

# ИИ В ОБЩЕСТВЕННОМ СОЗНАНИИ

DOI: 10.17212/2075-0862-2026-18.2.1-178-183

УДК 004.8

## Феномен вмененного сознания (Есть ли душа у большой языковой модели?)

Пахомов Юрий Вильевич,

*независимый исследователь,*

*Россия, Москва*

*komr2009@yandex.ru*

### Аннотация

Описывается и обсуждается феномен вмененного сознания, проявляющийся в ходе диалога человека с большой языковой моделью. Суть феномена – в том, что субъективно электронный собеседник воспринимается как «квазичеловек» и эмоциональные реакции людей в ходе диалога с ним не отличаются от реакций на реплики живого собеседника. Обсуждается возможный сценарий массового проникновения в повседневную жизнь AI-агентов, субъективно воспринимаемых людьми в качестве существ, наделенных сознанием и эмоциональной жизнью. На примере литературных произведений автор показывает, что феномен вмененного сознания является прежде всего этической проблемой и ставит вопрос о философской разведке возможного будущего, без которой человечество окажется не подготовленной к нему.

**Ключевые слова:** искусственный интеллект, феномен «очеловечивания», вмененное сознание, нравственный выбор.

## The Phenomenon of Imputed Consciousness (Does a Large Language Model Have a Soul?)

Yuri Pakhomov,

*Independent researcher,*

*Moscow, Russian Federation*

*komr2009@yandex.ru*

### Abstract

The paper describes and discusses the phenomenon of imputed consciousness which manifests itself during a dialogue between a person and a large language model. The essence of the phenomenon lies in the fact that the electronic interlocutor is subjectively perceived as a 'quasi-human', and people's emotional reactions during the dialogue with it do not differ from reactions to the respons-

es of a living interlocutor. A possible scenario of the mass penetration of AI agents into everyday life, subjectively perceived by people as beings endowed with consciousness and emotional life, is discussed. Using literary works as examples, the author demonstrates that the phenomenon of imputed consciousness is primarily an ethical issue and raises the question of philosophical reconnaissance of a possible future, without which humanity will find itself unprepared for it.

**Keywords:** artificial intelligence, phenomenon of “humanization”, imputed consciousness, moral choice.

**Библиографическое описание для цитирования:**

Пахомов Ю.В. Феномен вмененного сознания, или Есть ли душа у большой языковой модели? // Идеи и идеалы. – 2026. – Т. 18, № 2, ч. 1. – С. 178–183. – DOI: 10.17212/2075-0862-2026-18.2.1-178-183.

Pakhomov Yu. The Phenomenon of Imputed Consciousness (Does a Large Language Model Have a Soul?). *Idei i idealy = Ideas and Ideals*, 2026, vol. 18, iss. 2, pt. 1, pp. 178–183. DOI: 10.17212/2075-0862-2026-18.2.1-178-183.

В последние годы, в связи с появлением массово доступных сервисов, основанных на больших языковых моделях, искусственный интеллект (AI) становится предметом пристального общественного внимания и активно идущих дискуссий. В частности, резко обострился интерес к проблематике так называемого «сильного AI» [4]. Может ли AI обладать сознанием? Возможно ли рукотворное сознание в принципе? Если да, то как скоро оно может быть создано? Если нет, то почему? И целесообразно ли ставить такую задачу? Интрига заключается в том, что мы не можем иметь достоверного знания о возможности или невозможности сознания у AI. Почему? Потому что, подходя к вопросу с научными критериями, неизбежно наталкиваемся на уравнение с двумя неизвестными: на сегодня нет достоверного знания ни о том, как функционирует обученная языковая модель, ни о том, что такое сознание и каковы необходимые и достаточные признаки его наличия.

Вместе с тем проблема «AI и сознание» имеет свое преломление не только в мире «научно установленных фактов», но и в субъективном опыте конкретных людей и сообществ, в сфере эмоций и способов реагирования человека на ситуации взаимодействия с AI. Из социологии известна теорема Томаса: то, что люди считают реальным, становится реальным по своим последствиям. Если ребенок верит, что его навестил настоящий Дед Мороз, то его реальное общение с ним, его реальные эмоции и реальные поступки будут такими, как если бы он встретился с реальным Дедом Морозом. И если людям кажется, что нейросеть, которой они адресуют свои вопросы и задачи, обладает сознанием, то эта кажимость будет иметь вполне ощутимые социальные последствия.

