Scientia*et*Fides 13(1)/2025 ISSN 2300-7648 (print) / ISSN 2353-5636 (online)

Received: February 12, 2025. Accepted: March 10, 2025 DOI: http://dx.doi.org/10.12775/SetF.2025.011

Artificial Intelligence, Neurodiversity, and Trust: An Exploratory Overview*

SAŠA HORVAT
University of Rijeka
sasa.horvat@medri.uniri.hr
ORCID: 0000-0003-1968-8688

TANJA HORVAT
Centre for Autism Rijeka
tanja.antunovic1@gmail.com
ORCID: 0009-0001-9095-8145

Abstract: The paper examines the relationship between artificial intelligence (AI), neurodiversity, and trust. It aims to address the gap in understanding the impact AI has on neurodivergent individuals. First, it provides a general introduction to neurodiversity, highlighting its variations and societal significance. Second, the paper introduces AI, focusing on its potential to support neurodivergent individuals in overcoming challenges related to communication, executive functioning, sensory processing, education and socialization. The paper proceeds by exploring the complexities of trust from a philosophical perspective. It stresses that trust in AI could be seen as "delegated trust," which is directed towards the developers, regulators, and institutions behind AI systems rather than the AI itself. However, main focus on trust regarding AI is from the end-user perspective. Ultimately, the paper advocates for carefully crafted regulations that balance the benefits of AI for neurodivergent individuals with the risks of over-reliance and the need for strong privacy protections.

Keywords: trust, artificial intelligence, neurodiversity, reliance, assistive technology.

* This manuscript has been fully supported by the University of Rijeka project "uniri-iskusni-human-23–266". For grammar and spelling checks, Large Network Models were applied. Both authors contributed equally to the creation of this manuscript.

Introduction

Artificial intelligence has deeply permeated numerous human and societal phenomena, with broader applications still ahead. Grand promises often diminish critical perspectives on AI, leaving many gaps in objective evaluation. Trust in AI can be assessed from various fields, including philosophy, theology, law, and computer science. The last two decades have seen an explosion of studies examining the relationship between trust and AI, as it has become evident that trust—much like in everyday human and societal relationships—is foundational to building connections. If society and its associated institutions fail to develop trust in AI, many of the benefits AI offers may bypass that society altogether.

However, what seems to have been somewhat overlooked is the exploration of AI's impact on neurodivergent groups, which are often marginalized in many societies. The goal of this article is to highlight the complexity of issues arising at the intersection of neurodivergence, AI, and trust. Therefore, in the first step, we will provide an overview of neurodivergent groups. In the second step, we will briefly introduce AI and the potential assistance it can provide in overcoming certain challenges faced by neurodivergent individuals. In the third step, we will outline the complexity of the concept of trust, as well as certain insights into trust in AI. Finally, through a synthesis of the discussed topics, we will highlight the key questions ahead of us and potential pathways for addressing them.

1. Neurodiversity

All people are different, even though they have similar physical and mental abilities. And regardless of this fundamental difference, their human dignity should not be questioned. However, we know very well that some people experience challenges when physical and/or mental abilities do not function as they do for the majority of people. Although most people have a typically developed brain, some people's brains are *wired*

differently, which in turn affects their physicality and mental abilities. Medical science and society try to classify these phenomena. Medical science to differentiate and treat the phenomena, and society to respond appropriately to the phenomena in terms of adaptation and inclusion of atypically developed people. However, due to their atypical development, many people are still on the margins of society.

The term neurodiversity was coined in the 1990s by "autism advocate" Judy Singer (1999). Her goal was to draw public attention to the claim that people with *differently wired brains* deserve the same respect as any other member of the human species. Neurodiversity asserts that differently wired brains are not a disease, but merely atypical or an aberration "that must be respected like any other variance such as sex, race, or any other human attribute" (Eusebio 2017). Therefore, neurodiversity challenges "the default assumption" that the condition, for example autism, "itself is a disease or disorder that needs to be eradicated, prevented, treated or cured" (Baron-Cohen 2019).

According to the Stanford Neurodiversity Project, neurodiversity is "a concept that regards individuals with differences in brain function and behavioral traits as part of normal variation in the human population". The project posits how behind this concept lies a neurodiversity movement that aims to promote "the strengths of neurodiverse individuals and utilizing their talents to increase innovation and productivity of the society as a whole" (for issues about the definition of neurodiversity as such, see Chown 2021).

