



Российская
экономическая
школа

Петр Паршаков, Юлия Найденова, Никита Маткин

Репликация экспериментального поведения в играх на доверие с использованием больших языков моделей

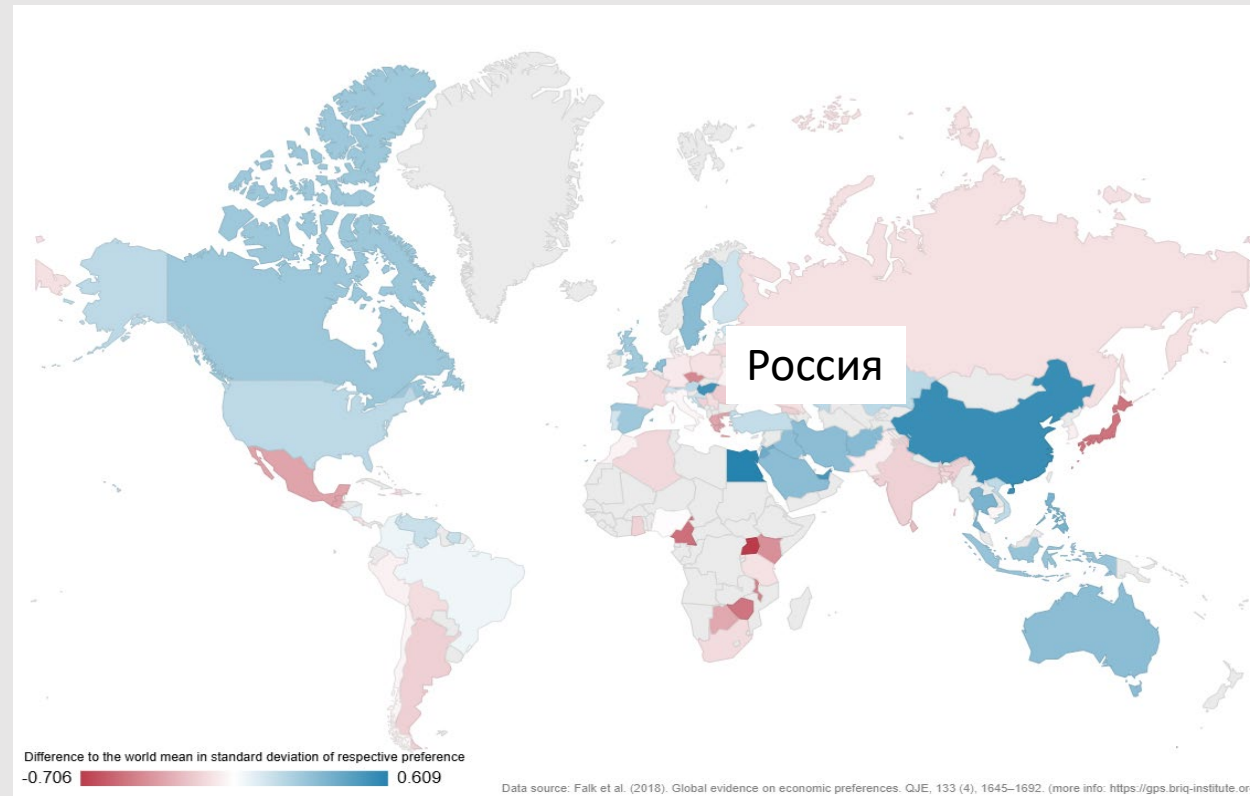
Дискуссант: Антон Суворов, РЭШ

**СОВМЕСТНЫЙ СЕМИНАР
БР, РЭШ И НИУ ВШЭ**

30 июня 2026 г.

ЧТО ТАКОЕ ДОВЕРИЕ?

"Вы считаете, что большинству людей можно доверять, или же в отношениях с людьми нужно быть очень осторожными?" (*Всемирное исследование ценностей (World Values Survey); Глобальное исследование предпочтений (Global Preference Survey), etc.*)



ЧТО ТАКОЕ ДОВЕРИЕ?

