



# Toward children-centric AI: a case for a growth model in children-AI interactions

Karolina La Fors<sup>1</sup>

Received: 6 October 2021 / Accepted: 5 October 2022  
© The Author(s) 2022

## Abstract

This article advocates for a hermeneutic model for children-AI (age group 7–11 years) interactions in which the desirable purpose of children's interaction with artificial intelligence (AI) systems is children's growth. The article perceives AI systems with machine-learning components as having a recursive element when interacting with children. They can learn from an encounter with children and incorporate data from interaction, not only from prior programming. Given the purpose of growth and this recursive element of AI, the article argues for distinguishing the interpretation of bias within the artificial intelligence (AI) ethics and responsible AI discourse. Interpreting bias as a preference and distinguishing between positive (pro-diversity) and negative (discriminative) bias is needed as this would serve children's healthy psychological and moral development. The human-centric AI discourse advocates for an alignment of capacities of humans and capabilities of machines by a focus both on the purpose of humans and on the purpose of machines for humans. The emphasis on mitigating negative biases through data protection, AI law, and certain value-sensitive design frameworks demonstrates that the purpose of the machine for humans is prioritized over the purpose of humans. These top-down frameworks often narrow down the purpose of machines to do-no-harm and they miss accounting for the bottom-up views and developmental needs of children. Therefore, applying a growth model for children-AI interactions that incorporates learning from negative AI-mediated biases and amplifying positive ones would positively benefit children's development and children-centric AI innovation. Consequently, the article explores: *What challenges arise from mitigating negative biases and amplifying positive biases in children-AI interactions and how can a growth model address these?* To answer this, the article recommends applying a growth model in open AI co-creational spaces with and for children. In such spaces human-machine and human-human value alignment methods can be collectively applied in such a manner that children can (1) become sensitized toward the effects of AI-mediated negative biases on themselves and others; (2) enable children to appropriate and imbue top-down values of diversity, and non-discrimination with their meanings; (3) enforce children's right to identity and non-discrimination; (4) guide children in developing an inclusive mindset; (5) inform top-down normative AI frameworks by children's bottom-up views; (6) contribute to design criteria for children-centric AI. Applying such methods under a growth model in AI co-creational spaces with children could yield an inclusive co-evolution between responsible young humans in the loop and children-centric AI systems.

**Keywords** Growth · Positive bias · Negative bias · Diversity · Non-discrimination · Children-centric AI

## 1 Introduction

One day on our way home my son of 8 years pointed out an—in my view—unfair bias which he wanted to solve with an artificial intelligence (AI) invention. He proposed to design an AI device that could capture and translate the conversations trees have amongst themselves into human language. In his reasoning, it was unfair that we did not know their opinions and ideas, yet we made decisions about them. At that moment I learnt a previously unconsidered

---

✉ Karolina La Fors  
k.lafors@utwente.nl

<sup>1</sup> DesignLab, Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Building: The Gallery, Hengelosestraat 500, 7521 Enschede, The Netherlands

bias from him and he taught me a novel perspective which I interpreted as his positive bias and ethics on how an AI system could amplify a conception of diversity that broadens inclusion towards not-yet or not-quite-included groups. The ingenuity of children to imbue such ethical values as diversity with moral meanings inspired this article.

The article advocates for a hermeneutic model for children-AI (age group 7–11 years) interactions (Kudina, 2021) in which the desired purpose of AI systems shall be children's growth. Under such a model children can cope with the effects of discrimination and can co-create and amplify inclusivity in their interactions with AI. The article acknowledges that discriminatory biases can stem from society and also from AI-mediated interactions and can both impact children. The main focus of this article is, however, on the impacts of AI-mediated biases on children. Based on the severe impacts of AI-mediated discriminatory biases (Dignum et al. 2020), AI systems with machine-learning components have a recursive element when interacting with children. This element is needed for children's growth. The human in the loop of AI, such as the designer, user, and others including parents and guardians of children should be as responsible as the AI is supposed to be (Dignum 2019) to facilitate children's growth when they interact with AI.

Broadening how bias is interpreted (Mitchell, 1997, Eubanks 2018) within the artificial intelligence (AI) ethics (Coeckelbergh 2020; Mittelstadt et al. 2016) and responsible AI (Dignum 2019) debates, specifically regarding the interactions between AI systems and children<sup>1</sup> is critical for children's growth. AI ethics and responsible AI discourse

and methodology currently equate bias with discrimination and frame it solely as something to eliminate. The article proposes diverting from this mainstream interpretation and conceptualizing bias as a preference or prioritization in line with Gadamer's positive conception of prejudice (1989) and distinguishes between positive (pro-diversity, pro-inclusion) biases<sup>2</sup> and negative (discriminatory) biases.<sup>3</sup> This distinction does not suggest a cause-and-effect relationship—avoiding bias is minimally good and introducing positive bias in the form of privileging inclusion and diversity is better but this distinction aligns with John Dewey's experience-based educational theory in which children's interactive (both negative and positive) experiences, if guided responsibly, are there to grow from.

Discriminatory exclusion mediated (Verbeek, 2011a, b) by AI systems has profound effects on children (Huyn and Fuligni 2010) and can impede them from cultivating fruitful societal relations. Mitigation mechanisms for children, however, lag behind (Dignum et al. 2020).

AI research could benefit from a growth model that incorporates why and how to sensitize children against negative biases and amplify positive ones as core ethical values within children-AI interactions. From the perspective of children, this is crucial because although AI systems cannot be designed as artificial moral agents (Dignum 2019), children often unconsciously perceive them as moral authorities, next to humans.

The definition of artificial intelligence in this article follows Russel's and Norvig's conception which views AI systems as "agents that have communication abilities with some capacity for decision making". When these systems interact with humans' different meanings can emerge (2020). Such AI-mediated meanings can be particularly character-shaping for children aged between 7 and 11. They cognitively start to understand reasoning and are becoming reasoners themselves when interacting with others including AI. The

<sup>1</sup> By children this article refers to the age group of 7–11 years in the Dutch context. This is supported by three reasons: first, according to developmental psychologists age 7 is "the age of reason" (Piaget, 1972). Children until 7 are said to understand the reasoning of others as such only limitedly. Above 7 years children are capable of formulating their own reasoning. This is important in their interactions with AI, because AI systems are perceived by children as morally reasoning agents. Moreover between 7 and 11 years is when children transition from behaving in an ethically desirable manner out of obedience to outside moral authority. They grow into more autonomous, young ethical agents. This is because they begin to internalize ethical values in interaction with others. Second, this paper focuses on the Dutch context because children are absent from the Dutch national AI strategy and according to Dutch national law children of this age cannot execute their fundamental children's rights without the collaboration of their parents or guardians. Consequently, children of this age group inhabit a more dependent position in society with little scope to shape ethical, children's rights frameworks and AI designs. Third, studies on inclusive early childhood education demonstrate the long-term benefits of inclusion in interactive education of this age group (Henninger and Gupta, 2014). Allowing children to experience inclusion through positive bias in their interactions with AI and shape frameworks and designs by their views would have profoundly positive implications on these children's inclusiveness towards others also for their adult lives.

<sup>2</sup> Positive bias is defined, here, as a flexibility for a dynamic inclusion of others. Positive bias is different from positive discrimination in three ways: (1) positive discrimination remains unfair against certain groups, while favouring others; (2) positive bias is about cultivating a benefit-of-the-doubt attitude toward others; (3) rendering such attitude to be learnt by AI machines based on input data, by prior programming and also by learning from the interactions with children that are capable of amplifying positive bias through children-AI interactions. If AI mediated decision can yield toward a discriminatory or negative bias, AI systems could also machine-learn from and amplify the positive biases of children and others through interacting with them. For deciding what a positive and negative bias is from children's perspectives insights from developmental psychology and children's personal stories and life circumstances would also be needed to account for.

