Program Evaluation Meets AI: An Integrative Literature Review on the Potential Implications of Artificial Intelligence in Agricultural Program Evaluation

Julysa A. Benitez[1]
abril.benitezbonilla@ag.tamu.edu

Miguel Diaz[2]
migueldiaz@vt.edu

Karissa Palmer[1]
karissapalmer@tamu.edu
727-858-7185

Emily Fuller[2]
emilyfuller@vt.edu

Dr. Rafael Landaverde[1]
rafael.q@ag.tamu.edu
440-476-8411

Dr. Holli Leggette[1]
hollileggette@tamu.edu
979-458-3039

Dr. Gary Wingenbach[1]
gary.wingenbach@ag.tamu.edu
979-321-5756

[1] **Texas A&M University**
Department of Agricultural Leadership, Education, and Communications

[2] **Virginia Tech University**
School of Public and International Affairs

Program Evaluation Meets AI: An Integrative Literature Review on the Potential Implications of Artificial Intelligence in Agricultural Program Evaluation

Introduction

Artificial Intelligence (AI) is set to revolutionize program evaluation methods across diverse sectors, including agriculture (Nielsen, 2023; Patel, 2023; Tilton et al., 2023). In response to global challenges such as food security, population growth, and climate change, agriculturalists are increasingly adopting program evaluation (Bamberger, 2024) to improve program practices and strategies (Thompson et al., 2021). Program evaluation's systematic approach aims to assess program effectiveness, identify areas for improvement, and understand underlying factors contributing to success or failure (Rossi et al., 2019). Further, recent advancements integrated AI into program evaluation across various industries, showcasing its potential in areas like natural language processing, machine learning, and predictive modeling (Marco et al., 2024). These technologies provide more accurate assessments and insights that have the potential to transform evaluation methods (Patel, 2023); however, the varying quality and specificity of sessions further emphasize the need for better guidance on AI use in evaluation (Raftree, 2023).

Thus, the literature we reviewed in our study highlights current AI implementations and their limitations, focusing on data visualization, processing, analysis, storage, and real-time reporting (Nielsen, 2023). By coupling AI with agricultural program evaluation, methods can be modernized to better equip evaluators and enhance existing evaluation processes (Tilton et al., 2023). Exploring AI's applications in program evaluation could uncover opportunities to adapt current integrations for agricultural contexts, potentially revolutionizing evaluation practices in this sector. Therefore, our study sought to answer one research question: How will the investigation into AI's implications for program evaluation underscore the need for further research in this evolving field?

Methods

Our study used an Integrative Literature Review (ILR) method to consolidate diverse research on AI's applications in program evaluation, encompassing scholarly articles, opinion pieces, and policy documents (Lubbe et al., 2020; Torraco, 2016). Adhering to Lubbe et al.'s (2020) structured ILR approach, we identified a literature gap regarding the underuse of AI in program evaluation, which shaped our research lens (Step 1). Next, we drew our sample using established inclusion criteria (i.e., scholarly publications, websites [e.g., Elizabeth Grim Consulting, Oxford Policy Management], working papers [e.g., Harvard Business School], blogs [e.g., Reform Works], and documents published between 2000-2024, excluding pre-2000 publications due to the recent emergence of AI (Chen et al., 2022; Step 2). Using Research Rabbit for enhanced search capabilities and AI software like ChatGPT to explore diverse perspectives, we completed the keyword-guided literature review (Step 2a) and comprehensive document analysis (Step 2b). Next, we conducted a critical review and data extraction, identifying themes such as Chat GPT, data mining, machine learning, generative AI, predictive analysis, AI in education evaluation, cost-efficiency, and associated challenges (Steps 3 and 4). We concluded with a team debrief to validate findings and ensure research integrity (Step 5).

Results

The ILR highlighted several key themes regarding AI's potential contributions to agricultural program evaluation. The 2023 American Evaluation Association (AEA) conference highlighted the growing interest and concerns about AI in evaluation with 18 presentations on AI, machine learning, and big data (Raftree, 2023). Sessions on ChatGPT attracted more attention than those on broader themes like big data or machine learning, indicating a need for clearer practical applications in evaluation (Raftree, 2023). AI-driven ChatGPT systems were identified as helpful tools for automating evaluation planning and enhancing stakeholder engagement through natural language processing capabilities (Patel, 2023). Data mining techniques were found effective in extracting valuable insights from complex agricultural datasets, facilitating evidence-based decision-making in program evaluation (Tilton et al., 2023). Machine learning algorithms emerged as powerful tools for predictive modeling of program outcomes, enabling evaluators to forecast impacts and optimize resource allocation (Rossi et al., 2019). Furthermore, generative AI technologies may enhance evaluation efficiency and scalability by automating routine tasks and reducing manual workload (Patel, 2023).

