

# *Amplifying positive bias: a case for positive biases in children-AI relations*

Dr. Karolina La Fors

Post-doc in Responsible Design, DesignLab, University of Twente

Broadening the interpretation of bias<sup>1</sup> within the artificial intelligence (AI) ethics<sup>2</sup> and responsible AI<sup>3</sup> discourse regarding the relations between machine-learning systems and children (age group 7-11 years) is pivotal. AI-mediated discriminatory exclusion has the most profound effects on children. However, mitigation mechanisms lag far more behind than concerning adults<sup>4</sup>. Developmental psychologists underlined that whereas discrimination can severely detriment children's psychological and physiological health and development<sup>5</sup>, allowing children to cultivate a strong sense of culture, gender<sup>6</sup>, identity<sup>7</sup>, and by this self-worth constitute the best long-term defense mechanisms against discriminatory effects. Broadening the interpretation of bias diverts from the mainstream discourse, which usually equates bias<sup>8</sup> with prejudice or discrimination and frames it (only) as something to eliminate. I propose alternatively to interpret bias as a preference or prioritization and distinguish between positive biases (pro-diversity, pro-inclusion) and negative biases (discriminative).

Research on why and how to amplify diversity and inclusion as core ethical values within children-AI relationships and against AI-mediated, discriminatory effects is scarce. This requires urgent addressing because, although AI technologies cannot be designed as fully artificial moral agents<sup>9</sup>, children can perceive them, next to humans, as fulfilling moral authority. Perceiving a moral authority to discriminate sets detrimental examples for children.

When defining artificial intelligence, this article follows Russel's and Norvig's conception which views AI systems as intelligent agents that can communicate with reasoning<sup>10</sup>. Such reasoning, however, cannot be completely bias-free<sup>11</sup>, because that would undermine AI's purpose. By following Gill's argument to approach AI as human-machine symbiosis in which human-machine and human-human relations are collectively understood:

---

<sup>1</sup> Mitchell, M. (1997) *Machine Intelligence*, McGraw-Hill, New York

<sup>2</sup> Coeckelbergh, M. (2020) *AI Ethics*, Cambridge: The MIT Press; Mittelstadt, B.D. et al. (2016). *The Ethics of Algorithms: Mapping the Debate*. *Big Data & Society*, December 10.1177/2053951716679679

<sup>3</sup> Dignum, V. (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, Springer

<sup>4</sup> Dignum, V. et. al. (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>>; La Fors, K. (2020) *Legal remedies for a forgiving society: Children's rights data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities*, *Comput. Law Secur. Rev.*, Vol. 38, Sep.

<sup>5</sup> Sawyer, P. J. et. al. (2012). *Discrimination and the stress response: Psychological and physiological consequences of anticipating prejudice in interethnic interactions*. *American Journal of Public Health*, 102(5), 1020–1026.

<sup>6</sup> Gilligan, C. (1982) *In A Different Voice*, Cambridge: Harvard University Press

<sup>7</sup> Piaget, J. (1972) *The Psychology of intelligence*, Totowa, NJ Littefield

<sup>8</sup> Lloyd, K. (2018) *Bias amplification in artificial intelligence systems*. Retrieved on 06/07/21: <<https://arxiv.org/abs/1809.07842>>

<sup>9</sup> See footnote 3, pp. 35-46

<sup>10</sup> Russel, S. J.; Norvig, P. (Eds.) (2020) *Artificial Intelligence A Modern Approach*; Fourth Edition; Pearson Series

