ValidAgent – An Open Repository of Validated Generative Agents for Behavioral Simulations and Multi-Agent Systems

Leon Lührs[1,2][0009-0004-5550-1387], Christian Stindt von Dohm[1,2][0009-0001-1818-7892], Anika Bittner[1][0009-0004-7090-1557], Timo Heinrich[1][0000-0001-6827-1374], and Matthias Meyer[1,2][0000-0003-2980-4670]

[1] Hamburg University of Technology, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany
[2] United Nations University Hub on Engineering to Face Climate Change, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany
leon.luehrs@tuhh.de

Abstract. Large Language Models are increasingly used to create generative agents for behavioral simulations and multi-agent systems, as they allow researchers to build entities that exhibit seemingly human-like behavior with relatively low effort. Although generative agents are frequently parameterized using empirical data, systematic empirical validation remains scarce: there is no established practice of benchmarking agent behavior against human reference data, nor a common standard for reporting validation in a way that allows comparison across studies. No dedicated infrastructure exists to support the collection, documentation, and reuse of validated agents. As a result, generative agents are difficult to transfer across studies, hard to reproduce, and lack comparable validation – ultimately limiting cumulative scientific progress. In response, we present ValidAgent, a prototype of an open, web-based repository that provides a common interface for researchers to browse, compare, and select from a curated collection of generative agent profiles. The platform emphasizes empirical validation, transparency, and reproducibility by requiring each agent profile to include structured documentation covering its design rationale, behavioral traits, underlying assumptions, intended scope of application, and validation results against human reference data. To demonstrate the platform's utility, we provide an initial set of empirically validated agents grounded in the die-roll honesty paradigm, a widely used experimental setup for studying dishonest behavior. Building on this, we aim to foster a researcher-driven ecosystem in which generative agents can be contributed, peer-reviewed, and benchmarked – ultimately establishing citable generative agent sets. In doing so, this initiative contributes to a more transparent and cumulative multi-agent simulation practice.

Keywords: Open Repository, Generative Agents, Large Language Models.

1 Introduction

With the emergence of Large Language Models (LLMs), researchers have increasingly begun to develop generative agents – autonomous entities designed to simulate human-like behavior with a level of flexibility that goes beyond many traditional agent architectures.[1] Unlike conventional agents, which typically rely on explicitly defined rules and fixed decision logic, generative agents leverage LLMs to enable contextual reasoning and adaptive decision-making [6, 7]. Their behavioral identity is typically defined through structured persona descriptions that guide the model's responses. Recent work emphasizes grounding this "homo silicus" [8] in empirical data such as interviews and experimental findings, thereby moving beyond purely synthetic modeling towards more behaviorally validated simulations [7]. As a result, generative agents hold considerable promise for simulating complex social interactions [6, 7, 9, 10], including legal reasoning [11, 12], economic decision-making [13], political negotiation [14, 15], or consumer behavior [8, 16].

Despite this potential, the development and application of generative agents remain highly fragmented. Most agents are created in isolation for specific projects. Consequently, it remains unclear to what extent agent configurations and behavioral outcomes can be replicated or transferred across studies. Recent literature recognizes the need for greater consistency in agent design and evaluation practices. Amin et al. [1] emphasize the multitude of design choices – from model family and version to hyperparameters and prompting strategies – and argue for standardized prompt templates to ensure consistency while maintaining flexibility. In terms of behavioral validity, Tseng et al. [17] highlight the need for expanded datasets and behavioral benchmarks grounded in real human data. However, approaches to empirically grounding generative agents vary considerably across studies and are often not documented in comparable ways. Without shared standards and a common infrastructure to address both design consistency and empirical grounding, generative agents remain difficult to compare, reproduce, and systematically build upon, ultimately limiting cumulative scientific progress in generative agent research.

Several tools and infrastructures address individual aspects of generative agent development, yet none provides a comprehensive solution for validated, reusable agent profiles. For instance, persona collections such as PersonaHub [18] offer large-scale, synthetically generated profiles at unprecedented scale, while DeepPersona [19] extends this approach with a more extensive attribute taxonomy, greater demographic diversity, and customizable profile generation. However, these resources focus on scalable synthetic persona generation rather than directly collected individual-level human datasets. At the same time, agent-oriented frameworks such as the GenAgents architecture [6], Concordia [20], or EDSL – commercially implemented as Polly by Expected Parrot [21] – provide powerful environments for constructing, simulating, and in some

