From Generic Guardrails to Pillars in the Structure of AI: Specificity in Paediatric Neurodisability & Neurodivergence

W.J. Farr[1] & I.S. Lee[2]

[1,2]*Faculty of Education, University of Cambridge, Hills Road, Cambridge, UK*

[1]wjf33@cam.ac.uk, [2]isl29@cam.ac.uk

DOI: 10.54337/aau.icdvrat26.18

Abstract

When so much of AI technological development is targeting strong and underlying elements of the human condition, mandatory markers in the sand are required. Market self-regulation does not work, as shown by the attention economy. What is needed are codified pillars, not just *guardrails*, to protect those with neurodisability and those who are neurodiverse. There are two required pillars: protection from social murder, where systemic inequity leads to preventable deaths, and protection from psychological attachment abuse arising from the emerging attachment economy. Existing AI regulatory frameworks raise questions about how vulnerabilities, disability, and dehumanised pattern recognition are recognised or obscured. This project maps AI regulations from this perspective.

Introduction

There is a gap between what regulatory approvals for Artificial Intelligence (EU Parliament and Council, 2024) should do and what they achieve. Regulation protects the public, especially the most vulnerable, such as children, from exploitation by AI tools. The EU act was the widest ranging in its limitations on the extent of AI applications, yet these regulations mention *vulnerable children* only inferentially, i.e., generic references to a wider and protected characteristic such as disability. Children with Neurodisability and Neurodivergence are unaccounted for. Other major adopters of regulations, (i.e., the United States and the United Kingdom) do even less. We explore artificial intelligence regulation through that lens, arguing that current AI regulatory efforts are cursory, only recognise "vulnerable populations" in generic terms, and are therefore not well prepared to address subtle, dehumanising, and pattern-based forms of harm (Woersdoerfer, 2025).

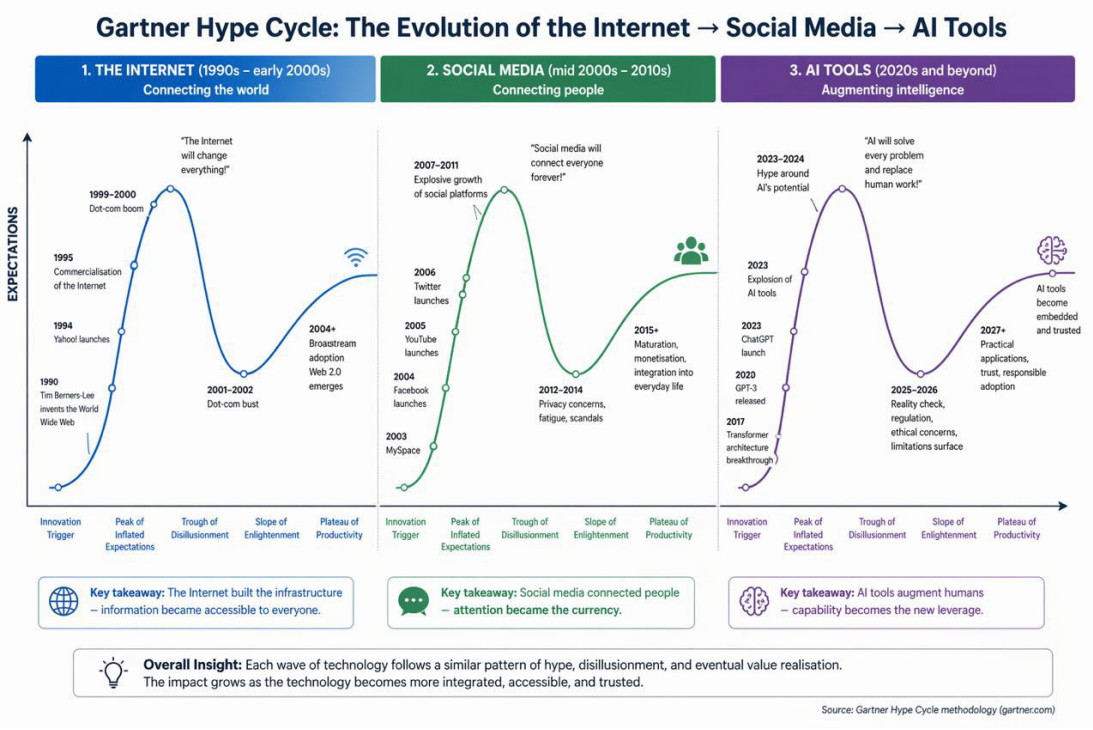

Figure 1: Gartner Hype Cycle across Internet, Social Media, and AI tools.

