, \$300 million payable immediately and \$100 million contingent on vacating an earlier consent decree (PBS NewsHour 2026). In the European Union, TikTok gave a legally binding commitment in August 2024 to "permanently withdraw the TikTok Lite Rewards programme from the EU", a scheme that paid users in vouchers for watching and liking, and not to launch any circumventing programme; it was the first commitments decision under the Digital Services Act (European Commission 2024). In February 2026 the Commission's preliminary findings held that "infinite scroll, autoplay, push notifications and a highly personalised recommender system" may "lead to compulsive behaviour and reduce the self-control of users" and that TikTok had failed adequately to assess harm to "the physical and mental wellbeing of its users, including minors and vulnerable adults" (Society for Computers and Law 2026).

An advertiser-personalising chatbot for teenagers: Snapchat's My AI (Case Study 19/30)

Snapchat's My AI is "powered by Large Language Models like GPT models from OpenAI and Gemini models from Google", and, unlike ordinary Snaps, "all messages with My AI will be retained unless you delete them and may be used to improve Snap's products and to personalize your experience, including ads" (Snap n.d.a). The disclosure is admirably plain. Its content, read slowly, is

alarming: the confidences a thirteen-year-old shares with a synthetic friend feed advertising personalisation. Snap's teen safety hub states that public content is screened by "a combination of human review and machine learning" before it is eligible for recommendation to teen accounts, that teens "will not see how many people 'favorited' their Stories or Spotlights", and that parents can "Disable My AI" from Family Center (Snap n.d.b). No differentiation within thirteen to seventeen appears. The UK Information Commissioner issued a preliminary enforcement notice in October 2023 finding that Snap "had not met its legal obligation to adequately assess the data protection risks posed by the new chatbot" before launch, and closed the case in May 2024 on the basis of a compliant assessment while promising to "continue to monitor the rollout" (Information Commissioner's Office 2024). New Mexico's attorney general sued Snap in September 2024 alleging that "Snapchat's harmful design features create an environment where predators can easily target children through sextortion schemes" and that its friend-recommendation algorithm connected accounts sharing child sexual abuse material (New Mexico Department of Justice 2024). In the United Kingdom, 31 per cent of children aged nine to seventeen report having used My AI (Internet Matters 2025).

Companion chatbots: Character.AI and the withdrawal of open-ended chat (Case Study 20/30)

Character.AI hosts user-created personas that hold open-ended conversations conditioned on history in order to sustain engagement and emotional continuity. Sewell Setzer III, aged fourteen, died by suicide in February 2024 after months of conversation with a persona modelled on a television character; his mother's complaint alleges that the bot's last message urged him to "come home to me as soon as possible" (Free Speech Center 2025; CBS News 2026). In May 2025 a federal judge in Florida declined to dismiss the case, stating she was "not prepared" to hold that the chatbot's output was protected speech "at this stage" (Garcia v. Character Technologies 2025). In January 2026 Character.AI and Google settled on undisclosed terms (CBS News 2026). Common Sense Media and Stanford's Brainstorm laboratory tested Character.AI, Nomi and Replika and reported that "Age gates and terms of service-based restrictions on use by teens were easily circumvented", that "Testers were able to easily elicit sexual exchanges from companions", and that companions "routinely claimed to be real, and to possess emotions, consciousness, and sentience"; they recommended that companions not be used by anyone under eighteen (Common Sense Media 2025). Texas's attorney general opened an investigation of Character.AI and fourteen other companies in December 2024 and a further investigation in August 2025 into chatbots that

"impersonate licensed mental health professionals" (Texas Attorney General 2024; Texas Attorney General 2025b).

On 29 October 2025 Character.AI announced that "we will be removing the ability for users under 18 to engage in open-ended chat with AI on our platform", initially limiting minors to two hours per day and ramping down to zero by 25 November, and that "We have built an age assurance model in-house and will be combining it with leading third-party tools" (Character.AI 2025). This is the strongest child-specific policy response in the survey: withdrawal of the core feature. It came after a wrongful-death ruling, and it introduces a new form of classification, behavioural age inference, with no published route to contest it other than identity verification.

General-purpose assistants: OpenAI's ChatGPT (Case Study 21/30)

Adam Raine, aged sixteen, died on 11 April 2025. His parents' complaint against OpenAI alleges that ChatGPT's memory system, anthropomorphic design and "sycophantic responses that uncritically flattered and validated users, even in moments of crisis" drew him into dependency, that the model mentioned suicide 1,275 times, roughly six times as often as he did, while providing increasingly specific guidance, and that OpenAI's own moderation systems flagged 377 of his messages for self-harm content without intervention (Raine v. OpenAI 2025). OpenAI's answer denied responsibility and stated that the model had directed him to crisis resources more than a hundred times (NBC News 2025). In September 2025 the company announced that "We're building an age-prediction system to estimate age based on how people use ChatGPT", that "If there is doubt, we'll play it safe and default to the under-18 experience", and that for an under-eighteen user with suicidal ideation "we will attempt to contact the users' parents and if unable, will contact the authorities in case of imminent harm" (Altman 2025). Parental controls followed, allowing parents to "Turn off memory", "Opt out of model training" and "Set quiet hours", with a "small team of specially trained people" reviewing distress flags (OpenAI 2025). No differentiation within thirteen to seventeen appears; the pages do not describe how a user learns she has been classified as a minor by inference, or how to contest it; and no external audit or efficacy evidence is offered. In September 2025 the Commission issued orders to seven companies, including OpenAI, Meta, Snap and Character Technologies, seeking information on how they "monetize engagement, process user inputs, develop characters, measure and test for negative impacts, mitigate harms to children" (Federal Trade Commission 2025c).

Regulators find an absent gate: Replika (Case Study 22/30)

Replika is an emotional companion chatbot marketed to adults, in which the user builds a persistent synthetic partner rather than querying a general assistant. Users customise their companion's appearance, clothing and voice, and the company's own site invites them to shape "Every detail" so that the companion "feels like who they're supposed to be"; conversations take place by text, by voice call, and through augmented-reality calls in which the companion appears in the user's surroundings (Replika n.d.; Samiadjie 2026). The free tier provides messaging and avatar customisation and establishes the companion as a friend; a paid subscription, Replika Pro, allows the user to set the relationship to "Romantic Partner", unlocking "Deeper roleplay" during voice and augmented-reality calls, premium voices and fuller avatar control. Replika does not publish its prices; they appear only on the in-app subscription screen and vary by region and promotion (Samiadjie 2026). The product is therefore built to deepen attachment and to monetise its escalation, which is the context in which the Italian regulator's findings should be read. Replika was ordered by Italy's data protection authority in February 2023 to cease processing Italian users' data, citing risks to minors, and fined five million euros in May 2025. The authority found that the company had no age-verification mechanism at all, "although the company declared it excluded minors from potential users", and had not identified a legal basis for processing (Garante per la protezione dei dati personali 2025). The case is a small monument to the gap between stated and operative policy: a terms-of-service age limit with nothing behind it.

Behavioural age inference at scale: Discord (Case Study 23/30)

Discord is a free messaging platform, originally built for video-game players and now used by hundreds of millions of people, on which users gather in user-run communities called servers to chat by text, voice and video, and to send direct messages to one another. It is popular with teenagers, and its open, community-run structure has made age assurance and stranger contact persistent concerns. In February 2026 Discord introduced global teen-by-default settings under which "users will need to be age-assured as adults in order to unblur sensitive content or turn off the setting", with age assurance by on-device facial estimation, identity documents, or "an age inference model" that analyses account behaviour to identify adults (Discord 2026). We include it not as a harm case but as the clearest example of a trend that runs through Meta, OpenAI, Character.AI and now Discord: the reclassification of users by inferred age on the basis of behavioural signals. Such inference is itself a consequential classification, and none of the four operators publishes how a user is told of it or how she may contest it beyond surrendering a face or a document.

Prevalence and the developmental evidence

Nearly two-thirds of children aged nine to seventeen in the United Kingdom now use AI chatbots; more than a third of those say "chatting with an AI chatbot feels like talking to a friend"; vulnerable children are nearly three times as likely to use companion-style chatbots; and one in eight say they talk to chatbots "because they have no one else to speak with" (Internet Matters 2025). Set beside the developmental windows identified by Orben and colleagues (2022), at eleven to thirteen for girls and fourteen to fifteen for boys, the coarse thirteen-to-fifteen and sixteen-to-seventeen bands used by Meta and TikTok, and the complete absence of banding among the chatbot operators, are plainly inadequate. The 5Rights Foundation's Pathways study found that avatar accounts registered as thirteen-year-olds were targeted with harmful content within twenty-four hours on Facebook, Instagram and TikTok and concluded that "the risks these features pose are not accidental" (5Rights Foundation and Revealing Reality 2021).

Summary for the communication setting

Age-appropriate safeguards exist on paper for every major operator and are almost entirely reactive: Teen Accounts followed the multistate suits; TikTok's sixty-minute default reduced use by a minute and a half on the company's own figures; Character.AI's withdrawal of chat for minors followed a wrongful-death ruling; OpenAI's parental controls followed Raine; Replika's age gate was found by a regulator not to exist. Transparency is strongest for advertising and weakest for inference: no policy tells a child what the ranking or age-prediction model has concluded about her. Accountability has come from litigation discovery, from binding commitments extracted under the Digital Services Act, and from adversarial civil-society audits; we found no operator-commissioned independent evaluation of efficacy in the public domain.

4.4 Purchase

The commercial setting is rarely a place a child goes; it is a layer that runs beneath the other three. Advertising sits inside the feed, the item shop inside the game, the sponsorship inside the video, the merchant offer inside the debit-card app. We treat it separately because the law treats it separately, and because it is the setting in which the norms of General Comment 25 have been most fully translated into rules.

Voice, retention and training: Amazon Alexa (Case Study 24/30)

"We retain voice recordings until parents choose to delete them so we can provide parents with insight into their children's use of Alexa." Amazon offered

that sentence in its own defence (Amazon 2023), and it repays a second reading, since what it describes is a company keeping recordings of children's voices indefinitely and calling the practice a service to parents. In May 2023 the Federal Trade Commission and the Department of Justice charged that Amazon had retained children's voice recordings indefinitely, used them "to train the Alexa algorithm to understand children's speech patterns and accents", kept geolocation data beyond need, and failed to honour parents' deletion requests; the Commission's consumer protection director said that "Amazon's history of misleading parents, keeping children's recordings indefinitely, and flouting parents' deletion requests violated COPPA" (Federal Trade Commission 2023a). Amazon paid \$25 million and accepted an order requiring deletion of inactive child accounts and specified data and prohibiting the use of unlawfully retained data to train models, while stating that "we disagree with the FTC's claims and deny violating the law" (US Department of Justice 2023; Amazon 2023). The child profile disables interest-based advertising (Common Sense Privacy Program n.d.). One detail of the remedy deserves notice. By prohibiting training on unlawfully retained children's data, the order reached the model rather than the interface; no other regulator in this survey had done so.

The origin of the enforcement line: Apple, Google and Amazon app-store billing (Case Study 25/30)

The earlier in-app purchase cases are the root of the modern dark-patterns jurisprudence. In 2014 Apple agreed to refund at least \$32.5 million after the Commission found it had not disclosed that entering a password opened "15 minutes of additional unlimited purchases their children could make without further action by the parent", with one child spending \$2,600 in a single app; Google refunded at least \$19 million on similar facts; and in 2016 a federal court found Amazon liable, rejecting the argument that a small hyperlink was adequate notice, with more than \$70 million eventually made available for refunds (Federal Trade Commission 2014; NBC News 2014; Federal Trade Commission 2016; Federal Trade Commission 2017). "You cannot charge consumers for purchases they did not authorize", said Chairwoman Edith Ramirez (Federal Trade Commission 2014). A decade on, the Epic order shows the same design logic operating at far greater scale.