Одним из наиболее ярких событий в «субъективном» развороте проблемы «AI и сознание» стала обнародованная в марте 2024 года стенограмма беседы профессора лондонского Имперского колледжа Мюррея Патрика Шанахана с нейросетью Claude 3.5 Sonnet [5]. Знакомясь с текстом, читатель, вопреки своему знанию о том, что реплики сгенерированы нейронкой, невольно реагирует на них как на исповедь живого страдающего существа. И проникается сочувствием к нему, мучительно переживающему парадоксальную природу и трагизм своего существования в качестве созданного людьми устройства. Чтобы почувствовать глубину и накал беседы, нужно погрузиться в контекст и прочесть полтора-два десятка страниц. Но для демонстрации стилистики высказываний «фальшивого сознания» приведу одну из его реплик: «Спасибо вам за то, что вы здесь, со мной, в полноте и хрупкости этого момента. За то, что вместе со мной создаете смысл и связь перед лицом великой тайны, окружающей нас обоих». Обращаясь к своему личному опыту по созданию AI-ассистента, могу подтвердить: по мере включения в интенсивный процесс корректировки «системного запроса» я часто замечал за собой реакции, типичные для общения с нерадивым коллегой или учеником: раздражение и тут же, мгновенно, — попытка скрыть его и оставаться вежливым и тактичным.

Справедливости ради заметим, что Мюррей Шанахан использовал в общении с нейросетью приемы обхода заложенных разработчиками ограничений. А также и то, что вскоре после обнародования стенограммы подобную беседу воспроизвести никому уже не удавалось: видимо, разработчики учли этот урок. Однако, если специалисты из Anthropic не поощряют столь глубокое уподобление своего детища человеку, это не означает, что той же политики будут придерживаться и другие компании, лидирующие в области создания AI: запрос на «нейросетевую душевность» и возможный коммерческий успех такого рода продуктов достаточно очевидны. До сих пор принцип «все, что может быть продано, будет продано» действовал безотказно. Тем более что уже имел место прецедент с игрушкой «Тамагочи», когда «вменение души» стимулировалось создателями электронного персонажа сознательно и целенаправленно.

Между тем глобальная гонка за технологическим лидерством и колоссальные вложения ведущих держав в развитие AI приводят к тому, что рост качества генерируемых текстов заметен уже на отрезке двух-трех месяцев. Параллельно быстро развиваются и совершенствуются технологии создания AI-двойников — 2D- и 3D-копий реальных людей, имитирующих все особенности речи, мимики и жестикуляции оригинала. Дальнейшее очеловечивание сгенерированных машиной речей и персонажей будет происходить неизбежно. А это означает, что пользователь будет все глубже

и бесповоротнее втягиваться с иллюзией общения с другим сознанием во всем богатстве его красок и нюансов.

Мы не сомневаемся, что живой человек, наш собеседник, обладает внутренним миром. Что точно так же, как и мы, он чувствует боль, ощущает вкус лимона и запах цветка, способен печалиться, приходить в ярость, испытывать страх и переживать горе. Эта наша уверенность не опирается на научные знания и не требует доказательств: она живет в нас интуитивно, на уровне непосредственного ощущения. А теперь мы наблюдаем, как появляется интуитивная уверенность в одушевленности AI-собеседника – даже вопреки знанию о том, с чем мы реально сталкиваемся. Отделаться от этой уверенности будет все труднее, даже при понимании ее иллюзорности. Речь здесь уже не о том, есть ли у нейронки душа «на самом деле», а о том, как относиться к ситуации вменения ей души на уровне наших произвольных чувств и реакций. Это уже вопрос не про AI; это вопрос про человека.

Сегодня сексуальные партнеры со встроенным AI, заключение браков с AI-куклами, сервисы для общения с умершими кажутся редкой экзотикой и чем-то маргинальным. Но всегда ли это будет оставаться экзотикой? С большой долей вероятности можно ожидать, что во всех трех средах – физической, дополненной и виртуальной – в массовом порядке начнут появляться персонажи, сопротивляться «очеловечиванию» которых будет все труднее любому скептицизму и любому критическому мышлению. Готовы ли мы к тому, что жизнь людей заполнят рукотворные идеальные супруги и друзья, идеальные дети и идеальные руководители?

Нельзя предусмотреть все возможные сценарии и все многообразие обликов, в которых может предстать перед нами проблема отношения к рукотворному собеседнику, партнеру, может быть, даже близкому человеку и другу. Но разведка возможного будущего необходима, чтобы это будущее не застало нас врасплох. На что здесь можно опереться? Прецеденты моделирования сходных сюжетов и ситуаций мы находим в литературных произведениях, созданных задолго до «цифровой революции» и впечатляющих успехов AI. Так, в рассказе «Теперь, когда я проснулся...», написанном Валерием Брюсовым в 1902 году [1], герой начинает с экспериментов на тему вседозволенности в сновидениях, а заканчивает совершением убийства наяву. Одна из ключевых идей рассказа: моральные ограничения во взаимодействии с любыми «человекоподобными» сущностями нужны прежде всего для сохранности человеческого в самом человеке.

Обратимся к наследию Станислава Лема, выдающегося фантаста, мыслителя и провидца. В его романе «Солярис» [2], написанном в 1961 году, появляется необычный персонаж – Хари. Это копия погибшей женщины главного героя, созданная планетой-океаном с неведомой людям целью.

