IntechOpen *Journals*

AI, Computer Science and Robotics Technology

NEWS & VIEWS

New Norms for AI: Zero Trust—Verify Then Trust

Allison Wylde[1,2,*]

[1] Cardiff University, Cardiff, UK
[2] Glasgow Caledonian University, London, UK
*Correspondence: E-mail: wyldea@cardiff.a.c.uk; allison.wylde@gcu.ac.uk

Abstract

This conceptual article highlights an important puzzle concerning the actions of prominent AI industry leaders calling for a six-month halt to the further development and training of AI, while some have already deployed and continue to deploy AI; a so-called arms race. As security risks increase at scale, managing the security of AI is an essential consideration. Although implementing standards for AI is currently a preferred solution for many, this article argues that, on their own, standards may not provide sufficient capability and flexibility. An argument is presented that cyber norms of zero trust for AI (ZTAI) alongside standards are essential for the security of AI, along with a call to action.

Keywords: AI, cyber security, trust, zero trust, conceptual framework

1. *Introduction*

Although many AI technology leaders have suggested that the development and training of AI should be stopped [1], some organisations, governments and individuals world-wide continue to expand their use of AI. Given important and increasing security concerns, management of the security of AI is essential for continued high-risk operations.

While implementing standards for AI is seen currently as an important and a preferred solution, this conceptual article argues that, on their own, standards may not provide sufficient capability. To achieve the fast and agile approach required to deal with AI, this article proposes leveraging cyber norms alongside standards for cyber security: zero trust for AI (ZTAI) [2].

2. *Controlling AI: standards*

The development of AI so far mirrors the ethos of the early internet, championed by Tim Berners-Lee (the father of the internet), as a space of openness and freedom. However, this approach to AI development has resulted in frameworks of

Citation
Allison Wylde (2023), New Norms for AI: Zero Trust—Verify Then Trust. *AI, Computer Science and Robotics Technology* 2(1), 1–4.

DOI
https://doi.org/10.5772/acrt.27

Published
12 September 2023

governance that seem polarised between some regions that have deployed standards, compared with others where there are no standards [3].

The focus of AI standards concerns issues including data quality and robustness, ethical considerations (privacy, transparency and accountability), and data trust frameworks for safe, secure, and equitable data transfers. Also, many other frameworks are in place, such as responsible business conduct and due diligence, but these may be not actioned. While some regions may not implement standards, initiatives and guidelines have been developed. These include ethical principles for military applications to ensure that AI systems are accountable, transparent, and consistent with human values, ethical principles for AI development focused on transparency, accountability and human oversight, and support to help promote AI research, as well as knowledge sharing through cooperation and partnerships [3]. Although a number of important questions remain concerning ethics in AI; as a conceptual paper, the scope here is limited to examining issues of trust in AI.

Although AI standards can ensure trustworthiness, transparency, and common definitions and frameworks, some limitations exist. Some standards may reduce flexibility, innovation or the time involved in developing standards may result in their redundancy. Therefore, deploying zero trust as a norm may address some limitations in standards through providing a fast and simple approach that can rapidly be taken up across multiple domains.

3. Trust in AI

Although some AI users appear to trust AI technology without question and with little regard to issues of safety, key challenges and questions arise concerning trust.

As the literature on trust is contested, this article views trust as relational, complex, and comprising separate processes often broken down into steps [4]. Trust-building is widely viewed as involving a trustor who holds a positive view of an outcome, followed by an assessment of trust and then a willingness to accept vulnerability and risk during trusting [4]. Although AI is trusted, the trust process appears to concentrate on the positive viewpoint element, fast-forwarding to trust [5]. As a result, potential security concerns are overlooked and in consequence, calls for a pause to the development of the AI technology increased [1].

One solution is offered through a zero trust approach, based on “verify first and trust later”. Zero trust rests on the premise of no presumptive trust and a risk-based approach to trust based on verification on a continuous basis [5]. A zero trust mindset focuses on the verification of identity prior to trust [6]. Identification is required for all entities, whether an individual, a device or a software [7].

Verification of identity overcomes some issues in AI, as models are often opaque, and the source information or algorithm may change. In sum, zero trust encourages users not to trust. Thus, this article suggests adding zero trust as a norm, alongside using standards to manage AI.