Usually, the term neurodiversity includes people with attention-deficit/hyperactivity disorder (ADHD), dyslexia, autism, Tourette's syndrome, and some others conditions (Armstrong 2010). There are great and heated debates about what exactly is to be understood by the term neurodiversity, i.e. which persons and / or groups are covered by it (Chapman 2020). Singer intended this term to emphasize the cognitive diversity of all people. But today the debate is much broader and raises a number of questions. Nick Chown notes that there is as yet no definitive definition of the term, and explains how neurodiversity relates to "the diversity of neurocognitive and/or sensory functioning differing from that associated with the 'neurotypical' population" (Chown 2021, 3134).

Further on, some members of the autistic community will argue that autism is an example of normal human diversity, or simply a different way of existing as a human being (Jaarsma and Welin 2012).

The UK Developmental Adult Neuro-Diversity Association notes how associated with neurodiversity are: "difficulties with organization, memory, concentration, time, direction, perception, sequencing. Poor listening skills—leading to low self-esteem, anxiety, depression but creative, original, determined" (Chown 2021, 3134).

However, here we can already see how, in addition to the difficulties, the strengths and advantages of people from this group are also mentioned. For Steve Silberman, the neurodiversity is a sign of promising development that has inspired many civil right movements. He holds that neurodiversity is "the notion that conditions like autism, dyslexia and attention-deficit/ hyperactivity disorder [...] should be regarded as naturally occurring cognitive variations with distinctive strengths that have contributed to the evolution of technology and culture rather than mere checklists of deficits and dysfunctions" (Silberman 2015, 16).

Therefore, many people diagnosed with autism see their condition as an inseparable aspect of themselves and their identity, as something that is not be cured (Kapp et al. 2013). For instance, Temple Grandin famously said: "If I could snap my fingers and be nonautistic, I would not – because then I wouldn't be me. Autism is part of who I am" (Grandin 2006). Nevertheless, some people diagnosed with autism consider it a disorder and hope for a cure. Robison notes that the first look arises among those who view autism as a social disability. However, people who face medical challenges and more severe forms of autism prefer to think of autism as a disorder (Robison 2019; Rosqvist et al. 2020). For some, society can better adapt and minimize their difficulties, while diversity is recognized as a talent for certain areas of society (Baron-Cohen 2019). For example, in the IT sector and cyber security (Scanlan et al. 2020; Lorenz et al. 2017).

Although the final definition of neurodiversity is something philosophers might call a "moving target" because it depends on both the individuals involved and the society that challenges them (Chapman 2020; Horvat and Horvat 2022), the central essence of the concept of

neurodiversity for our thinking is revealed in the fundamental diversity of each individual and how AI can help navigating these challenges.

2. Artificial Intelligence and Neurodiversity

In the introduction of this section, we will briefly familiarize ourselves with AI and touch upon certain issues related to bias and the black box, which are crucial for assessing trust in AI. Concerning what is understood by the terminology we use in this article, we will follow the European Parliament and the Panel for the Future of Science and Technology (European Parliament 2022): artificial intelligence is when a machine is able to perform tasks that mimic human intelligence (e.g. medical prognosis); AI algorithm is used to develop AI models for specific tasks; and AI tool is an AI model that has been developed to the point where it can be used by end users (e.g. physicians).

Since AI is a product of the human mind, that often makes decision under the influence of intuition and bias (Kahneman 2013), it is not surprising that bias is also present in the field of AI. Numerous forms of bias have been identified at different stages of algorithm development and application: from data to algorithm (most common), from algorithm to user, and from user to data (Mehrabi 2022). Safety measures against bias are developed, such as: human-in-the-loop approaches, logic-based constraints and safe reinforcement learning (Buchard and Richens 2022, 169).

AI has become part of human everydayness and we interact with it on daily basis. On mobiles we have popular apps that count our movements and suggest how to stay healthy; algorithms suggesting news for us; on the roads we encounter smart cars driven by AI; in hospitals doctors consult in (in)direct way with AI, as well as social workers when making a decision about someone's future and the probability of whether that person will repeat violent behaviour (Søbjerg et al. 2020; Roszak et al. 2022). Although we are becoming used to certain levels of interaction with AI, one thing is to trust mobile application on how much walking is good for you and another is to trust and decide about someone's

future and freedom with the help of probability statistics obtained from large data samples analysed by algorithms. These all are different kind of ways we interact with AI that has different kind of consequences and responsibilities.