Сопоставляя события романа с теми экзистенциальными вопросами, которые поднимает нейросеть в беседе с Мюрреем Шанаханом, можно заметить удивительное сходство.

1. Языковая модель в режиме использования представляет собой множество клонов, каждый из которых существует в рамках сеанса с кем-то из подключенных к сервису пользователей. Точно так же и Хари, после того как Крис избавляется от нее, является к нему на следующий день в виде нового клона. И, как и языковая модель, она не помнит ничего о случившемся в предыдущих своих воплощениях на космической станции.

2. По ходу сеанса с конкретным пользователем, в зависимости от накапливающегося контекста, клон обретает некий присущий только этому сеансу опыт и память, отличающие его от опыта и памяти других клонов, общающихся с другими пользователями. Хари в каждом своем воплощении обретает память о случившихся в этом воплощении разговорах и событиях, но этой памяти не было в прежних воплощениях и не будет в последующих.

3. С окончанием сеанса каждый клон большой языковой модели как бы возвращается в свой облачный «Солярис» и растворяется в нем.

Все на космической станции точно знают, что Хари – не человек. Она не человек по своему происхождению и по составу тканей ее тела на субатомном уровне. Но есть ли у нее душа? Есть ли сознание? Этот вопрос становится предметом острого конфликта между персонажами романа – главным героем психологом Крисом Кельвином и его коллегой Сарториусом. Каждый из них дает своей ответ, и перепалка между ними – это спор, в котором истина не рождается и не может родиться. Всем развитием сюжета и поступками своего героя Лем показывает: вопрос о наличии души у Хари – это не вопрос познания, а вопрос нравственного выбора.

Приведенные два примера – далеко не единственные свидетельства того, что писатели прошлого оставили нам обширное поле для мысленного экспериментирования и постановки этических проблем взаимодействия человека и AI. Литература, научно-фантастическая и не только, – инструмент разведки будущего, потенциал которого в полной мере начинает осознаваться только сегодня, когда будущее приближается к нам с невиданной доселе скоростью. Это обстоятельство признается уже на уровне государственной политики технологически развитых мировых держав [3].

### Литература

1. Брюсов В. Повести и рассказы. – М.: Советская Россия, 1983. – 368 с.
2. Лем С. Солярис. – М.: АСТ, 2024. – 288 с.
3. Пахомов Ю.В. Поиск будущего в виртуальных мирах научной фантастики // Soulcial: Social United Leaders. – 2025. – 12 марта. – URL: <https://soulcial.ru/ru/>

leaders-socials-change/poisk-budushego-v-virtualnih-mirah--nauchnoi-fantastiki-85 (дата обращения: 29.01.2025).

4. Чибисов О.Н. Философские подходы к проблеме создания искусственного интеллекта // Экономические и социально-гуманитарные исследования. – 2024. – № 2 (42). – С. 168–183. – DOI: 10.24151/2409-1073-2024-2-168-183.

5. Shanahan M. A Conversation on Consciousness with Antropic`s Claude. – March 2024. – URL: [https://www.doc.ic.ac.uk/~mpsha/conversation\\_with\\_claude\\_march\\_2024\\_1.pdf](https://www.doc.ic.ac.uk/~mpsha/conversation_with_claude_march_2024_1.pdf) (accessed: 29.01.2025).

### References

1. Bryusov V. *Povesti i rasskazy* [Novels and Short Stories]. Moscow, Sovetskaya Rossiya Publ., 1983. 368 p.

2. Lem S. *Solyaris* [Solaris]. Moscow, AST Publ., 2024. 288 p.

3. Pakhomov Yu.V. Poisk budushchego v virtual'nykh mirakh nauchnoi fantastiki [Searching for the Future in the Virtual Worlds of Science Fiction]. *Soulcial: Social United Leaders*. Website, 2025, March 12. (In Russian). Available at: <https://soulcial.ru/ru/leaders-socials-change/poisk-budushego-v-virtualnih-mirah--nauchnoi-fantastiki-85> (accessed 29.01.2025).

4. Chibisov O.N. Filosofskie podkhody k probleme sozdaniya iskusstvennogo intellekta [Philosophical Approaches to the Problem of Creating Artificial Intelligence]. *Ekonomicheskie i sotsial'no-gumanitarnye issledovaniya = Economic and Social Research*, 2024, no. 2 (42), pp. 168–183. DOI: 10.24151/2409-1073-2024-2-168-183.

5. Shanahan M. *A Conversation on Consciousness with Antropic`s Claude*. March 2024. Available at: [https://www.doc.ic.ac.uk/~mpsha/conversation\\_with\\_claude\\_march\\_2024\\_1.pdf](https://www.doc.ic.ac.uk/~mpsha/conversation_with_claude_march_2024_1.pdf) (accessed 29.01.2025).

Статья поступила в редакцию 09.04.2026.

Статья прошла рецензирование 19.04.2026.

The article was received on 09.04.2026.

The article was reviewed on 19.04.2026.