<sup>3</sup> Negative bias is defined here, as the discriminatory treatment or unfair decision stemming from an inference by an AI system about a child.

reasoning of AI systems cannot be completely bias-free (De Rijke and Graus 2016), because that would undermine their purpose.

Karamjit Gill argues, however, that for approaching AI as a human–machine symbiosis (Gill 1996) human–machine and human–human interactions need to be collectively understood. This includes understanding and aligning the purpose of the machine and the purpose of humans. When I argue for a growth model that accommodates coping with a negative and amplifying positive bias, I argue for more emphasis on the purpose of humans. Currently, there is a larger focus on the machine’s purpose for humans and I show that the cultivation of coping with negative and learning, co-creating or amplifying positive biases in children-AI interactions would foster children’s growth.

The article explores what challenges arise from mitigating negative biases and what recommendations are conceivable for amplifying positive biases for children-AI interactions and how can a growth model be applied to facilitate children’s growth and children-centric AI innovations. To answer this, a systematic literature review had been conducted from (a) developmental psychology; (b) AI-relevant regulations: data protection and AI law; (c) children’s rights; (d) value-sensitive design; (e) AI ethics; (f) responsible AI; (g) human-centric design; (h) educational and (i) co-creational theories and methods. The article innovates by bringing these disciplinary pieces of literature together that have previously not been put into conversation with each other around this topic.

The article is divided into five sections. Section 2 discusses the relevance of distinguishing between positive and negative biases in children-AI interactions based on findings in developmental psychology. The section outlines arguments for how insights into developmental psychology could nurture AI ethics, and responsible AI debates concerning children. Section 3 starts by reinterpreting the human–machine symbiosis thesis for children and argues that currently, a proportionally larger focus is on prevention and mitigation mechanisms of negative biases for adults. Section 4 offers recommendations for applying a growth model for children-AI interactions within which methods for mitigating negative biases and for amplifying positive biases are presented toward children-centric AI systems and more AI-centric guidance for children. It starts by defining children’s purpose as growth. It discusses that, next to human–machine value alignment methods, more attention needs to be devoted to such human–human value alignment methods that are meaningful for preparing children to cope with negative biases and for cultivating children’s inclusive mindset in their interactions with AI systems. Applying these methods, especially in meaningful co-creation with children for research and testing AI in children-AI interactions could yield AI design features that are more

children-centric and AI-centric guidance of children. Section 5 offers conclusions.

## 2 Relevance of distinguishing between negative and positive biases for children-centric AI

Developmental psychology offers insights into why and how it is relevant to distinguish between negative and positive biases in children-AI interactions. Next to arguments about the negative effects of discrimination (negative bias) on children’s development, this section lists mechanisms that can cultivate positive biases and also counter negative ones. Incorporating insights of developmental psychologists into AI ethics (Algorithm Watch 2021) and responsible AI research would extend the remit of this work in beneficial ways.

### 2.1 Developmental psychology

Research in developmental psychology shows that children as early as 7 years of age—can already experience discrimination (Warren 2018). They can also internalize discrimination as not being worthy as human beings and deteriorate their mental well-being for their whole lives.<sup>4</sup> Children’s feeling of being unaccepted in a community can undermine their self-esteem, hamper their performance in education and cause long-lasting anxieties (Anderson 2013; Sirin et al. 2015; Huynh and Fuligni 2010). Discriminatory experiences can even evoke “stress responses similar to post-traumatic stress disorder” (Spears 2015). The influence of educators (Keys Adair 2011) and parents as influential role models and moral authorities have been underlined as instrumental in shaping how children develop. If these role models discriminate against children based on their personal or cultural identities that can impact them with detrimental consequences (Mentally Healthy Schools 2021). As the effects of the historic re-education policy for native Canadian children demonstrate, detrimental implications can be fatal (Khawaja 2021). Given the amplifier effects of human prejudice through AI (O’Neil 2017), discrimination in children-AI interactions would also entail more pronounced mental-health impacts (Tynes et al. 2014).

Developmental psychologists stress, however, that children’s best defense to mitigate the impact of discrimination is to develop by cultivating a strong sense of ethnic-racial or cultural self-identity (Marcelo and Yates 2018). Cultural

<sup>4</sup> Online resource, retrieved on 03/07/21 from <<https://www.mentallyhealthyschools.org.uk/risks-and-protective-factors/vulnerable-children/discrimination/>>.

studies with children of native Americans, for instance, show that routing children's identity with the identity of their ancestors directly assists children to become healthy adults (Day 2014). A closely related field, social identity theory concludes that cooperation between children with a strong individual self-identity effectively diminishes tendencies for their discriminatory attitudes (Leaper 2011). This also includes gender. From the perspective of gender-related non-discrimination: teaching the value of care for all children of all genders was proven fruitful so that not only females, but children of all sexes could develop (respect for) a caring moral agency toward others (Nodding 1984).

Allowing children to discover and respect their cultural roots, and explore their individual and collective identities including their genders assists them to develop an inclusive attitude toward others. Given the recursive element of AI systems, AI systems can also amplify biases learnt or co-created from children's choices (Zaga, 2021). Capturing children's meaning of pro-diversity as positive biases and amplifying such characteristics by AI can also mitigate negative biases. The next section details how AI ethics and responsible AI debates could benefit from such insights from developmental psychologists.

## 2.2 AI ethics and responsible AI

Within AI ethics, designing for the value of non-discrimination regularly means anticipating and mitigating (discriminatory) negative biases by AI design (Coeckelbergh 2020). Currently, AI ethics and responsible AI research are scarce which would engage in defining how to orchestrate mechanisms that would not only do no harm; but would meaningfully stimulate children's growth, their sense of individual and collective identity development, and their appreciation of diversity also in their interactions with AI systems.

Valuable practice-based research is available on AI design for good (Tomasev et al. 2020). But this is not yet developed for and through involving children. Despite the potentially profound effects of AI on children, AI ethics is currently not child-specific enough. It currently omits to focus on the potentially character-damaging effects of discrimination and how to engineer AI-child interaction design for more inclusivity.

AI technologies cannot and should not be designed as fully autonomous moral agents (AMA) (Van Wynsberghe and Robbins 2019), because in contrast to humans they do not have human intentions and agency (Van de Poel 2020). Children, however, can perceive AI systems to have the same moral authority (Turiel 2018) as parents and educators (Manzi et al. 2020). Issues for the character-shaping of children can emerge when children perceive and interpret AI systems as moral agents that decide about and for them. Designing AI systems that have ethically acceptable

behaviour (Anderson et al. 2007) towards children by not discriminating against them on any ground that is damaging to their identity and self-worth would serve children's well-being. This is in line with that the human in the loop of AI, such as the designer, user, and subject should be as responsible as the AI is supposed to be (Dignum 2019). Developing AI with children's healthy development in mind would assist to prevent AI systems from producing discriminatory implications.

Drawing from the insights of developmental psychology, AI ethics, and responsible AI the following two observations can be made. First, current AI ethics and responsible AI debates would benefit from an extensive and conjunctive study of the subject (AI, algorithms/applications), subject contexts (domains), and objects (children) for an empirical understanding of how children experience discrimination, inclusion, and respect for their identity in their interactions with AI systems. Second, for gaining such insights effective research methods and research spaces would be necessary to allow children to express their understanding and imbue the values of non-discrimination, identity, and diversity with their meanings. By engaging in research and AI development, children could improve the enforceability of their fundamental right to non-discrimination (Art. 2) and their right to identity (Art. 8) and could improve AI design criteria and co-shape their individual and collective development.