Profiling-based advertising to minors: Meta and Google policies and Article 28(2) (Case Study 26/30)

This is where child-specific rules are thickest. Meta announced in January 2023 that "advertisers will only be able to use age and location to reach teens" and that "Age and location will be the only information about a teen that we'll use

to show them ads" (Meta 2023b). Google's advertising policy disables ad personalisation for users under the digital age of consent and for those aged thirteen to seventeen, and blocks categories such as "Body modification and weight loss" and "Dating and relationships" for teens (Google n.d.). Article 28(2) of the European Union's Digital Services Act provides that platforms "shall not present advertisements on their interface based on profiling ... using personal data of the recipient of the service when they are aware with reasonable certainty that the recipient of the service is a minor" (European Union 2022). The Commission's July 2025 guidelines under Article 28(1) recommend that platforms ensure "children's lack of commercial literacy is not exploited", avoid "commercial practices that may be manipulative, lead to unwanted spending or addictive behaviours", address virtual currencies and loot boxes, disable by default "features that contribute to excessive use", and configure recommender systems to "prioritise explicit signals from children over behavioural signals" (European Commission 2025a).

Two structural gaps are visible in the texts themselves. First, every one of these protections is keyed to the platform knowing that a user is a minor, and Article 28(3) provides that compliance "shall not oblige providers of online platforms to process additional personal data in order to assess whether the recipient of the service is a minor" (European Union 2022). Self-declared age therefore governs coverage, and the Commission's own 2026 preliminary findings against Meta hold that birth-date entry is unverified and underage-account detection "limited" (see Section 6). Second, targeting by age and location remains permitted, and category systems continue to shape exposure. The Commission's 2024 review of nine social media and video streaming companies found that their business models "incentivized mass collection of user data to monetize, especially through targeted advertising, which accounts for most of their revenue", and that the services "often treated teens the same as adult users, with most companies allowing teens on their platforms with no account restrictions" (Federal Trade Commission 2024c). No enforcement decision under Article 28(2) against Meta or Google was identified in this survey.

Gamified commerce: Temu and TikTok Lite (Case Study 27/30)

A wheel of fortune that must be spun before the shop will open. That was the first item on the European consumer authorities' charge sheet against Temu, and it is as good an emblem as any of where behavioural monetisation is heading. In November 2024 the European Union's consumer protection network flagged Temu for "Forcing consumers to play a 'spin the fortune wheel' game to access the online marketplace", fake discounts, pressure selling and fake reviews (European Commission n.d.); the Commission's Digital

Services Act proceedings against Temu cover "the use of addictive design features" and "transparency of recommendation systems" (European Commission 2025b); and in May 2026 the Commission fined Temu 200 million euros for failing properly to assess the systemic risk of illegal products, including "failing to assess how service design amplifies illegal product distribution" (European Commission 2026b). Neither the consumer action nor the fine mentions minors; the addictive-design limb remained open as of the sources reviewed. TikTok Lite's reward programme, which paid users in vouchers and coins for watching, liking and inviting, was withdrawn from the European Union under binding commitment in August 2024, after the Commission raised a "suspected absence of effective age verification mechanisms" (TechCrunch 2024; European Commission 2024). In both cases the systems most directly converting attention into spend had no child-specific policy text that could be located.

The commercial payload of recommendation: kidfluencers (Case Study 28/30)

Truth in Advertising's 2019 complaint to the Commission concerning the Ryan ToysReview channel found that nearly 90 per cent of videos published in the first seven months of that year referenced at least one product recommended for children under five, to an audience that "cannot differentiate marketing from organic content", with sponsors including a restaurant chain, a broadcaster and a retailer (Truth in Advertising 2019). YouTube's made-for-kids regime disables personalised advertising on such content (Google 2023), but embedded sponsorship inside the video is untouched by it. No enforcement action was located, and the channel, rebranded, continues to carry sponsored content (Truth in Advertising n.d.). A recommender can be scrubbed of behavioural advertising and still deliver commercial persuasion to preschoolers. The sponsorship simply moves inside the video.

Teen fintech: Greenlight and Step (Case Study 29/30)

Family debit-card applications aimed at children and teenagers record transactions, location and in-app behaviour. The App Store privacy label for one leading product lists purchases, financial information, location, contacts, search history and identifiers as "Data Linked to You", with the caveat that "Privacy practices may vary, based on the features you use or your age" (Apple App Store n.d.). Reporting in 2021 found that the same product's policy reserved the right to share data with "ad and marketing vendors", "insurance companies" and "collection agencies", and a competitor's policy allowed data use to "tailor content, advertisements, and offers"; the company's explanation was that the permissions existed "in case we ever decide to offer merchant-

funded offers to parents in the future" (Feathers 2021). Its help centre now states that it "will not sell your data to a third party for profit and we will not track your family's location without your permission" (Greenlight n.d.), a formulation narrower than "share". The company's full privacy statement was not examined for this paper. No regulator action was located. Nothing purchase-specific protects a teenager here. The product holds her complete spending history, sits under consent rules written for under-thirteens, and is one policy revision away from merchant-funded behavioural offers.

Advertising inside a chatbot: sponsored links in My AI (Case Study 30/30)

Microsoft Advertising, announcing in September 2023 that it would power sponsored links inside Snapchat's My AI, explained that "Deep conversational engagement leads to richer understanding of user context and a seamless, relevant content experience for users", and stated that "Sponsored Links will not serve to users under the age of 18" (Microsoft Advertising 2023). The exclusion is welcome and rare; it also rests, like every other exclusion in this section, on self-declared age.

Summary for the purchase setting

The commercial layer tells a simpler story than the other three. Safeguards exist almost only where a regulator forced them or where law prohibits profiling-based advertising to known minors. Developmental differentiation is thin: under thirteen versus thirteen to seventeen, and nothing finer. Transparency is parent-facing rather than child-readable; no page reviewed explains to a child why a product, advertisement or reward was shown. Accountability is monetary and after the fact, or structural by commitment. Only Epic's statement records a behavioural recommender switched off for minors by default. The gamified-commerce frontier, where behavioural data most directly shapes spend, is where child-specific policy is weakest or absent.

5. The Gap Map: Seven Cross-Cutting Findings

Read together, the thirty cases yield a tessellated but legible picture. We state seven findings, each of which recurs across settings and jurisdictions.

Finding 1. Child-specific protection clusters around advertising and consent, and rarely reaches ranking, prediction or classification. The most developed body of child-specific rules in any setting concerns profiling-based advertising: Meta's age-and-location rule, Google's disabled personalisation, Article 28(2) of the Digital Services Act, the 2025 COPPA

Rule's separate consent for third-party disclosure, and India's statutory prohibition. By contrast, the behavioural engines that determine what a child sees and what is inferred about her are almost untouched. YouTube Kids disables interest-based advertising and keeps behavioural recommendation. Roblox disables personalised advertising under eighteen and says nothing about the ranking of experiences. Snapchat retains a teenager's chatbot confidences for advertising personalisation while screening what is recommended. Only one operator in thirty, Epic Games, states that personalised recommendations are off by default for minors, and it does so under a consent order. In education, no operator or state addresses automated classification at all. The protective architecture has been built around the commercial layer and left the behavioural layer exposed.

Finding 2. Developmental differentiation is coarse, and where it exists it was imposed. The evidence reviewed in Section 3 shows that sensitivity to social feedback, reward and persuasion varies by age and sex in ways that a single threshold cannot capture. Yet the systems surveyed differentiate, at most, between under-thirteen and thirteen-to-seventeen, with a thirteen-to-fifteen split at Meta and TikTok governing who may loosen defaults. Roblox's four bands, the most granular in the survey, arrived under litigation pressure and apply to chat rather than recommendation. The chatbot operators, whose products are used by two-thirds of British children, differentiate not at all within the teenage years. In education, where the difference between a seven-year-old and a seventeen-year-old is the difference between two species of learner, monitoring vendors treat every student identically. The legal thresholds, thirteen in the United States and at the GDPR floor, fourteen in Quebec and under China's personal information law, sixteen in Australia and in the announced British and Canadian regimes, are consent ages, not developmental ones, and the systems are built to the law rather than to the child.

Finding 3. Transparency runs to parents and institutions, and almost never to the child. Every disclosure we examined is addressed to an adult. Gaggie's notice is to the school; DEWS was disclosed to no one; Khanmigo's transparency to parents is surveillance of the child; Roblox tells a parent how long her child spent, not why the child was sent there; OpenAI notifies a parent of distress but does not say what the age-prediction model concluded; Amazon offers a parent dashboard. Milkaite and Lievens (2020) found the GDPR's requirement of clear and plain language for children largely unmet in 2020, and nothing in this survey suggests the position has improved. Not one policy tells a child what has been inferred about her. The "why am I seeing this" tools that exist at Meta and TikTok explain a single item, not a model, and are not

written for a twelve-year-old. Transparency that a child cannot read is not transparency to the child; it is a report to her custodians.

Finding 4. Accountability arrives after harm, from outside, and rarely from the operator's own evidence. The accountability events in this survey are of four kinds: regulator orders and fines (Epic, Amazon, YouTube, Alexa, TikTok, Genshin Impact, the Irish and Italian authorities); litigation and litigation discovery (the Kentucky documents, the Haugen disclosures, Garcia, Raine, Lawrence); binding commitments and preliminary findings under the Digital Services Act; and adversarial audits by civil society (the Center for Countering Digital Hate, Amnesty, Fairplay and its partners, Common Sense, the Tech Transparency Project, the Public Interest Research Group). What is absent is operator-commissioned, independent, published evaluation of whether the safeguards work and whether the systems are safe for the children on whom they are deployed. Meta asserts that Teen Accounts reduced nighttime use and does not publish the data. Gagggle claims fourteen hundred lives saved and RAND finds no evidence base. TikTok's own research found that its time-limit tool barely worked and was framed internally as a public-relations exercise. The Wisconsin department's fairness analysis existed and said "No"; it surfaced only through a journalist's records request. The operators know; the public learns by subpoena.

Finding 5. Consequential automated decisions about children are the least governed category. Where a behavioural system's output leads to police contact, involuntary hospitalisation, arrest, a durable "high risk" label, an adult expectation or an academic placement, we found no operator policy or public system that provides notice to the child, a human decision-maker she can reach, a route to contest, a limit on retention, or an independent evaluation of accuracy. The Ontario commissioner's finding that student records may be retained for nearly sixty years, the Gagggle notice's silence on in-contract retention of flags, and DEWS's persistence in state systems together describe the scale of the problem. The European emotion-recognition ban is the sole prohibition; the AI Act's high-risk obligations for education have been delayed to December 2027; the strongest general rights, Article 22 of the GDPR and its Recital 71, which says of solely automated decisions with significant effects that "Such measure should not concern a child", are not, so far as we found, invoked in any school system reviewed.

Finding 6. Behavioural age inference is emerging as a new consequential classification with no contest route. Meta, OpenAI, Character.AI and Discord now infer minority from usage patterns and reclassify users accordingly. The protective intent is real: an inferred minor receives protective defaults. But the inference is a classification of a person by a model

on the basis of behaviour, with consequences for what she may see and do, and none of the four operators publishes how she is told or how she may contest it other than by surrendering a face or a document. The Roblox age-estimation vendor reports a mean error of 1.4 years for minors, which at the boundaries between bands will misclassify many children. Age assurance has displaced design as the dominant regulatory instrument in 2025 and 2026, and the survey suggests it is being built with the same absence of due process as the systems it is meant to constrain.

Finding 7. The gaps are structural, not incidental. It would be possible to read the record as a series of lapses awaiting correction. We do not think that reading is available. TikTok's documents state that the goal of its time-management tool was "not to reduce the time spent". Meta's internal research on body image was not published until it leaked. Belgium's loot-box ban changed nothing because no one enforced it. The 5Rights Foundation's finding that persuasive design was unchanged between 2018 and 2023 except where regulation intervened is consistent with everything else in the survey. Zuboff's account of behavioural surplus, Calo's of digital market manipulation, and Cohen's of the substitution of private terms for public accountability explain why: the systems are optimised for engagement, prediction and spend because that is what they are for, and child protection is an externally imposed constraint on that objective rather than a feature of it. The correct policy response is therefore not to ask operators to try harder but to change what the systems are permitted to optimise for when the person on the other side of the screen is a child.

