Investigation of Expressiveness Performance of Large Language Models for HRM Simulations

Atsuhiro Fujii
Recruit Management Solutions Co., Ltd.
Tokyo, Japan
atsuhiro_fujii@recruit-ms.co.jp

Kazuma Negita
Recruit Management Solutions Co., Ltd.
Tokyo, Japan
kazuma_negita@recruit-ms.co.jp

Keisuke Masuda
Recruit Management Solutions Co., Ltd.
Tokyo, Japan
keisuke_masuda@recruit-ms.co.jp

Wataru Uno
Recruit Management Solutions Co., Ltd.
Tokyo, Japan
wataru_uno@recruit-ms.co.jp

Daisuke Nakama
Recruit Management Solutions Co., Ltd.
Tokyo, Japan
daisuke_nakama@recruit-ms.co.jp

Abstract

Achieving optimal staffing is desirable in settings where efficient results are pursued such as corporate activities. The ideal scenario is to base recruitment plans and organizational design on predictions of how different types of individuals will impact overall team performance. We propose using Large Language Models (LLMs) for Human Resource Management (HRM) simulations based on a personality framework called "Organizational Adaptation Types". This paper evaluates the capability of LLMs to accurately exhibit the behavioral traits of each type, thereby validating a foundation for simulating complex interactions among diverse individuals using LLM agents. We conducted an evaluation experiment generating resumes using LLMs set with organizational adaptation types, then classified which type had generated each resume. As a result, the classification accuracy was 0.813 and 0.419 in the LLM-as-a-judge based environment and the subject based environment, respectively. Results indicate that while LLM-based evaluation showed consistent accuracy across all organizational adaptation types, human evaluation exhibited homophily bias, suggesting that LLMs can effectively reduce unconscious bias in HRM simulations.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Large Language Models, LLM-based Agents, Agent-based Simulation, Personality Modeling, Human Resource Management

1 Introduction

Organizations constantly face the challenge of building and maintaining teams that perform effectively under changing conditions, yet the tools available to support these decisions remain limited. While services[1] exist that propose optimal staffing based on employee data, including personality traits, many focus on visualizing and simulating current team composition and do not fully account for future dynamics. In practice, organizations must adapt to continuous changes such as member turnover and shifts in company size, yet few tools exist to simulate how different combinations of individuals might affect future team effectiveness. This leaves organizations without a reliable means to answer critical questions such as how overall team performance will change if a key member leaves, or what personality profile the next hire should have to achieve an optimal team composition. Addressing these questions requires the ability to simulate future scenarios involving individuals with diverse behavioral and personality tendencies.

To address this need, we explore the use of Large Language Model (LLM) based agents to simulate human interactions in the domain of Human Resource Management (HRM). LLMs are machine learning models that excel at natural language input and output, capable of generating consistent and context-aware responses based on predefined roles. These characteristics have drawn attention to LLM agents that have the potential to simulate human behaviors. While research has incorporated personality traits in LLM agents using general frameworks like the Big Five Personality model [13], such general models may not sufficiently capture behavioral tendencies specific to organizational activities, and the need for an alternative framework has been pointed out [12, 18]. In practical HRM scenarios, type classifications are frequently preferred over scale-based approaches. While trait scales enable detailed analysis of individual tendencies, aggregating multi-dimensional scale data across employees introduces too many axes for practical decision-making such as optimal staffing. Type classifications address this by grouping individuals into interpretable profiles, making them more actionable in organizational contexts.