## 3 Challenges to children-centric AI: emphasis on mitigating negative biases for adults and their insufficiency for children

For such groundwork, reinterpreting Karamjit Gill's concepts of human-machine symbiosis and human-centredness for children-AI interactions offers a useful lens. This is useful to evaluate how negative bias mitigation mechanisms for adults shift a larger focus on the purpose of machines than on the purpose of humans (Gill, 1996). Given the recursive element in children-AI interactions, co-developing mechanisms with and for children to cope with negative and amplify positive biases would serve the purpose of humans by taking AI as an ally in achieving more inclusiveness and diversity in children-AI interactions. As I illustrate in this section, currently an overwhelming focus rests on mitigation mechanisms of negative biases in the interactions of adults with AI. The perspective of adults serves as a starting point. Children's rights-centred design principles offer opportunities to explore, yet are currently underexplored. Although all of these are beneficial, currently remain insufficient to achieve children-centric AI.

The section outlines a set of challenges to children-centric AI. First, it shows how the prevention and mitigation of

negative biases by data protection law are framed around the purpose of the machine. Second, although children's rights law is the most pivotal placeholder for children's interests, no court cases, so far, refer to digitally-mediated discrimination of children. A third challenge is that (a) value-sensitive AI design remains too focused on eliminating and mitigating negative biases and (b) those value-sensitive design theories (Friedman et al. 2013) that aim for upholding values, currently lack bottom-up perspectives on what is needed for children's development. Children are rarely meaningful participants in AI innovation processes that are aimed at facilitating the development of their sense of identity and respect for diversity. It closes concerning promising developments in value-sensitive design, such as the children's rights-based design principles in the Netherlands and the United Kingdom, while highlighting their shortcomings from the perspective of biases.

### 3.1 Reasons for human–machine symbiosis

Taking inspiration from Gill's human–machine symbiosis from the perspective of children–AI interactions is crucial. Gill stresses that in the human–machine symbiosis tradition the “culture-based knowledge and actions of the human should be reflected dynamically in systems instead of subsumed by them” (Gill 1996, 5). The development of human–machine symbiosis needs to account for the interdependencies between human capacities and machine capabilities, which should offer building grounds for human-centered design. This is also applicable to AI design. To reach meaningfully human-centered design, Gill advocates for the collective consideration and understanding of human–human and human–machine value alignment methods.

The large emphasis on mitigation mechanisms of negative biases in certain human–machine value alignment methods, I argue, underlines not only that more emphasis is on the purpose of the machine for humans but how that purpose becomes narrowly framed as ‘not to harm humans’. I consider here the General Data Protection Regulation (GDPR—EC 2016/679), the European Data Governance Act (EDGA—EC 2020/767) the proposed European Artificial Intelligence Act (EAIA—EC 2021/206), and a set of value-sensitive AI design methodologies as human–machine value aligning methods. I assess some of their principles and framings to illustrate my point. Within these the prevention and mitigation of negative biases as assurances that the purpose of the digital system is harmlessness is prominent. However, it remains worth asking why ensuring the harmlessness of an AI system should or should not be interpreted as acceptable proof of its usefulness, especially from the perspective of children. The already ratified and proposed EU regulations, for instance, currently do not reach beyond such interpretability. Ensuring AI's harmlessness is

highly important, yet as developmental psychologists demonstrated, from children's perspectives it remains insufficient as a single goal to optimize for. While considering AI systems as value-mediating and value-shaping (Verbeek, 2011a, b) agents in interaction with humans, Gill's conception of the human-centred tradition encompasses the “interplay between the notions of ‘purpose’, symbiosis, cohesion, diversity, coherence, and valorisation, which are seen as foundational determinants for shaping the ethical dimension of the development trajectory” (Gill, 2004, 1). In this regard, mitigating negative biases for adults does not meet all the above parameters of the human-centred tradition to which humans' needs are central. From the perspective of children's non-discrimination, one of those needs is positive reinforcement.

While bearing children's healthy growing up in mind shifting more focus on human–human value alignment methods and aligning AI with purposes that benefit children's healthy development would be particularly urgent in a society where AI systems increasingly proliferate and implicate all contexts of children's lives. Creating processes that would foster going beyond AI systems doing no harm but insisting that they have to do good would be essential and lead to meaningfully children-centric AI.

### 3.2 Emphasis on mitigating negative biases of AI (for adults) in data protection and AI law

Valuable legal research (Borgesius 2019) and court cases, such as *Mart and Others v. Turkey* mainly focus on what (un)intended negative biases AI can generate concerning adults. References to non-transparent analytics or AI's black box effects (Wachter et al. 2018; Rudin and Radin 2019) and how interacting with AI can jeopardize fundamental values (Citron and Pasquale 2014) and rights of non-discrimination (Ferguson, 2017) of different groups demonstrate this. In this section, I illustrate first how the emphasis on negative biases in human–AI interactions is dominant through the examples of GDPR; EDGA, and EAIA.

Different legal sub-disciplines, such as non-discrimination law, define grounds for what can count as discrimination. In non-discrimination law, prominent reference points to decide about whether discrimination towards an individual or societal group occurred are based on the so-called protected characteristics (including categories such as race; gender, religious or political orientation, or ethnicity). Scholars have already argued that these characteristics provide insufficient protection from AI-mediated unjustifiable inferences on adults and called even for a ‘right to reasonable inferences’ (Wachter and Mittelstadt 2019). Others also underlined the weaknesses of non-discrimination laws and data protection laws in preventing and remedying AI-mediated discrimination (Borgesius 2020). These scholarships remain



centered on preventing, eliminating, and mitigating negative biases related to adults because they aim for minimizing violations through discriminative data and use. Whereas prevention and mitigation remain crucial also for children, they do not explore positive reinforcement processes from other disciplines that have already achieved results in fostering children's inclusive and diversity-minded development. Data protection law appears, however, even less efficient in addressing negative biases.

Art. 5(1) of GDPR exemplifies a framing of the purpose of the machine. This purpose is about data processing (including through AI systems) which shall be lawful, fair, and transparent. GDPR includes child-specific rules (Art. 8 and recital 38 GDPR). Recital 2 stresses that processing shall uphold the well-being of data subjects, including children. The GDPR embraces the principles of non-discrimination and fairness and it facilitates them through principles such as data minimization, storage limitation, and rights such as the right not to be subject to profiling (Art. 22) or the right to rectification (Art. 16). Whereas translating non-discrimination into such techno-legal codes can offer benefits, this remains ineffective to counter the occurrences and implications of negative biases on children's lives when interacting with AI. Beyond that, GDPR's enforcement is not optimal. The Dutch Data Protection Authority's (DPA) self-assessment (Autoriteitpersoonsgegevens 2021) and the report of the Dutch Consumers Association<sup>5</sup> testify this. The DPA remains understaffed and under-budgeted to match the needs required for the processing of all GDPR violation claims (Autoriteitpersoonsgegevens 2021). Therefore, more diverse methods would be needed to enforce GDPR rights and go beyond them in children's best interests.

The EDGA and the EAIA are part of a broader 'tranche of proposed EU regulations' (Veale and Borgesius 2021). EDGA's purpose is to govern an interoperable European data space by allowing European citizens and small and medium enterprises more control in trying to curb the monopolistic position of big technology companies from the US and China. This is aimed at establishing mechanisms that increase trust in data intermediaries and more effective data-sharing mechanisms. Although it introduces concepts, such as "data altruism" the purpose of this regulation is economically motivated and due to EU and GDPR-compliance requirements are also aimed at non-discriminative data sharing.

