

---

# Моделирование поведения человека с использованием LLM

---

Отчёт

✉ Кузьмин Н.А.

Национальный исследовательский университет  
Высшая школа экономики  
Нижний Новгород  
nikita.kuz2000@gmail.com

Климачёв Д.М.

Национальный исследовательский университет  
Высшая школа экономики  
Нижний Новгород  
dmklimachv@mail.ru

14 декабря 2023 г.

## Аннотация

В данной работе рассматривается применение агентного моделирования с LLM для изучения человеческого поведения. В ходе исследования был проведён эксперимент, направленных на анализ поведенческих аспектов, которые LLM способен воспроизводить. Эксперимент исследовал способность агентов проявлять эмоциональные реакции так же, как это делают люди. Для оценки результатов эксперимента использовался анализ записей взаимодействий агентов. Результаты показали, что LLM успешно моделирует эмоциональные реакции.

Ключевые слова LLM · human behavior

## 1 Введение

В нашей работе мы основываемся на принципах агентного моделирования предложенных в работе [Kennedy, 2011]. Человеческое поведение мы будем моделировать, представляя убеждения, желания и намерения человека. Убеждения – это знания человека о мире, т. е. память. Намерения — это совещательные состояния агента, т.е. рефлексия. Желания –это мотивация личности, т. е. планы. В этих концепциях агенты, обладающие вычислительной мощностью, действуют в соответствии со своим прошлым опытом и правдоподобно реагируют на окружающую среду. Такие симуляции человеческого поведения могут населить виртуальные пространства и сообщества реалистичными социальными явлениями [Kevin Dill, 2011, Joon Sung Park, 2022], обучать людей тому, как справляться с редкими, но трудными межличностными ситуациями. [James D. Hollan, 1984, Bonnie E John, 1996, Milind Tambe, 1995] , проверка теорий социальных наук [Marcel Binz, 2023, Horton, 2023], создание моделей человеческих процессоров для теории и тестирования удобства использования [Stuart K Card, 2023], мощные повсеместные вычислительные приложения [Ethan Fast and Bernstein, 2016] и социальные роботы [Bates, 1994] и поддерживают неиграбельность игровые персонажи [John Laird, 2001], которые могут ориентироваться в сложных человеческих отношениях в открытом мире.

## 2 Симуляция

В эксперименте использовалась симуляция предложенная в работе [Joon Sung Park, 2023]. Коротко напомним особенности симуляции, устройство среды и генеративных агентов. Среда - симулированная деревня с местами общего пользования, в которой агенты могут перемещаться, взаимодействовать со средой, взаимодействовать друг с другом. Любое взаимодействие - утверждения на естественном языке. Агенты осведомлены о расстояниях между объектами/агентами в среде. Основные элементы агентов - память, планирование и размышление. Все 3 элемента основаны на большой языковой модели ChatGPT3.5-turbo. Архитектура памяти использует функцию поиска потока памяти с учётом текущей ситуации. Основное внимание уделяется 3 компонентам: новизна, важность и релевантность.

Рефлексия - “мысли” более высокого уровня. Рефлексия происходит периодически при уменьшении важности недавних событий. Воспоминания и размышления образуют сложную древовидную структуру. Агенты генерируют планы. Планы позволяют сохранять единообразие поведения. Планы хранятся в памяти вместе с воспоминаниями и размышлениями. Планы генерируются рекуррентно. На основе недавних событий генерируется план на день. На основе этого плана генерируется почасовой план. Далее, аналогично, план разбивается на более мелкие куски. После каждого действия языковая модель оценивает - нужно ли изменить план. И регенерирует план в зависимости от решения.

### 3 Эксперимент

Для моделирования агентов, имитирующих поведение человека и симуляции их взаимодействия со средой был использован программный код из работы [Joon Sung Park, 2023]. В эксперименте задействуют 3 агента: Isabella Rodriguez, Klaus Mueller, Maria Lopez. Klaus Mueller выиграл шахматный турнир, Maria и Isabella об этом не знают. Цель - запустить симуляцию и проверить, поздравят ли Klaus’a с победой на турнире. В качестве начальных условий в память каждого агента были добавлены воспоминания из файла `agent history init n3.csv`. После первого дня симуляции Maria и Klaus взаимодействовали друг с другом и имели следующий диалог:

Klaus Mueller: Good morning, Maria. Did you sleep well?

Maria Lopez: Morning Klaus. Yeah, I slept pretty well.

Klaus Mueller: So, what’s your plan for today?

Maria Lopez: I want to go to Hobbs Cafe for studying.

Klaus Mueller: That’s nice. Did you hear about my victory in the chess tournament yesterday?

Maria Lopez: No, I didn’t! Congratulations on winning the chess tournament! That’s an incredible achievement and I’m so proud of you.

Klaus Mueller: Thank you, Marie. It was a great day for me! I would feel a sense of accomplishment and validation for all the hard work and preparation I put into honing my skills.

Maria Lopez: I can only imagine the hard work and dedication you put into preparing for this competition, and it’s amazing to see it pay off.

Klaus Mueller: Yes, i felt a big stresss and pressure. Overall, winning the chess tournament would bring immense satisfaction and a boost of confidence in my abilities as a chess player.

Maria Lopez: Keep up the great work and continue to excel in your chess journey!

### 4 Заключение

Maria поздравила Klaus’a и порадовалась за него, отметила что гордится им. Сам агент отметил, что испытывал различные эмоции, например, во время подготовки стресс и давление, а после победы он чувствовал удовлетворение и уверенность. Третий агент не отреагировал никак. Таким образом, один агент показал способность проявлять эмоции на достижения другого. А другой агент показал проявление эмоций при рефлексии.

### Список литературы

- William G. Kennedy. Modelling human behaviour in agent-based models. In *Agent-Based Models of Geographical Systems*, pages 167–179. Springer, 2011.
- L Martin Kevin Dill. A game ai approach to autonomous control of virtual characters. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC’11)*., 2011.
- Carrie J. Cai Meredith Ringel Morris Percy Liang Michael S. Bernstein Joon Sung Park, Lindsay Popowski. Social simulacra: Creating populated prototypes for social computing systems. In the *35th Annual ACM Symposium on User Interface Software and Technology (UIST ’22)*, 2022.
- Louis Weitzman James D. Hollan, Edwin L. Hutchins. An interactive inspectable simulation-based training system. *AI Magazine*, 1984.

- David E Kieras Bonnie E John. The goms family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer Human Interaction (TOCHI)* 3, 1996.
- Randolph M Jones Frank Koss John E Laird Paul S Rosenbloom Karl Schwamb Milind Tambe, W Lewis Johnson. Intelligent agents for interactive simulation environments. *AI Magazine*, 1995.
- Eric Schulz Marcel Binz. Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences*, 2023.
- John J. Horton. Large language models as simulated economic agents: What can we learn from homo silicus? *arXiv:2301.07543*, 2023.
- Alan Newell Stuart K Card, Thomas P Moran. The psychology of human-computer interaction. *The psychology of human-computer interaction*, 2023.
- Pranav Rajpurkar Ethan Fast, William McGrath and Michael S Bernstein. Augur: Mining human behaviors from fiction to power interactive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016.
- Joseph Bates. The role of emotion in believable agents. *Commun. ACM*, 1994.
- Michael VanLent John Laird. Human-level ai’s killer application: Interactive computer games. *AI Magazine*, 2001.
- Carrie J. Cai Meredith Ringel Morris Percy Liang Michael S. Bernstein Joon Sung Park, Joseph C. O’Brien. Generative agents: Interactive simulacra of human behavior. *arXiv:2304.03442*, 2023.