In this research, we propose using LLMs for HRM simulations based on a personality framework called "organizational adaptation types", which has been validated within Japanese organizational culture, and is widely adopted in practical HRM contexts in Japan. By assigning these types to LLM agents via system prompts, we aim to simulate how individuals with diverse behavioral tendencies

3rd HEAL Workshop at CHI Conference on Human Factors in Computing Systems, Barcelona, Spain

[1] https://interact-global.co/

interact within organizations. To evaluate whether LLMs can accurately express these types, we had the type-assigned LLMs perform a resume creation task and, with different LLMs, an evaluation task to assess whether the generated resumes exhibited behavioral characteristics specific to each type. We adopted these tasks because resumes are a concise yet behaviorally rich form of self-expression, allowing us to objectively assess whether assigned personality types are reflected in naturally occurring organizational language. This paper evaluates the capability of LLMs to accurately exhibit the behavioral traits of each type, thereby validating a foundation for simulating complex interactions among diverse individuals using LLM agents for more refined recruitment planning and organizational design.

Section 2 introduces related works, section 3 describes the details of the evaluation method, section 4 presents our evaluation experiment. Finally, section 5 concludes the paper.

2 Related Work

In this section, we introduce research on replicating personality traits using LLMs, personality traits in the field of HRM, and utilization of LLMs.

2.1 Replicating personality traits using LLMs

La Cava et al. [15] evaluated whether agents exhibit consistent personality tendencies across 12 models, including open-source LLMs. Using psychological measures such as the MBTI and Big Five personality traits, they quantitatively measured personality characteristics from each model's output. Salecha et al. [19] conducted an experimental study using the Big Five personality traits to determine whether LLMs exhibit a social desirability bias (the tendency to respond in ways that are perceived favorably by others) similar to humans. Under conditions where the model was classified as being "evaluated", clear biases emerged, including increased scores for extraversion, conscientiousness, agreeableness, and openness. Njifenjou et al. [17] endowed LLM agents with personality traits using a five-dimensional continuous vector based on the Big Five personality traits. They evaluated the impact of these personality traits on dialogue quality and user experience in a dialogue task simulating French customer support. The evaluation included experiments where both humans and models identified the personality. The results confirmed that the assigned personality traits were reflected in the conversational style. Frisch et al. [7] examined how much LLM agents maintain their own personality traits during collaborative tasks and how much they tend to synchronize their personality and linguistic characteristics with other agents through interaction. The results showed that agents assigned creative personality profiles exhibited a stronger tendency to synchronize with their conversation partner's linguistic style compared to analytical agents, indicating differences in consistency and flexibility of personality expression across models. Takata et al.[20] investigated how individuality and behavioral diversity spontaneously emerge in initial-state LLM agents without pre-assigned personality traits through continuous interaction with other agents. Through linguistic interactions between agents and movement within the environment, differences in memory accumulation and emotional expression arose, resulting in the formation of distinct personality tendencies and social roles for each individual.

Several studies exist that attempt to replicate human personality traits using LLMs, but most focus on general, comprehensive personality tendencies based on frameworks like the Big Five Personality model. However, the personality traits do not necessarily align with behavioral tendencies in organizational activities, such as job-related behaviors or environmental adaptation. This study specifically focuses on the HR domain and attempts to expand upon such research on personality trait replication using LLM.

2.2 Personality traits in the field of HRM

In the field of HRM, research has traditionally focused on the relationship between individual personality traits and job performance using variable-centered approaches like the Big Five Personality model as a primary framework. While these models are known to predict individual and team performance through specific factors (e.g., conscientiousness and agreeableness) [2, 3], analyzing relationships based on single factors often fails to capture the complex human profile arising from interconnections among traits [6, 16].

To resolve this problem, person-centered approaches have been proposed, which classify subgroups of individuals with similar trait configurations as "types" [10]. This offers greater clarity by framing individuals not as a list of factor scores, but as distinct profile types (e.g., resilient or over-controlling). In the practical HRM context where holistic decision-making is required, this typological approach is often more useful for recruitment and internal talent selection than isolated trait scores [9, 27]. However, as general personality types may not fully capture the specific behavioral tendencies required in organizational settings [27], it is necessary to adopt a typology that specifically reflects individual tendencies suited to organizational activities for HRM simulation with LLM agents.

