24 Beyond the radar

Addressing undeclared AI use in evaluation practice

Steve Jacob

What you will learn from this chapter

This chapter explores the impact of AI tools on professional practices within the field of evaluation. It focuses on the reporting phase, when findings are synthesized, drafted, and communicated. At this stage, the use of generative AI (GenAI) tools raises questions about authorship and transparency. Such issues become particularly salient in the context of Shadow AI, which refers to the use of AI outside formal authorization and regulatory frameworks (Puthal, 2025; Balogun et al., 2025). This phenomenon amplifies risks related to confidentiality and the professional credibility of evaluators. With no mechanisms in place to ensure oversight, Shadow AI also confronts evaluators with the challenge of aligning technological innovation with established professional standards and ethical guidelines grounded in principles such as transparency and accountability. Finally, readers are invited to consider the conditions under which AI use should be disclosed, the appropriate audiences for such disclosure, and the forms it may take.

Study context

The increasing accessibility of GenAI tools has begun to subtly and sometimes invisibly reshape the evaluation practice (Jacob, 2025). In one evaluation project conducted shortly after the release of ChatGPT, an evaluator tested the tool to synthesize a sizeable volume of documentation provided by the client. The use of AI in this evaluation was primarily exploratory, intended to assess the potential of the new AI tool in accelerating the data analysis process. Due to the evaluator's uncertainty regarding the success of this experiment with the project and the potentially negative reaction from the client, the decision was made to keep the use of AI in that evaluation a secret. This decision was made despite the confidential nature of the documents and the existence of a contractual agreement that would have protected their disclosure.

In a more recent case, as awareness of AI capabilities expanded, a team of evaluators contemplated the utilization of analogous tools for the processing

DOI: 10.4324/9781003799139-24

of data accumulated during the evaluation (McGuire & Martin, 2025). In this particular instance, the evaluators sought explicit permission from the client prior to proceeding. However, the client declined to authorize the use of AI, citing concerns regarding confidentiality, data protection, and the absence of clear institutional guidelines.

A notable incident occurred in Australia, where Deloitte was under scrutiny after delivering a 237-page government report containing AI-generated errors, including fabricated references. The firm subsequently acknowledged the use of GenAI in drafting the document, agreed to reimburse part of its 440,000-dollar consultancy fee, and faced financial and reputational consequences due to inadequate verification and lack of transparency in its reporting process (McGuirk, 2025).

What emerges from these situations is a shifting landscape in which the design of evaluations, the choice of analytical tools, and the transparency of reporting all intersect with the phenomenon of Shadow AI.

Rationale behind Shadow AI

The rise of Shadow AI can be explained by a combination of factors related to innovation, governance, organizational dynamics, and psychosocial dimensions.

The pursuit of effectiveness, efficiency, and expediency is a key drive of the adoption of AI tools, regardless of whether their use is formally regulated or carried out secretively. Many employees use AI tools to automate repetitive tasks, save time, and improve their performance, particularly in high-pressure or resource-constrained environments. Furthermore, some employees act proactively and innovatively, using these tools to experiment and develop new solutions, often with the intention of enhancing organizational performance (Chin et al., 2025). This desire to improve individual or organizational productivity is accentuated by the ease of access to GenAI tools, made possible by low-code or no-code platforms, and the popularization of tools such as ChatGPT, Copilot, and Gemini (Puthal et al., 2025).

At the same time, the constraints imposed by internal IT departments are motivating some employees to bypass established protocols. When IT departments' approval processes are considered too slow or their policies too restrictive, companies are unwittingly encouraging employees to adopt alternative solutions without prior approval. This phenomenon is exacerbated by the absence or ambiguity of internal AI governance policies, as regulatory frameworks remain underdeveloped or even non-existent in many organizations (Balogun et al., 2025).

In addition to these factors, there are underlying psychosocial elements that could explain the proliferation of Shadow AI. Some employees have admitted to using AI in secret, citing concerns about being perceived as replaceable or appearing lazy by delegating certain tasks to the machine (Silic et al., 2025).

Finally, the lack of awareness regarding legal, ethical, and cybersecurity risks is contributing to the clandestine use of AI in professional contexts (Balogun et al., 2025).

Risks and consequences of Shadow AI

The unauthorized use of AI tools generates multiple risks and may result in substantial consequences for evaluators, stakeholders, and the organizations they serve.

