

Book Review
**From Sandbox to Scale: What Scaling Responsible AI Offers Leaders
and Where the Guidance Runs Out**

Siri N Boggaram

Undergraduate Researcher, Bachelor of Biotechnology, Golden Gate University

Kshitij Taluja

Undergraduate Researcher, Bachelor of Computer Science, Golden Gate University

Dr. Smrite Goudhaman

Global Professor of Practice, Golden Gate University

Author: Noelle Russell

Title: *Scaling Responsible AI: From Enthusiasm to Execution*

Publisher: John Wiley & Sons, Inc.

Year: 2025

ISBN: 978-1-394-28964-6

Pages: 352

Abstract

Noelle Russell's *Scaling Responsible AI: From Enthusiasm to Execution* presents two frameworks, LEAD AI and SECURE AI, to help organizations move from initial AI experiments to responsible, large-scale deployment. Drawing on a two-decade career at IBM, AWS, Amazon Alexa, and Microsoft, Russell breaks the AI lifecycle into practical checklists, "Takeaways," and "Reflection Questions" for leaders, innovators, and practitioners. This review argues that while the frameworks offer a teachable, accessible language for launching a responsible-AI initiative, the book has three notable shortcomings: (1) its evidence base relies heavily on anecdote rather than peer-reviewed research; (2) its chapter on bias and fairness identifies problems effectively but does not equip readers to measure or address them; and (3) its guidance is oriented primarily toward large, U.S.-based enterprises, with little consideration for smaller organizations, the public sector, or the Global South. Overall, the book is a practical, workshop-ready guide to responsible AI at the

enterprise level, but readers seeking technical depth, documented evidence, or a more globally inclusive perspective will need to look beyond it.

Keywords: responsible AI; AI governance; AI scaling; enterprise leadership; algorithmic bias; AI risk management

This review differs from one based on academic or policy literature: it assesses a book written by someone who is inside the industry it discusses. The author, Noelle Russell, has drawn upon two decades of actual experience with AI implementations at IBM, AWS, Amazon Alexa, and Microsoft, and it is this practical perspective, rather than the reviewers' own backgrounds, that determines both the book's value and its shortcomings. The main metaphor in Chapter 2 compares AI to a "baby tiger": attractive while it is playing in the sandbox, but hazardous when it develops its teeth and claws. The narrative progresses from "Day One: The Hype Cycle" to "Day Two: The Road to Reality" and then to organizational resilience. The fifteen chapters are supported by two internal frameworks: LEAD AI (Leadership, Ethical foundations, AI governance, designing for inclusivity) in Chapter 1 and SECURE AI (Secure, Ethical, Compliant, Unbiased, Red Team, Explainable, Accountable, Iterative) in Chapter 7, together with checklists and end-of-chapter "Takeaways" and "Reflection Questions".

The main contribution of the book is also its greatest weakness. For a leadership team unfamiliar with AI governance, LEAD AI and SECURE AI are clear, memorable, and teachable in a single workshop session. Russell points out that SECURE AI "has been documented in the Responsible AI Standard which Microsoft released as early as 2018"

(Russell, 2025, p.131); its eight categories generally correspond to the four functions of the NIST AI Risk Management Framework (NIST, 2023): Secure corresponds to Govern, Red Team and Explainable resemble Map, Unbiased corresponds to Measure, and Ethical corresponds to Iterative Manage. The book never makes this comparison directly, so readers cannot assess how well it aligns with the standards against which their organization is likely to be audited.

The section on 'Bias and Fairness', Chapter 11, is one of the book's most effective. Russell looks at five kinds of bias: selection, algorithmic, confirmation, representation, and measurement. It uses well-known examples such as Amazon's discontinued resume-screening tool, which was found to disadvantage applications that included the word 'women's' (Dastin, 2018), and facial-analysis systems in which error rates increase for people with darker skin, echoing the findings of the 'Gender Shades' study, although Russell only gives a general credit to unnamed 'Stanford researchers' (Buolamwini & Gebru, 2018). Although the citations are not exact, the chapter does not explain to readers how to measure or reduce bias after it has been identified: it never refers to established fairness criteria such as demographic parity or equalized odds (Hardt et al., 2016), nor the audit methods a practitioner would need to operationalize their own checklist.