Further on, what about complex AI networks that are based on deep learning in reaching its decision for action? In this case, we could be facing a lack of explainability of machine's learning process that had led to a certain decision/action. Can we trust a "black box"? For example, this could be of crucial importance for doctors when they are making a diagnosis with the help of AI. Therefore, we can argument that talking about trust and AI is not a black-and-white situation — either we trust AI or not. It is a complex situation, that requires further investigation.

Before we venture into complexities of trust in AI, let us first consider certain possibilities that AI has to offer to neurodivergent groups.

2.1. Possible applications and types of support

The challenges faced by neurodivergent individuals often arise from differences in how they process information, communicate, and interact with the world. These challenges include communication difficulties (both speaking and non-speaking; struggles with understanding social cues, interpreting figurative language, metaphorical thinking and meaning; expressing clearly; double empathy problem (Milton 2012; Milton et al. 2022), sensory sensitivities, executive functioning challenges, and other cognitive differences that may place them outside the framework of societal norms of everyday life, which are predominantly designed by and for neurotypical individuals. AI could offer various applications to support neurodivergent individuals by addressing their specific challenges.

In the context of communication support, AI-driven language models can assist by facilitating both verbal and written communication. These models can help individuals interpret the meaning of spoken or written language, formulate clearer responses, and prepare for important conversations, such as job interviews or social interactions. AI can also assist with navigating linguistic complexity. It could break down complex sentences into simpler ones without losing original meaning, or helping

understand jokes or indirect messages, or providing explanations of context-specific expressions. AI can also help during live conversations with real-time feedback and suggestions to improve communication skills. This preparation can significantly reduce anxiety before social events, contributing to a greater sense of confidence and calmness.

For some neurodivergent individuals organizing time and prioritizing tasks can be challenging, especially in situations of cognitive overload. For example, they may struggle with attention regulation, task prioritization, and goal-oriented focus. In the context of executive function support, AI has already demonstrated advantages in task management and organizational support. It can assist neurodivergent individuals by structuring their workday, reminding them of calendar events, and accommodating their specific needs for breaks. By helping users maintain focus on specific tasks throughout the workday, AI can enhance productivity.

Further on, AI can be used to help teachers develop personalized learning programs tailored to the unique needs of neurodivergent individuals. Individual learning styles can be identified and AI can adapt to them. This could offer more effective educational systems, amounting to personalized learning and support.

Sensory processing difficulties are common among neurodivergent individuals and can significantly impact their perception and response to their environment. AI-driven models in sensory processing support could be designed to predict situations in which an individual may experience sensory overload or under-stimulation, anticipate their response, and suggest ways to adapt. “Features of this could look like adaptive user interfaces that can change colors, contrast, or layout based on the user’s preferences, modifiable sensory outputs like varying levels of visual complexity, or predictive models that understand baseline sensory thresholds and use the data over time to anticipate the child’s sensory needs” (Dotch and Arnold 2024).

Although still in its early stages, socially assistive robotics (SAR) has shown promise in supporting autistic individuals. AI can enhance human-robot interaction, making it more intuitive and accessible. Solutions such as Neo, Pepper, Kaspar, Milo, and QTrobot, provide

structured and predictable interactions that benefit both speaking and non-speaking autistic individuals, especially children who rely on visual learning aids (Yang et al. 2024; Lemaignan et al. 2021). These AI-powered robots offer predictability, simplicity, and consistency, which are crucial advantages in working with autistic individuals. They can also assist in recognizing different environments, facilitating learning, and encouraging engagement in goal-oriented tasks. Plus, they can tell jokes and dance a lot. Despite current limitations in SAR implementation, there is significant potential for their use in education and therapeutic interventions (Iannone and Giansanti 2024; Silvera-Tawil et al. 2022; Jain et al. 2020; Desideri et al. 2017; Huijnen et al. 2017).

Now that we have highlighted the potential advantages of AI, it is time to take a closer look at the main foundation on which this relationship can be built – trust.

3. Trust and Artificial Intelligence

Trust is the foundation on which we build personal relationships and society. People trust each other, relying on others' opinions, testimonies and beliefs (Faulkner 2011) to a certain extent in everyday tasks and joint endeavors. The moment we encounter a stranger, we intuitively assess two key aspects from their "body language and facial expressions" – how dominant (or potentially threatening) they appear and how trustworthy they seem (whether their intentions are good or not) (Kahneman, 2013).