The current evolution of AI tools follows the Gartner hype curve model (see figure 1 above), following the evolution of the internet and social media. In that space, over 80 different versions of AI regulation are currently in circulation, with none driving wholesale international action. Successful adoption appears limited primarily because of funding-related conflicts of interest. AI regulation adoption roughly doubles every two years. In 2017, only 1 country had active AI regulation, but as of 2025, 38 countries have such regulations, and half of all UN member states (70+) have an AI plan/equivalent policy trajectory. The 2021 UNESCO recommendation on AI ethics was aimed at its 190+ member states, becoming highly influential in establishing ethical principles and human rights safeguards. Yet the only binding agreement, akin to nuclear treaties (Harding, 2024) (revealing AI's level of importance) has been the Council of Europe AI Convention (EU Parliament and Council, 2024) which attempted to establish clear guard rails *first*, assuming that protections are needed. However, they do not go deep enough, as this project explores.

1. Race to the bottom of the brain stem

With artificial intelligence and related digital technologies, our traditional legal understandings are being tested. Over the last three decades, we have engaged in what has been called a "race to the bottom of the brainstem" (Harding, 2024; Shroff, 2025). This began with emails, then the growth of the internet, then the so-called Web 2.0 with platforms like Facebook and Yahoo!, then into social media, and finally, into artificial intelligence. This is a completely new kind of race, which has revealed some of humanity's worst parts. Collectively, digitally, and at scale, our legal and regulatory instincts have not kept pace. This has happened within roughly a generation. The pace of change is so fast that traditional boundary-creating modes are circumnavigated. While pattern recognition makes us human, it can also work against us. As humans, we look for similarities in other people, things, and ways of interacting (Baron-Cohen, 2020), but this same pattern recognition can drive prejudice, bias, and, in its worst forms, a complete lack of empathy for groups we do not understand, or those we "other". These tendencies will always be encoded into

large-scale AI systems, magnifying and automating structural harms, rather than merely encouraging isolated individual acts.

1.1. Mandatory Pillars and Guardrails

When so much of our technological development is targeting, whether consciously or not, fundamental and underlying elements of the human condition, then placing mandatory markers in the sand is a necessity. There is no point suggesting the market will regulate itself for humanity's betterment; it will not, as shown by the attention economy. Instead, we need a codified set of guardrails protecting those individuals who have i) neurodisabilities, that collection of neurological impairments from Cerebral Palsy, through to Profound (severe) Autism, and ii) those who are neurodiverse, like individuals with autism, dyslexia, dyspraxia, and ADHD. Mostly, neurodiverse individuals have needs, which can be sensory, but generally, communicate and use language effectively. Unlike them, individuals with neurodisabilities face lifelong and often congenital problems, often coupled with intellectual and/or physical impairment. In this light, required pillars should perform (at least) two functions: protecting those with neurodisability/disability itself from social murder, and protecting those with neurodiversity from psychological attachment abuse.

1.2. Neurodisability and Social Murder

Social murder (Ryan, 2026) encapsulates the notion that society acts through prejudice, e.g., policy that actively "others" disabled people, which makes them be seen as less than other people. As a result, the treatment of those with disabilities creates systemic inequity, which is known and can be ameliorated. Yet, nothing is done, particularly addressing preventable deaths of those with neurodisability, which is known as social murder. For example, it is well known (Mouridsen *et al*., 2008; Ryan, 2026) that one complication of profound autism is epilepsy, which is a leading cause of mortality through drowning. Yet, health, education, and social care policy and practice have not reflected this knowledge until recently (Xie et al., 2025). This system-level failure possibly excludes change, when instead, this knowledge should lead to proactive co-production, and therefore, a reduction in mortality rates (Ryan, 2026). Systemic patterns should be emphasised by AI and not overlooked.

1.3. Neurodiversity and Psychological Attachment Abuse

No less concerning is the worrying emergence in the technological move from the attention economy toward the emerging attachment economy. The former is well documented (Shu, Lai, and He, 2026), but the attachment economy is newly developing in terms of its breadth and individual impact. People are reportedly already in relationships with AI characters, and in some cases have "married" them, leading to negative effects on their well-being (Papadopoulos, 2025). Further, the slow creep of AI into pre-schoolers' toys places AI in a newly vulnerable space, and engaging with AI could potentially supersede and overshadow critical emergent and social skills like theory of mind (ToM) and self-regulation. If ToM is outsourced to a non-biological entity, the resulting developmental processes would no longer be shaped by human feedback loops (Center for Humane Technology, 2026). We aim to draw on existing AI regulatory frameworks (i.e., the EU AI Act and international principles from organisations like the OECD and UNESCO), as well as equality concepts, such as protected characteristics (sex, gender, ethnicity, disability), to ask: to what extent are disabled and neurodiverse children explicitly coded into AI regulation, and where are they simply assumed to be covered by generic references to "vulnerability" or non-discrimination?