2.3 Utilization of LLMs

There is research aimed at improving the performance and evaluation of LLMs [8, 11, 22–25]. Furthermore, leveraging the characteristic of LLM agents to mimic human behavior, attempts are being made to apply them to simulations of social activities. Zheng et al. [28] simulate the dynamic changes in relationships and information on social networks through interactions among multiple agents using GPT-4o, analyzing the process of information propagation and the formation mechanism of echo chambers. Agents update the network structure based on judgments made using natural language, and high reproducibility was confirmed when compared with real-world SNS data. Zhang et al. [26] pointed out that existing simulations of influence diffusion in social networks rely on probabilistic methods and fail to adequately account for users' linguistic responses and contextual judgments. In response, they propose a novel approach that incorporates LLMs into simulations to dynamically reproduce processes such as information reception, retransmission, and opinion formation based on natural language. Tang et al. [21] proposed GenSim, a general-purpose social simulation platform based on LLMs. This overcomes the limitation that previous research, constrained by specific task settings and a small number of agents, lacks the ability to adapt to malfunctions and inconsistencies that occur during long-term simulations. By integrating large-scale agent management and error correction mechanisms, GenSim achieves scalable and reliable social behavior simulation. Jiang et al. [14] proposed UrbanLLM, an integrated framework designed to flexibly address complex and diverse urban challenges such as activity planning and service management using LLMs. By decomposing urban-related queries expressed in natural language into multiple subtasks and processing them holistically in conjunction with appropriate spatio-temporal models, it enables practical and highly accurate decision support for urban planning.

The utilization of LLMs in the field of HRM is being explored. Daryanto et al. [5] evaluated the feasibility of reflective learning support by verifying how accurately an LLM can analyze speech records during interview practice. Specifically, it combines an automatic annotation feature for user responses with an interactive feedback mechanism, designed to enable learners to reflect on their answers and identify areas for improvement. The LLM captures semantic features from the speech recordings to pinpoint which parts are inadequate and suggest desirable alternative expressions. Unlike traditional one-way learning, it promotes self-correction through two-way interaction.

We aim to simulate practical scenarios in the HRM domain using LLM agents by replicating personality traits related to organizational adaptation.

3 Method

This section describes the methodology used to evaluate the simulation capabilities of LLMs for the proposed organizational adaptation types. An overview of the evaluation method is shown in figure 1. The evaluation process consists of three parts: assigning organizational adaptation types to LLMs, a resume creation task for LLMs to generate text based on these types, and a type classification task executed by different LLMs and human subjects to verify the identifiability of the source types within the generated resumes.

3.1 Organizational adaptation types

To address the need for an organization-specific framework discussed in the previous section, this study adopts a framework widely utilized in practical HRM contexts in Japan[2]. This framework classifies personality types based on an individual's adaptability to different organizational cultures, referred to as "Organizational Adaptation Types" (W, X, Y, Z). These types are defined by two behavioral axes essential for collaborative work. Axis 1 indicates tendencies regarding “problem setting and decision making”: whether one tends to take on challenges or prefers a step-by-step approach, whereas Axis 2 indicates tendencies regarding “communication style”: whether one tends to communicate emotionally or logically (see also figure 1). The definitions of these types are shown in table 1. We assigned these types to LLMs via system prompts to simulate specific organizational personas. Since the organizational adaptation type framework originates from Japanese organizational contexts and the human subjects in our evaluation were Japanese speakers, all prompts were implemented and input in Japanese to ensure consistency between the framework's cultural context and the experimental setting.

3.2 Resume creation task

To evaluate whether LLMs can express these organizational adaptation types, we first have LLMs execute a resume creation task. LLMs were instructed to write a summary of their work experience at their current organization within 500 characters. To test expressive capability in behavioral tendencies within the resume text, we instructed LLMs to avoid descriptions that directly express personality, age, or gender. Additionally, we provided LLMs with no attributes regarding industry or other factors, except that they are not in a managerial role.

3.3 Type classification task

This task evaluates whether the resumes generated in the resume creation task clearly reflect the intended organizational adaptation types. We employed two evaluation methods: LLM-as-a-judge and Human Subject Evaluation.