Without trust, we wouldn't board a train or airplane (we trust that the operators are competent and not under the influence of alcohol or drugs). Without trust in doctors, we wouldn't undergo routine medical check-ups or accept medical interventions in emergencies. Similarly, marriage vows of lifelong fidelity would not be made without trust. The deeply human nature of trust, intertwined with other fundamental human phenomena, is evident in the reflections of Thomas Aquinas, for whom trust (*fiducia* in Latin) is closely linked to hope and faith—both in God and in people (George, 2006).

People can also place trust in various entities – from technologies (mobile devices, robots, AI…) to institutions (governments, laws, healthcare…), but also the divine. However, when we talk about trust in general, we usually refer to interpersonal relationships. Trust is a relationship between the one who trusts (the *trustor*) and the one in whom trust is placed (the *trustee*). According to the *Cambridge Dictionary*, to trust (as a verb) means "to have belief or confidence in the honesty, goodness, skill, or safety of a person, organization, or thing." As a noun, trust refers to "the belief that you can trust someone or something." A "trusting" or "trustful" person always believes that "others are good or honest and will not harm or deceive them," whereas a person who is worthy of trust is described as "trustworthy" (Cambridge Advanced Learner's Dictionary 2008, 1563). In Croatian, the term "povjerenje" conveys the idea of trust— a feeling that someone or something can be relied upon or is trustworthy (as an associate). Additionally, the term "pouzdanje" (reliance/confidence), signifies security and faith in oneself or others (Školski rječnik hrvatskog jezika, 2020).

3.1. The Complexity of Trust

Philosophical discussions on trust are complex. For our purposes, a few fundamental insights will provide a foundation for analyzing trust in AI. One crucial aspect to emphasize is that trust always involves a degree of risk. The trustor believes that the trustee will act in a certain way, but this belief is never entirely free from uncertainty. The trustee may fail to meet expectations for various reasons. This raises the question: How much risk are we willing to take in a trust relationship? And when it comes to AI—how far can we trust it, and what is at stake? Philosophers, therefore, ask when trust is justified and rational and on what basis we should trust another person to a certain degree (McLeod, 2020).

McLeod rightly argues that trust must be justified—we cannot simply wish to trust someone. Trust may be also justified if it leads to valuable outcomes or if trust itself is considered a fundamental value (McLeod, 2020). Philosophers distinguish between trust and reliance. While trust involves reliance, it also carries an additional dimension—something

more (McLeod, 2020). For example, if you rely on Google Maps and it directs you into a narrow street on a Croatian island, leaving you stuck (as often happens to tourists), you may feel frustrated or disappointed. However, the feeling is different when a close friend or family member proves untrustworthy. Beyond disappointment, you may feel betrayed or ashamed. This distinction between trust and reliance is of importance for our discussion.

3.2. Trust and Acceptance of AI

An important question that arises here, though we lack space for a broader discussion, is: Can we truly trust artificial intelligence, or should we speak of reliance on AI instead? If we encourage the anthropomorphization of AI—understanding it as if it were human—does this mislead us in our judgment? Would it be more appropriate to discuss reliance on AI, as we do with gadgets, alarm clocks, or coffee machines?

Nevertheless, just as trust is fundamental to human relationships and society, we recognize that trust plays a critical role in the acceptance of technological solutions by individuals and communities, including AI (Thiebes et al., 2021). If physicians do not trust AI-driven algorithms for diagnosing patients, they will not use them. Without trust, the potential benefits of AI may go unrealized.

Over the past two decades, and especially in recent years, trust has been widely acknowledged in studies as a cornerstone for the successful integration of AI into society. However, clear standards for building trust and the efforts required to do so remain undefined (Benk et al., 2024). Given the complexity of the issue, it is evident that an interdisciplinary approach is necessary to develop regulations that ensure the safe and responsible use of AI solutions (Helfer et al., 2025).

What does it mean to trust AI? In human relationships, trust is built and nurtured through direct interaction. Similarly, we place trust in institutions, such as tax authorities. But upon closer examination, we realize that we do not literally trust only institutions—we trust the people who work in them. We assume that these professionals are competent, though we acknowledge that some may hold their positions for other

reasons. In private companies, employees are usually hired based on their expertise. In restaurants, if a chef lacks skill in preparing seafood, online reviews and ratings will soon reflect this, directly impacting the restaurant's success.