2. Methodology

2.1. Research design, sources, & strategy

This study employs Systematic Literature Review (SLR) combined with a scoping review approach to identify, analyse, and compare international artificial intelligence (AI) regulatory frameworks for children who have either a disability, neurodisability, or are neurodivergent. This hybrid design is due to the rapidly evolving and heterogeneous nature of AI governance, and both structured evidence synthesis (SLR) and broad exploratory mapping (scoping review) are required to capture academic, policy, and legal developments and reflect AI governance research (Batool, Zowghi and Bano, 2025). Data is going to be collected from three source categories: 1) Peer-reviewed academic literature (journals and conference proceedings), 2) Policy and regulatory documents (e.g., national AI strategies, supranational frameworks, and regulatory acts such as EU AI policy instruments), and 3) Institutional and international organisation reports (e.g., OECD, UNESCO, and regulatory agencies). A structured keyword search strategy will be applied around Boolean search terms on neurodisability, neurodivergence, and disability, using terms such as "AI regulation," "AI governance frameworks," "algorithm accountability," and "risk-based AI regulation." This multi-source approach reflects established practice in AI governance research, where policy and legal documents are treated as primary analytical material alongside academic literature due to the field's policy-driven nature (Morley *et al.*, 2022).

2.2. Initial Search

An initial search and review of peer-reviewed academic literature was conducted on June 11, 2026, using Cambridge University's online iDiscover service. The initial search was conducted using this search string and looking for articles containing these items, in any field: ("AI regulation" OR "AI governance" OR "algorithm* accountability" OR "risk-based AI" OR "artificial intelligence governance") AND (neurodisabilit* OR neurodiverg* OR disabilit* OR "neurodevelopmental disorder*" OR autis* OR ADHD). Search results were limited by date (from 2011 to 2026) and language (only English results were considered). In total, 39 articles were identified through this initial search. After 2 duplicates and 2 foreign-language articles were removed, 33 articles were screened based on title and abstract. Of that universe of screened literature, 15 articles were sought for retrieval and assessed for eligibility (18 being excluded). Of those articles that were sought for retrieval, 9 articles were removed because they did not address the topics being examined (AI regulation/governance/ through the lens of neurodisability/neurodivergence, etc. – 8 articles) or because they were unavailable (1 article). After the completion of this initial search (and article selection and appraisal process), 6 articles have been identified as the starting point that the remainder of the review will build upon.

3. Initial results

Each of the included initial articles help to underline the importance of the two pillars and validate the foci of the broader project. First, Curtis et al. (2026) help demonstrate that merely automated models of AI governance cause systemic inequities, restricting the provision of key resources and, at least implicitly, risks social murder. Marland & Arantes (2026) offer a useful conceptual balance to the Curtis et al. (2026) article by positing that the social model of disability should mean that educators, administrators, and policymakers change their focus from attempting to "fix" whatever vulnerabilities the individual has to instead restructuring the environment (in this case, the AI architecture) around the individual, through legal means, and in doing so, obviating institutional barriers that previously existed. In addition, the analysis of Sousa et al. (2024) regarding affective computing demonstrates how it, arguably as part of the attachment economy, can read, manipu-

late, and exploit human emotions, which poses particularly acute risks to paediatric populations with neurodivergence. Block (2026)'s analysis of Article 50(1) of the EU AI Act provides a significant critique of existing legislative guardrails, which, for example, only have simplistic disclosure requirements, which helps to show why standard transparency laws are not effective or strong enough to protect children, particularly children with neurodivergence, from deep psychological attachment manipulation. More broadly, Bolarín-Miró et al. (2026) explicitly identify specificity as being a key requirement for global AI policies if they are to move beyond the generic category of "vulnerability" and therefore, helpfully and clearly promoting the flourishing of neurodivergent individuals. Finally, Kemigisha (2025) shows that truly inclusive governance of AI should not just be another regulatory requirement, but rather necessary in moving toward a macro-sustainable future. These articles will be used as a baseline for the broader analytical approach, the results of which will be presented in full at the conference.