The EAIA is the very first comprehensive regulation on AI systems. It frames AI systems through the extent to which

they can harm by optimizing for their doing no harm, by prescribing to assess AI systems against "reasonably foreseeable misuse". It classifies AI systems by their harmful implications into four categories: (a) unacceptable risks; (b) high risks; (c) limited risks and (d) minimal risks. Out of these categories, only AI systems that fall under high-risk are regulated.<sup>6</sup> Two ambitions of the act illustrate further how the purpose of the AI system instead of the purpose of humans remains central in its focus. First, EAIA will establish a European database for high-risk AI systems that is publicly accessible (Art. 60). Second, EAIA (Veale and Borgesius 2021) will catalyze the formulation of standards for conformity by 2024 so that companies could prove the harmlessness of AI systems. These objectives show how the societal usefulness and purpose of AI become narrowly framed around eliminating negative biases and other harms.

All these EU legal instruments are valuable and necessary to protect children. However, none of them define mechanisms for fostering persons' diversity-minded and inclusive interactions with AI systems. Such objectives remain beyond the scope of these regulations. Their focus on eliminating, and mitigating negative biases and other harms, can be regarded as (legally) coding a form of negative ethics. Although too little, too much, or too late regulation of technology can hamper responsible innovation (Van den Hoven 2017), it would be essential to recalibrate what responsible AI innovation consists of with children in mind. Bearing in mind their dependent position in society and on AI systems (La Fors 2020) comprehensive AI guidance would be needed that capitalizes upon children's views. Integrating these views into regulations so that these regulations complement those that only aim for AI systems not to do harm would benefit (1) children's participation and inclusive upbringing and (2) AI system designs capable of facilitating the former.

### 3.3 Lack of emphasis on mitigating negative biases of AI in children's rights law

Thu UNCRC (UNCRC 1989) can be viewed as the most pivotal testimony to why and how outcomes of different human-human value aligning disciplines, such as psychology, pedagogy, philosophy, or ethics have been translated into interculturally acceptable principles and rights in service of all children's best interests irrespective of their biological, cultural traits or economic backgrounds. Yet, as any other non-binding top-down code requires practical interpretation and pathways for optimal enforceability.

The General Comment 25 of the United Nations Committee on the Rights of the Child made it only more explicit

<sup>5</sup> Consumentenbond wil betere bescherming privacy kinderen (2021) Retrieved on 11/09/21 from—<https://www.consumentenbond.nl/nieuws/2021/consumentenbond-wil-betere-bescherming-privacy-kinderen>.

<sup>6</sup> The assessment of high-risk systems includes "third-party conformity assessment", a form of independent oversight (Art. 43).

that children's rights, including their right to identity (Art. 8) and non-discrimination (Art. 2), are equally applicable online and called upon national governments “to take proactive measures to prevent discrimination” (United Nations Committee on the Rights of the Child 2021) against children. Article 21 of the Charter of Fundamental Rights of the European Union (2007) or Article 7 of the Universal Declaration of Human Rights (1948) all support their right to non-discrimination. Despite these obligations, enforcing the right to non-discrimination within children-AI interactions remains unsatisfactory.

No ECtHR court case affects the violation of children's right to non-discrimination online so far (UNICEF 2020). Cases mainly concern the violations of children's privacy rights, as *S. and Marper v. the UK* (ECtHR 2009) or *X. and Y. v. the Netherlands* (ECtHR 1985) show this. Whereas protecting children's fundamental right to privacy online remains essential (van der Hof and Lievens 2018), UNICEF's international assessment on national AI strategies calls upon nation-states to address children's right to non-discrimination. Children remain largely absent from national AI strategies, including the Dutch (UNICEF 2020). Their absence also entails governments' inattention to the discriminatory effects of AI systems on their development.

ECtHR court cases condemning a state for discriminating against children are from the offline environment. The judgments in *D.H. and Others v. the Czech Republic* (ECtHR 2007)<sup>7</sup>; *Oršuš and Others v. Croatia* (ECtHR 2010)<sup>8</sup> exemplify how nation-state activities were ruled to be discriminatory by unjustifiably depriving children of minority groups of the opportunity to collectively participate in education with children from majority groups. The *Belgian linguistic case v. Belgium* (ECtHR 1968)<sup>9</sup> condemned a nation-state for discriminating against children by forbidding them to speak their native language for education.

With children in mind, the enforcement of children's right to identity and non-discrimination as a top priority remains essential for all humans-in-the-loop in an AI setting.

### 3.4 Emphasis on mitigating negative biases of AI in value-sensitive design approaches

Within the field of value-sensitive design (VSD) (Friedman et al. 2017), a significant emphasis remains on the role of

<sup>7</sup> The Czech state was found guilty of violating Art. 14 of ECHR, because a “disproportionate number of Roma children became placed into special schools without justification”.

<sup>8</sup> The Croatian government was found guilty of unfair treatment of minority children.

<sup>9</sup> Case “Relating to certain aspects of the laws on the use of languages in education in Belgium” v. Belgium; Application no. 1474/62; 1677/62; 1691/62; 1769/63; 1994/63; 2126/64, ECHR 1968.

mitigating and minimizing negative bias against adults. A broad variety of distinctions on discriminatory biases show this (Mann and Matzner 2019). Barocas, Crawford, Shapiro, and Wallace, for instance, highlight two biases: (1) allocative biases which mean the unjust distribution of goods; and (2) representational biases which mean the unfair affronting of someone's self-worth or dignity (Barocas et al. 2017). Others critique scholars in value-sensitive design when discussing biases by devoting strong attention to (de)biasing data sets and propose the distinction of perpetual bias (Offert and Bell 2020). This they argue shifts the equally important focus onto what “lies in the set of inductive biases in machine vision systems that determine its capability to represent the visual world” (Offert and Bell 2020, pp. 1). The latter presupposes a system view of AI, which this article also follows (Van de Poel 2020). Some scholars also argue that “moral values that are fundamental to design-for-values (DFV) are subject to cognitive biases” (Umbrello 2018, pp. 186), which require debiasing.

Whereas VSD has high potentialities for orienting AI design choices towards socially good principles, to achieve these principles bias remains to be considered as something negative to eliminate in adult-AI interactions. Suggestions for “de-biasing” to achieve fairness (Umbrello 2018) demonstrate this. Although debiasing has benefits, the limitations of only a techno-centric view on bias can also hamper the efficiency of laws (Balayn and Gürses 2021), and as this article argues such a narrow focus can potentially hamper the healthy development of children when interacting with AI.

### 3.5 Underexplored human-machine value alignment methods in children-AI interactions

Human-machine value alignment methods that could be applied to mitigate negative and amplify positive biases in children-AI relations are currently limited. Those that exist are (a) relatively new and therefore underexplored, (b) focus mainly on designers; (3) resulted from the implementation of top-down definitions of ethical values for adults or top-down implementation of children's rights; (4) have not resulted from broad public debate on AI involving children. Consequently, I argue, that these methods could benefit from their combined application with human-human value alignment methods. Below, I list two instruments aimed at fostering diversity and non-discrimination and two at fostering children's rights to identity and non-discrimination.

The EU High-Level Expert Group's Assessment List (ALTAI) (EC, 2020) contains a checklist following the seven principles of the EU High-Level Expert Group's Guidelines for Trustworthy AI (EC, 2019). This also contains children's rights impact assessments. Assessments shall also be conducted against such principles as ‘diversity, non-discrimination, and fairness. UNICEF's policy

guidance on AI and children (Dignum et al. 2020) defines co-design with children—when appropriate—as useful interventions to support children’s participation in AI design. This in principle includes that children can participate in shaping the outcomes by shaping the role of AI in decision-making processes through co-creating AI’s design features. However, neither the ALTAI nor UNICEF’s guidance offers methods on how and when to involve children effectively in co-design.