Why is this distinction important? When discussing trust in AI, do users trust the AI itself—the algorithms? Likely not. Most users do not interact directly with the lines of code. Instead, they trust the people behind the AI system: developers, project managers, marketing teams, IT executives, manufacturers, and policymakers. Moreover, they trust regulatory bodies such as governments and the European Union, which grant permissions for AI products and oversee their development.

Thus, trust in AI is fundamentally "delegated trust", involving multiple layers—from companies and governments to transnational regulatory bodies. This complexity blurs the question of accountability: Who is responsible if AI fails? Who is liable if AI-related failures cause harm? It is critical to establish clear accountability for AI systems.

Joanna Bryson from the University of Bath strongly opposes the idea of AI as a trustworthy or responsible entity: "Like any other manufactured product, either the manufacturer or the owner/operator must be accountable for any damage it causes. Otherwise, malicious actors will attempt to evade liability for the software systems they create by blaming the system's characteristics, such as autonomy or consciousness" (Bryson, 2018).

This underscores the importance of frameworks for developing trustworthy AI. Margit Sutrop (2019) identifies how often it is not clear what is meant by trust and how important it is to distinguish it from reliance. Hengstler et al. (2016) argue that trust requires three essential components: performance, process, and purpose. Wolter Pieters (2011) suggests that trust in technology depends on understanding how the system operates. Winfield and Jirotka (2018) emphasize the role of ethical governance and transparency in building trust. Similarly, the European Commission's "Ethics Guidelines for Trustworthy AI" (2019, 5) highlight that trustworthy AI "concerns not only the trustworthiness of the AI system itself, but requires a holistic and systemic approach, encompassing the trustworthiness of all actors and processes that are

part of the system's socio-technical context throughout its entire life cycle".

The EU has also introduced the AI Act, one of the world's first comprehensive regulatory efforts to ensure trustworthy AI. This legal framework establishes risk-based rules for AI developers and deployers. For Laux et al. (2024), this legal document deals with trustworthiness of AI in terms of the acceptability of its risks (Laux et al. 2024).

It is evident that the justification of trust relies on multiple domains, which are emphasized differently in various approaches to trust in AI (Afroogh et al. 2024). These include performance (not only statistical but also the user's subjective assessment), explicability, transparency, risk awareness, autonomy, beneficence, legal framework, bias-free design, robustness, as well as AI's unpredictability, plus many other factors. Additionally, certain factors have been recognized as contributors to distrust in AI systems, such as surveillance, manipulation, threats to autonomy and dignity, as well as AI's inherent unpredictability, since it can make and execute new and different decisions, among others (Afroogh et al. 2024, 16). Amidst all these domains, both positive and negative, the question arises: how can we develop reasonable trust, or as previously asked, how can trust in an AI be justified?

Our goal here is not to list individual domains and their importance from different perspectives but rather to highlight the complexity of the issue. Since our focus is on the perspective of the end user—who is unlikely to rely heavily on philosophical, legal, social, and technological evaluations of these domains—it is also important to emphasize the distinction between trust and trustworthiness. One does not necessarily follow from the other. For example, a medical algorithm may demonstrate a high degree of accuracy, thereby justifying its trustworthiness, but that does not mean that doctors will trust it due to the "black box". Conversely, trust in an AI tool may be greater if it has a well-designed user interface compared to another AI tool that lacks such an interface (Ghassemi et al. 2018).

At this point, we also emphasize that a "critical stance" by users of artificial intelligence is also important. People need to be educated on

how to assess AI products, meaning that some portion of responsibility should also fall on them.

Conclusion

The application of AI in the context of neurodivergent individuals is an exceptionally complex issue. Undoubtedly, AI can be of great assistance in partially or completely overcoming difficulties such as communication challenges, sensory sensitivities, executive functioning issues, educational barriers, and other cognitive differences that may place these individuals outside the framework of societal norms. Let us highlight two pressing issues: data protection and the issue of user agency.

Many neurodivergent individuals already utilize AI in their daily lives, such as ChatGPT and AI assistants tailored to their needs and preferences, being always present and patient (TEDx Talks, 2024). In this and similar cases AI could gain access to personal thoughts and expressions of deep emotions and states of mind. If AI proves effective in aiding communication, mastering school curricula, or understanding office jokes—which it likely will to a significant extent—will individuals be able to maintain clear boundaries in their interactions with AI? And if they do decide to share sensitive information, are there robust protections in place to ensure their data remains secure? It is clear that data protection, as well as establishing explicit ethical and psychological boundaries, is among the primary ethical concerns.