A primary concern is the maintenance of data confidentiality. Through Shadow AI practices, evaluators may, whether intentionally or inadvertently, expose sensitive information to external systems. For example, since confidential documents can be uploaded into publicly accessible platforms, such as ChatGPT or Gemini, bypassing organizational data-protection mechanisms is now possible, thereby increasing the risk of unauthorized access or misuse. The consequences of such breaches can include the leakage of sensitive information, the exposure of personal data from participants, or the disclosure of financial and contractual details linked to evaluation mandates. Documented cases, such as the incident at Samsung where internal meeting notes and source code were shared via ChatGPT, illustrate the potential for serious breaches of information security (Sidorkin, 2025). In such contexts, organizational intellectual property may also be compromised when internal data are absorbed into public AI models that are not under institutional control (Renieris et al., 2023). These concerns over confidentiality and intellectual property are echoed in recent work explaining how local AI systems can mitigate these risks by keeping sensitive data within organizational boundaries (Bustamente Aragonés, 2027).

Shadow AI also introduces cybersecurity vulnerabilities. The use of unauthorized tools expands the digital attack surface by creating additional entry points into sensitive systems and datasets. When evaluators rely on publicly accessible GenAI platforms, they may unconsciously bypass institutional firewalls, encryption protocols, and access controls that are specifically designed to safeguard confidential information. In the absence of adequate pre-defined cybersecurity measures, Shadow AI can facilitate data exfiltration, malware injection, or phishing attempts, particularly when evaluators copy and paste confidential and personal information into online platforms. These risks are further amplified when AI platforms retain data for model training or store prompts in ways that may be accessible to third parties (Puthal et al., 2025).

A third area of concern involves legal and regulatory compliance. Shadow AI can expose evaluators, consulting firms, or public bodies to legal and reputational risks, including the loss of client trust, contract termination, or litigation. The professional credibility of evaluators is at stake since clients entrust them with sensitive information, assuming it will be handled securely. Failure to comply with frameworks such as the European Union's General Data Protection

Regulation (GDPR, governing data privacy) or the U.S. Health Insurance Portability and Accountability Act (HIPAA, regulating health data) can result in sanctions or lawsuits. For example, an American hospital was sued for using an unapproved diagnostic AI tool that contributed to a medical error, while the U.S. Securities and Exchange Commission fined a bank for employing an unauthorized AI-based credit scoring system that produced discriminatory outcomes (Balogun, 2025). Such cases underscore the issue of professional responsibility, as evaluators who integrate AI secretively and without approval may ultimately be held accountable for errors generated by these systems.

Beyond these technical and legal aspects, Shadow AI generates a form of "governance drift," a situation where employees' practices diverge from established organizational rules and standards (Silic et al., 2025). Unequal adoption of Shadow AI within teams can also introduce biases in performance assessment, thereby undermining fairness and equity in organizations (Dolci & Aguiar, 2025).

Overall, Shadow AI constitutes a multidimensional threat situated at the intersection of technological, human, ethical, and legal risks. For evaluators, such dynamics are particularly problematic, as the credibility of evaluation depends on adherence to principles of fairness, transparency, and accountability.

Discussion: how evaluators should position themselves regarding Shadow AI

As GenAI tools rapidly spread, evaluators have to revisit existing ethical frameworks and adapt them to this new reality. This discussion situates the phenomenon of Shadow AI within the broader themes of evaluation standards, ethics, and competencies. It presents a set of guiding questions to assist evaluators in reflecting on how to integrate AI into their professional practice in ways that uphold established standards of transparency, accountability, and integrity. It also highlights the importance of cultivating digital and ethical competencies to effectively address the challenges posed by AI-assisted evaluation.

Over the past decades, many Voluntary Organizations for Professional Evaluation (VOPEs) have adopted evaluation standards or ethical guidelines to help evaluators navigate ethical challenges and highlight the core values of the profession. While these standards and guidelines were adopted prior to the "ChatGPT tsunami," they nevertheless provide a foundation for guiding disclosure of AI use in evaluation. When applied to AI, these principles emphasize a key requirement: the use of AI must be disclosed in a transparent, proactive, and contextualized manner, aligning with the core professional values of evaluation.

This expectation of transparency extends beyond the field of evaluation to other knowledge workers as well. Notably, many major scientific publishers now require disclosure of AI involvement.

When to disclose

Disclosure should begin at the earliest stages of an evaluation (e.g., during the planning, contracting, and designing) and continue throughout the implementation, reporting, and dissemination of results. Early and sustained communication ensures that clients, participants, and other stakeholders understand from the outset how AI is being used and can make informed decisions about its appropriateness. Disclosure may also be formalized through a contractual clause specifying whether AI use is authorized, for which tasks, and under what conditions AI is authorized (e.g., human validation, confidentiality).