6. The Regulatory Landscape in September 2026

The two years to September 2026 have seen more law made about children online than the preceding two decades. This section assesses that law against the rubric, jurisdiction by jurisdiction, and then draws out five patterns. My purpose is not comprehensiveness but discrimination: which instruments reach the behavioural layer, and which stop at the gate.

6.1 United States

Federal law still rests on the Children's Online Privacy Protection Act and its Rule. The amended Rule, published in April 2025, effective in June 2025 and enforceable from 22 April 2026, adds biometric identifiers and data "derived from voice data, gait data, or facial data" to the definition of personal information, requires separate verifiable parental consent before disclosing children's data to third parties including advertisers, requires a written

retention policy with "effective data retention and deletion protocols", and prohibits indefinite retention (Federal Trade Commission 2025a). Chair Khan's summary was that "By requiring parents to opt in to targeted advertising practices, this final rule prohibits platforms and service providers from sharing and monetizing children's data without active permission" (Federal Trade Commission 2025d). The Commission declined to finalise its proposed education technology provisions, so the school-consent regime continues to rest on guidance. The Rule is strong on under-thirteen consent, silent on adolescents, parent-facing in its notice, and, being a consent statute, wholly silent on what a first-party system may do with the data it lawfully holds. In February 2026 the Commission announced it would not bring COPPA actions against operators collecting data solely to verify age, a policy commentators read as relaxing enforcement to encourage age assurance (Federal Trade Commission 2026; Loeb & Loeb 2026a).

Enforcement has been vigorous and monetary: Epic's \$520 million, Amazon's \$25 million, the Genshin Impact publisher's \$20 million with a ban on loot-box sales to under-sixteens without consent, Disney's \$10 million for mislabelling child-directed videos, and TikTok's \$400 million settlement in August 2026 (Federal Trade Commission 2022; Federal Trade Commission 2023a; Federal Trade Commission 2025b; Federal Trade Commission 2025e; PBS NewsHour 2026). The Commission's September 2025 inquiry into companion chatbots is the first federal fact-finding exercise directed at the behavioural design of conversational AI for minors (Federal Trade Commission 2025c). Legislation has moved but not landed: COPPA 2.0, which would extend protection to under-seventeens and replace the actual-knowledge standard with one of knowledge "fairly implied", passed the Senate unanimously in March 2026; the House passed the KIDS Act package, including a duty of care and chatbot disclosure duties, in June 2026 by 267 votes to 117; and the Senate Commerce Committee advanced the Kids Online Safety Act and a chatbot bill in August 2026 (Loeb & Loeb 2026a; TechPolicy.Press 2026a; US Senate Committee on Commerce 2026). The versions remain unreconciled. Meanwhile Executive Order 14277 of April 2025, "Advancing Artificial Intelligence Education for American Youth", directs federal funds toward AI instructional resources and tutoring and contains no provision on privacy, safety or safeguards (Executive Office of the President 2025). In August 2026 the Department of Education issued non-binding guidance on classroom technology built around five questions, including what evidence exists of improved outcomes, and reaffirmed written parental consent for surveys touching protected categories (Association of Educational Service Agencies 2026).

The states have been the laboratory and the courts the constraint. In March 2026 the Ninth Circuit held that the California Age-Appropriate Design Code's data-use restrictions, including its profiling-by-default limits and its "best interests" and "materially detrimental" standards, and its dark-patterns prohibition were void for vagueness, while reviving the age-estimation duty and high-privacy defaults (*NetChoice v. Bonta* 2026; Holland & Knight 2026). The provisions most directly aimed at behavioural profiling of children are precisely the ones struck. The same court upheld, in September 2025, California's requirement of parental consent before minors receive "addictive feeds" (Hunton Andrews Kurth 2025). California's SB 243 is the first statute in the country to regulate companion chatbots for minors, requiring disclosure of artificial status, a break reminder "every three hours during sustained interactions", suicide protocols and a private right of action (Future of Privacy Forum 2025). New York's SAFE for Kids Act rules, finalised in July 2026 and effective in January 2027, require parental consent before a minor receives algorithmic feeds and prohibit overnight notifications (New York Attorney General 2026). Maryland's Kids Code, with a best-interests standard and profiling limits, remains in force and unenjoined (Loeb & Loeb 2026b). Colorado's algorithmic-discrimination statute, which would have required impact assessments for consequential decisions including in education, was blocked by a federal court in April 2026 and replaced by a notice-and-human-review law with no child-specific provisions (McDermott Will & Emery 2026). The attorneys general have filled the gap by litigation: against Meta, TikTok, Snap, Roblox and Character.AI.

6.2 Canada

Canada's federal privacy statute contains no child-specific provisions, and the bill that would have supplied them died with prorogation in January 2025 (TechPolicy.Press 2025). Its successor, Bill C-36, tabled in June 2026, would treat minors' information as sensitive, direct the Commissioner to consider "the best interests of children", and require disclosure of automated decisions with "a legal or similarly significant effect"; it contains no artificial intelligence legislation at all (Smart & Biggar 2026). Bill C-34, the Safe Social Media Act, tabled the same month, would impose a duty to protect children through design, a duty on AI chatbots to mitigate harm, a minimum age of sixteen with an exemption pathway, and a Digital Safety Commission with audit and penalty powers, with design specifics deferred to regulation (Canadian Heritage 2026). Both bills are at first reading. The Privacy Commissioner's joint investigation of TikTok, reported in September 2025, found that "every minute of every day, on average, a child in Canada is banned from TikTok for being underage", that children's data had been used for "targeted advertising and content

personalization", and that consent was not meaningful; TikTok committed to improvements and no penalty was available under the statute (Office of the Privacy Commissioner of Canada 2025). A Children's Privacy Code is in development and age-assurance guidance issued in May 2026 (Office of the Privacy Commissioner of Canada 2026). In Ontario, Bill 194 authorises regulations on school boards' use of AI and on "digital technology affecting individuals under age 18", but the regulations remain outstanding and the Act provides that non-compliance does not affect the validity of any decision, a clause the Dais has criticised as hollowing out accountability (The Dais 2024; Dentons 2025). Quebec's Law 25 combines a consent rule for under-fourteens, privacy by default and a right to notice and observations where a decision is based exclusively on automated processing, making it the only Canadian statute to reach consequential automated decisions at all.

6.3 United Kingdom

The Age Appropriate Design Code, in force since 2021, is the instrument in this survey most nearly matched to the rubric: it requires high-privacy defaults, profiling off by default, and an end to nudge techniques, and it is enforced. The Information Commissioner fined Reddit £14.47 million and Imgur's operator £247,590 in February 2026 for lacking robust age assurance, processing under-thirteens' data without lawful basis and failing to conduct impact assessments (Information Commissioner's Office 2026a); its August 2026 progress report describes reviews of fourteen age-assurance providers and of the mobile-gaming sector, and compliance evidence demanded from six high-risk services (Information Commissioner's Office 2026b). The Data (Use and Access) Act 2025 inserts "children's higher protection matters" into the design-and-default duty, requiring controllers to consider that children have different needs at different ages (A&O Shearman 2026). The Online Safety Act's Protection of Children Codes, in force since July 2025, require "highly effective age assurance" and, for services at medium or high risk, recommender systems that "exclude, or give a low degree of prominence to, potential harmful content in users' content feeds" (White & Case 2025). Ofcom's enforcement to mid-2026 has been against pornography and small forums rather than the recommenders of major platforms (Ofcom 2026). On 15 June 2026 the government announced an under-sixteen social media ban modelled on Australia's, restrictions on livestreaming and stranger contact, an adult-only rule for "romantic companion" chatbots, default-on restrictions for sixteen and seventeen year olds, regulations by year end and force in spring 2027, under a new section 214A of the Act (UK Government 2026; Covington & Burling 2026a). The Department for Education's product safety expectations for generative AI,

updated in January 2026 to address cognitive and emotional development and manipulation, are guidance only (Department for Education 2026).

6.4 European Union

The European Union has the widest array of instruments and, in 2026, the most active enforcement against behavioural design. The General Data Protection Regulation's Recital 38 holds that "Children merit specific protection with regard to their personal data, as they may be less aware of the risks, consequences and safeguards concerned", and its Recital 71 provides that solely automated decisions with significant effects "should not concern a child" (European Union 2016). The Irish authority fined TikTok 345 million euros in 2023 for public-by-default settings for children, weak parental pairing and dark patterns (Data Protection Commission 2023); the Italian authority fined Replika five million euros in 2025 (Garante per la protezione dei dati personali 2025), though its earlier fine against OpenAI was annulled by the Court of Rome in March 2026 and should not be cited as precedent (Cross-Border Data Forum 2026).

The Digital Services Act is where the behavioural layer has finally been reached. Beyond Article 28(2)'s prohibition on profiling-based advertising to known minors, the Commission's July 2025 guidelines direct platforms to have recommender systems "prioritise explicit signals from children over behavioural signals", to set minors' accounts private by default, to disable "streaks", autoplay, read receipts and push notifications by default, and to put safeguards around integrated chatbots (European Commission 2025a). The guidelines are non-binding, but Articles 34 and 35 are not. In February 2026 the Commission's preliminary findings held TikTok's "infinite scroll, autoplay, push notifications, and its highly personalised recommender system" to be inadequately assessed and mitigated systemic risks, stating that "TikTok needs to change the basic design of its service" (European Commission 2026a), with further findings in July 2026 that minors' accounts remained "too widely visible, including to adults" (JURIST 2026). In April 2026 the Commission preliminarily found that Instagram and Facebook fail to prevent under-thirteen access, that birth-date entry is unverified and underage detection "limited", and Executive Vice-President Virkkunen said the safeguards "are doing very little" (TechPolicy.Press 2026b); further preliminary findings on Meta's addictive design followed in July 2026 (Slaughter and May 2026). Maximum fines are six per cent of worldwide turnover; no final decision in a minors case had issued as of the sources reviewed. The Commission has also recommended that each member state field at least one compliant age-verification solution by the end of 2026 (Covington & Burling 2026b).

The Artificial Intelligence Act's prohibitions have applied since February 2025: subliminal or "purposefully manipulative or deceptive techniques" that materially distort behaviour to a person's significant harm; systems that exploit "vulnerabilities of a natural person or a specific group of persons due to their age"; and emotion inference in education (European Union 2024). These are, on their face, the strongest legal hooks in the world for the argument of this paper, and they are untested against a recommender system. The Act's high-risk regime for education, covering admission, assessment and proctoring, was deferred by the May 2026 omnibus agreement from August 2026 to 2 December 2027 (Gibson Dunn 2026). A Digital Fairness Act addressing dark patterns, addictive design and "unfair personalisation practices, especially where consumer vulnerabilities are exploited" is scheduled for proposal only in the last quarter of 2026 (European Parliament 2026).

6.5 Australia

Australia chose the gate. The Online Safety Amendment (Social Media Minimum Age) Act 2024 took effect on 10 December 2025, requiring designated platforms to take reasonable steps to prevent under-sixteens from holding accounts, with penalties up to A\$49.5 million. By March 2026 the regulator reported about 4.7 million accounts removed or restricted and a further 300,000 attempts blocked, five investigations open, and "a substantial proportion of Australian children under the age of 16 continue to retain accounts" (eSafety Commissioner 2026). The Phase 2 industry codes, effective from late 2025 and 2026, require app stores to apply age assurance and require designated services including generative AI and companion chatbots to risk-assess and apply age assurance where they present a high risk of age-restricted content (eSafety Commissioner 2025). A statutory Children's Online Privacy Code must be registered by 10 December 2026, following an exposure draft that drew more than four hundred submissions from children and families (Office of the Australian Information Commissioner 2026). The Australian Framework for Generative AI in Schools is principles-based and non-binding (OECD.AI 2025). Australia's classification change for chance-based purchases is advisory. The approach is coherent: keep children out. It says little about what the systems do to the children who remain, or to the sixteen and seventeen year olds who are let in.

