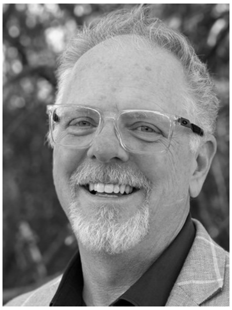
Received: 29.09.2025 | Revised: 11.01.2026 | Accepted: 15.01.2026

Trust and Responsibility in AI: An Interdisciplinary Social-Sector Perspective



Nathan Chappell

Abstract: The rapid adoption of artificial intelligence has intensified debates about responsibility, ethics, and trust. While regulatory frameworks and organizational ethics statements are proliferating, responsible AI is too often treated as compliance or reputation management rather than an organizing principle of practice. This perspective argues that social-sector organizations - including nonprofits, NGOs, and other mission-driven institutions - offer an instructive lens for rethinking responsible AI because they operate with structural vulnerability and high trust dependence. Drawing on business ethics, organizational theory, nonprofit and social-sector management, and human flourishing scholarship, the paper proposes shifting from harm-avoidance toward trust-centered, flourishing-oriented AI integration.

Keywords: Responsible AI; Trust; Social sector; Nonprofit management; Human flourishing

Vertrauen und Verantwortung in der KI: Eine interdisziplinäre Perspektive aus dem sozialen Sektor

Zusammenfassung: Die rasante Verbreitung künstlicher Intelligenz hat die Debatten über Verantwortung, Ethik und Vertrauen intensiviert. Während regulatorische Rahmenbedingungen und ethische Leitbilder von Organisationen immer zahlreicher werden, wird verantwortungsvolle KI allzu oft als Compliance- oder Reputationsmanagement behandelt und nicht als organisierendes Prinzip der Praxis. Diese Perspektive argumentiert, dass Organisationen des sozialen Sektors - darunter gemeinnützige Organisationen, NGOs und andere missionsorientierte Institutionen - eine lehrreiche Perspektive für ein Umdenken in Bezug auf verantwortungsvolle KI bieten, da sie mit struktureller Vulnerabilität und hoher Vertrauensabhängigkeit arbeiten. Auf der Grundlage von Wirtschaftsethik, Organisationstheorie, Management im gemeinnützigen und sozialen Sektor sowie wissenschaftlichen Erkenntnissen zur menschlichen Entfaltung schlägt der Artikel vor, von der Vermeidung von Schaden hin zu einer vertrauenszentrierten, auf menschliches Wohlergehen ausgerichteten KI-Integration überzugehen.

Stichwörter: Verantwortungsvolle KI; Vertrauen; Sozialer Sektor; Nonprofit-Management; Menschliche Entfaltung

Introduction - Responsibility or Reputation?

The rapid adoption of artificial intelligence across industries has sparked urgent debates about responsibility. In the corporate world, however, responsible AI too often risks collapsing into reputation management. Declarations of ethical intent, ESG statements, and glossy cause marketing campaigns abound, yet responsibility is frequently treated as a safeguard for brand image rather than an organizing principle of practice. The consequences are not hypothetical. Biased algorithms and opaque systems already show how easily efficiency eclipses fairness when responsibility is an afterthought.

The social sector offers a strikingly different lens. For nonprofits, responsibility is not optional, nor is it a reputational exercise. It is existential. Their missions depend on fragile bonds of trust with donors, beneficiaries, and communities. Without that trust, funding dries up, partnerships dissolve, and impact is curtailed. This makes nonprofits a vital proving ground for what responsible AI should look like.

As both a nonprofit practitioner and an architect of AI solutions, my vantage point is shaped by years of seeing how technology collides with mission, values, and human dignity. From this perspective, the nonprofit sector is both a warning system and a blueprint for what responsible AI must become.

Structural Vulnerability as Clarity: A Social-Sector Lens

Many organizations adopt AI with powerful incentives to scale quickly, improve efficiency, and capture competitive advantage. In large enterprises, formal strategies and secured software landscapes can accelerate adoption, sometimes allowing responsibility to drift into a form of “risk management” handled after value has been extracted. At the same time, in small and medium-sized enterprises (SMEs), family businesses, start-ups, and social-sector organizations, adoption is often far less centralized - employees bring their own devices, experiment with generative tools, and use software without an official strategy or license. This “shadow AI” reality can increase both productivity and exposure, especially when data governance, privacy expectations, and accountability are unclear.

