

Toward Decision Support that Respects Personal Preferences: Human Weighting of LLM-Generated Evaluation Axes

1st Yuna Kanai
Meiji University
Tokyo, Japan
cs262009@meiji.ac.jp

2nd Homei Miyashita
Meiji University
Tokyo, Japan
homei@homei.com

Abstract—As a step toward AI-assisted decision support that better respects personal preferences, we examine delegation boundaries by restricting user input to weighting evaluation axes while delegating axis generation and option scoring to an LLM. We present a Simple Additive Weighting (SAW) method in which the LLM produces evaluation axes and numeric scores, and users express their preferences by weighting those axes.

In a small exploratory paper-based study on movie selection and company choice, participants described weighting as easy to understand and low-effort. For participants with complete ranking data, similarity analyses suggest that the method sometimes diverged from direct rankings, especially in the movie scenario; interviews and an exploratory follow-up suggest that some divergences prompted articulation of criteria not stated beforehand, although this observation remains preliminary.

We also frame weighting as a reflective process that makes value prioritization explicit, which may inform interactive systems for interpretive or creative tasks. Pairwise comparison was perceived as more reliable but also more effortful, suggesting an efficiency-reliability trade-off. Repeated LLM runs under fixed settings produced different axes and scores, suggesting that delegation choices can reshape the decision space even when users retain control over weighting.

Rather than claiming effectiveness, we contribute an exploratory design perspective on when weighting may and may not be sufficient.

Index Terms—Decision-making support, Personalization, AI-assisted.

I. INTRODUCTION

In everyday decision-making and creative activities, people evaluate subjects based on multiple perspectives and make selections or expressions by internally adjusting the importance of these perspectives. For example, in movie appreciation, multiple evaluation axes—such as narrative originality, visual aesthetics, and emotional resonance—are involved, and the weighting assigned to these axes ultimately shapes one’s overall evaluation and preferences.

Much prior work on AI-based decision support emphasizes efficiency [1], [2]. Here, we focus on a narrower design question: which part of the decision-making process should remain under human control when other parts are delegated to AI. Particularly in creative contexts, interpretations and evaluations can vary significantly depending on which aspects

creators or audiences prioritize. The framework proposed in this study aims to make this process of “weighting” explicit, thereby offering a means for individuals to reflect on and reconsider their own values and preferences.

This research aims to explore up to which stage humans should remain involved in order to retain control over decision-making, without transferring decision authority to AI. We take the position that human decision-making should be respected, and that AI should be positioned as a supporting component rather than a replacement. Based on this perspective, we examine whether allowing humans to determine the importance of evaluation axes—namely, the weighting of evaluation axes—can serve as a minimal yet meaningful locus of human control in decision-making. To investigate this hypothesis, we conduct experiments using a decision-support method based on the simple additive weighting method. In the proposed approach, an AI system (Gemini 2.5 Pro) is responsible for generating evaluation axes and assessing each option with respect to those axes, which primarily focus on subjective and experiential aspects that are difficult to quantify. Humans, in contrast, determine only the weights of the evaluation axes according to their own values and preferences. By delegating other components of the process to AI, the method aims to reduce human cognitive and time burdens while preserving a degree of human control over the decision process. This framing allows us to ask whether weighting alone can serve as a minimal yet meaningful locus of human control in AI-assisted decision support.

II. RELATED WORK

Multi-criteria decision-making methods have long been studied as a way to support complex choices involving multiple, often conflicting criteria. Among them, the Simple Additive Weighting (SAW) method is one of the most fundamental approaches, combining scores for each evaluation axis with human-defined weights to generate a ranking [3]. Prior studies have shown that even approximate or simplified weighting strategies often produce practically acceptable results, suggesting that exact weight specification may not always be necessary [4]. Another widely used approach is the Analytic

Hierarchy Process (AHP), which derives weights through pairwise comparisons and provides consistency checks [5]. While AHP can generate highly reliable rankings, it requires substantial effort as the number of options increases, making it unsuitable for lightweight or everyday decision-making scenarios.

More recent studies have explored the use of AI to automate parts of the decision-making pipeline, such as extracting evaluation axes or conducting pairwise comparisons. These approaches demonstrate that AI can approximate or even replicate human judgments [1], [2]. However, many of them implicitly transfer decision authority to AI by automating not only evaluation but also the determination of importance across criteria. In contrast, this work focuses on explicitly preserving human control by fixing the human role at a specific and minimal point in the decision process: determining the weights of evaluation axes.

III. METHOD

We use a Simple Additive Weighting (SAW) framework in which AI performs evaluation while humans assign weights.