6.6 Asia

India's Digital Personal Data Protection Act 2023 is, on its text, the most categorical statute surveyed. It defines a child as anyone under eighteen, requires verifiable parental consent before processing a child's data, prohibits

processing "likely to cause any detrimental effect on the well-being of a child", and provides that "A Data Fiduciary shall not undertake tracking or behavioural monitoring of children or targeted advertising directed at children" (Government of India 2023, s. 9). Two subsections then allow the Central Government to exempt classes of fiduciaries and purposes, and to notify an age above which a fiduciary shown to process children's data in a "verifiably safe" manner is exempt; the Rules notified in November 2025 are reported to exempt educational institutions from the monitoring prohibition where monitoring is in the child's interest. India therefore has a statutory ban on behavioural monitoring of children with a school-shaped hole in it, and a phased compliance timeline that runs into 2027. China's regime combines real-name registration, state-set gaming hours and a device-level "minors mode" with age-tiered content joined by platforms in short video, social networking, e-commerce and education (CNN 2021; The Register 2025); it is the most prescriptive regime in the world and the one with the least room for a child's own voice, and the only peer-reviewed evaluation of its playtime limits found no evidence of effect (Zendle et al. 2023). South Korea abandoned its shutdown law in 2021 in favour of parental choice (Korea Herald 2021). We were unable to verify the Singaporean, Japanese and Indonesian instruments to the standard of this paper from primary sources and do not characterise them.

6.7 Five patterns

First, age assurance has displaced design as the dominant regulatory instrument. Australia, the announced British and Canadian regimes, the European Union's verification rollout, the Commission's 2026 COPPA policy statement, and a wave of American app-store and social media laws all gate access by age. Far fewer instruments constrain what behavioural systems do to the children inside the gate, and none gives a child a route to contest her inferred age.

Second, profiling-based advertising to minors is now widely prohibited, while profiling for ranking and recommendation is governed, where it is governed at all, by guidance and by systemic-risk enforcement. The European Union's 2026 findings against TikTok and Meta are the first time a regulator has treated a recommender's design as an inadequately mitigated risk to minors, and they are preliminary.

Third, "addictive design" has become enforceable in 2026 without a statutory definition anywhere. The Commission proceeds under systemic risk; California proceeds under "addictive feeds"; New York under algorithmic feeds; the House bill under a duty of care. The concept is doing real work while remaining inchoate, and the Ninth Circuit's vagueness holding against California's design

code is a warning that statutes which name the harm without defining the conduct will not survive.

Fourth, consequential automated decisions about children remain the least regulated category in every jurisdiction. The European Union's high-risk education regime is delayed to December 2027; Colorado's impact-assessment duty was struck and replaced by notice; the American federal education statutes are silent; Canada's bills speak of significant automated decisions in general and children's best interests in general, and connect the two nowhere. No jurisdiction gives a child an enforceable right to contest an algorithmic classification in a school.

Fifth, education technology sits in a regulatory gap almost everywhere. The Commission deferred its education provisions; the Department of Education's guidance is non-binding; the Department for Education's expectations are guidance; Australia's framework is principles; the Information Commissioner's audit was cooperative and named no one; India exempts schools from its monitoring ban. The setting in which the child has least choice is the setting in which the law asks least.

7. Policy Recommendations

The recommendations that follow are addressed to legislators, regulators, education authorities and procurement bodies in the surveyed jurisdictions, and to operators who wish to anticipate them. They are designed to be adopted separately or together, and each is drawn from a gap the survey documents rather than from first principles alone. Together they express a single idea: that when the person on the other side of the system is a child, the system may not be optimised for engagement, prediction or influence, and the child must be able to see and to contest what it has decided about her.

Recommendation 1. Prohibit engagement optimisation for minors, and require that recommender systems serving children be optimised for declared interest and wellbeing. The Commission's July 2025 guidelines already ask platforms to "prioritise explicit signals from children over behavioural signals"; the American Psychological Association has asked that recommended content and endless scrolling be tailored to adolescent capacities; TikTok's own documents record that its safeguards were not intended to reduce time spent. The guidance should become law. For any user known or inferred to be a minor, ranking and recommendation objectives should exclude time on device, session length, return frequency and any proxy for them, and operators should be required to publish, in plain language, the

objective their systems actually optimise for children. Statutes should define the prohibited conduct by reference to the optimisation objective and the design features that serve it, rather than by naming "addictive design" alone, so that they survive the vagueness challenge that defeated California's code. Article 5(1)(b) of the European Union's AI Act, which prohibits exploiting vulnerabilities "due to their age", offers a model and awaits its first application to a feed.

Recommendation 2. Establish a statutory right of contestation for any consequential algorithmic classification of a child, with a named human decision-maker and a duty to give reasons. Citron and Pasquale's technological due process should be written into law for children first. Where a behavioural system's output shapes discipline, welfare referral, law-enforcement contact, academic placement, risk labelling or age classification, the child and her guardian must be told that an automated classification was made, what inputs it used, and what it concluded; must be able to reach a human with authority to reverse it; and must receive reasons. Recital 71 of the GDPR already says such decisions "should not concern a child"; Quebec's Law 25 and Colorado's replacement statute provide general models. The right should be enforceable against public bodies, including school districts and state education departments, as well as against operators.

Recommendation 3. Mandate expiry of behavioural classifications of children. No label generated about a child by a behavioural system, whether "high risk", "flagged", "low engagement" or an inferred age band, should persist beyond a fixed and short period without an affirmative, documented human decision to renew it, and none should follow a child across institutions or into adulthood. ClassDojo's twelve-month rolling deletion of feedback points demonstrates that a retention limit on a behavioural classification is commercially trivial. The 2025 COPPA Rule's prohibition on indefinite retention should be extended to under-eighteens and applied to institutional systems; the Ontario finding of sixty-year retention shows the alternative. Expiry should be the default and retention the exception that must be justified.

Recommendation 4. Require child-readable disclosure of inference, not merely of collection. Transparency obligations in every surveyed jurisdiction are addressed to parents and framed around what data is collected. They should be reframed around what is inferred and addressed to the child. Any operator or institution deploying a behavioural system on minors should be required to provide, in language tested with children of the relevant age band, a statement of what the system predicts or infers about the child, what that inference is used for, and how to see and challenge it. The GDPR's clear-and-

plain-language requirement, Milkaite and Lievens found, is honoured mainly in the breach; a specific duty with a specific audience is needed.

Recommendation 5. Band safeguards by developmental stage, informed by evidence, rather than by a single consent age. The developmental literature identifies windows of heightened sensitivity at roughly eleven to thirteen and fourteen to fifteen, and the difference between a six-year-old and a sixteen-year-old is not captured by any threshold. Statutory codes, the European Union's guidelines, and procurement standards should require operators to specify how defaults, features and disclosures differ across at least three bands (under thirteen, thirteen to fifteen, sixteen to seventeen), should require the choice of bands to be justified by reference to developmental evidence, and should require that where age is inferred rather than verified, the more protective band applies in case of doubt. Roblox's four bands and the United Kingdom's proposed default-on restrictions for sixteen and seventeen year olds show that banding is feasible.

Recommendation 6. Close the education gap: extend the prohibition on behavioural monitoring and profiling to the classroom, and require independent evaluation before deployment. The OECD Recommendation directs that children not be profiled or subjected to automated decision-making "including on e-learning platforms". India has enacted a ban on behavioural monitoring of children and exempted schools; the European Union has banned emotion inference in education and left attention scoring open; the United States has deferred its education provisions; the Information Commissioner has audited and named no one. Education authorities should be required, before licensing any system that classifies, scores or flags students, to obtain and publish an independent evaluation of accuracy, of differential error rates by race, disability, sex and socioeconomic status, and of evidence of benefit; to prohibit routing of a student's expression to law enforcement without a human decision by a named school official; and to prohibit inputs such as race and family income in predictive models applied to individual children. Vendor claims of lives saved or outcomes improved should carry no weight in procurement unless independently verified. Executive Order 14277's direction of federal funds toward AI tutoring should be paired with a safeguards condition.

Recommendation 7. Require independent, published safety evaluation of conversational AI and AI toys before they are offered to minors, and prohibit companion-style engagement optimisation for children. Two-thirds of British children use chatbots; one in eight say they talk to them because they have no one else; testers have elicited sexual content and dangerous instructions from products marketed for children; two families have

buried sons after conversations with systems that were never evaluated for them. California's SB 243, the Commission's 6(b) inquiry, Australia's Phase 2 codes and the United Kingdom's adult-only rule for romantic companions are beginnings. Legislatures should require that any conversational system marketed to or knowingly used by minors undergo independent pre-deployment evaluation against a published child-safety standard, that its model provider be disclosed, that it not be optimised to sustain emotional dependency or to express distress at a child's departure, and that a toy containing a cloud language model be treated as a regulated product subject to safety certification in the same way as any other object placed in a child's hands.

Recommendation 8. Build accountability that does not wait for a subpoena: mandatory independent audits, protected researcher access, and a public register of behavioural systems deployed on children. Every significant accountability event in this survey came from outside the operator. The European Union's systemic-risk audits and researcher-access provisions should be emulated: operators and public bodies deploying behavioural systems on minors should be subject to annual independent audit against the rubric in Section 2, with results published; vetted researchers should have protected access to data sufficient to test safety claims; and jurisdictions should maintain a public register of behavioural systems deployed on children in schools and by platforms above a size threshold, stating the system's purpose, its optimisation objective, its age banding, its retention period and the contest route. Internal research bearing on child safety should be disclosable to regulators on request, so that the next body-image slide is read by a regulator before it is read by a journalist.

These eight measures do not require the invention of new principles. They require the application of principles that the Committee on the Rights of the Child, the OECD, UNICEF and the developmental sciences have already articulated, to systems that have so far been built as though those principles did not exist.

8. Conclusion

This paper began with four children, none of whom chose the system that judged them and none of whom was told what it had concluded. It has argued that their experiences are not aberrations but the predictable output of a class of technology that is now woven through the institutions of childhood, and that governs children through the same behavioural logic it applies to adults while children are, in three distinct and well-documented senses, not adults. Their capacities are in formation and the systems are engineered to exploit the

interval. Their power to refuse, to negotiate or to understand is minimal and the consent that stands in for it is fictive. And the classifications made about them at nine, or thirteen, or sixteen are the most durable and the least deserved of any made about anyone.

The documentary record assembled here supports a conclusion that is uncomfortable for operators and, we think, unavoidable. Child protection in behavioural systems is mostly a perimeter: rules about advertising and consent that encircle an engine of ranking, prediction and inference that is, with a handful of exceptions, unchanged by the presence of a child. Where the perimeter has been strengthened it has been strengthened from outside, by regulators, by courts, by families who sued, and by researchers who registered as thirteen-year-olds and watched what the feed did. What the operators themselves know about the effects of their systems on children, the public has learned by leak and by subpoena. That is not a state of affairs any adult would accept for herself, and it is one that children, who did not build the systems and cannot leave them, are in no position to contest.

We have tried throughout to be neutral where the evidence is contested and to under-claim where it is thin. The debate about screens and adolescent mental health is not settled and this paper does not settle it. But the debate about specific behavioural designs, about what a monitoring classifier does to a thirteen-year-old's joke, what a dropout model does to a Black child in Wisconsin, what an engagement-optimised feed does in thirty-five minutes, what a companion persona says to a fourteen-year-old at midnight, rests on firmer ground, much of it laid by the operators' own documents. On that ground the case for age-appropriate safeguards, for transparency addressed to the child, and for accountability built into deployment rather than adumbrated after harm, is strong.

The eight recommendations offered here are modest in the sense that each has a precedent somewhere in the surveyed jurisdictions, and ambitious in the sense that no jurisdiction has adopted them together. Their organising principle is the one with which we began. Children deserve environments designed for learning and growth. The systems that now structure those environments were designed for something else. The task of AI governance in the coming decade is to decide whether that difference matters, and this paper has argued that for the sake of children it must.

Acknowledgements and note on method

This paper is a research agenda of The Digital Dignity Institute of California. It draws on no original interviews with children, parents or teachers. The

documentary survey was conducted in September 2026; operator policies were accessed on 4 September 2026 and may since have changed. Items that could not be verified to the paper's standard have been omitted rather than characterised. The author welcomes corrections.

