

## Evaluation in the Era of Artificial Intelligence: The Roles of AI and Human Evaluators

---

Akinori Matsuzaka

Japan University Accreditation Association

### Abstract

Revolutionary changes to human life are emerging as artificial intelligence (AI) brings multiple benefits to society. However, it also poses critical challenges to conventional practices. Higher education is not exempt from its effects, and quality assurance (QA) is now at a transitional phase. This paper explores external quality assurance (EQA) in the era of AI, highlighting the roles of AI and human evaluators. Glancing at a set of evaluator competencies which the American Evaluation Association (AEA) has stipulated, it poses two questions: (i) what roles should (and can) AI assume? And (ii) what responsibilities should human evaluators maintain? By discussing these two key questions, it is expected to gain insights into what human roles can and should not be replaced by AI.

**Keywords:** external quality assurance (EQA), artificial intelligence (AI), evaluator, context-responsiveness, holistic view

### Introduction

Revolutionary changes to human life are emerging as artificial intelligence (AI) brings multiple benefits to society. However, it also poses critical challenges to conventional practices. Higher education is not exempt from its effects, and quality assurance (QA) is now at a transitional phase. This paper explores external quality assurance (EQA) in the era of AI, highlighting the roles of AI and human evaluators. By reviewing some prior research and evaluative practices by the Japan University Accreditation Association (JUAA), the paper explores new approaches and perspectives in evaluation.

## 1. Changed Circumstances for Evaluation

Many practitioners and researchers have discussed over evaluators including their behaviours or thinkings in evaluative operations. For example, Patton (2018) speaks of

<p><b>1.0</b></p> <p>DOMAIN <b>PROFESSIONAL PRACTICE</b></p> <p>focuses on what makes evaluators distinct as practicing professionals</p> <p>Professional practice is grounded in AEA’s foundational documents, including the Program Evaluation Standards, the AEA Guiding Principles, and the AEA Statement on Cultural Competence.</p>	<p>The competent evaluator . . .</p> <ol style="list-style-type: none"> <li>1.1 Acts ethically through evaluation practice that demonstrates integrity and respects people from different cultural backgrounds and indigenous groups.</li> <li>1.2 Applies the foundational documents adopted by the American Evaluation Association that ground evaluation practice.</li> <li>1.3 Selects evaluation approaches and theories appropriately.</li> <li>1.4 Uses systematic evidence to make evaluative judgments.</li> <li>1.5 Reflects on evaluation formally or informally to improve practice.</li> <li>1.6 Identifies personal areas of professional competence and needs for growth.</li> <li>1.7 Pursues ongoing professional development to deepen reflective practice, stay current, and build connections.</li> <li>1.8 Identifies how evaluation practice can promote social justice and the public good.</li> <li>1.9 Advocates for the field of evaluation and its value.</li> </ol>
<p><b>2.0</b></p> <p>DOMAIN <b>METHODOLOGY</b></p> <p>focuses on technical aspects of evidence-based, systematic inquiry for valued purposes</p> <p>Methodology includes quantitative, qualitative, and mixed designs for learning, understanding, decision making, and judging.</p>	<p>The competent evaluator . . .</p> <ol style="list-style-type: none"> <li>2.1 Identifies evaluation purposes and needs.</li> <li>2.2 Determines evaluation questions.</li> <li>2.3 Designs credible and feasible evaluations that address identified purposes and questions.</li> <li>2.4 Determines and justifies appropriate methods to answer evaluation questions, e.g., quantitative, qualitative, and mixed methods.</li> <li>2.5 Identifies assumptions that underlie methodologies and program logic.</li> <li>2.6 Conducts reviews of the literature when appropriate.</li> <li>2.7 Identifies relevant sources of evidence and sampling procedures.</li> <li>2.8 Involves stakeholders in designing, implementing, interpreting, and reporting evaluations as appropriate.</li> <li>2.9 Uses program logic and program theory as appropriate.</li> <li>2.10 Collects data using credible, feasible, and culturally appropriate procedures.</li> <li>2.11 Analyzes data using credible, feasible, and culturally appropriate procedures.</li> <li>2.12 Identifies strengths and limitations of the evaluation design and methods.</li> <li>2.13 Interprets findings/results in context.</li> <li>2.14 Uses evidence and interpretations to draw conclusions, making judgments and recommendations when appropriate.</li> </ol>
<p><b>3.0</b></p> <p>DOMAIN <b>CONTEXT</b></p> <p>focuses on understanding the unique circumstances, multiple perspectives, and changing settings of evaluations and their users/stakeholders</p> <p>Context involves site/location/ environment, participants/stakeholders, organization/structure, culture/diversity, history/traditions, values/beliefs, politics/economics, power/privilege, and other characteristics.</p>	<p>The competent evaluator . . .</p> <ol style="list-style-type: none"> <li>3.1 Responds respectfully to the uniqueness of the evaluation context.</li> <li>3.2 Engages a diverse range of users/stakeholders throughout the evaluation process.</li> <li>3.3 Describes the program, including its basic purpose, components, and its functioning in broader contexts.</li> <li>3.4 Attends to systems issues within the context.</li> <li>3.5 Communicates evaluation processes and results in timely, appropriate, and effective ways.</li> <li>3.6 Facilitates shared understanding of the program and its evaluation with stakeholders.</li> <li>3.7 Clarifies diverse perspectives, stakeholder interests, and cultural assumptions.</li> <li>3.8 Promotes evaluation use and influence in context.</li> </ol>