How to disclose

The manner of disclosure is as important as its timing. Information about AI use must be intelligible, accessible, complete, and adapted to the digital literacy and cultural context of the intended audiences. Reports, notices, presentations, and direct conversations are all viable means to this end. Regardless of the chosen format, the objective should be clarity and inclusion. Disclosure should avoid overly technical language or opaque explanations, instead empowering interested parties to understand clearly the implications of AI for the evaluation process. For example, when AI tools are used to assist in coding qualitative data, the evaluator could include a short note in the methodology section explaining which tool was employed, what type of data it processed, and how the outputs were verified. This simple clarification helps maintain transparency without overburdening readers with technical details.

To whom to disclose

Disclosure must reach all parties directly or indirectly affected by the evaluation process. It is important that clients and funders are well informed of any potential use of AI in an evaluation so as to enable them to assess whether the proposed methods align with contractual obligations and ethical considerations. Evaluation participants also deserve clear and detailed information, as their data, testimonies, or survey responses may be processed through AI systems. They have a right to know how their contributions are being managed, stored, and interpreted. Decision-makers who base policy adjustments or resource allocations on evaluation results require explicit disclosure as well so they can judge the reliability, limitations, and potential biases introduced by AI-supported methods.

What to disclose

The content of disclosure must be comprehensive and explicit, ensuring that no relevant information is omitted or obscured. Evaluators should specify the role of AI in the evaluation, whether it is used for text generation, data coding,

analysis, or drafting recommendations. It is also a professional responsibility to clearly delineate the limitations and risks associated with the tools, including potential biases, errors, opacity, and so-called "hallucinations." Furthermore, it is beneficial to outline the safeguards in place, such as human oversight, verification mechanisms, and recourse options. This information could lead to a broader discussion regarding the potential impact of AI use on evaluative findings and recommendations, particularly regarding accuracy and interpretation.

Emerging guidelines from evaluation providers

Some evaluation firms are developing their own guidelines for using and deploying AI tools in order to regulate the conduct of their employees. For example, a recent document emphasizes the importance of transparency and states the following:

> Generative AI should not be used without the knowledge of those exposed to the generated content. Users should clearly communicate to their mission supervisor, whether internal or external, the uses of AI systems and, to the best of their knowledge, disclose information about the data sources and algorithms used.
>
> (Pluricité, 2025:13, translation)

This statement is supported by the following general principles: "Always disclose the use of generative AI to your superiors" and "inform stakeholders of AI applications in accordance with current professional sector practices" (Pluricité, 2025:13, translation). Similar initiatives are emerging beyond the private sector as research institutes and universities increasingly adopt internal guidelines for GenAI use, particularly with regard to transparency, data protection, and research integrity.

These guidelines illustrate a growing commitment to transparency, aimed at preventing Shadow AI and establishing AI governance policy. Anuj et al. (2027) detail institutional practices that align with the governance recommendations advanced here. It is noteworthy that the movement toward AI transparency and governance is currently being driven more by evaluation providers than by evaluation commissioners. This reflects the longstanding asymmetry in the management of ethical responsibilities across the field.

Evaluator competencies in the digital era

The implementation of ethical values and professional standards necessitates competencies in digital and AI literacy. Evaluators must not only understand how AI tools process data and generate results but also how to determine when AI use is appropriate and what safeguards are necessary to uphold professional

Table 24.1 Shadow AI in evaluation: what to do and what to avoid

Do's	*Don'ts*
Disclose early and often: communicate AI use from planning through to reporting	**Don't use AI secretly**: hidden or late disclosure undermines trust
Use plain and accessible language: explain AI's roles and limits clearly	**Don't rely on technical jargon**: avoid vague or overly complex references that confuse stakeholders
Adapt communication: take into account stakeholders' literacy and context (digital, cultural, linguistic)	**Don't assume** not all stakeholders share the same understanding of AI tools
Protect confidentiality: apply safeguards and check approval before uploading sensitive data	**Don't upload confidential or personal data** into public AI tools without safeguards. (E.g., HR files, contracts, participant data, etc.)
Acknowledge risks and limitations such as biases, errors, opacity, and hallucinations	**Don't present AI-assisted findings** as neutral, objective, or infallible
Frame AI use within ethical frameworks to meet core professional values such as transparency, fairness, and accountability	**Don't treat disclosure as a checkbox** while ignoring ethical and professional guidelines

integrity. Digital and AI literacy is an evolving capacity that requires continuous learning and adaptation (Rinaldi & Nielsen, 2024). By developing and strengthening these competencies evaluators are able to adapt their methods and ensure the responsible and transparent use of AI in evaluation.

In summary

Disclosing the use of AI in evaluation must be done in a timely and consistent manner. It has to be transparent, complete, and comprehensible. Table 24.1 presents a set of recommendations that illustrate how evaluators can integrate AI tools into their practice while remaining aligned with professional and ethical standards.