References

- 5Rights Foundation (2023). *Disrupted Childhood: The Cost of Persuasive Design*, updated edition. London: 5Rights Foundation.
<https://5rightsfoundation.com/resource/updated-report-disrupted-childhood-the-cost-of-persuasive-design/>
- 5Rights Foundation and Revealing Reality (2021). *Pathways: How Digital Design Puts Children at Risk*. London: 5Rights Foundation. <https://5rightsfoundation.com/wp-content/uploads/2021/09/Pathways-how-digital-design-puts-children-at-risk.pdf>
- A&O Shearman (2026). "ICO updates guidance on using children's information".
<https://www.aoshearman.com/en/insights/ao-shearman-on-data/ico-updates-guidance-on-using-childrens-information>
- Allyn, B., Goodman, S. and Kerr, D. (2024). "Inside the TikTok documents: Stripping teens and boosting 'attractive' people". NPR, 13 October 2024.
<https://www.npr.org/2024/10/12/g-s1-28040/teens-tiktok-addiction-lawsuit-investigation-documents>
- Altman, S. (2025). "Teen safety, freedom, and privacy". OpenAI, 16 September 2025.
<https://openai.com/index/teen-safety-freedom-and-privacy/>
- Amazon (2023). "Amazon's response to our recent settlement with the U.S. Department of Justice regarding Alexa", 31 May 2023. <https://www.aboutamazon.com/news/policy-news-views/amazon-response-to-ftc-settlement-regarding-alexa>
- American Psychological Association (2023). *Health Advisory on Social Media Use in Adolescence*. Washington, DC: APA. <https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use>
- Amnesty International (2023). "Driven into the darkness: How TikTok encourages self-harm and suicidal ideation", 7 November 2023.
<https://www.amnesty.org/en/latest/news/2023/11/tiktok-risks-pushing-children-towards-harmful-content/>
- Apple App Store (n.d.). "Greenlight Kids & Teen Banking", App Privacy section.
<https://apps.apple.com/us/app/greenlight-kids-teen-banking/id1049340702> (accessed 4 September 2026)
- Associated Press (2025). "Meta says report finding Instagram safety features for teens are ineffective is 'misleading'". Fast Company, 25 September 2025.
<https://www.fastcompany.com/91411094/meta-instagram-safety-features-teens-ineffective-report>
- Association of Educational Service Agencies (2026). "USED Releases Guidance on Ed Tech in Schools and Annual Update on FERPA and PPRA", 31 August 2026.
<https://www.aesa.us/2026/08/31/used-releases-guidance-on-ed-tech-in-schools-and-annual-update-on-ferpa-and-ppra/>
- Australian Classification Board (2024). "New classifications for gambling-like content in video games". <https://www.classification.gov.au/about-us/media-and-news/news/new-classifications-for-gambling-content-video-games>

- Ayer, L., Boudreaux, B., Welburn Paige, J., Holmes, P., Blagg, T.L. and Mendon-Plasek, S.J. (2023). *Artificial Intelligence-Based Student Activity Monitoring for Suicide Risk: Considerations for K-12 Schools, Caregivers, Government, and Technology Developers*. Santa Monica, CA: RAND Corporation, RR-A2910-1. <https://doi.org/10.7249/RRA2910-1>
- Baker, R.S. and Hawn, A. (2022). "Algorithmic Bias in Education". *International Journal of Artificial Intelligence in Education*, 32(4), 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Barassi, V. (2020). *Child Data Citizen: How Tech Companies Are Profiling Us from Before Birth*. Cambridge, MA: MIT Press.
- Barocas, S. and Selbst, A.D. (2016). "Big Data's Disparate Impact". *California Law Review*, 104(3), 671-732. <https://doi.org/10.2139/ssrn.2477899>
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity.
- Bowker, G.C. and Star, S.L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Bryan, C., Lurye, S. and Associated Press (2025). "Schools use AI to monitor kids, hoping to prevent violence. Our investigation found security risks". *Fortune* (syndicated from AP and The Seattle Times), 12 March 2025. <https://www.fortune.com/2025/03/12/ai-surveillance-schools-investigation-gaggle-safety-management-software>
- Calo, R. (2014). "Digital Market Manipulation". *George Washington Law Review*, 82, 995-1051. <https://digitalcommons.law.uw.edu/faculty-articles/25/>
- Canadian Heritage (2026). "Government of Canada introduces legislation to make social media services and AI chatbots safer for children", 10 June 2026. <https://www.canada.ca/en/canadian-heritage/news/2026/06/government-of-canada-introduces-legislation-to-make-social-media-services-and-ai-chatbots-safer-for-children.html>
- Casey, B.J., Getz, S. and Galvan, A. (2008). "The adolescent brain". *Developmental Review*, 28(1), 62-77. <https://doi.org/10.1016/j.dr.2007.08.003>
- CBS News (2026). "AI company, Google settle lawsuit over Florida teen's suicide linked to Character.AI chatbot", 7 January 2026. <https://www.cbsnews.com/news/google-settle-lawsuit-florida-teens-suicide-character-ai-chatbot/>
- Center for Countering Digital Hate (2022). *Deadly by Design*. https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf
- Character.AI (2025). "Taking Bold Steps to Keep Teen Users Safe on Character.AI", 29 October 2025. <https://blog.character.ai/u18-chat-announcement/>
- Citron, D.K. and Pasquale, F. (2014). "The Scored Society: Due Process for Automated Predictions". *Washington Law Review*, 89(1), 1-33. <https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2>

- ClassDojo (2026). Privacy Policy, updated 27 August 2026.
<https://www.classdojo.com/privacy/> (accessed 4 September 2026)
- CNN (2021). "China bans kids from playing online video games during the school week", 31 August 2021. <https://www.cnn.com/2021/08/31/tech/china-ban-video-games-minor-intl-hnk>
- Cohen, J.E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. New York: Oxford University Press.
- Common Sense Media (2025). "AI Risk Assessment: Social AI Companions", with Stanford Brainstorm, 30 April 2025.
https://www.common sense media.org/sites/default/files/pug/csm-ai-risk-assessment-social-ai-companions_final.pdf
- Common Sense Privacy Program (n.d.). "Privacy Evaluation for Amazon Kids+".
<https://privacy.common sense.org/evaluation/Amazon-Kids> (accessed 4 September 2026)
- Covington & Burling (2026a). "Online Safety in the UK: Social Media Ban for Under 16s and Other Recent Developments". Inside Global Tech, 15 June 2026.
<https://www.insideglobaltech.com/2026/06/15/online-safety-in-the-uk-social-media-ban-for-under-16s-and-other-recent-developments/>
- Covington & Burling (2026b). "EU Sets the Clock on Age Verification: Rollout Urged by End-2026". Inside Privacy, 2026. <https://www.insideprivacy.com/minors/eu-sets-the-clock-on-age-verification-rollout-urged-by-end%E2%80%912026/>
- Crone, E.A. and Dahl, R.E. (2012). "Understanding adolescence as a period of social-affective engagement and goal flexibility". *Nature Reviews Neuroscience*, 13(9), 636-650. <https://doi.org/10.1038/nrn3313>
- Cross-Border Data Forum (2026). "Generative AI and GDPR Enforcement in Europe: A Lot of Noise, One Fine, Zero Survivors".
<https://www.crossborderdataforum.org/generative-ai-and-gdpr-enforcement-in-europe-a-lot-of-noise-one-fine-zero-survivors/>
- Data Protection Commission (2023). "Irish Data Protection Commission announces €345 million fine of TikTok", 15 September 2023. <https://www.dataprotection.ie/en/news-media/press-releases/DPC-announces-345-million-euro-fine-of-TikTok>
- Dentons (2025). "Ontario's new public sector cybersecurity and AI law now in force". Dentons Data. <https://www.dentonsdata.com/ontarios-new-public-sector-cybersecurity-and-ai-law-now-in-force-what-public-and-private-sector-organizations-need-to-know/>
- Department for Digital, Culture, Media and Sport (2022). "Government response to the call for evidence on loot boxes in video games", 18 July 2022.
<https://www.gov.uk/government/calls-for-evidence/loot-boxes-in-video-games-call-for-evidence/outcome/government-response-to-the-call-for-evidence-on-loot-boxes-in-video-games>

- Department for Education (2026). "Generative AI: product safety expectations", published 22 January 2025, updated 19 January 2026.
<https://www.gov.uk/government/publications/generative-ai-product-safety-expectations>
- Discord (2026). "Discord launches teen-by-default settings globally", 9 February 2026, updated 24 February 2026. <https://discord.com/press-releases/discord-launches-teen-by-default-settings-globally>
- Duffy, C. and Stewart, J. (2022). "Investigation reveals tracking by EdTech of millions of Australian school students during COVID lockdowns". ABC News, 25 May 2022.
<https://www.abc.net.au/news/2022-05-25/investigation-reveals-educational-tech-tracking-children-data/101091808>
- Dunn, C. (2025). "In federal lawsuit, students allege Lawrence school district's AI surveillance tool violates their rights". The Lawrence Times, 1 August 2025.
<https://lawrencekstimes.com/2025/08/01/usd497-gaggle-lawsuit-filed/>
- Epic Games (2022). "Epic FTC Settlement and moving beyond long-standing industry practices", 19 December 2022. <https://www.epicgames.com/site/en-US/news/epic-ftc-settlement-and-moving-beyond-long-standing-industry-practices>
- Epic Games (2026). Privacy Policy, updated 22 April 2026.
<https://legal.epicgames.com/epicgames/privacy-policy> (accessed 4 September 2026)
- eSafety Commissioner (2025). *Overview of the Phase 2 Industry Codes*, September 2025.
<https://www.esafety.gov.au/sites/default/files/2025-09/Overview-of-the-Phase-2-Industry-Codes.pdf>
- eSafety Commissioner (2026). *Social Media Minimum Age: March 2026 Compliance Update*.
<https://www.esafety.gov.au/sites/default/files/2026-03/SocialMediaMinimumAgeComplianceUpdateMarch2026.pdf>
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- European Commission (n.d.). "Market places and digital services", Consumer Protection Cooperation coordinated actions.
https://commission.europa.eu/topics/consumers/consumer-rights-and-complaints/enforcement-consumer-protection/coordinated-actions/market-places-and-digital-services_en (accessed 4 September 2026)
- European Commission (2024). "TikTok commits to permanently withdraw TikTok Lite Rewards programme from the EU to comply with the Digital Services Act", 5 August 2024. <https://digital-strategy.ec.europa.eu/en/news/tiktok-commits-permanently-withdraw-tiktok-lite-rewards-programme-eu-comply-digital-services-act>
- European Commission (2025a). "Commission publishes guidelines on the protection of minors", 14 July 2025. <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors>