The absence of endnotes or a bibliography is part of a wider approach; incidents such as the 2017 Equifax breach and Amazon's hiring tool are presented as examples rather than being supported by documentation, and regulations such as the GDPR and CCPA are mentioned without being examined. While this is not really a deficiency of a scholarly kind

but rather a purposeful decision in the context of a business book aimed at practitioners, it leaves readers wishing for sources they could cite or wanting to verify the claims for themselves.

The checklists included in the book: an "AI capability assessment framework", an "AI vendor due diligence checklist", and a "Responsible AI audit", are based on organisational structures Russell has encountered as a consultant, such as dedicated compliance teams, separate cloud budgets, and in-house ethicists. The book provides little advice for organisations that have limited resources, such as non-profits or municipal governments, or for an organisation outside the United States and Western Europe. In the section on "Inclusive Testing" in Chapter 7, cultural background, ability, and socioeconomic status are identified as factors that should be taken into account but nothing is said about language coverage, low-bandwidth deployment, or the global data-annotation workforce; these omissions are in line with the wider criticisms of governance frameworks that have been developed for well-resourced, Western-based institutions (Birhane, 2020). Furthermore, the companion website, which is referred to numerous times as a source of templates, also shows that the book is closely tied to Russell's AI Leadership Institute consulting practice, a point that should be considered when evaluating the independence of the advice given.

Russell speaks in a friendly, first-person, motivational style, uncommon for this genre, about her journey from working with IBM mainframes to her time at Amazon Alexa, where she developed more than one hundred Alexa skills while being the only woman, the only Latina, and the only mother on the team. Although this tone gives the impression of

authenticity, it at times replaces analysis with affirmation, as in the case of stating that "scaling responsibly is a critical factor" with certainty rather than with qualification.

The book seems more like a corporate training program than a monograph. The Takeaways and Reflection Questions at the end of each chapter, together with nine criteria for evaluating an AI initiative, ten pitfalls associated with 'AI pilot purgatory', and nine Red Team examples, turn every chapter into a kind of workshop appropriate for group onboarding.

As an onboarding guide for enterprise leaders, the book is mostly successful in that it presents governance in an accessible, checklist style and gives a thorough treatment of the entire process, including sandbox experimentation through to production monitoring, bias auditing, and organizational resilience. However, the three limitations mentioned above determine its extent: it is based on anecdotal evidence which cannot be verified, its chapter on bias and fairness identifies problems but does not provide readers with the means to solve them, and the advice given is tailored to large companies in the United States. For those in such a position the book will prove to be a truly useful starting point; but for those who need technical depth, documented evidence, or advice appropriate for smaller organisations, the public sector, or the Global South, they should combine it with the NIST AI Risk Management Framework (NIST, 2023), the current regulation, or the Gartner (2024) hype-cycle research that Russell herself references.

References

- Birhane, A. (2020). *Algorithmic colonization of Africa*. *Script-ed*, 17(2), 389–409.
<https://doi.org/10.2966/scrip.170220.389>
- Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. *Proceedings of Machine Learning Research*, 81, 77–91.
- Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
- Gartner, Inc. (2024). *Hype cycle for emerging technologies*.
<https://www.gartner.com/en/articles/hype-cycle-for-emerging-technologies>
- Hardt, M., Price, E., & Srebro, N. (2016). *Equality of opportunity in supervised learning*. *Advances in Neural Information Processing Systems*, 29, 3315–3323.
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- Russell, N. (2025). *Scaling responsible AI: From enthusiasm to execution*. John Wiley & Sons.