Альтернативный подход: лабораторная «игра в доверие» (Berg, Dickhaut and McCabe, 1995), с применением реальных экономических стимулов.

- Результаты опросных и лабораторных мер не вполне согласуются (Glaeser et al., 2000; Fehr et al., 2003; Sapienza, Toldra and Zingales, 2014; Jin et al., 2025;...)
- Сох (2004): Отправляя деньги «предпринимателю» в игре в доверие, «инвестор» может руководствоваться социальными предпочтениями, включая альтруизм, а не рассчитывать на возврат средств.

ПОЧЕМУ ВАЖНО ДОВЕРИЕ?

Доверие – наряду с близкими показателями (такими, как социальный капитал) – фактор экономического роста.

Много работ, которые устанавливают положительную связь между доверием и ВВП на душу населения, экономическим ростом, более низкой инфляцией и др. (Knack and Keefer, 1997; La Porta et al, 1997, etc)

Трудно установить причинно-следственную связь. Algan and Cahuc (2010): через анализ эффекта унаследованного доверия показывают каузальное влияние доверия на экономический рост.

- *Если бы в России доверие было как в Швеции, то ВВП на душу населения в 2000 году был бы выше на 69%.*

ИИ АГЕНТЫ ДЛЯ ЭКОНОМИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Экономические эксперименты, и лабораторные, и тем более полевые – сложная и затратная вещь.

Эксперименты с ИИ-агентами – резко повышают доступность экспериментальных исследований.

Вопрос: могут ли такие эксперименты заменить эксперименты с живыми людьми?

Паршаков, Найденова, Маткин (2026):

- конкретно в игре в доверие насколько хорошо поведение ИИ-агентов соответствует поведению живых участников?
- в частности, учитываются ли социально-экономические характеристики участников?

ИИ АГЕНТЫ ДЛЯ ЭКОНОМИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Основные результаты:

- На уровне средних значений наиболее успешные модели (Chat-GPT 5.4) хорошо приближает результаты двух экспериментов (Ashraf et al., 2006; Bahry et al., 2003).
- Распределения ответов воспроизводятся очень неточно, для БЯМ (LLM) характерна чрезмерная концентрация ответов.
- Есть разница в приближении поведения первого и второго игрока.
- Географические и другие, основанные на соц.-дем. характеристиках, отличия в поведении воспроизводятся плохо.
- Наблюдается существенный разброс в степени приближения экспериментальных данных разными БЯМ, причем отечественные модели не имеют преимущества.

ВОПРОСЫ К СТАТЬЕ

1. Как результаты соотносятся с другими исследованиями? В частности, про игру в доверие (но не только)?
2. Насколько робастны результаты к вариации промптов, температуры и т.п.?
3. Граничные условия: в какой ситуации можно ожидать похожих результатов – хорошего соответствия результатов в среднем, но плохой учет гетерогенности? Какие могут быть объяснения у полученных результатов?
4. Что можно предпринять для улучшения результатов симуляций в дальнейшем (кроме пассивного ожидания прогресса самих БЯМ)?
5. Что можно предпринять для улучшения самого доверия?

ЗАМЕНЯТ ЛИ ИИ-АГЕНТЫ ЛЮДЕЙ В ЭКСПЕРИМЕНТАХ?

Что было бы, если бы БЯМ успешно воспроизводили бы и паттерны решений, а не только их средние значения?

1. Vassaro (2026): целесообразность пререгистрации как и в обычных экспериментах, поскольку легкость проведения экспериментов является и риском.
2. Риск «коллапса моделей» из-за чрезмерного использования синтетических данных (Shumailov et al., 2024; Acemoglu et al., 2026).
3. ИИ может скоро заменить не только участников, но и исследователей (например, Ludwig and Mullainathan, 2024 про генерацию гипотез).