<p><b>4.0</b></p> <p>DOMAIN <b>PLANNING &amp; MANAGEMENT</b></p> <p><b>focuses on determining and monitoring work plans, timelines, resources, and other components needed to complete and deliver an evaluation study</b></p> <p>Planning and management include networking, developing proposals, contracting, determining work assignments, monitoring progress, and fostering use.</p>	<p><b>The competent evaluator . . .</b></p> <ul style="list-style-type: none"> <li>4.1 Negotiates and manages a feasible evaluation plan, budget, resources, and timeline.</li> <li>4.2 Addresses aspects of culture in planning and managing evaluations.</li> <li>4.3 Manages and safeguards evaluation data.</li> <li>4.4 Plans for evaluation use and influence.</li> <li>4.5 Coordinates and supervises evaluation processes and products.</li> <li>4.6 Documents evaluation processes and products.</li> <li>4.7 Teams with others when appropriate.</li> <li>4.8 Monitors evaluation progress and quality and makes adjustments when appropriate.</li> <li>4.9 Works with stakeholders to build evaluation capacity when appropriate.</li> <li>4.10 Uses technology appropriately to support and manage the evaluation.</li> </ul>
<p><b>5.0</b></p> <p>DOMAIN <b>INTERPERSONAL</b></p> <p><b>focuses on human relations and social interactions that ground evaluator effectiveness for professional practice throughout the evaluation</b></p> <p>Interpersonal skills include cultural competence, communication, facilitation, and conflict resolution.</p>	<p><b>The competent evaluator . . .</b></p> <ul style="list-style-type: none"> <li>5.1 Fosters positive relationships for professional practice and evaluation use.</li> <li>5.2 Listens to understand and engage different perspectives.</li> <li>5.3 Facilitates shared decision making for evaluation.</li> <li>5.4 Builds trust throughout the evaluation.</li> <li>5.5 Attends to the ways power and privilege affect evaluation practice.</li> <li>5.6 Communicates in meaningful ways that enhance the effectiveness of the evaluation.</li> <li>5.7 Facilitates constructive and culturally responsive interaction throughout the evaluation.</li> <li>5.8 Manages conflicts constructively.</li> </ul>

Figure 1. Source: AEA (2018)

‘evaluative thinking’ historically. Reviewing many prior discussions in the past, he describes that critical thinking, reasoning, argumentation, and telling a coherent, evidence-based story have been regarded as essential skills for evaluators (p. 20). The other type of discussions we can see in the ones like the AEA (2018), which defines the expected competencies among evaluators. It articulates a set of competencies across five domains (see Figure 1).

As the Figure 1 indicates, competencies defined by the AEA (2018) range so wide, from, e.g., acting ethically (1.1), determining evaluation questions (2.2) to, e.g., facilitating constructive and culturally responsive interaction (5.7). Here, however, a question arises. Do all these competencies still have meaning in the era of AI?