The method consists of four steps: (1) the AI generates evaluation axes, including subjective and experiential aspects; (2) users assign weights to each AI-generated axis based on their preferences, without editing, deleting, or merging the axes; (3) the AI assigns numerical scores to each option on each axis; and (4) final rankings are computed by weighted summation.

This division of labor is intended to make weighting the sole human intervention point while keeping the procedure lightweight. In this study, Gemini 2.5 Pro generated evaluation axes and scores using fixed prompts and parameters, and a single set of ten axes was used across all participants to control variability.

We conducted a small exploratory user study with five participants in two scenarios: movie selection and company choice. We compared four conditions: (a) direct ranking, (b) direct ranking informed by AI-generated axes, (c) the proposed weighted-SAW method, and (d) pairwise comparison (AHP-based). All conditions were completed in the same fixed order (a→b→c→d), which may have introduced learning, fatigue, and anchoring effects.

Participants completed all conditions on paper, and we collected task completion time, ranking similarity, and subjective feedback. To compare rankings, we used Kendall's τ_b (structural agreement) and Spearman's footrule (positional similarity), providing complementary perspectives on ranking differences [6]–[8].

IV. FINDINGS AND DISCUSSION

Participants generally described weighting as easy to understand and low-effort, suggesting that it can serve as a lightweight form of user input. However, the proposed method did not consistently match participants' direct rankings, particularly in the movie-selection scenario. Some discrepancies were not perceived as simple errors; participants reported that

some results prompted them to articulate aspects they had not explicitly considered beforehand.

Pairwise comparison produced rankings closer to participants' direct rankings and was perceived as more reliable, although it required greater effort. Given the small sample size, these findings should be interpreted as exploratory rather than conclusive.

This study explored whether restricting human involvement to weighting evaluation axes can function as a lightweight form of user control in AI-assisted decision-making.

The results suggest that weighting can support preference reflection, but outcomes remain strongly shaped by AI-generated axes and scores, which partially define the decision space. If these axes do not align with user perspectives, the effectiveness of weighting may be limited.

Additionally, we observed variability in AI-generated axes and scores across repeated runs, indicating that the evaluation process is not fully stable. Such variability may reshape the decision space itself, raising questions about consistency and control.

Rather than claiming that the present prototype already respects personal preferences, this work treats such respect as a design goal and uses this study to probe when minimal human intervention may support reflection and user control.

Future work should examine a broader range of users and scenarios, particularly in creative contexts, and should test richer interventions such as editing or restructuring evaluation axes.

REFERENCES

- [1] X. Lu, J. Li, K. Takeuchi, and H. Kashima, "AHP-powered LLM reasoning for multi-criteria evaluation of open-ended responses," in *Findings of the Association for Computational Linguistics: EMNLP 2024*. Miami, Florida, USA: Association for Computational Linguistics, 2024, pp. 1847–1856. [Online]. Available: <https://aclanthology.org/2024.findings-emnlp.101/>
- [2] I. Svoboda and D. Lande, "Enhancing multi-criteria decision analysis with AI: Integrating analytic hierarchy process and GPT-4 for automated decision support," *arXiv preprint arXiv:2402.07404*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.07404>
- [3] H. Taherdoost, "Analysis of simple additive weighting method (SAW) as a multi-attribute decision-making technique: A step-by-step guide," *Journal of Management Science & Engineering Research*, vol. 6, no. 1, pp. 21–24, 2023. [Online]. Available: <https://journals.bilpubgroup.com/index.php/jmser/article/view/5400>
- [4] J. M. Keisler, "The value of assessing weights in multi-criteria portfolio decision analysis," *Journal of Multi-Criteria Decision Analysis*, vol. 15, no. 5-6, pp. 111–123, 2008.
- [5] T. L. Saaty, "How to make a decision: The analytic hierarchy process," *European Journal of Operational Research*, vol. 48, no. 1, pp. 9–26, 1990. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0377221790900571>
- [6] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938. [Online]. Available: <https://academic.oup.com/biomet/article-abstract/30/1-2/81/176907>
- [7] P. Diaconis and R. L. Graham, "Spearman's footrule as a measure of disarray," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 2, pp. 262–268, 1977. [Online]. Available: <https://academic.oup.com/jrsssb/article/39/2/262/7027404>
- [8] T. Kamishima, "Order distance and probabilistic models," *Proceedings of the Second-Class Research Meeting of the Japanese Society for Artificial Intelligence*, vol. 2009, no. DMSM-A902, p. 07, 2009, (in Japanese). [Online]. Available: https://www.jstage.jst.go.jp/article/jsaisigtwo/2009/DMSM-A902/2009_07/_article-char/ja/