4. ZTAI for cyber security?

Although standards for AI enable trustworthiness and transparency along with simplified operations, this approach takes time. In the case of AI for cyber security, where immediate verification and monitoring is required, zero trust offers a rapid solution. A ZTAI approach could be implemented across the operations of a wide range of sectors and domains to manage permissions and access control to reduce threats. Indeed, zero trust is already deployed in cyber security operations [8] and offers a knowledge base that can be leveraged to provide a simple framework that can be deployed without further delay.

5. Call to action

To take zero trust forward, what is needed now is support and cooperation among members of the AI ecosystem and beyond. Cooperation is essential to help develop the AI body of knowledge. AI models are perfectly suited to this task, and it may be that researchers can harness the capability of AI to work on this task. Checking the output, given the scale, is a daunting task, well beyond human capabilities. What is required are ideas beyond simple retrospective checks, along with adequate time and space to consider and concentrate on the larger issues and problems, not least imagining the possible avenues of further development of the models and potential capabilities alongside risks.

6. Concluding remarks

Alongside the double-exponential rise in the capabilities of AI models [9], and an audience split between unbridled trust and acceptance, and those who call for caution; questions of trust are crucial [10]. This article recommends framing questions not simply as trust and/or distrust in AI, but rather leveraging zero trust thinking to help address the ZTAI puzzle. Indeed, future scenarios could explore both the possible and the impossible. As a final word, as all technology relies on availability, in high-risk applications such as cyber security there can be no errors. No one wants to receive the message: “The server is currently overloaded with other requests; sorry about that! Please try again later or contact us through our help center if the error persists” [11]. Yet another example of the need for zero trust?

Conflict of interest

The author declare no conflict of interest.

Acknowledgments

The author thanks University and Industry colleagues and friends for their time and support. No funding was received for this work.

References

1 Futureoflife.org. *Pause giant AI experiments: An open letter [Internet]*; 2023 Mar 22. Available from: https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

2 Internet Governance Forum, IGF 2022. *Best practice forum on cyber security output document [Internet]*; 2023 Jan. Available from: https://www.intgovforum.org/en/filedepot_download/56/24125.

3 Galindo L, Perset K, Sheeka F. An overview of national AI strategies and policies. *Going Digital Toolkit Note, No. 14*. Paris: OECD Publishing; 2021. Available from: https://goingdigital.oecd.org/data/notes/No14_ToolkitNote_AIStrategies.pdf.

4 Mayer R, Davis J, Schoorman F. An integrative model of organizational trust. *Acad Manag Rev*. 1995;**20**(3):709–734.

5 Kindervag J. No more chewy centers: Introducing the zero trust model of information security. *Forrester [Internet]*; 2010 Sept 17. Available from: https://media.paloaltonetworks.com/documents/Forrester-No-More-Chewy-Centers.pdf.

6 National Cyber Security Centre (NCSC). *Zero trust architecture design principles [Internet]*; 2021 Nov 20. Available from: https://www.ncsc.gov.uk/collection/zero-trust-architecture.

7 Rose S, Borchert O, Mitchell A, Connelly S. *Zero trust architecture, NIST special publication 888-207*. Gaithersburg, MD: NIST; 2020.

8 *The White House. 2022. Moving the US Government towards zero trust cybersecurity principles [Internet]*; 2022 Jan 26. Available from: https://www.whitehouse.gov/omb/briefing-room/2022/01/26/office-of-management-and-budget-releases-federal-strategy-to-move-the-u-s-government-towards-a-zero-trust-architecture/.

9 Harris T, Raskin A. *The A.I. dilemma*, YouTube [Internet]; 2023 Mar 9. Available from: https://www.youtube.com/watch?v=xoVJKj8lcNQ.

10 Wylde A. Zero trust: Never trust, always verify. In: *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. Piscataway, NJ: IEEE; 2021 Jun 14. p. 1–4, doi:10.1109/CyberSA52016.2021.9478244.

11 OpenAI. *Chat GPT-3 Language model [Internet]*. Output based on questions concerning AI standards; 2023 Mar 23.