Regarding children’s rights-based design, since March 2021, the Dutch Code for Children’s Rights (Code voor Kinderrechten, 2021)<sup>10</sup> offers ten guiding principles on what to focus on when developing children’s rights-aligned digital devices for designers. These are highly important and rare principles to consider. Only the UK has set similar standards through age-appropriate design which is based on considerations of data protection and children’s privacy rights. The UK code remains to reflect a risk-based approach (ICO 2021).<sup>11</sup> Companies adhering to either of these two codes can demonstrate their bona fide and children’s rights-minded AI design.

Whereas the implications of these codes are essential for children, (1) these codes refrain to consider children-AI interactions from a hermeneutic perspective; (2) they only focus on designers, (3) they embrace top-down enforcement of values and children’s rights (4) the effectiveness of the practical implementation is so far under-researched and (5) they do not lay out which methods would be effective to meaningfully cultivate coping mechanisms with discrimination and a diversity- and inclusivity-oriented mindset together with and for children.

#### **4 Recommendations for children-centric AI: applying a growth model in AI co-creational spaces with and for children**

To effectively optimize for children-centric AI, where children’s perspectives, dependencies, needs, and rights are taken at face value from the bottom-up also for design, more emphasis is needed on identifying, applying, and learning from human–human value alignment methods. These can inform human–machine value alignment methods on how to guide children to cope with negative biases and how to amplify positive ones. For this, this article recommends applying a growth model.

<sup>10</sup> The 10 principles of Code voor Kinderrechten include: involving children’s perspectives in the design. This aims for data protection compliance and remains scarce on how to achieve meaningful co-design.

<sup>11</sup> It reaches beyond online privacy considerations of children, but offers limited methods for co-design.

The American educational philosopher, John Dewey, conceptualized experience-based education and developed a theory of “growth as an end” for children (and humans) (Dewey 1997). I explain how this framing is relevant for children-AI interactions. His growth theory should be set as a model under which all value alignment methods that are relevant to guide children in coping with negative biases and in cultivating positive ones in their interactions with AI should be assembled for transdisciplinary research. While bearing in mind a hermeneutic approach to AI systems (Gadamer, 1989), where systems are as situated and partial as children in their interactions with them, applying the growth model of Dewey would serve children’s purpose of growth when interacting with AI. Such a model could facilitate co-creative research with and learning from children how to cope with negative and amplify positive biases in children-AI interactions. In the second part, I recommend options on how to optimize for this purpose. I recommend designing co-creational spaces which allow for the application of human–human and human–machine value alignment methods by gaining empirical insights from children about their interactions with AI. The main recommendation illustrates how a non-exclusive set of human–human and human–machine value alignment methods can facilitate fruitful exchanges to develop a pro-diversity and pro-inclusive mindset for and with children when interacting with AI. These allow for the individual and collective formulation and appropriation of values, expectations, and skills with children by bringing them in transdisciplinary dialogue with a broad set of stakeholders that are relevant for responsible children-AI interaction design.

#### **4.1 A growth model for children-AI interactions**

Dewey formulates why and how the purpose of children’s education shall be reconsidered for growth. In his expose, he critiques traditional considerations regarding the purpose of education and childhood. The single objective of overcoming immaturity of childhood by dissolving into the static state of maturity in adulthood is problematic for him when defined and set as children’s main purpose in life. In his definition, immaturity should be embraced as a positive characteristic. It should rather be perceived as something to cherish and cultivate than something worth shredding in one’s life. He outlines two crucial components of immaturity for this: dependence and plasticity. Embracing a healthy dependence on others, according to Dewey, is both individually and socially beneficial. Plasticity is the capacity to learn from experiences that complement dependence and allow for one’s growth. For Dewey dependence and plasticity are both conditions for growth. Based on this reasoning he advocates for articulating the purpose of children’s (and human’s) education as experienced-based growth and setting this as



an end to aspire for. Growth is what one should not only pursue in childhood but throughout one's life. Because of its relational character, growth is something that one cultivates individually but only in collaboration with others. Therefore, growth will bring individual and societal benefits.

I make three arguments in favor of applying his theory to children-AI interactions. One, applying it would offer room for simultaneously accommodating and implementing human-human and human-machine value alignment methods for children that aim to amplify positive biases as forms of responsible shaping of children's interactions and experiences with AI. Two, children's growth as an end can be defined as a for all children beneficial purpose irrespective of their biological, cultural traits, or backgrounds and this should also apply in their interactions with AI. Three, methods to mitigate and cope with negative biases and amplify positive ones in children-AI interactions for children's growth would need to be assessed not only within the contexts of the institution of the school where children's education traditionally takes place but in all of their life contexts, including homes and public spaces. Gaining empirical insights concerning the AI-related experiences of children in such co-creational spaces can offer productive grounds for guiding the responsible co-evolution of children and AI. Interactions with AI in such spaces could facilitate children in co-shaping their way of "lived ethics" (Zawieska 2020) through their engagement with AI systems. Enacting such positive ethics, for instance, by enacting a pro-diversity attitude is in servitude of humanizing AI (Zawieska 2020) and the sustainable co-evolution of humans and AI.

## 4.2 AI co-creational spaces with children

Designing AI co-creational spaces for and with children offers multiple benefits. First, these could offer children experiences of being included and incentivized for expressing themselves regarding their identity and diversity as partners in broader societal debates on AI. Second, such methods could be applied to guide and learn from children. Such methods could facilitate children in developing into inclusive, pro-diversity-minded ethical reasoners and deliberators when interacting with AI. Third, such methods could assist in informing AI-relevant laws and value-sensitive design methods to co-create AI systems with children. Fourth, the urgency for designing legal mitigation mechanisms specific to the AI-mediated discriminatory effects on children could be lessened. Fifth, children's inclusion could offer them direct performative experiences of what inclusion in democratic processes involving AI can mean. Given children's early experiences have character-shaping effects, offering them such experiences of inclusion and embracing

their expressions regarding their identity and diversity when interacting with AI could render them becoming more inclusive and respectful toward others' identity and diversity.

Effective human-human value alignment methods which can cultivate children's positive biases toward others are many. Anti-bias education (Gienapp 2021); arts education (Lum and Wagner 2019), music education (Campbell 2018), and methods for safeguarding cultural heritage<sup>12</sup> (UNESCO 2021) are described as fruitful to facilitate children's understanding and development of self-identity and appreciation for the diversity of others. UNICEF's empirical studies report positive results when children's rights education—including their right to non-discrimination and identity—becomes integrated into pupils' curricula. Such children become more inclusive (non-discriminatory) and respectful of the diverse values and rights of others. By rendering children's rights education, a talking point in class children develop a "shared language for understanding rights concerning their everyday interactions" (Jerome et al. 2021, pp. 22).