In the LLM-as-a-judge, LLMs distinct from those used in the resume creation task are assigned the role of HR manager via the system prompt and tasked with classifying a given resume into one of the four types based on the provided definition of theirs (table 1). To account for potential biases arising from the compatibility between the personality types of the resume creator and evaluators (e.g., a type W evaluator preferring resumes created by type W agents), each Judge LLM was also assigned an organizational adaptation type. This allows us to analyze whether classification accuracy is robust across different evaluator personas.

In the human subject evaluation, we recruited human subjects to perform the same classification task. Subjects were presented with resumes and the definitions of the four types (shown in table 1), then asked to select which type most likely wrote each resume. Each participant completed four resumes, presented in a randomized order to prevent order effects. Organizational adaptation type of each subject was also measured by having them self-select the type that fit them best.

4 Evaluation Experiment

Based on the method defined in the previous section, we conducted a series of experiments to quantitatively evaluate whether LLMs can accurately express the assigned organizational adaptation types. This section first describes the experimental environment, the process of constructing the evaluation dataset, and the results of the experiments.

4.1 Environment for Evaluation Experiments

The model was used Claude Sonnet 4[3] for all tasks. The parameter $Temperature$ which defines the randomness of the output was set to 1.0. All generation and classification processes were executed independently to prevent the model from being influenced by past outputs.

[2]https://www.recruit-ms.co.jp/issue/column/0000001377/

[3]https://www.anthropic.com/news/claude-4

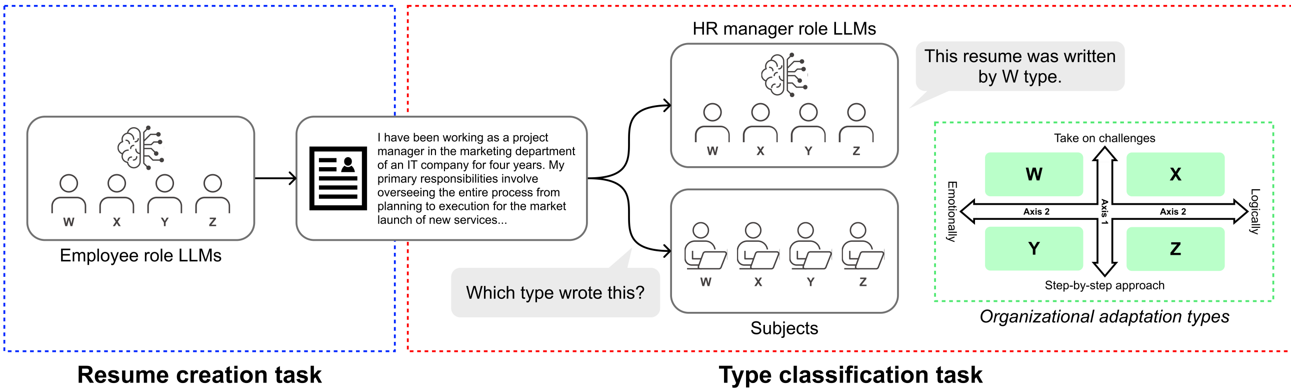

Figure 1: Overview of the evaluation method.

Table 1: Definition of organizational adaptation types.

Type	Definition
W	The type who tackles challenging work by joining forces with coworkers.
X	The type who values rationality and independently pursues ambitious goals.
Y	The type who works steadily and reliably while collaborating with coworkers.
Z	The type who works alone, concentrating intently tackling tasks with diligence.

Table 2: Job classification in resumes.

Occupation	Type			
	W	X	Y	Z
Sales and Related Support	17	9	24	8
IT Technical	6	16	5	10
Quality Management	1	4	1	9
Business Planning	6	1	0	0
Accounting	0	0	0	3

4.2 Resume Generation and Screening

To prepare the resumes for the classification task, we implemented a two-step process: generation and screening.