Conclusion

AI has the potential to automate routine tasks and achieve notable efficiency gains for evaluators. However, due to the accessibility and apparent ease of using GenAI, evaluators may be at risk of neglecting the core values underpinning evaluation practice.

Shadow AI introduces potential risks, including confidentiality breaches, biased results, and reputational damage. More fundamentally, it undermines the trust that is the cornerstone of the evaluator-client relationship and, ultimately, the legitimacy of evaluation itself.

Addressing Shadow AI in evaluation requires more than technical vigilance; it necessitates an explicit commitment to transparency through systematic disclosure of AI use. This disclosure is not merely procedural. It demonstrates adherence to established professional standards and ethical guidelines, thereby reinforcing the credibility and integrity of the evaluation process.

Conflict of interest statement

The author declares no conflicts of interest regarding this manuscript.

AI use declaration

In the preparation of this chapter, the author used generative artificial intelligence tools. ChatGPT and PDF AI were used to assist in the extraction and formulation of text once referenced sources had been read and analyzed. Additionally, DeepL was used for linguistic refinement. All AI-generated outputs were reviewed, validated, and verified by the author.

References

Anuj, H., Raimondo, E., & Den Boer, H. (2027). Balancing innovation and rigor at the World Bank Independent Evaluation Group: Thoughtful integration of artificial intelligence for evaluation. In K. Bruce, V. J. Gandhi, & S. B. Nielsen (Eds.), *From algorithms to evidence: Using generative AI in evaluation practice* (pp. 90–99). New York, NY: Routledge. https://doi.org /10.4324/9781003799139

Balogun, A. Y., Metibemu, O. C., Olutimehin, A. T., Ajayi, A. J., Babarinde, D. C., & Olaniyi, O. O. (2025). The ethical and legal implications of Shadow AI in sensitive industries: A focus on healthcare, finance and education. *Journal of Engineering Research and Reports, 27*(3), 1–22. https://doi.org/10.9734/jerr/2025/v27i31414

Bustamante Aragonés, S. (2027). Leveraging self-hosted AI solutions for data-secured research and evaluation. In K. Bruce, V. J. Gandhi, & S. B. Nielsen (Eds.), *From algorithms to evidence: Using generative AI in evaluation practice* (pp. 262–271). New York, NY: Routledge. https://doi.org /10.4324/9781003799139

Chin, T., Li, Q., Mirone, F., & Papa, A. (2025). Conflicting impacts of shadow AI usage on knowledge leakage in metaverse-based business models: A Yin-Yang paradox framing. *Technology in Society, 81*. https://doi.org/10.1016/j.techsoc.2024.102793

Dolci, P. C., & Aguiar, M. S. (2025). Governance and generative artificial intelligence: Challenges and risks of shadow AI in business environment. *AMCIS 2025 Proceedings*. https://aisel.aisnet.org/amcis2025/lacais/lacais/1

Jacob, S. (2025). Artificial intelligence and the future of evaluation: From augmented to automated evaluation. *Digital Government: Research and Practice, 6*(1), Article 10. https://doi.org/10.1145/3696009

Mazzeo Rinaldi, F., & Nielsen, S. B. (2024). Artificial intelligence. Challenges for evaluators. In S. Bohni Nielsen, F. Mazzeo Rinaldi, & G. J. Petersson (Eds.), *Artificial*

intelligence and evaluation. Emerging technologies and their implications for evaluation (pp. 287–307). Routledge. https://doi.org/10.4324/9781003512493-14

McGuire, M., & Martin, L. (2025, June 19). *AI reality check: Navigating AI use in evaluation practice* [Webinar]. Canadian Evaluation Society.

McGuirk, R. (2025, October 7). Deloitte to partially refund Australian government for report with apparent AI-generated errors. *AP News*. https://apnews.com/article/ab54858680ffc4ae6555b31c8fb987f3

Pluricité (2025). *Charte pour l'usage et le déploiement de systèmes d'intelligence artificielle de Pluricité et Emoha*.

Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., & Yeun, C. Y. (2025). Shadow AI: Cyber security implications, opportunities and challenges in the unseen frontier. *SN Computer Science, 6*(5), 405. https://doi.org/10.1007/s42979-025-03962-x

Renieris, E. M., Kiron, D., & Mills, S. (2023, June 20). Building robust RAI programs as third-party AI tools proliferate. *MIT Sloan Management Review*.

Sidorkin, A. M. (2025). AI platforms security. *AI-EDU Arxiv*. https://doi.org/10.36851/ai-edu.vi.5444

Silic, M., Silic, D., & Kind-Trüller, K. (2025). From Shadow IT to Shadow AI–Threats, Risks and Opportunities for Organizations. *Strategic Change*. https://doi.org/10.1002/jsc.2682