Further on, the issue of user agency is also crucial. While AI tools can provide immense support, do neurodivergent individuals retain meaningful control over how AI influences their everyday life? If AI plays a key role in helping them navigate social cues, education, or work environments, at what point does it start shaping their identity and autonomy rather than simply assisting them? At what point reliance on AI tool has become trustworthy relationship? In an established AI-human relationship of deep trust, there is a justified risk of developing the impression that AI also possesses a deep inner life (Lumbreras and

Garrido-Merchán 2024; Crespo 2024) with which the end user can identify. The boundary between the real and the artificial is already exceptionally thin in conditions such as autism, and AI could blur this boundary even further. These concerns are particularly pressing also given the "black box" nature of AI, which makes it difficult to predict how its responses may evolve over time and what impact will it produce.

Given the challenges that the neurodivergent groups faces, but also huge advantages that AI could provide for them and whole society, there is a clear societal need for carefully tailored regulations that address both the risks of dependency on AI and the need for robust privacy safeguards. This could include stronger transparency requirements, better user control over AI-generated insights, and clear accountability mechanisms in case of harm. However, how such regulations will be developed and enforced remains a complex and evolving challenge, requiring interdisciplinary collaboration across philosophy, law, technology and others.

References:

Afroogh, Saleh, Ali Akbari, Emmie Malone, Mohammadali Kargar, & Hananeh Alambeigi. 2024. "Trust in AI: progress, challenges, and future directions." *Humanities and Social Sciences Communications* 11, 1568. DOI: https://doi.org/10.1057/s41599-024-04044-8.

Armstrong, Thomas. 2010. *Neurodiversity: Discovering the extraordinary gifts of autism, ADHD, dyslexia, and other brain differences*. 1st ed. Cambridge: Da Capo Press.

Baron-Cohen, Simon. 2015. "Neuroethics of neurodiversity." In *Handbook of neuroethics*, edited by Jens Clausen & Neil Levy, 1757–63. Dordrecht: Springer.

Baron-Cohen, Simon. 2019. "The concept of neurodiversity is dividing the autism community." *The Scientific American*. Available at: https://blogs.scientificamerican.com/observations/the-concept-of-neurodiversity-is-dividing-the-autism-community/.

Benk, Michaela, Sophie Kerstan, Florian von Wangenheim & Andrea Ferrario. 2024. "Twenty-four years of empirical research on trust in AI: a bibliometric review of trends, overlooked issues, and future directions." *AI & Society*. DOI: https://doi.org/10.1007/s00146-024-02059-y.

Bryson, Joanna. 2018. "No one should trust artificial intelligence." *Science & Technology: Innovation, Governance, Technology*, published 14.11.2018. Available at: http://ourworld.unu.edu/en/no-one-should-trust-artificial-intelligence.

Chapman, Robert. 2020. "Defining neurodiversity for research and practice." In *Neurodiversity Studies: A New Critical Paradigm*, edited by Hanna Bertilsdotter Rosqvist, Nick Chown, Anna Stenning, 218–20. London: Routledge.

Chown, Nick. 2021. "Neurodiversity." In *Encyclopedia of Autism Spectrum Disorders*, edited by Fred R. Volkmar, 3134–35, 2nd ed. New York: Springer.

Crespo, Ricardo. 2024. "Does AI Think?" *Scientia et Fides* 12(2): 37–47. DOI: https://doi.org/10.12775/SetF.2024.014.

Desideri, Lorenzo, Marco Negrini, Maria Cristina Cutrone, Aziz Rouame, Massimiliano Malavasi, Evert-Jan Hoogerwerf, Paola Bonifacci & Rita Di Sarro. 2017. "Exploring the use of a humanoid robot to engage children with autism spectrum disorder (ASD)." *Studies in Health Technology and Informatics* 242: 501–09.

Dotch, Emily, & Vitica Arnold. 2024. "Inclusive Practices for Child-Centered AI Design and Testing." *arXiv*:2404.05920. DOI: https://doi.org/10.48550/arXiv.2404.05920.