References

Baron-Cohen, S. (2020) *The pattern seekers: How autism drives human invention*. Basic Books.

Batool, A., Zowghi, D. and Bano, M. (2025) "AI governance: a systematic literature review," *AI and Ethics*, 5(3), pp. 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Block, M. J. (2026) "Transparency in human-AI interaction – An analysis of Article 50(1) AI Act," *Computer Law & Security Review*, 61, 106327. https://doi.org/10.1016/j.clsr.2026.106327

Bolarín-Miró, E., Carmona Martínez, M.M., and López-Arranz, A. (2026) Inclusive Governance of Artificial Intelligence: Towards an Ethical Framework for Neurodivergence. In Alvarez, I., Arias-Oliva, M., Dediu, AH., & Silva, N (Eds.), *Ethical and Social Impacts of Information and Communication Technology, Lecture Notes in Computer Science, vol 15939* (pp. 60–68). Springer. https://doi.org/10.1007/978-3-032-01429-0_6

Curtis, H., *et al.* (2026) "Computer Says No": Disabled Welfare Experiences and Envisioned Futures Under AI Governance. In Oliver, N., Shamma, D. A., Candello, H., Cesar, P., Lopes, P., Bozzon, A., Kosch, T., Liao, V., Ma, X., Artizzu, V., Draxler, F., Lopez, G., Reinschluessel, A. V., Tong, X., & Toups Dugas, P. O. (Eds.), *CHI '26: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (pp. 1–25). Association for Computing Machinery. https://doi.org/10.1145/3772318.3791025

EU Parliament and Council (2024) *European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union.* http://data.europa.eu/eli/reg/2024/1689/oj (Accessed: April 23, 2026).

Harding, V. (2024) *AI Needs You*. Princeton and Oxford University Press.

Kemigisha, S. (2025) "Reimagining Sustainable Futures: Integrating AI-Driven Solutions and Inclusive Communication Strategies for Accelerated Progress toward the SDGS," *African Multidisciplinary Journal of Research*, 2(2), pp. 288–307. https://doi.org/10.71064/spu.amjr.2.2.2025.439

Marland, B., & Arantes, J. (2026) "Reframing AI governance in education: insights from the social model of disability," *Learning, Media, and Technology*, pp. 1–16. https://doi.org/10.1080/17439884.2025.2595443

Morley, J. *et al.* (2022) "Governing Data and Artificial Intelligence for Health Care: Developing an International Understanding," *JMIR Formative Research*, 6(1), p. e31623. https://doi.org/10.2196/31623

Mouridsen, S.E. *et al.* (2008) "Mortality and causes of death in autism spectrum disorders: An update," *Autism*, 12(4), 403–414.

Papadopoulos, C. (2025) "The Use of AI Chatbots for Autistic People: A Double-Edged Sword of Digital Support and Companionship," *Neurodiversity*, 3, pp. 1–7. https://doi.org/10.1177/27546330251370657

Ryan, S. (2026) *Critical Health and Learning Disabilities: An Exploration of Erasure and Social Murder.* London, UK: Wiley.

Shroff, L. (2025) *The People Outsourcing Their Thinking to AI, The Atlantic.* https://www.theatlantic.com/technology/2025/12/people-outsourcing-their-thinking-ai/685093/ (Accessed: February 4, 2026).

Sousa, J., *et al.* (2024) Converging Affective Computing and Ethical Challenges: The Quest for Universal Access in Human-Machine Cooperation. In: Antona, M., & Stephanidis, C (Eds.), *Universal Access in Human-Computer Interaction, Lecture Notes in Computer Science, vol 14696* (pp. 106–121). Springer. https://doi.org/10.1007/978-3-031-60875-9_8

Technology, C. for H. (2026) "Attachment Hacking and the Rise of AI Psychosis," *[Center for Humane Technology]*, 21 January. https://centerforhumanetechnology.substack.com/p/attachment-hacking-and-the-rise-of?utm_medium=web (Accessed: February 4, 2026).

Xie, K. *et al.* (2025) " Unintentional Drowning Incidents Involving Children with Autism Spectrum Disorder Treated in US Emergency Departments, 2016–2020," *Journal of Autism and Developmental Disorders.* https://doi.org/10.1007/s10803-025-06769-0

Woersdoerfer, M. (2025) "Ten reasons why – the case for more and better AI regulation," *AI and Ethics,* 6(1), p. 62. https://doi.org/10.1007/s43681-025-00865-8