Social-sector organizations, by contrast, often operate with structural vulnerability. Resource scarcity is not unique to nonprofits - many private companies, especially SMEs and start-ups, also face tight margins. What differs is how quickly trust loss becomes existential for mission-driven institutions. A single ethical lapse - such as a biased allocation of resources or a misstep in data privacy involving vulnerable communities - can permanently damage credibility with donors, beneficiaries, partners, and regulators. In many social-sector contexts, there is no practical second chance to “pivot” after a failed experiment. The risks are immediate and relational.

Yet scarcity is not only a constraint; it functions as a kind of ethical clarity. Decisions about whether and how to adopt AI are filtered through a sharper lens: Will this tool uphold or erode trust? Will it enhance dignity or compromise it? With limited resources, nonprofits must scrutinize technology not for how it scales operations, but for how it aligns with mission. The absence of financial incentives enables nonprofits to see risks and opportunities more clearly.

By being forced to ask first-order ethical questions rather than defaulting to market-driven metrics, nonprofits reveal both the dangers of unchecked AI adoption and the possibility of aligning innovation with responsibility. Scarcity produces not paralysis but vision - the ability to see the forest through the trees.

Trust as the True Currency

If corporations often measure success in profits, social-sector organizations ultimately measure success in impact - with trust as the enabling condition that makes sustained impact possible. Every donation and partnership rests on confidence that resources are stewarded responsibly. In the context of AI, this means that ethical lapses are not minor reputational setbacks; they are existential threats that can unravel years of relationship-building in an instant.

Nonprofit adoption of AI - whether in donor engagement, program delivery, or operations - is therefore filtered through the trust imperative. Tools must not only function effectively but also be explainable, transparent, and aligned with mission values. A predictive system that improves efficiency while undermining fairness or privacy is unacceptable because it erodes nonprofit legitimacy.

Nonprofits rarely have that luxury. Once trust is broken, donors move on, beneficiaries lose confidence, and the organization may never recover. This asymmetry makes nonprofits uniquely attuned to the ethical stakes of AI.

From this vantage point, responsible AI is not merely about avoiding harm. It is about sustaining the fragile bonds of confidence between institutions and the people they serve. For corporations seeking to embed ethics into practice, the nonprofit example offers a living model of trust-centered AI integration - an approach where responsibility is not a public-relations veneer but an existential requirement.

From Standards to Culture

The European Union's AI Act (European Union, 2024), with its tiered risk classifications, represents the most comprehensive regulatory effort to date. UNESCO's Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021) and the OECD's AI Principles (Organisation for Economic Co-operation and Development [OECD], 2019) further articulate global commitments to fairness, transparency, and accountability. Alongside these, ESG frameworks and corporate declarations of AI ethics proliferate. Together, these instruments provide an important set of guardrails, creating a shared vocabulary for what responsible AI should mean in practice.

Yet a persistent tension remains: standards are necessary but insufficient. A company can comply fully with existing regulations while still treating responsibility as a box to check rather than a value to embody. The danger is compliance theater - satisfying reporting requirements without cultivating a culture of responsibility. In such cases, ethics is outsourced, and responsibility reduced to paperwork.

Their limited resources rarely allow for elaborate compliance structures or dedicated ethics offices. Instead, responsibility must be integrated into the day-to-day culture of the organization because survival demands it. Decisions about whether to use AI for something as simple as scheduling volunteers are not purely operational; they are ethical. Is it accessible, fair, and transparent? These questions cannot be outsourced; they must be lived daily.

For businesses navigating the complexities of responsible AI, this social sector perspective is instructive. The challenge is not merely harmonizing diverse standards or reporting systems. It is embedding responsibility into the cultural DNA of organizations, where ethical reflection informs not just what companies report, but how they operate. Responsibility, in this sense, must move from external compliance to internal conviction.

Beyond Ethics: Human Flourishing as the Horizon

Ethics should be the minimum expectation for AI. The more urgent and transformative question is whether AI contributes to human flourishing. Nonprofits are uniquely positioned to surface this broader horizon because their missions are already aligned with advancing human well-being.