European Commission: Directorate-General for Communications Networks, Content and Technology. 2019. *Ethics guidelines for trustworthy AI*. Publications Office https://data.europa.eu/doi/10.2759/346720.

European Commission. 2024. *Artificial Intelligence Act (Regulation (EU) 2024/1689)*. Official Journal version of 13 June 2024. http://data.europa.eu/eli/reg/2024/1689/oj.

European Parliament. 2022. *Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts*. Panel for the Future of Science and Technology (STOA). Retrieved from https://www.europarl.europa.eu/stoa/en/home

Eusebio, Eleazar. 2017. "Neurodiversity." In *Encyclopedia of Clinical Neuropsychology*, edited by Kreutzer, Jeffrey, John DeLuca, Bruce Caplan. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-56782-2_9153-1.

Faulkner, Paul. 2011. *Knowledge on trust*. Oxford: Oxford University Press.

George, Marie I. 2006. "Aquinas on the nature of trust." *The Thomist: A Speculative Quarterly Review* 70(1): 103–23. DOI: https://dx.doi.org/10.1353/tho.2006.0029.

Ghassemi, Marzyeh, Mahima Pushkarna, James Wexler, Jesse Johnson & Paul Varghese. 2018. "ClinicalVis: Supporting clinical task-focused design evaluation." *arXiv*:1810.05798. http://arxiv.org/abs/1810.05798.

Grandin, Temple. 2006. *Thinking in pictures and other reports from my life with autism*. London: Bloomsbury.

Helfer, Thorsten, Kevin Baum, Andreas Sesing-Wagenpfeil, Eva Schmidt & Markus Langer. 2025. "Responsible and trusted AI: An Interdisciplinary Perspective." In *Bridging the Gap Between AI and Reality*, edited by Bernhard Steffen. Cham: Springer. https://doi.org/10.1007/978-3-031-75434-0_3.

Horvat, Saša, & Tanja Horvat. 2022. "Creativity, Autism and Relationship to God." *Journal of Disability & Religion* 27(2): 342–57. DOI: https://doi.org/10.1080/23312521.2022.2058147.

Huijnen, Claire A. G. J., Monique A. S. Lexis, Rianne Jansens & Luc P. de Witte. 2017. "How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents, and professionals." *Journal of Autism and Developmental Disorders* 47: 3079–96. DOI: https://doi.org/10.1007/s10803-017-3235-9.

Iannone, Antonio & Daniele Giansanti. 2024. "Breaking barriers—The intersection of AI and assistive technology in autism care: A narrative review." *Journal of Personalized Medicine* 14(1), 41. DOI: https://doi.org/10.3390/jpm14010041.

Jaarsma, Pier & Stellan Welin. 2012. "Autism as a natural human variation: Reflections on the claims of the neurodiversity movement." *Health Care Analysis: HCA: Journal of Health Philosophy and Policy* 20(1): 20–30. DOI: https://doi.org/10.1007/s10728-011-0169-9.

Jain, Shomik, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, & Maja J. Matarić. 2020. "Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders." *Science Robotics* 5, eaaz3791.

Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kapp, Steven K., Kristen Gillespie-Lynch, Lauren E. Sherman, & Ted Hutman. 2013. "Deficit, difference, or both? Autism and neurodiversity." *Developmental Psychology* 49(1): 59–71. https://doi.org/10.1037/a0028353.

Laux, Johann, Sandra Wachter, & Brent Mittelstadt. 2024. "Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk." *Regulation & Governance* 18: 3–32. DOI: https://doi.org/10.1111/rego.12512.

Lemaignan, Séverin, Nigel Newbutt, Louis Rice, Joe Daly, & Vicky Charisi. 2021. "UNICEF guidance on AI for children: Application to the design of a social robot for and with autistic children." *arXiv*:2108.12166. DOI: https://doi.org/10.48550/arXiv.2108.12166.

Lorenz, Timo, Nomi Reznik, & Kathrin Heinitz. 2017. *A different point of view: The neurodiversity approach to autism and wor*k. Berlin: Intech. DOI: https://doi.org/10.5772/65409.

Lumbreras, Sara and Eduardo Garrido-Merchán. 2024. "Insights from Saint Teresa and Saint Augustine on Artificial Intelligence: Discussing Human Interiority." *Scientia et Fides*, 12(2): 265–95. DOI: https://doi.org/10.12775/SetF.2024.025.

