Designing and Defining Children's Agency in the Age of AI (CHAILD): A Two-year Roadmap

By Jun Zhao, Manolis Mavrikis, Carina Prunkl, Kaska Porayska-Pomsta, Wayne Holmes, Nigel Shadbolt and Beeban Kidron

1. INTRODUCTION

AI systems are increasingly central to children's digital environments, encompassing connected toys, apps, voice assistants, and online learning platforms. While these systems offer valuable opportunities for children's development and learning, they also pose significant risks, such as screen time addiction and various forms of online harms. Current approaches to children's digital experiences are largely dominated by restrictive and protective measures, neglecting children's potential to exercise control and make age-appropriate, informed choices themselves. While it may be challenging for younger children (especially those under 5) to make meaningful decisions, such approaches make little contribution to fostering children's

autonomous agency in a digital context. There is an urgent need for new frameworks, guidelines, and strategies for policymakers, developers, parents, and educators to better support children's development, well-being, and safety in an increasingly datafied environment.

The two-year UKRI-funded project **CHAILD** — *Children's Agency in the Age of AI: Leveraging Interdisciplinarity* — aims to establish a foundational understanding of children's agency in the age of AI. This white paper outlines the critical challenges we plan to address and provides a roadmap for our research. We invite collaboration and expressions of interest from all potential stakeholders.

http://chaild.org..

2. An Agency-by-Design Approach

Fostering children's agency in digital contexts is a complex task due to the lack of a unified understanding of agency across various disciplines. In philosophy, autonomous *agency* is broadly defined as the effective capacity to make decisions of their own that are of practical import to their lives. Philosophers emphasise the need for certain competencies, some of which only develop with age. Conversely, in learning sciences, agency is viewed as essential for children's knowledge acquisition and is defined as 'the capacity of individuals to act independently and make choices of their own', emphasising the central role in supporting learning. Additionally, psychology literature highlights that children's agency is influenced by factors such as *goals, epistemic abilities* (prior knowledge about a specific topic), and *cognitive development stages*. This multifaceted conceptualisation across disciplines indicates the complexity of agency and the need for an interdisciplinary investigation. In this white paper, we propose a working definition of "agency" as the following:

> *Children's agency in the digital context* refers to their ability to make choices of their own and act independently within algorithmic systems.

Re-examining children's agency in the digital context is essential for three key reasons:

1. First, a clear definition of children's agency is essential to understand how it is influenced by factors such as age, cognitive development, and the socio-technical environments they experience both online and offline.
2. Second, the definition is critical for developing mechanisms that foster children's agency, which can be undermined by data-driven algorithmic systems, through mechanisms such as personalisation of learning pathways, automated content recommendations, and pressure for social validation.

3. Finally, this mission is especially urgent in light of emerging AI-related legislation, which has not yet fully accounted for the impact of these AI-based systems on children's rights and agency.

While several international guidelines, such as the [United Nations General Assembly Resolution on the Rights of the Child](#) (2023), the [OECD Recommendation on Children in the Digital Environment](#) (2021) and the Council of Europe's [Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018),](#) emphasise children's rights in the digital context, **the concept of supporting children's autonomous agency in the digital context has not been explicitly addressed.** These documents primarily focus on children's *fundamental rights* in the digital contexts, including respect for children's evolving capabilities, non-discrimination, the rights to be heard, and privacy. However, they provide limited guidance on empowering children to exercise control or make informed choices, often confining the discussions of agency to narrower contexts, such as privacy settings.

We argue that the increasing prevalence of AI poses significant challenges to children's autonomy. Future digital designs and innovations **must take a more proactive approach in fostering children's critical agency and autonomy** by embedding these considerations into the design process from the onset. Rather than treating children as passive recipients or merely vulnerable individuals in need of protection, digital systems should be developed to enable them to navigate these spaces independently and critically.

An agency-by-design approach focuses on empowering children as active participants in both the design and use of AI systems. The approach aims to foster children's autonomy, creativity, and decision-making ability by involving them in the development process and ensuring that AI systems cater to their needs, capabilities, and rights. It is guided by three key principles:

1. **Participation**, with children being co-designers and their input playing a central role in shaping how AI interacts with them. This approach emphasises children's perspectives and ensures that technology aligns with their cognitive and emotional development.
2. **Autonomy** should be the cornerstone for empowering children to make choices and take control over their interactions with AI systems, through mechanisms such as positive scaffolding, encouraging children to make independent decisions while learning in safe environments and developing self-efficacy.
3. **Rights-Centred**, ensures that AI systems respect and protect children's agency rather than diminish it. This aligns closely with frameworks like the "[Child Rights by Design](#)" principle, which emphasises the creation of equitable, inclusive, and age-appropriate technologies.