Mason (2023) argues that some evaluation competencies are quite likely to be impacted by AI. Drawing from the AEA competencies list shown above, she argues that competencies that are asocial and non-creative, and those that are less strategy-based, are likely to be replaced by AI. This includes competencies like determining evaluation questions, using systematic evidence, and selecting evaluation theories and approaches (Mason, 2023, pp. 16–20). The other researchers also make similar discussion. Ferreti (2023) contends that conventional paperwork and ‘by-the-book’ approaches can be substituted by

AI (pp. 76–83). These discussions make it imperative to re-examine evaluators' roles, raising critical questions about what roles AI should (and can) assume; and what responsibilities human evaluators should maintain.

## **2. Discussion: The Roles of AI and Evaluators**

Delineating these roles requires us to take into consideration the purposes of evaluations. Threshold evaluations entail distinct procedures from enhancement-oriented evaluations. An exploration of the answer leading to new ideas is, therefore, required to examine various approaches in different context. The following represents merely an initial idea for starting discussion, derived from the context of the JUAA and from some other studies.

### **2.1 Assigning Roles to AI**

Some possible roles of AI include:

- Data collection and analysis
- Assistance in paperwork
- Dialogical partner
- Assistance in professional development of evaluators

#### **2.1.1 Data Collection and Analysis, and Assistance in Paperwork**

The JUAA is an EQA agency in Japan, evaluating and accrediting higher education institutions (HEIs) and some programmes. All the practices of the JUAA adopt 'fitness-for-purpose' approaches, respecting each HEI's mission, purpose, and the policies derived therefrom (JUAA, 2018). For this reason, evaluators are required to be context-responsive and engage in clear reasoning when they identify problematic areas and conveying their findings. Clear reasoning makes it central in evaluators' work to gather and interpret a wide range of the qualitative and quantitative evidence which is collected from official documents and via dialogue with HEIs. The qualitative evidence here refers to, for example, any written materials which demonstrate, for example, student intellectual and personal growth, or any narrative underlining positive or negative effects. Under the quantitative evidence, for example, numerical data related to financial states is included.

Of the possible four roles shown above, the first two are most basic but greatly beneficial ones. While the JUAA has not employed AI in evaluation yet, utilizing AI for data collection and analysis is a possible way. As is widely acknowledged, AI excels at processing vast amounts of data, and as documented in González-Pérez et al. (2025), this is also the case in higher education domain. Moreover, AI's processing ability can be demonstrated in drafting,

summarising, and simplifying evaluators' reports. This will increase operational efficiency and increased efficiency could enable better allocation of evaluator time.

### **2.1.2 Dialogical Partner**

Given the potential of AI, it can help evaluators to deepen their thinking. Shibu (2024) reports how a US-based business school employed AI to prepare the documents for accreditation. She reports that a school has a Chatbot-type AI evaluate how well scholarly articles and journals align with the UN's sustainable development goals, like ending poverty and ensuring sustainable consumption. This alignment is necessary for business schools to obtain accreditation from the AACSB (the Association to Advance Collegiate Schools of Business). This AI program works as a dialogical partner, acting like a real peer reviewer in responses to individual queries. The program creates a custom report that includes possible answers to the question, like "what is the societal impact of the article?" Considering such answers, the business school reexamine and improve the article, when it hasn't been published yet, and as a result, the school can prepare for successful accreditation.

This AI use case may indicate a potential in quality assurance activities. The AI is capable not only of responding to evaluators' questions but also of generating new ones by reviewing diverse kinds of written materials and datasets. If evaluators do not uncritically accept such answers and questions as they are but use a kind of "insights", then they can leverage to obtain new viewpoints or can identify any overlooked risk factors. Context-responsive evaluation, like that conducted by the JUAA, is not a standardised and formulaic process. It requires situation-specific inquiry and understanding. AI can help evaluators in such work by becoming a dialogical partner.

### **2.1.3 Assistance in Professional Development of Evaluators**

AI may have utility in improving evaluators' professional development. Some researchers have acknowledged the benefits of integrating AI into professional training, because it enables the design of tailored training to suit individual needs and proficiency levels (Fakhar et al., 2024). In terms of the JUAA, evaluators are trained using case studies, yet the cases do not cover all the likely risk areas and are not sufficient for diverse evaluators with different levels of expertise. A personalised tutor system utilising AI may enable the delivery of training that is specifically customised to meet the learning needs of evaluators. If evaluation is to be done in context-responsive manners, this tailored training must have crucial meaning. Evaluators can be trained through case studies prepared according to their level, enhancing their skills in context-responsive evaluation.