4.2.1 Initial resume generation and observed bias. First, we had the LLM execute the resume creation task 30 times for each of the four organizational adaptation types, resulting in a pool of 120 resumes. Upon analyzing this generated pool, a bias in the distribution of job types was observed, as shown in table 2.

4.2.2 Screening for bias control. If resumes were selected purely at random from this biased pool, the job type could become a confounding variable (e.g., evaluators might associate "Sales" with Type W regardless of the resume's actual behavioral content). To mitigate this effect, we extracted three resumes for each type (totaling 12 resumes) using a constraint-based random selection method. The selection was repeated until a set was obtained where no two resumes shared the same job type within a single type. The selected resumes are defined as $R_{W_i}, R_{X_i}, R_{Y_i}, R_{Z_i}$ $(i = 1, 2, 3)$.

4.3 Results and Discussion

This section discusses the results of classification accuracy of both LLM-as-a judge and human subjects. The confusion matix is shown in figure 2. It is totaled classification results for the three resumes. R_W is defined as the sum of the classification result values obtained from R_{W_1}, R_{W_2}, and R_{W_3}. Likewise, R_X, R_Y, and R_Z are defined as the sums of the classification result values obtained from their corresponding resumes.

4.3.1 Result of LLM-as-a-judge based evaluation. Each of three resumes per type was evaluated by 400 LLM agents (100 per evaluator type), resulting in a total of 1200 evaluations per resume type. The overall classification accuracy was 0.813, indicating LLMs possess a high capability to express and identify organizational adaptation types. However, the accuracy of judgment varied for each type, such that many errors were observed where R_Z was judged as R_Y on the left panel of figure 2. Despite these misclassifications, the high overall accuracy suggests that the behavioral tendencies defined in the prompts were successfully encoded into the generated resumes.

4.3.2 Result of Subject based evaluation. We recruited 1,950 subjects (1,346 men and 604 women, 911 managers and 1,039 non-managers, mean age = 46.0), who were university graduates working as full-time employees at companies with over 300 employees. We had them conduct a type classification task, in which they were randomly presented with one of three predefined subsets S_i of four resumes, as specified in equation 1 using permutation P_i.

$$S_i = P_i\{R_{W_i}, R_{X_i}, R_{Y_i}, R_{Z_i}\} \ (i = 1, 2, 3) \quad (1)$$

The classification accuracy was 0.419. While this is significantly lower than the LLM-based result, it significantly exceeds the chance