[1] There is no standardized terminology for these agents. We use "generative agents" as a working label, but note that terms such as "LLM-based agents," "generative AI," "AI agents," or "agentic AI" are also in use and may partially overlap [see, e.g., 1–5]. We conceptualize generative agent behavioral simulations as part of the broader multi-agent systems field.

cases benchmarking generative agents against human responses. Yet they are primarily designed for agent execution and experimentation rather than for sharing and reuse of validated agent profiles across studies. Dedicated repository infrastructures in other research domains demonstrate the value of structured sharing for cumulative and comparable research [22]. For example, platforms such as CoMSES Net [23], AGENTBLOCKS [24], OpenAI Gym [25], or Sotopia [26] demonstrate how curated repositories and benchmarking ecosystems advance research in their respective fields. Similarly, general-purpose repositories such as Zenodo [27] or OSF [28, 29] could technically host agent artifacts, yet lack domain-specific standards, structured documentation requirements, and community-driven quality control mechanisms. Platforms such as GitHub [30] or Hugging Face [31] provide versioned hosting for code, models, and datasets, but are not designed to curate generative agent profiles according to domain-specific validation standards or to treat them as structured, reviewable research objects.

In response, we propose a web-based, open-access repository dedicated to empirically validated generative agents – ValidAgent.[2] This platform offers a common interface for researchers to share, browse, compare, and select from a curated collection of agent profiles. Each profile integrates detailed documentation encompassing design rationale, behavioral characteristics, underlying assumptions, intended application scope, and validation outcomes against human reference data. Our approach prioritizes transparency, reproducibility, and reusability to foster a research ecosystem where generative agents can be contributed, peer-reviewed, benchmarked, and cited [32] – laying the foundation for cumulative scientific progress in multi-agent behavioral simulations. As a first application, we provide an initial set of agents validated within the well-established die-roll honesty paradigm [33], a widely used experimental framework for studying dishonest behavior.

2 A New Repository Design and Architecture – ValidAgent

The ValidAgent repository is implemented as a web-based, open-access platform that enables researchers to discover, evaluate, and export validated generative agents with minimal technical knowledge. Its design is structured around two core entities: Agent Sets and Agent Cards.

Agent Sets constitute the primary unit of organization. Each set represents a curated collection of agents that were jointly validated against, e.g., a shared human reference dataset or experimental paradigm. The repository allows users to browse and filter Agent Sets by application context, agent characteristics, or empirical support. Each entry provides a structured overview of the collection, including the number of agents, key persona attributes (e.g., demographic attributes or Big Five dimensions), intended application scope, validation outcomes, provenance, versioning, and licensing information (see Figure 1 in the Supplemental Material).

Within each Agent Set, Agent Cards provide detailed access to individual agent profiles – the structured persona configurations that define agent behavior in simulations.

[2] https://tuhh-maccs.github.io/validagent/

Each Agent Card contains two sections: persona modules and validation. The modules section displays the agent's persona configuration as modular components that can be toggled on or off, allowing the user to tailor the exported agent configuration to their needs. The interface dynamically updates both the generated persona text and token count estimates, making the trade-off between behavioral alignment and computational cost and efficiency transparent. The validation section visualizes how closely the selected agent's behavior aligns with human reference data, updating live as modules are adjusted. This enables informed decisions about the appropriate level of detail for different simulation contexts (see Figure 2 in the Supplemental Material).

Configured agent profiles can be exported in structured formats such as JSON and YAML, ready to be instantiated as agents within simulation environments. To ensure interoperability, the export layer adapts configurations to the input requirements of established frameworks, including GenAgents [6], Concordia [20], EDSL [21], and Sotopia [26]. In these frameworks, agent behavior emerges from the combination of a statically initialized agent profile defining personality, behavioral traits, or background, and the dynamic interaction and coordination logic provided by the framework itself. ValidAgent focuses on the former: providing empirically validated agent profiles that can be directly imported into existing environments. To this end, internal representations of profiles are mapped to framework-specific import formats, ensuring that validated configurations are readily deployable across established simulation environments (see Table 1 in the Supplemental Material).

New Agent Sets can be contributed through a structured submission interface requiring a standardized format of agent data, documentation, and accompanying metadata on validation procedures and human reference datasets. Each submission undergoes a community review process to ensure completeness of documentation, traceability of the validation process, and technical functionality. Integrated version control guarantees that each published set remains citable, while subsequent updates are handled through a transparent re-review process, ensuring long-term reproducibility [32].