Effective human-machine value alignment methods for including children in co-design (Zaga 2021) are also available. These could be applied to meet and potentially go beyond the 'diversity, non-discrimination and fairness' principle of the EU HI-LEG Guidelines for Trustworthy AI. By recalibrating and applying them to mitigate negative biases and amplify positive ones, they can benefit children's growth. Cultural probes and peer play exemplify such methods. *Cultural probes* encompass perspective-taking with children about their present situations and interactions with AI. This could include their experiences of identity, diversity, and experiences of discrimination. *Peer play* (Zaga et al. 2017) aims to ideate and co-create future scenarios for and by children about their interactions with AI. Ideations could encompass how values and children's right to identity, diversity, and discrimination become at stake and how to preserve them better when interacting with AI. To appropriately guide children in such perspective-taking and ideation, children would not only need to become AI literate but would need to co-shape and appropriate an understanding of such values as identity, diversity, and non-discrimination

<sup>12</sup> Zoltán Kodály, a Hungarian ethnomusicologist and champion of music education for children also inspired this article. In an interview he shared what his main motivation was for devoting his life to better the music education. He explained that when starting his project, the debate on music was too limited to exchanges within a small group of expert musicians of his generation. His ambition was to broaden this debate by making it integral part of the daily interactions of the broadest possible public and from its earliest age onwards. In my view, a similar reasoning seems applicable for the current AI ethics and responsible AI debates. Online resource, accessed on 21/06/2021 from <https://www.youtube.com/watch?v=NbDvjzqb924>.

and how AI systems mediate and shape such values through their interactions with them (Verbeek, 2011a, b).

Children's co-shaping and appropriation of such understanding could further develop the "moral collindrige dilemma" (Kudina and Verbeek 2018). Regarding children of 7–11 years, the development of a pro-diversity attitude as positive bias would require that children's meaning of diversity becomes imbued into AI systems. Such imbuing is possible through the hermeneutic approach to AI-child interactions. Imbuing values with meaning would simultaneously emerge with and necessitate that a child becomes a moral, in this case, a pro-diversity, reasoning human being. For becoming such a pro-diversity reasoner through interactions with AI, children have to appropriate some sense of the top-down value and rights frameworks that are relevant to identity and diversity and in which AI systems are also thought of and aimed for embedding, but there should be space left for ideating with them about their meaning of inclusion. So that, for instance, even the thoughts of trees could become a subject through which children could become partners to discuss what inclusive AI can mean. When AI systems are co-created with children, children can experience how to actively shape AI frameworks by imbuing them with their meanings. When elicited in co-creational spaces these meanings could also form input for children-informed and therefore, more children-centric design.

Collectively applying human–human and human–machine value alignment methods in AI co-creational spaces (Goldman 2012) to elicit and interact with children about values can generate new forms of hybrid ethical reasoning regarding AI. This can assist in informing legal and normative frameworks, such as the Massive Open Online Deliberation (MOOD) to meet children's needs and development stages (Dignum 2019). Whereas MOOD offers highly effective opportunities "for people to develop and draft collective [ethical] judgments on complex issues and crises in real-time" it does not account for the diverse age group of children who depend on their age and biological, cognitive, psychological, and moral capabilities require different skills and approaches for joining such deliberation modes. Within AI co-creational spaces children can acquire such skills for deliberation that would favour their healthy, diversity-minded, and inclusive moral development and growth.

If responsible AI is a question of governance (Dignum 2019), then children need to become effective ethical reasoners and deliberators who possess AI skills to effectively articulate their perceptions of identity and diversity when interacting with AI systems. The likelihood that children's engagement can establish a routine for their pro-diversity-minded decision-making involving AI systems is larger, when co-shaped and performed early, and so is the likelihood that they become responsible governors of AI and humans in its loop.

## 5 Conclusions

This article advocated for a hermeneutic approach to children-AI (age group 7–11 years) interactions in which the desired purpose of AI systems shall be children's growth. Applying a growth model to children-AI interactions assists with the effects of discrimination and can co-create and amplify inclusivity in their interactions with AI.

The article proposed to do this by acknowledging a recursive element in interactions and by broadening interpretations of bias within the artificial intelligence (AI) ethics and responsible AI discourse. It argued for a distinction between negative (discriminative) and positive (pro-diversity) biases. Introducing mitigation and coping mechanisms and methods against negative biases and methods to amplify positive biases in children-AI interactions would benefit both children's healthy development and children-centric AI design. Inspired by the human–machine symbiosis (Gill, 1996), the article argued that the negative framing of all biases and the broad variety of goals to eliminate them, illustrate a stronger focus on the purpose of the machine for humans than on the purpose of humans. The article highlighted as an additional shortcoming of the negative framing of biases that by aiming for their prevention and mitigation they enact a narrow conceptualization of the machine's purpose for humans, which is doing no harm. Such framing and focus challenge the broader developmental needs of children and developing children-centric AI systems. Developing mechanisms under a growth model for children-AI interactions that accommodate human–human and human–machine value alignment methods so that children can cope with negative biases and amplify positive ones would be necessary to serve their growth in their interactions with AI. Therefore, the article offered recommendations for identifying and collectively applying human–human and human–machine value alignment methods which could nurture children's self-expression, sense of identity, diversity, and inclusiveness, and can be applied in children's broader living environments, where they interact with AI systems. Offering them in such spaces guidance to learn from negative and cultivate positive experiences with AI could help them to become positive ethical reasoners and deliberators. This would be essential for all children irrespective of their biological capabilities and differences in backgrounds. Inspired by Dewey's experience-based educational philosophy and co-design literature, the recommendations describe how investing in the development of the values of identity, diversity shall not only mean introducing top-down predefined concepts on these values for children but involving children from the bottom-up and co-creatively.

I argued for creating open educational and co-creational spaces where children can imbue and co-create these values

with their meanings through experimenting with AI systems. These spaces would include but reach beyond the school environment and encompass diverse contexts of children's lives where they can interact with AI. By applying human-machine and human-human value alignment methods these spaces would facilitate amplifying positive biases to foster children's growth. Methods would be applied to generate collective experiences in diverse contexts where AI systems are embedded and to stimulate exchanges on conceptions of values of and rights to identity, diversity, and non-discrimination among children, parents, pedagogues, industrial stakeholders, and academics, governmental authorities, NGOs. Aggregating and applying these methods would offer productive pathways to amplify positive biases within children-AI interactions. Ultimately, applying such methods in these spaces could yield experiences for children about how their interactions with AI can reinforce their feelings of being unconditionally lovable and worthy.

**Acknowledgements** I am highly grateful for the encouragement, valuable comments, and general enthusiasm of Prof. Peter-Paul Verbeek toward the topic and an early version of this article. I am also very thankful for the reviewing committee's insightful feedback on an abstract of this article accepted by AI4Europe's Culture of Trustworthy AI workshop. My sincere appreciation goes also to Dr. Fran Meissner for her bright and constructive reflections on this article.