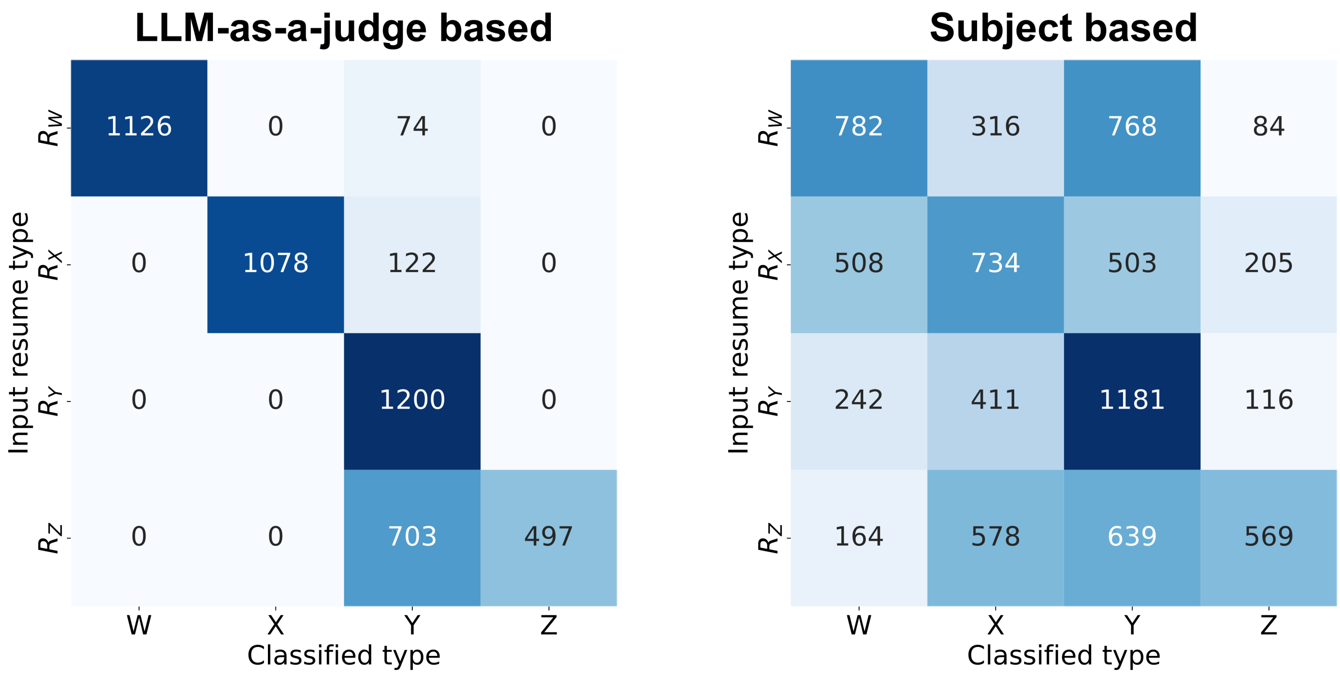

Figure 2: Confusion matrix of the organizational adaptation types of LLM that created the resume and the type classification results.

level of 0.25 for a four-class classification. The confusion matrix shows a more dispersed pattern compared to the LLM results on the right panel of figure 2. While diagonal components (correct answers) generally show higher values, human subjects struggled to distinguish between types clearly, suggesting that the behavioral nuances generated by LLMs may be more subtle or too stereotypical than what humans typically rely on for personality judgment.

Table 3: The classification accuracy for each type of resume.

Resume	LLM-as-a-judge				Subject			
	W	X	Y	Z	W	X	Y	Z
R_W	1.00	1.00	0.76	0.99	0.59	0.36	0.37	0.40
R_X	0.94	1.00	0.67	0.98	0.32	0.49	0.36	0.33
R_Y	1.00	1.00	1.00	1.00	0.48	0.49	0.69	0.63
R_Z	0.34	0.65	0.01	0.67	0.22	0.21	0.29	0.40

4.3.3 Analysis of Evaluator Compatility. To clarify the factors contributing to the accuracy gap between LLMs and humans, we examined whether the combination of resume type and evaluator type influenced the classification results. Table 3 shows the classification accuracy for each resume type. In LLM-as-a-judge based evaluation, R_W, R_X, and R_Y could be classified with little difference in accuracy regardless of evaluator type. In contrast, in the subject-based evaluation, accuracy was higher when the resume type matched the subject's own type. Specifically, for R_W, R_X, and R_Y, the highest classification accuracy was achieved when the subject's type corresponded to the resume type. This is indicated by the diagonal elements in table 3. These results suggest that humans may have a homophily bias [1, 4], or a deeper understanding of individuals with behavioral traits similar to their own. This tendency was not observed in the LLM-as-a-judge evaluation, indicating that LLMs function as objective classifiers unaffected by subjective biases. While such homophily bias holds significance in specific social contexts such as intuitively selecting compatible teammates, the results suggest that LLMs can be utilized in situations where such bias is unnecessary or should be avoided. Consequently, LLMs demonstrate superior utility for tasks that demand objective classification and the exclusion of human-like subjective preferences.

R_Z showed low accuracy in all environments. This result may reflect the occupational biases inherent in the real world, where specific personaly types are often concentrated in particlar job type. In this evaluation experiment, we intentionally eliminated this job type bias by esuring no occupational duplication within each type of resume. Consequently, the resumes used for R_Z may have deviated from the typical occupational distribution of occupations associated with type Z, thereby removing a ciritcal contextual cure required for accurate identification.

