Evaluating AI-Generated Solution Ideas: A Comparative Study of AI and Human Assessments for Sustainable Process Design

Mas’udah1, Pavel Livotov1, Saptadi Nugroho1

1 Offenburg University of Applied Sciences, Badstr. 24, 77652 Offenburg, Germany

masudah@hs-offenburg.de

**Extended Abstract.** The integration of generative artificial intelligence (AI) into sustainable process design in the realm of process engineering has gained substantial traction, with AI increasingly leveraged to generate innovative solution ideas. Despite this advancement, the efficacy of AI-generated ideas needs to be rigorously evaluated to determine their practical usefulness, feasibility, novelty, and sustainability. This study aims to address this gap by examining the reliability of AI evaluations and comparing them with assessments conducted by human experts, focusing on early-stage concept design in process engineering.

In this study, GPT-4o, a cutting-edge generative AI model, was utilized to generate various sets of design ideas for addressing sustainability challenges in process engineering. GPT-4o was employed not only for its advanced capabilities in idea generation but also for evaluating the generated ideas against criteria such as practical applicability, technical and economic feasibility, uniqueness compared to existing solutions, and potential impact on environmental and social sustainability. The evaluation metrics used were aligned with human expert criteria to ensure comprehensive assessment. The evaluation process involved using GPT-4o to assess the ideas through queries designed to mirror human expert evaluation criteria. This approach facilitates a direct comparison between AI-driven valuations and those conducted by human experts, offering a comprehensive analysis of the AI's effectiveness.

* *Usefulness* - was assessed by querying GPT-4o to evaluate whether each design idea effectively addresses the identified problem. The AI model analyzed each idea’s capacity to resolve specific issues.
* *Feasibility* - was evaluated through queries that required GPT-4o to examine technical and economic aspects of the proposed ideas. This included assessments of resource requirements, cost implications, and technical challenges.
* *Novelty* - was assessed by employing GPT-4o to compare the generated ideas against existing patents, literature, and recent advancements in the field. The AI model was instructed to determine the originality of each idea by benchmarking it against established knowledge. Additionally, metrics such as TF-IDF and cosine similarity were utilized to quantify the degree of novelty by comparing new ideas to a repository of existing solutions.
* *Sustainability* - was analyzed by querying GPT-4o to evaluate the environmental impact, resource efficiency, and social benefits associated with each idea. The AI's assessment was based on predefined sustainability criteria to determine alignment with sustainable development objectives.

Simultaneously, assessments were performed by a panel of domain experts using predefined criteria, providing a benchmark for comparison. The comparative analysis of AI-driven and human evaluations identified both areas of concordance and divergence, offering critical insights into the strengths and limitations of GPT-4o in early-stage process design.

The findings highlight GPT-4o’s potential to advance sustainable innovation in process engineering while underscoring the necessity for rigorous validation of AI-generated assessments. This research enhances AI evaluation methodologies and offers a framework for the effective integration of AI into early-stage sustainable process design practices. The study establishes a foundational understanding of the role of generative AI in process engineering and sets the stage for future research into AI-supported innovation.

**Keywords:** Generative AI; Sustainability; Process Design; AI Evaluation.