Human flourishing can be understood across seven empirically grounded dimensions - character, health, relationships, finances, happiness, faith, and meaning (VanderWeele, 2017). Together, they reflect a view of well-being that extends beyond compliance. Each dimension offers a lens for evaluating AI not only for its risks but also for its capacity to enhance life. AI that reinforces fairness strengthens character; systems that protect access to care or reduce unnecessary barriers advance health; technologies that foster inclusion enrich relationships. When AI reduces inequality, it supports financial dignity; when it lowers stress or creates space for joy, it contributes to happiness.

These dimensions sharpen the contrast between “do no harm” and “do good.” Corporate responsibility often frames ethics in terms of preventing harm - avoiding bias, ensuring compliance, protecting privacy. While necessary, this bar is too low. Flourishing asks a deeper question: does this technology make life more worth living? For nonprofits, the answer resonates naturally because their missions already align with these outcomes. For corporations, adopting a flourishing-centered framework represents a paradigm shift, moving responsible AI out of the realm of marketing or regulation and into the realm of purpose.

This shift changes governance conversations from “Are we compliant?” to “Are we worthy of trust, and does this system make life more worth living for the people touched by it?”

Implications for Future Research

Responsible AI in trust-dependent and mission-driven contexts raises several research questions that merit further inquiry:

1. How can organizations operationalize trust as a governance and performance measure without reducing it to an instrumental variable?
2. Which governance models best support flourishing-oriented AI adoption across sectors, including SMEs and social-sector organizations?
3. How do AI-related ethical failures differ structurally and relationally across organizational types, and what recovery pathways exist?
4. What practices most effectively embed responsibility into organizational culture rather than delegating it to compliance functions?
5. How can human flourishing be operationalized as an evaluative standard for AI systems, and what indicators or metrics could credibly assess an AI model’s contribution to individual and collective well-being?

Conclusion – Learning from the Canaries

Nonprofits are the canaries in the coal mine of AI ethics. Their scarcity forces clarity, their survival depends on trust, and their missions are aligned with flourishing. Unable to treat ethics as marketing or an add-on, nonprofits reveal both the perils of neglect and the promise of alignment. They show how easily dignity is compromised when technology is careless, and how powerfully well-being grows when responsibility is embedded.

For corporate actors, the lesson is unmistakable. Responsible AI cannot remain an external exercise in reputation or compliance. It must become a cultural commitment - a way of doing business inseparable from identity. Standards and regulations may prevent the worst abuses, but they cannot create authentic responsibility. That requires conviction, leadership, and a willingness to adopt new measures of success.

If ethics is the floor, flourishing is the ceiling. The nonprofit sector demonstrates how this shift is possible: by making trust the central metric, embedding responsibility into culture, and aligning technology with humanity's deepest aspirations. Businesses that learn from these canaries will not only avoid catastrophe; they will help chart a path toward an AI future where innovation and flourishing reinforce one another.

The coal mine of AI adoption is fraught with risk, but the canaries are already singing.

References

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities and risks. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

Organisation for Economic Co-operation and Development. (2019). OECD AI principles. Adopted May 2019; updated May 2024. OECD. https://oecd.ai/en/ai-principles

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. Adopted 23 November 2021. United Nations Educational, Scientific and Cultural Organization.

VanderWeele, T. J. (2017). On the promotion of human flourishing. Proceedings of the National Academy of Sciences of the United States of America, 114(31), 8148-8156. https://doi.org/10.1073/pnas.1702996114

Nathan Chappell is Chief AI Officer at Virtuous and a leading voice at the intersection of generosity, responsible AI, and social impact. He is the author of Nonprofit AI: A Comprehensive Guide to Implementing Artificial Intelligence for Social Good and The Generosity Crisis: The Case for Radical Connection to Solve Humanity's Greatest Challenges. His work focuses on trust-centered AI adoption, ethical innovation, and the application of advanced technologies to strengthen human connection and social good.

Address: Virtuous Software, 1 N 1st St, Phoenix, AZ 85004, USA.
E-mail: nathan.chappell@virtuous.org
ORCID: https://orcid.org/0009-0006-2487-6332