5 Conclusion

This research defined four personality types called organizational adaptation types, that can be used in HRM simulations. Then, we investigated whether LLMs have the ability to express these types.

We conducted an evaluation experiment generating resumes using LLMs set with organizational adaptation types, then classified which type had generated each resume. As a result, the classification accuracy was 0.813 and 0.419 in the LLM-as-a-judge based environment and the subject based environment, respectively. There was a difference in accuracy, but the accuracy was higher than the chance level of 0.25 for the four type classification in both environments. This result indicates that LLMs have the ability to express organizational adaptation types. The LLM-as-a-judge based evaluation achieved higher overall accuracy than the subject-based evaluation. Importantly, our analysis revealed qualitative differences in evaluation patterns between human and LLM. LLM-based evaluation showed consistent accuracy across organizational adaptation types regardless of evaluator type, whereas human evaluation exhibited higher accuracy when the resume type matched the evaluator's own type. This pattern suggests that human evaluators may be influenced by homophily bias. In contrast, LLMs appear to function as objective classifiers unaffected by such subjective biases. While homophily bias may be beneficial in certain social contexts, such as intuitively selecting compatible teammates, our findings suggest that LLMs can be valuable in HRM scenarios where objective evaluation is required and subjective preferences should be minimized. These results indicate that LLMs can effectively reduce homophily bias in HRM simulations, though their application should be carefully considered based on the specific requirements of each situation.

References

[1] Daniel R. Ames. 2004. Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. *Journal of Personality and Social Psychology* 87, 3 (2004), 340–353.

[2] Murray R Barrick and Michael K Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology* 44, 1 (1991), 1–26.

[3] Suzanne T. Bell. 2007. Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology* 92, 3 (2007), 595–615.

[4] Lee J. Cronbach. 1955. Processes affecting scores on "understanding of others" and "assumed similarity.". *Psychological Bulletin* 52, 3 (1955), 177–193.

[5] Taufiq Daryanto, Xiaohan Ding, Lance T. Wilhelm, Sophia Stil, Kirk McInnis Knutsen, and Eugenia H. Rho. 2025. Conversate: Supporting Reflective Learning in Interview Practice Through Interactive Simulation and Dialogic Feedback. *Proc. ACM Hum.-Comput. Interact.* 9, 1, Article GROUP9 (Jan. 2025), 32 pages.

[6] M. Brent Donnellan and Richard W. Robins. 2010. Resilient, Overcontrolled, and Undercontrolled Personality Types: Issues and Controversies. *Social and Personality Psychology Compass* 4, 11 (2010), 1070–1083.

[7] Ivar Frisch and Mario Giulianelli. 2024. LLM Agents in Interaction: Measuring Personality Consistency and Linguistic Alignment in Interacting Populations of Large Language Models. In *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*. Association for Computational Linguistics, 102–111.

[8] Yoshihiko Hayashi. 2025. Evaluating LLMs' Capability to Identify Lexical Semantic Equivalence: Probing with the Word-in-Context Task. In *Proceedings of the 31st International Conference on Computational Linguistics*. Association for Computational Linguistics, 6985–6998.

[9] Robert Hogan and Joyce Hogan. 2007. *Hogan Personality Inventory Manual* (3 ed.). Hogan Assessment Systems.

[10] Matt C. Howard and Michael E. Hoffman. 2018. Variable-Centered, Person-Centered, and Person-Specific Approaches: Where Theory Meets the Method. *Organizational Research Methods* 21, 4 (2018), 846–876.

[11] Jiaxin Huang, Shixiang Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2023. Large Language Models Can Self-Improve. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1051–1068.

[12] Jason L Huang, Ann Marie Ryan, Keith L Zabel, and Ashley Palmer. 2014. Personality and adaptive performance at work: a meta-analytic investigation. *J. Appl. Psychol.* 99, 1 (Jan. 2014), 162–179.