Conclusions & Recommendations

Integrating AI into program evaluation offers significant advantages but presents significant challenges as well, including ensuring factual accuracy, maintaining cultural sensitivity, and avoiding biases, especially regarding marginalized groups (Ferretti, 2023; Head et al., 2023; Patel, 2023). AI tools reflect a partial view of the world, influenced by their developers and the data they are trained to use (Ferretti, 2023). Evaluators must critically assess AI outputs and collaborate with AI scientists to create ethical frameworks for AI use in program evaluation (Head et al., 2023; Patel, 2023).

Observation at the AEA conference about evaluators becoming alienated from evaluative aspects when over-relying on AI underscores the importance of understanding AI's capabilities and limitations (Raftree, 2023). Evaluators must discern AI's impact on their profession, including potential mechanization of activities, intellectual property issues, and legal concerns (Ferretti, 2023; Linden & Yarnold, 2016; Patel, 2023), while recognizing that the technology cannot replace the human component involved in program evaluation.

To effectively integrate AI into agricultural program evaluation, evaluators should familiarize themselves with AI, conduct pilot projects, and gradually implement AI into evaluation processes (Patel, 2023). Comprehensive research and testing are needed to identify AI's benefits and limitations, ensuring ethical standards and maintaining the human element in evaluations. Ongoing research and dialogue are crucial for leveraging AI's potential while mitigating risks and ensuring evaluation integrity. Agricultural program evaluators must update evaluation methods as more efficient and innovative techniques arise (Tilton et al., 2023). More research is necessary to assess AI's efficiency in improving evaluator productivity within agricultural program evaluation.

References

Bamberger, M. (2024). Chapter 34: Using big data to strengthen evaluation. In K. Newcomer & S. Mumford (Eds.), *Research Handbook on Program Evaluation*. Edward Elgar Publishing Limited.

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, *25*(1), 28–47. https://www.jstor.org/stable/48647028

Cvetkovic, A., Savela, N., Latikka, R., & Oksanen, A. (2024). Do we trust artificially intelligent assistants at work? An experimental study. *Human Behavior and Emerging Technologies*, *2024*, e1602237. https://doi.org/10.1155/2024/1602237

Ferretti, S. (2023). Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators. *New Directions for Evaluation*, *2023*(178–179), 73–84. https://doi.org/10.1002/ev.20557

Head, C. B., Jasper, P., McConnachie, M., Raftree, L., & Higdon, G. (2023). Large language model applications for evaluation: Opportunities and ethical implications. *New Directions for Evaluation*, *2023*(178–179), 33–46. https://doi.org/10.1002/ev.20556

Linden, A., & Yarnold, P. R. (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *22*(6), 868–874. https://doi.org/10.1111/jep.12592

Lubbe, W., ten Ham-Baloyi, W., & Smit, K. (2020). The integrative literature review as a research method: A demonstration review of research on neurodevelopmental supportive care in preterm infants. *Journal of Neonatal Nursing*, 26(6), 308–315. https://doi.org/10.1016/j.jnn.2020.04.006

Marco, G., Evertsson, E., Riley, D. J., Tyrchan, C., & Rathi, P. C. (2024). Augmenting DMTA using predictive AI modelling at AstraZeneca. *Drug Discovery Today*, 103945. https://doi.org/10.1016/j.drudis.2024.103945

Nielsen, S. B. (2023). Disrupting evaluation? Emerging technologies and their implications for the evaluation industry. *New Directions for Evaluation*, 2023(178–179), 47-57. https://doi.org/10.1002/ev.20558

Patel, D. (2023). Revolutionizing program evaluation with generative AI: An evidence-based methodology. *International Journal for Multidisciplinary Research*, 5(3), 1–11. https://www.ijfmr.com/papers/2023/3/4105.pdf

Raftree, L. (2023, November 8). What's next for emerging AI in evaluation? Takeaways from the 2023 AEA conference. MERL Tech. https://merltech.org/emerging-ai-for-evaluation/

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach*. SAGE Publications.

Thompson, N. M., Reeling, C. J., Fleckenstein, M. R., Prokopy, L. S., & Armstrong, S. D. (2021). Examining intensity of conservation practice and adoption: Evidence from cover crop use on U.S. Midwest farms. *Food Policy, 101*, 1–10. https://doi.org/10.1016/j.foodpol.2021.102054

Tilton, Z., LaVelle, J. M., Ford, T., & Montenegro, M. (2023). Artificial intelligence and the future of evaluation education: Possibilities and prototypes. *New Directions for Evaluation*, 2023(178-179), 97–109. https://doi.org/10.1002/ev.20564

Torraco, R. J. (2016). Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future. *Human Resource Development Review*, 15(4), 404–428. https://doi.org/10.1177/1534484316671606