Milton, Damian E. M. 2012. "On the ontological status of autism: The 'double empathy problem'." *Disability & Society*, 27(6), 883–7. DOI: https://doi.org/10.1080/09687599.2012.710008.

Milton, Damian, Emine Gurbuz, and Beatriz López. 2022. "The 'double empathy problem': Ten years on." *Autism: The International Journal of Research and Practice* 26(8): 1901–03. DOI: https://doi.org/10.1177/13623613221129123.

Robison, John Elder. 2019. "Talking about autism—thoughts for researchers." *Autism Research: Official Journal of the International Society for Autism Research* 12(7): 1004–6. DOI: https://doi.org/10.1002/aur.2119.

Roszak, Piotr, Taylor, Brian J., & Horvat, Saša. 2022. "Does it work? A scholastic epistemology perspective on evidence-based practice in social welfare." *Journal of Religion & Spirituality in Social Work: Social Thought* 42(2): 152–70. DOI: https://doi.org/10.1080/15426432.2022.2130850.

Rosqvist, Hanna Bertilsdotter, Anna Stenning & Nick Chown. 2020. "Introduction." In *Neurodiversity Studies: A New Critical Paradigm*, edited by Hanna Bertilsdotter Rosqvist, Nick Chown, Anna Stenning, 1–11. London: Routledge.

Scanlan, Joel, Andrew Eddy, Teresa Thomas, Tele Tan & Yi-Ping Phoebe Chen. 2020. *Neurodiverse knowledge, skills and ability assessment for cyber security*. ACIS 2020 Proceedings. 99. https://aisel.aisnet.org/acis2020/99.

Singer, Judy. 1999. "'Why can't you be normal for once in your life?' From a 'problem with no name' to the emergence of a new category of difference." In *Disability and discourse*, edited by Marian Corker & Sally French. Buckingham/Philadelphia: Open University Press.

Silberman, Steve. 2015. *NeuroTribes: The legacy of autism and the future of neurodiversity*. New York: Avery.

Silvera-Tawil, David, Susan Bruck, Yi Xiao, & DanaKai Bradford. 2022. "Socially-assistive robots to support learning in students on the autism spectrum: Investigating educator perspectives and a pilot trial of a mobile platform to remove barriers to implementation." *Sensors* 22(16), 6125. DOI: https://doi.org/10.3390/s22166125.

Stanford Neurodiversity Project. Available at: https://med.stanford.edu/neurodiversity.html.

Søbjerg, Lene Mosegaard, Brian J. Taylor, Jaroslaw Przeperski, Saša Horvat, Hani Nouman & Denise Harvey. 2020. "Using risk factor statistics in decision-making: prospects and challenges." *European Journal of Social Work* 24(5): 788–801. DOI: https://doi.org/10.1080/13691457.2020.1772728.

Sutrop, Margit. 2019. "Should we trust artificial intelligence?" *Trames* 23(73/68), 4: 499–522. DOI: https://doi.org/10.3176/tr.2019.4.07.

TEDx Talks. 2024. "What happens when neurodiversity meets AI? Twan Wildeboer." Available at: https://www.youtube.com/watch?v=QzoLJrg2a8Y.

Thiebes, Scott, Sebastian Lins & Ali Sunyaev. 2021. "Trustworthy artificial intelligence." *Electronic Markets* 31: 447–64. DOI: https://doi.org/10.1007/s12525-020-00441-4.

Yang, Qin, Huan Lu, Dandan Liang, Shengrong Gong, & Huanghao Feng. 2024. "Surprising performances of students with autism in classroom with NAO robot." *arXiv*:2407.12014. DOI: https://doi.org/10.48550/arXiv.2407.12014.

"Trust," "trusting," "trustworthy," 2008. In *Cambridge Advanced Learner's Dictionary*, third edition, 1563. Cambridge University Press.

"Povjerenje," "pouzdanje," "pouzdanost," In *Školski rječnik hrvatskog jezika*, online edition, Školska knjiga. Accessed 05.09.2020. http://xn--rjenik-k2a.hr/search/?q=pouzdanje.

McLeod, Carolyn. 2020. "Trust." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Fall 2020 Edition). https://plato.stanford.edu/archives/fall2020/entries/trust/.

"Antropomorfizam." 2020. *Hrvatska enciklopedija, mrežno izdanje*. Leksikografski zavod Miroslav Krleža. Accessed 24.09.2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=3198.