3 Conclusion

Generative agents hold significant promise for advancing behavioral multi-agent simulations, yet their current development landscape is fragmented, lacking shared standards for validation, documentation, and reuse. In this paper, we introduce ValidAgent, a web-based, open-access repository for empirically validated generative agents that treats agent profiles as structured, versioned, and reviewable research objects. By organizing agents into validated sets, providing modular and transparent documentation, enabling interoperability with existing agent frameworks, and establishing quality assurance mechanisms, the platform lays the groundwork for reproducible, comparable, and citable generative agent resources. In doing so, we aim to shift generative agents from isolated project-specific implementations toward a sustainable research infrastructure that supports cumulative scientific progress.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amin, D., Salminen, J., Ahmed, F., Tervola, S.M.H., Sethi, S., Jansen, B.J.: How is generative AI used for persona development?: A systematic review of 52 research articles (2025). https://doi.org/10.48550/arXiv.2504.04927.
2. Rio-Chanona, R.M. del, Pangallo, M., Hommes, C.: Can generative AI agents behave like humans? Evidence from laboratory market experiments (2025). https://doi.org/10.48550/arXiv.2505.07457.
3. Sapkota, R., Roumeliotis, K.I., Karkee, M.: AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. Information Fusion. **126**(Part B), Article 103599 (2026). https://doi.org/10.1016/j.inffus.2025.103599.
4. Sun, G., Zhan, X., Such, J.: Building better AI agents: A provocation on the utilisation of persona in LLM-based conversational agents. In: Proceedings of the 6th ACM Conference on Conversational User Interfaces. Article no. 35. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3640794.3665887.
5. Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Qin, W., Zheng, Y., Qiu, X., Huang, X., Zhang, Q., Gui, T.: The rise and potential of large language model based agents: A survey. Sci. China Inf. Sci. **68**(2), Article 121101 (2025). https://doi.org/10.1007/s11432-024-4222-0.
6. Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. Article no. 2. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3586183.3606763.
7. Park, J.S., Zou, C.Q., Shaw, A., Hill, B.M., Cai, C., Morris, M.R., Willer, R., Liang, P., Bernstein, M.S.: Generative agent simulations of 1,000 people (2024). https://doi.org/10.48550/arXiv.2411.10109.
8. Horton, J.J.: Large language models as simulated economic agents: what can we learn from homo silicus? (2023). https://doi.org/10.3386/w31122.
9. Li, Y., Sun, L., Zhang, Y.: MetaAgents: Large language model based agents for decision-making on teaming. In: Proceedings of the ACM on Human-Computer Interaction. Article no. CSCW134 (2025). https://doi.org/10.1145/3711032.
10. Park, J.S., Popowsky, L., Cai, C., Morris, M.R., Liang, P., Bernstein, M.S.: Social simulacra: Creating populated prototypes for social computing systems. In: UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. Article no. 74 (2022). https://doi.org/10.1145/3526113.3545616.
11. Dong, Y.R., Hu, T., Collier, N.: Can LLM be a personalized judge? (2024). https://doi.org/10.48550/arXiv.2406.11657.
12. Guha, N., Nyarko, J., Ho, D., Ré, C., Chilton, A., K, A., Chohlas-Wood, A., Peters, A., Waldon, B., Rockmore, D., Zambrano, D., Talisman, D., Hoque, E., Surani, F., Fagan, F., Sarfaty, G., Dickinson, G., Porat, H., Hegland, J., Wu, J., Nudell, J., Niklaus, J., Nay, J., Choi, J., Tobia, K., Hagan, M., Ma, M., Livermore, M., Rasumov-Rahe, N., Holzenberger, N., Kolt, N., Henderson, P., Rehaag, S., Goel, S., Gao, S., Williams, S., Gandhi, S., Zur, T., Iyer, V., Li, Z.: LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models. Advances in Neural Information Processing Systems. **36**, 44123–44279 (2023).
13. Li, N., Gao, C., Li, M., Li, Y., Liao, Q.: EconAgent: Large language model-empowered agents for simulating macroeconomic activities (2024). https://doi.org/10.48550/arXiv.2310.10436.