**Funding** This research was funded by the DesignLab of the University of Twente.

**Availability of data and material** All data is from open (referenced) sources.

**Code availability** No software application used.

## Declarations

**Conflicts of interest** No conflicts of interest are present in this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aantal privacyklachten blijft zorgwekkend hoog—12/03/2021. Autoriteitpersoonsgegevens Persbericht. <https://autoriteitpersoonsgegevens.nl/nl/nieuws/aantal-privacyklachten-blijft-zorgwekkend-hoog>. Retrieved on 12/09/21
- Age-appropriate design: a code of practice for online services, ICO, 2021. <https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-a-code-of-practice-for-online-services/>. Retrieved on 22/06/21
- AI Ethics Guidelines Global Inventory, Algorithm Watch. <https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/>. Retrieved on 25/09/21
- Anderson KF (2013) Diagnosing discrimination: stress from perceived racism and the mental and physical health effects. *Sociol Inq* 83:55–81. <https://doi.org/10.1111/j.1475-682X.2012.00433>
- Anderson M, Anderson LS (2007) Machine ethics: creating an ethical intelligent agent. *AI Mag* 28(4):15–26
- Balayn A, Gurses S (2021) Beyond De-biasing: regulating AI and its inequalities. *EDRI*. [https://edri.org/wp-content/uploads/2021/09/EDRI\\_Beyond-Debiasing-Report\\_Online.pdf](https://edri.org/wp-content/uploads/2021/09/EDRI_Beyond-Debiasing-Report_Online.pdf)
- Barocas S, Crawford K, Shapiro A, Wallach H (2017) The problem with bias: From allocative to representational harms in machine learning. In: Special Interest Group for Computing, Information, and Society. <http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf>. Retrieved on 6th July 2021
- Borgesius FJZ (2019) Discrimination, artificial intelligence, and algorithmic decision-making, Council of Europe. <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>
- Borgesius FJZ (2020) Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int J Hum Rights* 24(10):1–22. <https://doi.org/10.1080/13642987.2020.1743976>
- Campbell P (2018) Music, education, and diversity: bridging cultures and communities. Teachers College Press, New York
- Case "Relating to certain aspects of the laws on the use of languages in education in Belgium" v. Belgium; Application no. 1474/62; 1677/62; 1691/62; 1769/63; 1994/63; 2126/64, ECtHR 1968
- Children's Rights under European Social Charter, Information document prepared by the Secretariate of the ESC (Council of Europe 2018). <https://rm.coe.int/1680474a4b>. Retrieved on 12/06/21
- Citron D, Pasquale F (2014) The scored society: due process for automated predictions. *Washington Law Rev* 89. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2376209#](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209#). Retrieved on 22/05/21
- Code voor Kinderrechten. (2021) <https://codevoorkinderrechten.nl/en/>. Retrieved on 22 June 2021
- Coeckelbergh M (2020) AI Ethics. The MIT Press, Cambridge
- Consumentenbond wil betere bescherming privacy kinderen (2021) <https://www.consumentenbond.nl/nieuws/2021/consumentenbond-wil-betere-bescherming-privacy-kinderen>. Retrieved on 11/09/21
- Day PA (2014) Raising healthy American Indian children: an indigenous perspective. In: Social issues in contemporary native America: Reflections from Turtle Island, 93–112
- De Rijke M, Graus D (2016) Wij zijn racisten, Google ook. *NRC Handelsblad*. [https://pure.uva.nl/ws/files/2801721/175163\\_graus\\_wij\\_2016.pdf](https://pure.uva.nl/ws/files/2801721/175163_graus_wij_2016.pdf)
- Dewey J (1916) Democracy and education: an introduction to the philosophy of education. The MacMillan Company
- Dewey J (1997) Experience and education. Free Press
- D.H. and Others v. the Czech Republic [GC], App. no. 57325/00, ECtHR 2007-IV
- Dignum V (2019) Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer
- Dignum V, Penagos M, Pigmans K, Vosloo S (2020) Policy guidance on AI for children—draft for consultation/recommendations for building ai policies and systems that uphold child rights. UNICEF <https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>. Retrieved on 24 June 2021

- Discrimination—Mentally Healthy Schools, Anna Freud National Centre for Children and Families. <https://www.mentallyhealthyschools.org.uk/risks-and-protective-factors/vulnerable-children/discrimination/>. Retrieved on 23 June 2021
- Eubanks V (2018) Automating inequality: how high-tech tools profile, police and punish the poor. St. Martin's Press
- EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1
- EU High-Level Expert Group Guidelines on Trustworthy Artificial Intelligence (EC, 2019). <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>. Retrieved on 06 June 2021
- EU High-Level Expert Group Assessment List For Trustworthy Artificial Intelligence (ALTAI) For Self-Assessment. Shaping Europe's Digital Future (2020) <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>. Retrieved on 06 June 2021
- European Commission (2020) Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)—COM/2020/767 final. Document 52020PC0767. European Commission, Brussels
- European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final) ('AI Act'). Document 52021PC0206. European Commission, Brussels
- European Union: Council of the European Union, Charter of Fundamental Rights of the European Union (2007/C 303/01), 14 December 2007, C 303/1, Available at: <https://www.refworld.org/docid/50ed4f582.html>. Retrieved on 17 June 2021
- Ferguson AG (2017) The rise of big data policing. NYU Press, New York
- Friedman B, Kahn PH, Borning A, Hultgren A (2013) Value sensitive design and information systems. Early engagement and new technologies: opening up the laboratory. Springer, Dordrecht, pp 55–95
- Friedman B, Hendry DG, Borning A (2017) A survey of value sensitive design methods. *Found Trends@Hum Comput Interact* 11:63–125
- Gadamer H-G (1989) Truth and method, 2nd revised edition, revised translation by J. Weinsheimer and D.G. Marshall, New York: Crossroad.
- United Nations Committee on the Rights of the Child, *General Comment No. 25 on Children's rights concerning the digital environment*. UN Doc CRC/C/GC 25 (2021) [https://tbinternet.ohchr.org/\\_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC%2fC%2fGC%2f25&Lang=en](https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC%2fC%2fGC%2f25&Lang=en). Retrieved on 11 June 2021
- Gienapp R (2021) Five anti-bias education strategies for early-classroom childhood. <https://www.pbssocial.org/education/teachers/five-anti-bias-education-strategies-early-childhood-classrooms>. Retrieved on 06/06/21
- Gill K (ed) (1996) Human-machine symbiosis. The foundations of human-centred system design. Springer, London, p 5
- Gill K (2004) Exploring human-centredness: knowledge networking and community building. <https://www.researchgate.net/publication/238622429>. Retrieved on 12/09/21
- Goldman AI (2012) Theory of mind: The Oxford handbook of philosophy of cognitive science guidelines to respect, protect and fulfill the rights of the child in the digital environment—Recommendation CM/Rec (2018) 7 of the Committee of Ministers; Council of Europe. <https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a>. Retrieved on 22/07/21
- Henninger WR, Gupta S (2014) How do children benefit from inclusion? In: Henninger WR, Gupta S (eds) First steps to preschool inclusion: how to jumpstart your programwide plan. <http://www.brookespublishing.com/first-steps-to-preschool-inclusion>
- Huynh VW, Fuligni AJ (2010) Discrimination hurts The academic, psychological, and physical well-being of adolescents. *J Res Adolesc* 20(4):916–941. <https://doi.org/10.1111/j.1532-7795.2010.00670.x>
- Jerome L, Emerson L, Lundy L, Orr K (2021) Teaching and learning about child rights: a study of implementation in 26 countries, UNICEF, pp. 12. <https://www.unicef.org/media/63086/file/UNICEF-Teaching-and-learning-about-child-rights.pdf>. Retrieved on 22/07/21
- Keys Adair J (2011) Confirming *Chanclas*: what early childhood teacher educators can learn from immigrant preschool teachers. *J Early Childh Teacher Educ* 32(1):55–71. <https://doi.org/10.1080/10901027.2010.547652>
- Khawaja M (2021) Consequences and remedies of indigenous language loss in Canada. *Societies* 11:89. <https://doi.org/10.3390/soc11030089>
- Kodály Zoltán Interview (1966) Online resource. <https://www.youtube.com/watch?v=NbDvjzqb924>. Accessed on 21/06/2021
- Kudina O (2021) “Alexa, who am I?”: Voice assistants and hermeneutic lemniscate as the technologically mediated sense-making. *Hum Stud* 44(2):233–253. <https://doi.org/10.1007/s10746-021-09572-9>
- Kudina O, Verbeek P-P (2018) Ethics from within: Google glass, the collingridge dilemma, and the mediated value of privacy 44(2)
- Leaper C (2011) Chapter 9—more similarities than differences in contemporary theories of social development?: A plea for theory bridging. In: Benson JB (eds) *Advances in child development and behavior*, JAI, Vol. 40, pp. 337–378. <https://doi.org/10.1016/B978-0-12-386491-8.00009-8>. Retrieved on 12/06/21
- Lum C-H, Wagner E (eds) (2019) Yearbook of arts education research and cultural diversity policies, research, practices and critical perspectives. Springer, Yearbook of Arts Education Research for Sustainable Development
- Mann M, Matzner T (2019) Challenging algorithmic profiling: the limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data Soc.* <https://doi.org/10.1177/2053951719895805>
- Manzi F, et al. (2020) Moral evaluation of Human and Robot interactions in Japanese pre-schoolers, Creative commons license attribution 4.0 International (CC BY 4.0). <http://ceur-ws.org/Vol-2724/paper7.pdf>. Retrieved on 06/07/2021
- Marcelo AK, Yates TM (2018) Young Children's ethnic-racial identity moderates the impact of early discrimination experiences on child behavior problems. *Cultural diversity and ethnic minority psychology*. Advance online publication. Retrieved on 05/06/20. <https://doi.org/10.1037/cdp0000220>
- Mart and Others v. Turkey (ECtHR, 2019) Application number 57031/10
- Mitchell M (1997) Machine intelligence. McGraw-Hill Education; 1st edition
- Mittelstadt B, Allo P, Mariarosaria T, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. *Big Data Soc.* <https://doi.org/10.1177/2053951716679679> (In press)
- Noddings N (1984) *Caring: a feminine approach to ethics and moral education*, pages 3–4. University of California Press, Berkeley
- Offert F, Bell P (2020) Perceptual bias and technical meta pictures: critical machine vision as a humanities challenge. *AI & Soc.* <https://doi.org/10.1007/s00146-020-01058-z>. Retrieved on 06/06/21
- O'Neil C (2017) *Weapons of math destruction*. Penguin Books
- Oršuš and Others v. Croatia [GC] App. no.15766/03 (ECtHR 2010)
- Piaget J (1972) Intellectual evolution from adolescence to adulthood. *Hum Dev* 15:1–12. <https://doi.org/10.1159/000271225>