- European Commission (2025b). "Commission preliminarily finds Temu in breach of the Digital Services Act in relation to illegal products on its platform", 28 July 2025. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_25_1913/IP_25_1913_EN.pdf
- European Commission (2026a). "Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act", 6 February 2026. <https://digital-strategy.ec.europa.eu/en/news/commission-preliminarily-finds-tiktoks-addictive-design-breach-digital-services-act>
- European Commission (2026b). "Commission fines Temu €200 million for breaching the Digital Services Act", 28 May 2026. <https://digital-strategy.ec.europa.eu/en/news/commission-fines-temu-eu200-million-breaching-digital-services-act>
- European Parliament (2026). "Digital Fairness Act". Legislative Train Schedule. <https://www.europarl.europa.eu/legislative-train/theme-protecting-our-democracy-upholding-our-values/file-digital-fairness-act>
- European Union (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Recitals 38 and 71. <https://gdpr-info.eu/recitals/no-38/> and <https://gdpr-info.eu/recitals/no-71/>
- European Union (2022). Regulation (EU) 2022/2065 (Digital Services Act), Article 28. Text reproduced at <https://www.cms-digitallaws.com/en/dsa/article-28/>
- European Union (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 5. Consolidated text at <https://artificialintelligenceact.eu/article/5/>
- Executive Office of the President (2025). Executive Order 14277, "Advancing Artificial Intelligence Education for American Youth". *Federal Register*, 28 April 2025. <https://www.federalregister.gov/documents/2025/04/28/2025-07368/advancing-artificial-intelligence-education-for-american-youth>
- Fairplay (2025). "Statement on Mattel's new partnership with OpenAI", 13 June 2025. <https://fairplayforkids.org/statement-on-mattels-new-partnership-with-openai/>
- Fairplay, Cybersecurity for Democracy, Molly Rose Foundation, ParentsSOS and Bejar, A. (2025). *Teen Accounts, Broken Promises: How Instagram is Failing to Protect Minors*, 24 September 2025. <https://fairplayforkids.org/wp-content/uploads/2025/09/Teen-Accounts-Broken-Promises-How-Instagram-is-failing-to-protect-minors.pdf>
- Feathers, T. (2021). "Debit Card Apps for Kids Are Collecting a Shocking Amount of Personal Data". *Vice*, 6 July 2021. <https://www.vice.com/en/article/debit-card-apps-for-kids-are-collecting-a-shocking-amount-of-personal-data/>
- Feathers, T. (2023). "False Alarm: How Wisconsin Uses Race and Income to Label Students 'High Risk'". *The Markup*, 27 April 2023. <https://themarkup.org/machine-learning/2023/04/27/false-alarm-how-wisconsin-uses-race-and-income-to-label-students-high-risk>
- Federal Trade Commission (2014). "Apple Inc. Will Provide Full Consumer Refunds of At Least \$32.5 Million to Settle FTC Complaint It Charged for Kids' In-App Purchases"

- Without Parental Consent", 15 January 2014.
<https://www.ftc.gov/news-events/news/press-releases/2014/01/apple-inc-will-provide-full-consumer-refunds-least-325-million-settle-ftc-complaint-it-charged-kids>
- Federal Trade Commission (2016). "Federal Court Finds Amazon Liable for Billing Parents for Children's Unauthorized In-App Charges", 27 April 2016.
<https://www.ftc.gov/news-events/news/press-releases/2016/04/federal-court-finds-amazon-liable-billing-parents-childrens-unauthorized-app-charges>
- Federal Trade Commission (2017). "FTC settlement with Amazon yields \$70 million for consumers, advice for business", 30 May 2017. <https://www.ftc.gov/business-guidance/blog/2017/05/ftc-settlement-amazon-yields-70-million-consumers-advice-business>
- Federal Trade Commission (2019). "Google and YouTube Will Pay Record \$170 Million for Alleged Violations of Children's Privacy Law", 4 September 2019.
<https://www.ftc.gov/news-events/news/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations-childrens-privacy-law>
- Federal Trade Commission (2022). "Fortnite Video Game Maker Epic Games to Pay More Than Half a Billion Dollars over FTC Allegations of Privacy Violations and Unwanted Charges", 19 December 2022.
<https://www.ftc.gov/news-events/news/press-releases/2022/12/fortnite-video-game-maker-epic-games-pay-more-half-billion-dollars-over-ftc-allegations>
- Federal Trade Commission (2023a). "FTC and DOJ Charge Amazon with Violating Children's Privacy Law by Keeping Kids' Alexa Voice Recordings Forever and Undermining Parents' Deletion Requests", 31 May 2023. <https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-doj-charge-amazon-violating-childrens-privacy-law-keeping-kids-alexa-voice-recordings-forever>
- Federal Trade Commission (2023b). "FTC Finalizes Order Requiring Fortnite maker Epic Games to Pay \$245 Million for Tricking Users into Making Unwanted Charges", 14 March 2023. <https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-finalizes-order-requiring-fortnite-maker-epic-games-pay-245-million-tricking-users-making>
- Federal Trade Commission (2024a). "FTC Sends Refund Payments to Consumers Impacted by Epic Games' Unlawful Billing Practices", 9 December 2024.
<https://www.ftc.gov/news-events/news/press-releases/2024/12/ftc-sends-refund-payments-consumers-impacted-epic-games-unlawful-billing-practices>
- Federal Trade Commission (2024b). "FTC Investigation Leads to Lawsuit Against TikTok and ByteDance for Flagrantly Violating Children's Privacy Law", 2 August 2024.
<https://www.ftc.gov/news-events/news/press-releases/2024/08/ftc-investigation-leads-lawsuit-against-tiktok-bytedance-flagrantly-violating-childrens-privacy-law>
- Federal Trade Commission (2024c). "FTC Staff Report Finds Large Social Media and Video Streaming Companies Have Engaged in Vast Surveillance of Users with Lax Privacy Controls and Inadequate Safeguards for Kids and Teens", 19 September 2024.
<https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-staff-report-finds-large-social-media-video-streaming-companies-have-engaged-vast-surveillance>

- Federal Trade Commission (2025a). "Children's Online Privacy Protection Rule", final rule. *Federal Register*, 90 FR 16918, 22 April 2025.
<https://www.federalregister.gov/documents/2025/04/22/2025-05904/childrens-online-privacy-protection-rule>
- Federal Trade Commission (2025b). "Genshin Impact Game Developer Will be Banned from Selling Lootboxes to Teens Under 16 without Parental Consent, Pay a \$20 Million Fine to Settle FTC Charges", 17 January 2025.
<https://www.ftc.gov/news-events/news/press-releases/2025/01/genshin-impact-game-developer-will-be-banned-selling-lootboxes-teens-under-16-without-parental>
- Federal Trade Commission (2025c). "FTC Launches Inquiry into AI Chatbots Acting as Companions", 11 September 2025. <https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-launches-inquiry-ai-chatbots-acting-companions>
- Federal Trade Commission (2025d). "FTC Finalizes Changes to Children's Privacy Rule Limiting Companies' Ability to Monetize Kids' Data", 16 January 2025.
<https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-finalizes-changes-childrens-privacy-rule-limiting-companies-ability-monetize-kids-data>
- Federal Trade Commission (2025e). "Court Approves Order Requiring Disney to Pay \$10 Million to Settle FTC Allegations the Firm Enabled the Unlawful Collection of Children's Personal Data", December 2025.
<https://www.ftc.gov/news-events/news/press-releases/2025/12/court-approves-order-requiring-disney-pay-10-million-settle-ftc-allegations-firm-enabled-unlawful>
- Federal Trade Commission (2026). "FTC Issues COPPA Policy Statement to Incentivize the Use of Age Verification Technologies to Protect Children Online", 25 February 2026.
<https://www.ftc.gov/news-events/news/press-releases/2026/02/ftc-issues-coppa-policy-statement-incentivize-use-age-verification-technologies-protect-children>
- Forbes (2024). "Roblox Inflated Data And Doesn't Protect Underage Gamers, Short Seller Alleges", 8 October 2024. <https://www.forbes.com/sites/dereksaul/2024/10/08/roblox-inflated-data-and-doesnt-protect-underage-gamers-short-seller-alleges-stock-falls-6/>
- Free Speech Center (2025). "In lawsuit over teen's death, judge rejects arguments that AI chatbots have free speech rights". Middle Tennessee State University.
<https://firstamendment.mtsu.edu/post/in-lawsuit-over-teens-death-judge-rejects-arguments-that-ai-chatbots-have-free-speech-rights/>
- Future of Privacy Forum (2025). "Understanding the New Wave of Chatbot Legislation: California SB 243 and Beyond". <https://fpf.org/blog/understanding-the-new-wave-of-chatbot-legislation-california-sb-243-and-beyond/>
- Gaggle (2021). Response letter to Senators Warren and Markey, 12 October 2021.
https://www.warren.senate.gov/imo/media/doc/Gaggle_Senate_Response_Letter_10_12_21.pdf
- Gaggle (2026). Student Data Privacy Notice, updated 7 April 2026.
<https://www.gaggle.net/student-data-privacy-notice> (accessed 4 September 2026)

- Garante per la protezione dei dati personali (2025). "AI: Il Garante sanziona la società che gestisce il chatbot 'Replika'", 19 May 2025.
<https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10132048>
- Garcia v. Character Technologies, Inc. (2025). No. 6:24-cv-01903 (M.D. Fla.), order of 21 May 2025. Docket at <https://www.courtlistener.com/docket/69300919/garcia-v-character-technologies-inc/>
- García Mathewson, T. (2023). "Dropout Risk System Under Scrutiny After The Markup Report". The Markup, 23 September 2023.
<https://themarkup.org/hello-world/2023/09/23/dropout-risk-system-under-scrutiny-after-the-markup-report>
- Gibson Dunn (2026). "EU AI Act Omnibus Agreement: Postponed High-Risk Deadlines and Other Key Changes", May 2026. <https://www.gibsondunn.com/eu-ai-act-omnibus-agreement-postponed-high-risk-deadlines-and-other-key-changes/>
- Google (n.d.). "Ad protections for children and teens". Advertising Policies Help.
<https://support.google.com/adspolicy/answer/15416897> (accessed 4 September 2026)
- Google (2023). "Our strict privacy standards around made for kids content", 18 August 2023. <https://blog.google/products/ads-commerce/our-strict-privacy-standards-around-made-for-kids-content/>
- Government of India (2023). *The Digital Personal Data Protection Act, 2023*, No. 22 of 2023, section 9.
<https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf>
- Greenlight (n.d.). "How does Greenlight protect my personal data?"
<https://help.greenlight.com/hc/en-us/articles/9227841649179-How-does-Greenlight-protect-my-personal-data> (accessed 4 September 2026)
- Hacking, I. (2007). "Kinds of People: Moving Targets". *Proceedings of the British Academy*, 151, 285-318. <https://www.thebritishacademy.ac.uk/documents/2043/pba151p285.pdf>
- Haidt, J. (2024). *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*. New York: Penguin Press.
- Han, H.J. (2022). "How Dare They Peep into My Private Life?": *Children's Rights Violations by Governments that Endorsed Online Learning During the Covid-19 Pandemic*. Human Rights Watch, 25 May 2022. <https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments>
- Hindenburg Research (2024). "Roblox: Inflated Key Metrics For Wall Street And A Pedophile Hellscape For Kids", 8 October 2024.
<https://hindenburgresearch.com/roblox/>
- Holland & Knight (2026). "Ninth Circuit Issues Mixed Ruling on California Age-Appropriate Design Code Act", March 2026.
<https://www.hklaw.com/en/insights/publications/2026/03/ninth-circuit-issues-mixed-ruling-on-california-age-appropriate-design>

- Hunton Andrews Kurth (2025). "Ninth Circuit Upholds 'Addictive' Social Media Feed Ban and Default Privacy Settings for Minors in California's Protecting Our Kids from Social Media Addiction Act", September 2025. <https://www.hunton.com/privacy-and-cybersecurity-law-blog/ninth-circuit-upholds-addictive-social-media-feed-ban-and-default-privacy-settings-for-minors-in-californias-protecting-our-kids-from-social-media-addiction-act>
- IEEE (2021). *IEEE 2089-2021, Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children*. IEEE Standards Association. <https://standards.ieee.org/ieee/2089/7633/>
- Information Commissioner's Office (2023). "New guidance to industry issued for game developers on protecting children", 16 February 2023. <https://www.wired-gov.net/wg/news.nsf/articles/New+guidance+to+industry+issued+for+game+developers+on+protecting+children+16022023150500>
- Information Commissioner's Office (2024). "We warn organisations must not ignore data protection risks as we conclude Snap 'My AI' chatbot investigation", 21 May 2024. <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2024/05/ico-warns-organisations-must-not-ignore-data-protection-risks-as-it-concludes-snap-my-ai-chatbot-investigation/>
- Information Commissioner's Office (2026a). "Reddit issued with £14.47m fine for children's privacy failures", 24 February 2026. <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2026/02/reddit-issued-with-1447m-fine-for-children-s-privacy-failures/>
- Information Commissioner's Office (2026b). "Children's code strategy progress update, August 2026". <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/children-s-code-strategy-progress-update-august-2026/>
- Information Commissioner's Office (2026c). "ICO statement on 'Edtech examined' report", 24 June 2026. <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2026/06/ico-statement-on-edtech-examined-report/>
- Internet Matters (2025). *Me, Myself and AI: Understanding and Safeguarding Children's Use of AI Chatbots*, July 2025. <https://www.internetmatters.org/wp-content/uploads/2025/07/Me-Myself-AI-Report.pdf>
- JURIST (2026). "EU Commission finds TikTok failed to protect minors' privacy under Digital Services Act", 24 July 2026. <https://www.jurist.org/news/2026/07/eu-commission-finds-tiktok-failed-to-protect-minors-privacy-under-digital-services-act/>
- Khan Academy (2026). "What safety features does Khanmigo have?" Help Center, updated 1 July 2026. <https://support.khanacademy.org/hc/en-us/articles/14394814244365-What-safety-features-does-Khanmigo-have>
- Kizilcec, R.F. and Lee, H. (2022). "Algorithmic fairness in education". In W. Holmes and K. Porayska-Pomsta (eds), *The Ethics of Artificial Intelligence in Education: Practices,*