14. Baker, Z.R., Azher, Z.L.: Simulating the U.S. senate: An LLM-driven agent approach to modeling legislative behavior and bipartisanship (2024). https://doi.org/10.48550/arXiv.2406.18702.
15. Hua, W., Fan, L., Li, L., Mei, K., Ji, J., Ge, Y., Hemphill, L., Zhang, Y.: War and peace (WarAgent): Large language model-based multi-agent simulation of world wars (2024). https://doi.org/10.48550/arXiv.2311.17227.
16. Li, Y., Liu, Y., Yu, M.: Consumer segmentation with large language models. Journal of Retailing and Consumer Services. **82**, Article 104078 (2025). https://doi.org/10.1016/j.jretconser.2024.104078.
17. Tseng, Y.-M., Huang, Y.-C., Hsiao, T.-Y., Chen, W.-L., Huang, C.-W., Meng, Y., Chen, Y.-N.: Two tales of persona in LLMs: A survey of role-playing and personalization (2024). https://doi.org/10.48550/arXiv.2406.01171.
18. Ge, T., Chan, X., Wang, X., Yu, D., Mi, H., Yu, D.: Scaling synthetic data creation with 1,000,000,000 personas (2025). https://doi.org/10.48550/arXiv.2406.20094.
19. Wang, Z., Zhou, Y., Luo, Z., Ye, L., Wood, A., Yao, M., Mansour, S., Pan, L.: DeepPersona: A generative engine for scaling deep synthetic personas (2025). https://doi.org/10.48550/arXiv.2511.07338.
20. Vezhnevets, A.S., Agapiou, J.P., Aharon, A., Ziv, R., Matyas, J., Duéñez-Guzmán, E.A., Cunningham, W.A., Osindero, S., Karmon, D., Leibo, J.Z.: Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia (2023). https://doi.org/10.48550/arXiv.2312.03664.
21. Expected Parrot: EDSL Documentation: https://www.expectedparrot.com/, last accessed 2026/02/14.
22. Berger, U., Bell, A., Barton, C.M., Chappin, E., Dreßler, G., Filatova, T., Fronville, T., Lee, A., van Loon, E., Lorscheid, I., Meyer, M., Müller, B., Piou, C., Radchuk, V., Roxburgh, N., Schüler, L., Troost, C., Wijermans, N., Williams, T.G., Wimmler, M.-C., Grimm, V.: Towards reusable building blocks for agent-based modelling and theory development. Environmental Modelling & Software. **175**, Article 106003 (2024). https://doi.org/10.1016/j.envsoft.2024.106003.
23. Rollins, N.D., Barton, C.M., Bergin, S., Janssen, M.A., Lee, A.: A computational model library for publishing model documentation and code. Environmental Modelling & Software. **61**, 59–64 (2014). https://doi.org/10.1016/j.envsoft.2014.06.022.
24. Filatova, T., Verbeek, L., Warnier, M., Ghorbani, A., Nikolic, I., Grimm, V., Berger, U., Barton, M., Bell, A., Lee, A., Magliocca, N.R., Wagenblast, T.: AGENTBLOCKS: A community platform for sharing, comparing, and improving reusable building blocks for (agent-based) models. Journal of Artificial Societies and Social Simulation. **28**(4), Article 11 (2025). https://doi.org/10.18564/jasss.5831.
25. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W.: OpenAI Gym (2016). https://doi.org/10.48550/arXiv.1606.01540.
26. Zhou, X., Zhu, H., Mathur, L., Zhang, R., Yu, H., Qi, Z., Morency, L.-P., Bisk, Y., Fried, D., Neubig, G., Sap, M.: SOTOPIA: Interactive evaluation for social intelligence in language agents (2024). https://doi.org/10.48550/arXiv.2310.11667.
27. Zenodo: https://zenodo.org/, last accessed 2026/02/14.
28. Soderberg, C.K.: Using OSF to share data: A step-by-step guide. Advances in Methods and Practices in Psychological Science. **1**(1), 115–120 (2018). https://doi.org/10.1177/2515245918757689.
29. Sullivan, I., DeHaven, A., Mellor, D.: Open and reproducible research on open science framework. Current Protocols Essential Laboratory Techniques. **18**(1), Article e32 (2019). https://doi.org/10.1002/cpet.32.

30. GitHub: https://github.com/, last accessed 2026/02/20.
31. Hugging Face: https://huggingface.co/, last accessed 2026/02/20.
32. Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR guiding principles for scientific data management and stewardship. Sci Data. **3**(1), Article 160018 (2016). https://doi.org/10.1038/sdata.2016.18.
33. Fischbacher, U., Föllmi-Heusi, F.: Lies in disguise: An experimental study on cheating. Journal of the European Economic Association. **11**(3), 525–547 (2013). https://doi.org/10.1111/jeea.12014.