### 3. Rethinking the Roles of Evaluators

As has been argued so far, AI possesses significant potential to benefit evaluative works. Is it, however, sufficient to consider only the advantageous aspects? Is not there anything to be considered more? Is there nothing else that should be taken into account?

In this regard, it is important to note two pivotal points before proceeding with the discussion. Firstly, the limitations of AI. Despite its potential value in facilitating increased efficiency and accuracy, AI raises concerns about the possibility of malfunction or incorrect behaviour. For example, it can produce biased or inaccurate outcomes. AI hallucinations are increasingly recognized as a significant issue even among higher education area (e.g. Elsayed, 2024). Secondly, the changed landscape. In Japan, a situation is emerging in which HEIs are leveraging AI to process extensive data and provide services like tailored learning support (Matsuzaka & Kato, 2023, p. 34). Although this is not yet ubiquitous, a recent surge in institutional research (IR) by HEIs suggests that AI's data processing may soon be utilised for QA in many HEIs. Student-related data—such as classroom performance, survey responses, and broader indicators of learning and students' lives—are widely gathered and utilized by Japanese higher education institutions as a foundation for quality assurance (Torii et al., 2023). The implications of this are by no means insignificant, because such recent development will accelerate the transformation of the QA landscape—specifically, it will shift toward a more data-driven approach, enabling a broader and more diverse range of evidence to be utilized in external evaluations.”

Considering such circumstances, discerning the relevance and validity of a wide array of evidence will be more crucial. Moreover, providing a holistic view in stead of fragmented findings will become more important. “Holistic” refers here to an approach that emphasize the whole or entity, rather than focusing on isolated parts, seeking to understand how each element or fragmented fact interacts with each other. Of course, the holism–reductionism debate has persisted throughout intellectual history, and holistic views are not acknowledged as universally effective (Harris & Alderman, 2025, pp. 16–17). However, their role in evaluation is nonetheless meaningful, especially when considering the need of the context-responsiveness.

These insights are supported by prior research, including the work of Stensaker and Leiber (2015). In their pursuit of a framework capable of assessing the organizational impact of EQA, they draw upon a wide range of existing studies. One of the key points identified in their work implies the necessity of a holistic perspective to ensure the impact of evaluation. They take note of that in large-size HEIs of today, many different interests coexist, making agreements on priorities and decisions difficult. This means in the context of QA that defining quality and implementing activities to assure quality may be difficult and

there is high likelihood of organizational fragmentation, which may impede the implementation of effective QA. In other words, under such conditions, EQA without holistic-view may reinforce organisational fragmentation and measures and actions to the problems identified by evaluators may be done with no or less awareness on how quality is taken care of at the whole organisational level (Stensaker and Leiber, 2015, 334–335). Other insights can be obtained from the work by P. Ewell. Using the term “quality game”, he brought it to the light that EQA in U.S. higher education had evolved into a strategic and symbolic “game” played between HEIs and EQA bodies (Ewell, 2007, pp.136–138). In other words, EQA may easily drift from its original purpose and devolve into a reactive and fragmented exercise. All these prior studies underline the necessity of the holistic view.

Here lies a domain where evaluators can — or must — play an essential role. It is the human evaluator who can grasp the essence through various interactions and synthesize a final judgment from a holistic perspective. Of course, AI can provide any information or data that is not fragmented but integrated. Yet, it cannot fully interpret anything unsaid or nuanced — especially within narratives and documents. And, although evaluation needs to be based on the clear evidence, the validity of it is highly context-dependent, and there exists the necessity of “interpretation” of context. The notion of “context” in this setting encompasses not only textual information but also the information derived from human communication which sometime indicates the elements beyond words. This is the reason why that the AEA places a set of competencies related to human relations and social interactions on its list (AEA, 2018). And therefore, many EQA bodies engaged in accreditation or the other type of evaluation incorporate on-site-visit procedure, emphasizing oral and visual evidence along with written one. Consequently, the enduring responsibilities of human evaluators center on their ability to interpret complex contexts, synthesize diverse forms of evidence, and engage meaningfully with stakeholders. These highly-context-dependent or interpersonal competencies are essential for constructing a comprehensive and valid evaluative judgment.