3. Five challenges

To realise this agency-by-design approach is not without challenges.

Defining agency

The diverse conceptualisations of agency across disciplines present a significant challenge in designing agency for children in the digital context. However, this should not undermine the importance of establishing a clear conceptual framework for children's agency. Addressing this challenge requires a multidisciplinary and cross-sector approach that integrates developmental, social, and ethical perspectives and possibly more. Such an approach is essential not only to recognise the significance of agency for children but also to unpack the complexities involved in fostering it.

Assessing (or measuring) agency

Until we define what agency means for children, it will be challenging to determine how to measure it effectively or identify the specific aspects to address. Existing approaches often rely on quantitative measurements, such as self-reported experiences or the monitoring of behavioural and neurocognitive responses; however, many of these methods have yet to be adapted specifically for children in digital contexts. Assessing agency requires not only a clear definition to guide the development of benchmarks but also a multifaceted approach that considers how AI technologies may impact various aspects of children's agency, including their motivation, imitation, and moral judgement.

Designing for agency

While the Human-Computer Interaction (HCI) community has a long history of designing for and with children, the concept of agency is often overlooked in the development of AI systems aimed at children. Moreover, although the HCI community has extensive experience in involving children as active stakeholders or co-designers, there is limited consensus on how best to prioritise children's values, especially when innovations may introduce conflicting interests. This challenge is further exacerbated by the increasing sophistication and unpredictability of AI technologies, such as the rapid development of generative AI and its integration into children's digital lives. A principled approach is essential to effectively embed children's values, ensuring their need for agency is both recognised and addressed. These challenges are further compounded by the diversity in children's cognitive and developmental stages, highlighting the importance of developing age-appropriate and context-sensitive solutions.

Supporting agency with technology

Current approaches to supporting children's agency have explored various mechanisms, including fostering motivation, developing habits, encouraging decision-making, providing

personalised or adaptive instructions, and promoting peer support. However, the effectiveness of these approaches has rarely been evaluated through long-term studies or in real-world settings. ***Longitudinal studies are necessary to understand the sustained effects on children's agency*** and to track changes in their (sense of) autonomy over time. Such studies would offer deeper insights into how agency develops and evolves in response to different interventions. Supporting children's agency with technology also requires a careful blend of developmental insights, safe and adaptive design, and consideration of parents and educators' roles. It involves challenges at the intersection of technology, education, psychology, and child development.

Increasing public awareness

Children's agency in the digital context has not been sufficiently discussed in existing legislation, policies, or design guidelines. This has created a significant knowledge gap among the general public, including parents, educators, and practitioners involved in designing AI systems for children. While some educators are aware of risks such as exposing children to harmful content, or using ineffective pedagogic practices , many lack clear guidance or training, which amplifies concerns about children's agency in digital learning environments. Engaging parents is another crucial aspect of promoting the development of children's agency. Raising public awareness has often been neglected or approached without adequately understanding the specific support required. Therefore, it is crucial to not only recognise the urgent need for a more focused effort but also adopt a stakeholder-led approach to ensure relevance and effectiveness.

4. The mission of CHAILD

Given the landscape and open challenges described above, the CHAILD project aims to achieve the following in the next two years:

- **Defining children's critical agency in the digital context:** This will be achieved by integrating interdisciplinary expertise, drawing on the knowledge of computer scientists, learning scientists, social scientists, and philosophers.
- **Establishing approaches to assess the impact on agency:** We will leverage existing assessment methods from psychology and education, incorporating insights from longitudinal studies and observing developmental changes over time.
- **Developing design principles for agency:** Grounded empirical evidence and collaborative co-development with industry partners will form the basis for creating guidelines that promote children's agency in technological innovations.
- **Creating new guidelines for parental and educational support:** By actively listening to the voices of parents and educators, we will co-create materials and resources that directly address their needs.
- **Formulating policy recommendations to prioritise children's agency:** We will collaborate with a strong and highly experienced policy team to develop

recommendations and ethical guidelines for creating AI systems for children, engaging with relevant global policy audiences and forums, guided by a clear policy delivery roadmap.

Closing remarks

CHAILD is funded by the first round of [UKRI's new cross research council responsive mode (CRCRM) pilot scheme](). The CRCRM scheme has been developed to support emerging ideas from the research community that transcend, combine or significantly span disciplines, to ensure all forms of interdisciplinary research have a home within UKRI. This provides unique opportunities for interdisciplinary research projects like CHAILD.

If you'd like to help CHAILD's mission of building children's agency, as a parent, educator, researcher, innovator or policymaker, please join the CHAILD community by clicking [here].