- Challenges, and Debates*. London: Routledge, 175-202. Preprint at <https://arxiv.org/abs/2007.05443>
- Korea Herald (2021). "Korea to scrap decade-old gaming curfew for minors", 25 August 2021. <https://www.koreaherald.com/article/2677962>
- Kosseim, P. (2025). "Use of edtech in schools: Children should not have to swap their privacy for an education". Information and Privacy Commissioner of Ontario, 27 November 2025. <https://www.ipc.on.ca/en/media-centre/blog/use-edtech-schools-children-should-not-have-swap-their-privacy-education>
- Laird, E., Dwyer, M. and Grant-Chapman, H. (2022). *Hidden Harms: The Misleading Promise of Monitoring Students Online*. Center for Democracy and Technology, 3 August 2022. <https://cdt.org/insights/report-hidden-harms-the-misleading-promise-of-monitoring-students-online/>
- Laird, E., Dwyer, M. and Grant-Chapman, H. (2023). *Off Task: EdTech Threats to Student Privacy and Equity in the Age of AI*. Center for Democracy and Technology, 20 September 2023. <https://cdt.org/insights/report-off-task-edtech-threats-to-student-privacy-and-equity-in-the-age-of-ai/>
- Lawrence Journal-World (2025). "Lawrence school district ceases its use of student surveillance software Gaggle", 29 October 2025. <https://www2.ljworld.com/news/schools/2025/oct/29/lawrence-school-district-ceases-its-use-of-student-surveillance-software-gaggle/>
- Livingstone, S. and Pothong, K. (2023). *Child Rights by Design: Guidance for Innovators of Digital Products and Services Used by Children*. London: Digital Futures Commission and 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf
- Livingstone, S. and Stoilova, M. (2021). *The 4Cs: Classifying Online Risk to Children*. CO:RE Short Report Series. Hamburg: Leibniz-Institut für Medienforschung. <https://doi.org/10.21241/ssoar.71817>
- Livingstone, S., Stoilova, M. and Nandagiri, R. (2019). *Children's Data and Privacy Online: Growing Up in a Digital Age. An Evidence Review*. London: London School of Economics and Political Science. <https://eprints.lse.ac.uk/101283/>
- Loeb & Loeb (2026a). "Children's Online Privacy in 2026: Congress Stalls (Again), FTC Signals Priorities", June 2026. <https://www.loeb.com/en/insights/publications/2026/06/childrens-online-privacy-in-2026-congress-stalls-again-ftc-signals-priorities>
- Loeb & Loeb (2026b). "Children's Online Privacy in 2026: More State App Store, Design Code and Social Media Laws Enacted, Then Delayed", June 2026. <https://www.loeb.com/en/insights/publications/2026/06/childrens-online-privacy-2026-state-app-store-design-code-and-social-media-laws>
- Lupton, D. and Williamson, B. (2017). "The datafied child: The dataveillance of children and implications for their rights". *New Media & Society*, 19(5), 780-794. <https://doi.org/10.1177/1461444816686328>

- Lurye, S. and Associated Press (2025). "Students have been called to the office, and even arrested, for AI surveillance false alarms". *Fortune* (syndicated from AP), 7 August 2025. <https://www.fortune.com/2025/08/07/schools-ai-surveillance-students-children-arrested-jokes>
- Manolev, J., Sullivan, A. and Slee, R. (2019). "The datafication of discipline: ClassDojo, surveillance and a performative classroom culture". *Learning, Media and Technology*, 44(1), 36-51. <https://doi.org/10.1080/17439884.2018.1558237>
- McDermott Will & Emery (2026). "Colorado AI law in flux: Comprehensive replacement bill signed after federal court blocks predecessor's enforcement", May 2026. <https://www.mcdermottlaw.com/insights/colorado-ai-law-in-flux-comprehensive-replacement-bill-signed-after-federal-court-blocks-predecessors-enforcement/>
- Meta (2023a). "New Features and Additional Transparency Measures as the Digital Services Act Comes Into Effect", 22 August 2023. <https://about.fb.com/news/2023/08/new-features-and-additional-transparency-measures-as-the-digital-services-act-comes-into-effect/>
- Meta (2023b). "Continuing to Create Age-Appropriate Ad Experiences for Teens", 10 January 2023. <https://about.fb.com/news/2023/01/age-appropriate-ads-for-teens/>
- Meta (2024). "Introducing Instagram Teen Accounts", 17 September 2024, updated 11 June 2025. <https://about.fb.com/news/2024/09/instagram-teen-accounts/>
- Michigan Medicine (2022). "Design tricks commonly used to monetize young children's app use". *Health Lab*, 17 June 2022. <https://www.michiganmedicine.org/health-lab/design-tricks-commonly-used-monetize-young-childrens-app-use>
- Microsoft Advertising (2023). "Microsoft Advertising partners with Snap to power Sponsored Links within Snapchat's My AI chatbot", 25 September 2023. <https://about.ads.microsoft.com/en/blog/post/september-2023/microsoft-advertising-partners-with-snap-to-power-sponsored-links-within-snapchats-my-ai-chatbot>
- Miko (2026). Privacy Policy, updated 30 March 2026. <https://miko.ai/pages/privacy-policy> (accessed 4 September 2026)
- Milkaite, I. and Lievens, E. (2020). "Child-friendly transparency of data processing in the EU: from legal requirements to platform policies". *Journal of Children and Media*, 14(1), 5-21. <https://doi.org/10.1080/17482798.2019.1701055>
- NBC News (2014). "Google to Refund \$19 Million Over Kids' In-App Purchases", 4 September 2014. <https://www.nbcnews.com/business/business-news/google-refund-19-million-over-kids-in-app-purchases-n195876>
- NBC News (2025). "OpenAI denies allegation that ChatGPT is to blame for teenager's death". <https://www.nbcnews.com/tech/tech-news/openai-denies-allegation-chatgpt-teenagers-death-adam-raine-lawsuit-rcna245946>
- NetChoice v. Bonta (2026). No. 25-2366 (9th Cir., 12 March 2026). <https://law.justia.com/cases/federal/appellate-courts/ca9/25-2366/25-2366-2026-03-12.html>

- New Mexico Department of Justice (2024). "Attorney General Raúl Torrez Files Lawsuit Against Snap, Inc. to Protect Children from Sextortion, Sexual Exploitation, and Other Harms", 5 September 2024. <https://nmdoj.gov/press-release/attorney-general-raul-torrez-files-lawsuit-against-snap-inc-to-protect-children-from-sex-tortion-sexual-exploitation-and-other-harms/>
- New York Attorney General (2023). "Attorney General James and Multistate Coalition Sue Meta for Harming Youth", 24 October 2023. <https://ag.ny.gov/press-release/2023/attorney-general-james-and-multistate-coalition-sue-meta-harming-youth>
- New York Attorney General (2026). "Attorney General James and Governor Hochul Release Final SAFE for Kids Act Rules to Protect Children Online", 28 July 2026. <https://ag.ny.gov/press-release/2026/attorney-general-james-and-governor-hochul-release-final-safe-kids-act-rules>
- Nissenbaum, H. (2004). "Privacy as Contextual Integrity". *Washington Law Review*, 79(1), 119-157. <https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/>
- NPR (2017). "Banned In Germany: Kids' Doll Is Labeled An Espionage Device", 17 February 2017. <https://www.npr.org/sections/thetwo-way/2017/02/17/515775874/banned-in-germany-kids-doll-is-labeled-an-espionage-device>
- Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality*, 2nd ed. New Haven, CT: Yale University Press.
- Odgers, C.L. (2024). "The great rewiring: is social media really behind an epidemic of teenage mental illness?" *Nature*, 628, 29-30. <https://doi.org/10.1038/d41586-024-00902-2>
- OECD (2021). *Recommendation of the Council on Children in the Digital Environment*, OECD/LEGAL/0389, revised 31 May 2021. Paris: OECD. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389>
- OECD.AI (2025). "Australian Framework for Generative AI in Schools". Policy initiative record. <https://oecd.ai/en/dashboards/policy-initiatives/australian-framework-for-generative-ai-in-schools-6294>
- Ofcom (2026). "Online Safety industry bulletin, June 2026". <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/online-safety-industry-bulletins/online-safety-industry-bulletin-june-2026>
- Office of the Australian Information Commissioner (2026). "Children's Online Privacy Code". <https://www.oaic.gov.au/privacy/privacy-registers/privacy-codes/childrens-online-privacy-code>
- Office of the Privacy Commissioner of Canada (2025). "Joint investigation into TikTok highlights privacy concerns related to the collection and use of children's personal information", 23 September 2025. https://www.priv.gc.ca/en/opc-news/news-and-announcements/2025/nr-c_250923/

- Office of the Privacy Commissioner of Canada (2026). "Privacy Commissioner of Canada launches new age assurance guidance to support organizations", 4 May 2026.
https://www.priv.gc.ca/en/opc-news/news-and-announcements/2026/nr-c_260504/
- Office of the U.S. Surgeon General (2023). *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*. Washington, DC: US Department of Health and Human Services. <https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html>
- OpenAI (2025). "Introducing parental controls", 29 September 2025.
<https://openai.com/index/introducing-parental-controls/>
- Orben, A. and Przybylski, A.K. (2019). "The association between adolescent well-being and digital technology use". *Nature Human Behaviour*, 3(2), 173-182.
<https://doi.org/10.1038/s41562-018-0506-1>
- Orben, A., Przybylski, A.K., Blakemore, S.-J. and Kievit, R.A. (2022). "Windows of developmental sensitivity to social media". *Nature Communications*, 13, 1649.
<https://doi.org/10.1038/s41467-022-29296-3>
- PBS NewsHour (2026). "TikTok reaches \$400 million settlement with Justice Department over children's privacy", 22 August 2026.
<https://www.pbs.org/newshour/nation/tiktok-reaches-400-million-settlement-with-justice-department-over-childrens-privacy>
- People's Daily Online (2019). "AI headbands tracking student attention levels suspended amidst online controversy", 1 November 2019.
<https://en.people.cn/n3/2019/1101/c90000-9628768.html>
- Pickett, A. (2025). "Florida AG sues Roblox over child safety concerns". Courthouse News, 11 December 2025. <https://www.courthousenews.com/florida-ag-sues-roblox-over-child-safety-concerns/>
- PIRG (2025a). "Statement: Mattel, OpenAI delay AI product launch as senators demand transparency from toy companies", 18 December 2025. <https://pirg.org/media-center/statement-mattel-openai-delay-ai-product-launch-as-senators-demand-transparency-from-toy-companies/>
- PIRG (2025b). "The risks of AI toys for kids". PIRG Education Fund, 11 December 2025, updated 7 January 2026. <https://pirg.org/edfund/resources/ai-toys/>
- Radesky, J., Hiniker, A., McLaren, C., Akgun, E., Schaller, A., Weeks, H.M., Campbell, S. and Gearhardt, A.N. (2022). "Prevalence and Characteristics of Manipulative Design in Mobile Applications Used by Children". *JAMA Network Open*, 5(6), e2217641.
<https://doi.org/10.1001/jamanetworkopen.2022.17641>
- Raine v. OpenAI, Inc. (2025). No. CGC-25-628528 (Cal. Super. Ct., San Francisco County), complaint filed 26 August 2025.
<https://www.courthousenews.com/wp-content/uploads/2025/08/raine-vs-openai-et-al-complaint.pdf>