<sup>11</sup> De Rijke, M., Graus, D. (2016). *Wij zijn racisten, Google ook*. *NRC Handelsblad*. Retrieved on 06/07/21: <[https://pure.uva.nl/ws/files/2801721/175163\\_graus\\_wij\\_2016.pdf](https://pure.uva.nl/ws/files/2801721/175163_graus_wij_2016.pdf)>

when I argue for positive bias, I also argue for more emphasis on the purpose of humans<sup>12</sup>. Regarding children, Dewey views “growth as an end”<sup>13</sup> and the purpose of children and their education. Therefore, adding growth to the principles for human-AI value alignment<sup>14</sup> by cultivating children’s sense of identity and diversity as positive biases would not only be fair. This would holistically foster: 1) children’s psychological well-being; 2) their capacity to imbue these values with their meanings; 3) children’s right to identity (Art. 8) and non-discrimination (Art. 2) as defined by the United Nations Convention on the Rights of the Child<sup>15</sup>; 4) the “diversity, non-discrimination and fairness” principle of the EU High-Level Expert Group’s Guidelines on Trustworthy AI<sup>16</sup> and the recommendations of UNICEF’s Policy Guidance on AI and children<sup>17</sup>.

Consequently, this research asks: *What challenges and opportunities are there to cultivate positive biases in children-AI relations in order to foster children’s diversity-minded and inclusive moral development, and what recommendations can be formulated so that AI can amplify positive biases for children and society?*

To answer this, the article synthesizes theoretical discussions on discrimination and identity within developmental psychology, AI ethics, responsible AI, and children’s rights to specify challenges and opportunities in mitigating negative and amplifying positive biases. It recommends open educational and co-creational spaces to amplify positive biases as requirements for children’s growth.

Such spaces would generate the following activities: 1) co-creative meaning-making<sup>18</sup> sessions with children on the values of identity, diversity, and non-discrimination; 2) aligning these meanings in children-adult; children-children and children-AI relations through interactive ethical, educational reflections and co-design scenarios. All these activities would include, but not be limited to: dialogue; art; craft; play sessions to discover and align children’s value meanings. These activities would be based on Dewey’s experience-based educational methods<sup>19</sup>, Verbeek and Tjink’s guidance-ethics methods<sup>20</sup>, and responsible co-design<sup>21</sup>. The gained transdisciplinary insights would offer input for non-formal and formal educational curricula for school teachers and children; ethical, legal, and design frameworks<sup>22</sup>; and children-centric trustworthy AI solutions. Amplifying positive biases in such spaces would ultimately cultivate children’s development of self-worth and capabilities to worthy others in their AI-mediated relations.

---

<sup>12</sup> Gill, K. (Ed.) (1996) *Human-Machine Symbiosis, The Foundations of Human-Centred System Design*, Springer-Verlag, London

<sup>13</sup> Dewey, J. (1997) *Experience and Education*, Free Press, New York

<sup>14</sup> Gabriel, I. (2020) Artificial Intelligence, Values, and Alignment. *Minds & Machines* 30, 411–437, pp. 432

<sup>15</sup> Convention on the Rights of the Child (1989) Treaty no. 27531. *United Nations Treaty Series*, 1577, pp. 3-178. Retrieved on 13/07/2021: <[https://treaties.un.org/doc/Treaties/1990/09/19900902%2003-14%20AM/Ch\\_IV\\_11p.pdf](https://treaties.un.org/doc/Treaties/1990/09/19900902%2003-14%20AM/Ch_IV_11p.pdf)>

<sup>16</sup> European Commission (2019): *EU High-Level Expert Group Guidelines on Trustworthy Artificial Intelligence* - Retrieved on 7/22/2021: <<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>>

<sup>17</sup> See footnote 4.

<sup>18</sup> Kegan, R. (1982) *The evolving self: problem and process in human development*. Cambridge, MA: Harvard University Press

<sup>19</sup> See footnote 13.

<sup>20</sup> Verbeek, P-P.; Tjink, D. (2020): *Guidance ethics approach. An ethical dialogue about technology with perspective on actions*. Platform voor de Informatie Samenleving. Enschede - Retrieved on 8/22/2021 <<https://ecp.nl/publicatie/guidance-ethics-approach/>>

<sup>21</sup> Zaga, C. (2021) *The Design of Robotings: Non-anthropomorphic and Non-verbal Robots to Promote Children's Collaboration Through Play*. Doctoral Thesis, University of Twente

<sup>22</sup> Code of Children Rights - <<https://codevoorkinderrechten.nl/en/>>