## **Conclusion**

This paper seeks to initiate an ongoing discussion of the roles of AI and human evaluators. Different EQA in different contexts generate diverse suggestions and require different skillsets. However, as argued above, it is evident that thinking about AI necessitates reconsidering the role of human evaluators. More exactly, it can be said that thinking about the roles of AI makes us aware again what is inherent to evaluation and, thus, what kind of responsibilities evaluators must continue to bear. In other words, highly-context-dependent or interpersonal competencies must be highlighted today. Of course, in order to advance the

discussion further, it will be necessary to conduct a more multifaceted examination based on qualitative research and empirical data. This is a task that could not be accomplished within the scope of this paper and is, therefore, left for future work.

## References

- The American Evaluation Association (AEA). 2018. *The 2018 AEA Evaluator Competencies*.
- Elsayed, Hassan. 2024. "The Impact of Hallucinated Information in Large Language Models on Student Learning Outcomes: A Critical Examination of Misinformation Risks in AI-Assisted Education." *Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity* 9 (8): 11-23. <https://northernreviews.com/index.php/NRATCC/article/view/2024-08-07>.
- Ewell, Peter. 2007. "The 'Quality Game': External Review and Institutional Reaction over Three Decades in the United States." In *Quality Assurance In Higher Education: Trends in Regulation, Translation and Transformation*, edited by Don F. Westerheijden, Bjørn Stensaker and Maria João Rosa, 119-153. Dordrecht: Springer Netherlands.
- Fakhar, Hamza, Mohammed Lamrabet, Noureddine Echantoufi, Khalid El Khattabi, and Lotfi Ajana. 2024. "Towards a New Artificial Intelligence-based Framework for Teachers' Online Continuous Professional Development Programs: Systematic Review." *International Journal of Advanced Computer Science and Applications* 15: 480-493. <https://doi.org/10.14569/IJACSA.2024.0150450>.
- Ferretti, Silva. 2023. "Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators." *New Directions for Evaluation* 2023 (178-179): 73-84. <https://doi.org/https://doi.org/10.1002/ev.20557>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20557>.
- González-Pérez, L. I., F. J. García-Peñalvo, and A. J. Argüelles-Cruz. 2025. "Data-Driven Learning Analytics and Artificial Intelligence in Higher Education: A Systematic Review." *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*: 1-1. <https://doi.org/10.1109/RITA.2025.3615512>.
- Harris, Benjamin, and Lyn Alderman. 2025. "Reasoning in Evaluation: Why Does It Matter?" *New Directions for Evaluation* 2025 (185-186): 15-23. <https://doi.org/https://doi.org/10.1002/ev.20631>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20631>.
- The Japan University Accreditation Association (JUAA). 2018. *University Accreditation Handbook*.
- Mason, Sarah. 2023. "Finding a safe zone in the highlands: Exploring evaluator competencies in the world of AI." *New Directions for Evaluation* 2023 (178-179): 11-22. <https://doi.org/https://doi.org/10.1002/ev.20561>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20561>.
- Matsuzaka, Akinori, and Miharu Kato. 2023. "Digital Technologies and Ethical Approaches in Quality Assurance: A Case of the JUAA." In *Anthology of APQN Academic Conference 2023*, edited by the Asia-Pacific Quality Network (APQN), 34-37. the APQN.
- Patton, Michael Quinn. 2018. "A Historical Perspective on the Evolution of Evaluative Thinking." *New Directions for Evaluation* 2018 (158): 11-28. <https://doi.org/https://doi.org/10.1002/ev.20325>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20325>.
- Shibu, Sherin. 2024. "A New AI Chatbot Is Revolutionizing Business School Curriculum and Accreditation — Here's What It Could Change." *Entrepreneur* April 23, 2024.
- Stensaker, Bjørn, and Theodor Leiber. 2015. "Assessing the organisational impact of external quality assurance: hypothesising key dimensions and mechanisms." *Quality in Higher Education* 21 (3): 328-342. <https://doi.org/10.1080/13538322.2015.1111009>. <https://doi.org/10.1080/13538322.2015.1111009>.
- Torii, Tomoko, Nobuhiko Kondo, and Koichi Yamamoto. 2023. "Holistic Approach to Successful Institutional Research and Institutional Effectiveness Based on Local Intelligence in Japanese Universities: Required Conditions for Bridging IR and IE." *Asia-Japan Research Academic Bulletin* 4. [https://doi.org/10.34389/asiajapanbulletin.4.0\\_11](https://doi.org/10.34389/asiajapanbulletin.4.0_11).