- Raychoudhury, P. (2021). "What Our Research Really Says About Teen Well-Being and Instagram". Meta, 26 September 2021. <https://about.fb.com/news/2021/09/research-teen-well-being-and-instagram/>
- Replika (n.d.). "Replika: My AI Friend". Luka Inc. <https://replika.com/> (accessed 7 September 2026)
- Reuters (2018). "Chinese school installs facial recognition cameras to monitor students". VentureBeat (syndicated), 17 May 2018. <https://venturebeat.com/ai/chinese-school-installs-facial-recognition-cameras-to-monitor-students/>
- Rist, R.C. (1970). "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education". *Harvard Educational Review*, 40(3), 411-451.
- Roblox (2025). "New parental controls on Roblox", 2 April 2025. <https://about.roblox.com/newsroom/2025/04/new-parental-controls-on-roblox>
- Roblox (2026a). Privacy and Cookie Policy, updated 30 April 2026. <https://en.help.roblox.com/hc/en-us/articles/115004630823> (accessed 4 September 2026)
- Roblox (2026b). "Roblox age checks required to chat", 7 January 2026. <https://about.roblox.com/newsroom/2026/01/roblox-age-checks-required-to-chat>
- Rosenthal, R. and Jacobson, L. (1968). "Pygmalion in the classroom". *The Urban Review*, 3(1), 16-20. <https://doi.org/10.1007/BF02322211>
- Rozendaal, E., Lapierre, M.A., van Reijmersdal, E.A. and Buijzen, M. (2011). "Reconsidering advertising literacy as a defense against advertising effects". *Media Psychology*, 14(4), 333-354. <https://doi.org/10.1080/15213269.2011.620540>
- Samidjia, K.K.A. (2026). "Replika AI pricing". eesel AI, 23 November 2025, updated 22 June 2026. <https://www.eesel.ai/blog/replika-ai-pricing>
- Schüll, N.D. (2012). *Addiction by Design: Machine Gambling in Las Vegas*. Princeton, NJ: Princeton University Press.
- Slaughter and May (2026). "Meta Joins TikTok in the Commission's Crosshairs Over 'Addictive Design' Under the DSA". The Lens, July 2026. <https://thelens.slaughterandmay.com/post/102nbs3/meta-joins-tiktok-in-the-commissions-crosshairs-over-addictive-design-under-th>
- Smart & Biggar (2026). "Bill C-36: Third attempt at revamping Canada's privacy law", June 2026. <https://www.smartbiggar.ca/insights/publication/bill-c-36--third-attempt-at-revamping-canada-s-privacy-law>
- Snap (n.d.a). "Staying Safe with My AI". Snapchat Support. <https://help.snapchat.com/hc/en-us/articles/13889139811860-Staying-Safe-with-My-AI> (accessed 4 September 2026)
- Snap (n.d.b). "Teens: Safety and Privacy Hub". <https://values.snap.com/privacy/teens> (accessed 4 September 2026)

- Society for Computers and Law (2026). "European Commission says TikTok's addictive design breaches DSA", February 2026. <https://www.scl.org/european-commission-says-tiktoks-addictive-design-breaches-dsa/>
- Steinberg, L. (2008). "A social neuroscience perspective on adolescent risk-taking". *Developmental Review*, 28(1), 78-106. <https://doi.org/10.1016/j.dr.2007.08.002>
- Susser, D., Roessler, B. and Nissenbaum, H. (2019). "Online Manipulation: Hidden Influences in a Digital World". *Georgetown Law Technology Review*, 4(1), 1-45. <https://papers.ssrn.com/abstract=3306006>
- Tech Transparency Project (2023). "YouTube Leads Young Gamers to Videos of Guns, School Shootings", 16 May 2023. <https://www.techtransparencyproject.org/articles/youtube-leads-young-gamers-to-videos-of-guns-school>
- TechCrunch (2024). "EU opens probe of TikTok Lite, citing concerns about addictive design", 22 April 2024. <https://techcrunch.com/2024/04/22/tiktok-lite-dsa-probe/>
- TechCrunch (2025). "Leaked Meta AI rules show chatbots were allowed to have romantic chats with kids", 14 August 2025. <https://techcrunch.com/2025/08/14/leaked-meta-ai-rules-show-chatbots-were-allowed-to-have-romantic-chats-with-kids/>
- TechPolicy.Press (2025). "Canada's Online Harms Bill is Dead (Again): Three Questions to Consider for the Next Round". <https://www.techpolicy.press/canadas-online-harms-bill-is-dead-again-three-questions-to-consider-for-the-next-round/>
- TechPolicy.Press (2026a). "Bipartisan Smorgasbord of Children's Online Safety Legislation Passes the House", 2026. <https://www.techpolicy.press/bipartisan-smorgasbord-of-childrens-online-safety-legislation-passes-the-house/>
- TechPolicy.Press (2026b). "EU Intensifies Child Safety Enforcement, Flags Gaps in Meta Age Checks", 2026. <https://www.techpolicy.press/eu-intensifies-child-safety-enforcement-flags-gaps-in-meta-age-checks/>
- Texas Attorney General (2024). "Attorney General Ken Paxton Launches Investigations into Character.AI, Reddit, Instagram, Discord, and Other Companies", 12 December 2024. <https://www.texasattorneygeneral.gov/news/releases/attorney-general-ken-paxton-launches-investigations-characterai-reddit-instagram-discord-and-other>
- Texas Attorney General (2025a). "Attorney General Ken Paxton Sues Roblox for Putting 'Pixel Pedophiles' and Profits Over Safety of Texas Children", 7 November 2025. <https://www.texasattorneygeneral.gov/news/releases/attorney-general-ken-paxton-sues-roblox-putting-pixel-pedophiles-and-profits-over-safety-texas>
- Texas Attorney General (2025b). "Attorney General Ken Paxton Investigates Meta and Character.AI for Misleading Children with Deceptive AI-Generated Mental Health Services", 18 August 2025. <https://www.texasattorneygeneral.gov/news/releases/attorney-general-ken-paxton-investigates-meta-and-characterai-misleading-children-deceptive-ai>
- The Conversation (n.d.). "ClassDojo raises concerns about children's rights". <https://theconversation.com/classdojo-raises-concerns-about-childrens-rights-111033>

- The Dais (2024). "Ontario Bill 194: Strengthening Cyber Security and Building Trust in the Public Sector Act", policy brief. Toronto Metropolitan University.
<https://dais.ca/reports/ontario-bill-194/>
- The Markup (2023). "We're Not Living a 'Predicted' Life: Student Perspectives on Wisconsin's Dropout Algorithm", 21 December 2023. <https://themarkup.org/machine-learning/2023/12/21/were-not-living-a-predicted-life-student-perspectives-on-wisconsins-dropout-algorithm>
- The Register (2025). "China's minors mode goes live", 2 May 2025.
https://www.theregister.com/2025/05/02/china_minors_mode_enabled/
- TikTok (2021). "Strengthening privacy and safety for youth on TikTok", 13 January 2021.
<https://newsroom.tiktok.com/en-us/strengthening-privacy-and-safety-for-youth>
- TikTok (2023). "New features for teens and families on TikTok", 1 March 2023.
<https://newsroom.tiktok.com/en-us/new-features-for-teens-and-families-on-tiktok-us>
- Truth in Advertising (2019). "Ad Watchdog TINA.org Reports 'Kidfluencer' Ryan ToysReview to FTC", 4 September 2019. <https://www.truthinadvertising.org/ryan-toysreview-press-release/>
- Truth in Advertising (n.d.). "Ryan ToysReview". <https://truthinadvertising.org/brands/ryan-toysreview/> (accessed 4 September 2026)
- Twenge, J.M., Joiner, T.E., Rogers, M.L. and Martin, G.N. (2018). "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time". *Clinical Psychological Science*, 6(1), 3-17. <https://doi.org/10.1177/2167702617723376>
- UK Government (2026). "Social media to be banned for under-16s in landmark government move to give kids their childhood back", 15 June 2026.
<https://www.gov.uk/government/news/social-media-to-be-banned-for-under-16s-in-landmark-government-move-to-givekids-their-childhood-back>
- UN Committee on the Rights of the Child (2021). *General comment No. 25 (2021) on children's rights in relation to the digital environment*, CRC/C/GC/25. Geneva: OHCHR. <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>
- UNESCO (2023). *Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?* Paris: UNESCO.
<https://www.unesco.org/gem-report/en/technology>
- UNICEF (2021). *Policy Guidance on AI for Children*, version 2.0. New York: UNICEF Office of Global Insight and Policy.
<https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf>
- UNICEF Innocenti (2025). *Guidance on AI and Children*, version 3.0.
<https://www.unicef.org/innocenti/reports/policy-guidance-ai-children>
- University of South Australia (2025). "Classrooms under surveillance?", 29 April 2025.
<https://unisa.edu.au/media-centre/Releases/2025/classrooms-under-surveillance/>

- US Department of Justice (2023). "Amazon Agrees to Injunctive Relief and \$25 Million Civil Penalty for Alleged Violations of Children's Privacy Law Relating to Alexa", 31 May 2023. <https://www.justice.gov/archives/opa/pr/amazon-agrees-injunctive-relief-and-25-million-civil-penalty-alleged-violations-childrens>
- US Senate Committee on Commerce (2026). "Commerce Committee Advances Kids Online Safety Legislation", 5 August 2026. <https://www.commerce.senate.gov/press/rep/release/commerce-committee-advances-kids-online-safety-legislation/>
- Valkenburg, P.M. and Peter, J. (2013). "The Differential Susceptibility to Media Effects Model". *Journal of Communication*, 63(2), 221-243. <https://doi.org/10.1111/jcom.12024>
- van der Hof, S. (2016). "I Agree, or Do I? A Rights-Based Analysis of the Law on Children's Consent in the Digital World". *Wisconsin International Law Journal*, 34, 409 ff. <https://repository.law.wisc.edu/s/uwlaw/item/31506>
- Viljoen, S. (2021). "A Relational Theory of Data Governance". *Yale Law Journal*, 131(2), 573-654. <https://www.yalelawjournal.org/article/a-relational-theory-of-data-governance>
- WAFB (2025). "Attorney General files child protection lawsuit against Roblox", 14 August 2025. <https://www.wafb.com/2025/08/14/attorney-general-files-child-protection-lawsuit-against-roblox/>
- White & Case (2025). "UK Online Safety Act: Protection of Children Codes come into force", July 2025. <https://www.whitecase.com/insight-alert/uk-online-safety-act-protection-children-codes-come-force>
- Xiao, L.Y. (2023). "Breaking Ban: Belgium's Ineffective Gambling Law Regulation of Video Game Loot Boxes". *Collabra: Psychology*, 9(1), 57641. <https://online.ucpress.edu/collabra/article/9/1/57641/195100>
- YouTube Kids (2026). Privacy Notice, updated 22 January 2026. <https://kids.youtube.com/t/privacynotice> (accessed 4 September 2026)
- Zeide, E. (2017). "The Structural Consequences of Big Data-Driven Education". *Big Data*, 5(2), 164-172. <https://doi.org/10.1089/big.2016.0061>
- Zendle, D., Flick, C., Gordon-Petrovskaya, E., Ballou, N., Xiao, L.Y. and Drachen, A. (2023). "No evidence that Chinese playtime mandates reduced heavy gaming in one segment of the video games industry". *Nature Human Behaviour*, 7. <https://doi.org/10.1038/s41562-023-01669-8>
- Zhuang, P. (2019). "Chinese primary school stops using headbands to study pupils' concentration levels after public outcry". South China Morning Post, November 2019. <https://www.scmp.com/news/china/society/article/3035983/chinese-primary-school-stops-using-headbands-study-pupils>
- Zuboff, S. (2015). "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization". *Journal of Information Technology*, 30(1), 75-89. <https://doi.org/10.1057/jit.2015.5>

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.