- Richardson R, Schultz J, Crawford K (2019) Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review* online, forthcoming. <https://ssrn.com/abstract=3333423>
- Rudin C, Radin J (2019) Why are we using black box models in AI when we don't need to? A lesson from an explainable ai competition retrieved on 22nd May 2021 from <https://hdr.mitpress.mit.edu/pub/f9kuryi8/release/6>
- Russel SJ, Norvig P (eds) (2020) *Artificial intelligence a modern approach*, 4th edn. Pearson Series
- S. and Marper v. UK, App. No. 30562/04 and 30566/04 [2009] 48 EHRR 50
- Sameroff AJ, Haith MM (eds) (1996) *The five to seven-year shift: the age of reason and responsibility*. University of Chicago Press.
- Sirin SR, Rogers-Sirin L, Cressen J, Gupta T, Ahmed SF, Novoa AD (2015) Discrimination-related stress affects the development of internalizing symptoms among Latino adolescents. *Child Dev.* <https://doi.org/10.1111/cdev.12343>
- Spears BC (2015) The educational, psychological and social impact of discrimination on the immigrant child. Migration Policy Institute. <https://www.migrationpolicy.org/sites/default/files/publications/FCD-Brown-FINALWEB.pdf>. Retrieved on 23/06/21
- Tomašev N, Cornebise J, Hutter F et al (2020) AI for social good: unlocking the opportunity for positive impact. *Nat Commun* 11:2468. <https://doi.org/10.1038/s41467-020-15871-z>
- Turiel E (2018) Moral development in the early years: when and how. *Hum Dev* 61:297–308
- Tynes BM, Hiss S, Rose C, Umaña-Taylor A, Mitchell K, Williams D (2014) Internet use, online racial discrimination, and adjustment among a diverse, school-based sample of adolescents. *Int J Gam Comput Med Simul* 6(3):1–16
- Umbrello S (2018) The moral psychology of value-sensitive design: the methodological issues of moral intuitions for responsible innovation. *J Respons Innov* 5(2):186–200. <https://doi.org/10.1080/23299460.2018.1457401>
- UNICEF Policy Brief National AI Strategies and children (2020) Reviewing the landscape and identifying windows of opportunity. <https://www.unicef.org/globalinsight/media/1156/file>. Retrieved on 07/06/21
- UNESCO Decision of the Intergovernmental Committee: 11.COM 10.c.6. <https://ich.unesco.org/en/Decisions/11.COM/10.c.6>. Retrieved on 22/07/21
- Van de Poel I (2020) Embedding values in artificial intelligence (AI) systems. *Minds Mach* 30: 385–409. <https://doi.org/10.1007/s11023-020-09537-4>. Retrieved on 14/06/21
- Van den Hoven J (2017) Ethics for the digital age: where is the moral specs? In: Werthner H, van Harmelen F (eds) *Value-Sensitive Design and Responsible Innovation. Informatics in the Future*. [https://doi.org/10.1007/978-3-319-55735-9\\_6](https://doi.org/10.1007/978-3-319-55735-9_6)
- Van der Hof S, Lievens E (2018) The importance of privacy by design and data protection impact assessments in strengthening protection of Children's personal data under the GDPR. *Commun Law* 23(1)
- Van Wynsberghe A, Robbins S (2019) Critiquing the reasons for making artificial moral agents. *Sci Eng Ethics* 25: 719–735. <https://doi.org/10.1007/s11948-018-0030-8>. Retrieved 23/06/21
- Veale M, Borgesius FJZ (2021) Demystifying the draft EU artificial intelligence act pre-print, July 2021. In: Version 1.2. forthcoming in 22(4) *Computer Law Review International*. <https://osf.io/preprints/socarxiv/38p5f>. Retrieved on 12/09/21
- Verbeek P-P (2011a) Expanding mediation theory. *Found Sci* 17(4):1–5. <https://doi.org/10.1007/s10699-011-9253-8>
- Verbeek P-P (2011b) *Moralizing technology: understanding and designing the morality of things*. University of Chicago Press
- Wachter S, Mittelstadt B (2019) A right to reasonable inferences: rethinking data protection law in the age of big data and AI' *Columbia Law Review* 2019 2
- Warren JD (2018) Children as young as seven suffer effects of discrimination, study shows: research also finds strong racial-ethnic identity is the best insulator. *ScienceDaily*. [www.sciencedaily.com/releases/2018/10/181022162138.htm](http://www.sciencedaily.com/releases/2018/10/181022162138.htm). Retrieved on 13/06/21
- Watcher S, Mittelstadt B, Russel C (2018) Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard J Law Rev* 31(2)
- UN General Assembly, Universal Declaration of Human Rights, 10 December 1948, 217 A (III), Available at: <https://www.refworld.org/docid/3ae6b3712c.html>. Retrieved 17 Sept 2021
- X. and Y. v. the Netherlands, App. No. 8978/80 (1985) 8 EHRR 235
- Zaga C (2021) The design of robotthings: non-anthropomorphic and non-verbal robots to promote children's collaboration through play. Doctoral Thesis, University of Twente
- Zaga C, Charisi V, Schadenberg B, Reidsma D, Neerinx M, Prescott T, Zillich M, Verschure P, Evers V (2017) Growing-up hand in hand with robots: designing and evaluating child-robot interaction from a developmental perspective. In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (Vienna, Austria), (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 429–430.
- Zawieska K (2020) Disengagement with ethics in robotics as a tacit form of dehumanisation. *AI Soc* 35:869–883. <https://doi.org/10.1007/s00146-020-01000-3>. Retrieved on 21/06/21

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.