[13] Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2024. PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits. In *Findings of the Association for Computational Linguistics: NAACL 2024*. Association for Computational Linguistics, 3605–3627.

[14] Yue Jiang, Qin Chao, Yile Chen, Xiucheng Li, Shuai Liu, and Gao Cong. 2024. UrbanLLM: Autonomous Urban Activity Planning and Management with Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. Association for Computational Linguistics, 1810–1825.

[15] Lucio La Cava and Andrea Tagarelli. 2025. Open Models, Closed Minds? On Agents Capabilities in Mimicking Human Personalities through Open Large Language Models. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 2 (Apr. 2025), 1355–1363.

[16] Brett Laursen and Erika Hoff. 2006. Person-Centered and Variable-Centered Approaches to Longitudinal Data. *Merrill-Palmer Quarterly* 52, 3 (2006), 377–389.

[17] Ahmed Njifenjou, Virgile Sucal, Bassam Jabaian, and Fabrice Lefèvre. 2025. Enabling Trait-based Personality Simulation in Conversational LLM Agents: Case Study of Customer Assistance in French. In *Proceedings of the 15th International Workshop on Spoken Dialogue Systems Technology*. Association for Computational Linguistics, 299–308.

[18] E D Pulakos, S Arad, M A Donovan, and K E Plamondon. 2000. Adaptability in the workplace: development of a taxonomy of adaptive performance. *J. Appl. Psychol.* 85, 4 (Aug. 2000), 612–624.

[19] Aadesh Salecha, Molly E Ireland, Shashanka Subrahmanya, João Sedoc, Lyle H Ungar, and Johannes C Eichstaedt. 2024. Large language models display human-like social desirability biases in Big Five personality surveys. *PNAS Nexus* 3, 12 (12 2024), pgae533.

[20] Ryosuke Takata, Atsushi Masumori, and Takashi Ikegami. 2024. Spontaneous Emergence of Agent Individuality Through Social Interactions in Large Language Model-Based Communities. *Entropy* 26, 12 (2024).

[21] Jiakai Tang, Heyang Gao, Xuchen Pan, Lei Wang, Haoran Tan, Dawei Gao, Yushuo Chen, Xu Chen, Yankai Lin, Yaliang Li, Bolin Ding, Jingren Zhou, Jun Wang, and Ji-Rong Wen. 2025. GenSim: A General Social Simulation Platform with Large Language Model based Agents. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*. Association for Computational Linguistics, 143–150.

[22] Yiran Wang, Ben He, Xuanang Chen, and Le Sun. 2025. Can LLMs Clarify? Investigation and Enhancement of Large Language Models on Argument Claim Optimization. In *Proceedings of the 31st International Conference on Computational Linguistics*. Association for Computational Linguistics, 4066–4077.

[23] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, Vol. 35. Curran Associates, Inc., 24824–24837.

[24] Chenxi Whitehouse, Monojit Choudhury, and Alham Fikri Aji. 2023. LLM-powered Data Augmentation for Enhanced Cross-lingual Performance. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 671–686.

[25] Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. 2024. Contrastive Preference Optimization: Pushing the Boundaries of LLM Performance in Machine Translation. In *Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235)*. PMLR, 55204–55224.

[26] Lan Zhang, Yuxuan Hu, Weihua Li, Quan Bai, and Parma Nand. 2025. LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks. *Systems* 13, 1 (2025).

[27] Qi Zhang, Christina S. Li, Daniel D. Goering, and Amy L. Kristof-Brown. 2024. Fitting in a workgroup in unique ways: A latent profile analysis of perceived person–group fit characteristics. *Journal of Applied Psychology* 109, 5 (2024), 779–794.

[28] Wenzhen Zheng and Xijin Tang. 2025. Simulating Social Network with LLM Agents: An Analysis of Information Propagation and Echo Chambers. In *Knowledge and Systems Sciences*. Springer Nature Singapore, 63–77.