



<https://doi.org/10.22363/2313-2302-2024-28-4-1050-1066>

EDN: JVEONZ

Научная статья / Research Article

## Методологическое взаимодействие экспериментальной и компьютерной философии

Е.А. Алексеева  

*Государственный академический университет гуманитарных наук, Москва, Россия*

 [eaalekseeva@gaugn.ru](mailto:eaalekseeva@gaugn.ru)

**Аннотация.** Исследование посвящено проблеме взаимосвязи методов двух современных философских направлений: экспериментальной и компьютерной философии. Оба эти направления отличаются тем, что они отходят от принципов «кабинетной» философии, хотя и не порывают к ней полностью, обращаются к эмпирическим данным и т.п. Рассматриваются ключевые принципы экспериментальной философии, которые позволяют интегрировать методологическую структуру экспериментальных философских исследований с подходами в области компьютерной философии. Также демонстрируется, что с помощью компьютерной философии могут быть трансформированы такие классические философские методы, как концептуальный анализ и мысленные эксперименты, что в том числе позволяет частично объединить между собой экспериментальную и компьютерную философию. Выделяются четыре группы компьютерных технологий, которые могут быть использованы в связке с философскими методами. Это технологии анализа данных, компьютерного моделирования (многоагентные системы и т.п.), генеративные ИИ-технологии, компьютерные игры. В контексте философских исследований результаты, полученные с помощью этих ИКТ-инструментов, требуют особой смысловой интерпретации. В данной работе рассматривается проблематика потенциально возможного экспериментального философского исследования, посвященного проблеме подтверждения/доказательства в области доказательной медицины и медицинской деятельности. Предлагаются варианты того, каким образом можно было бы использовать методы компьютерной философии для проведения такого исследования.

**Ключевые слова:** метафилософия, философские методы, концептуальный анализ, концептуальный инжиниринг, мысленный эксперимент, философский эксперимент, анализ данных, многоагентные системы, генеративные ИИ-технологии, компьютерные игры

**Информация о финансировании.** Исследование выполнено в Государственном академическом университете гуманитарных наук в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № FZNF-2023-0004 «Цифровизация и формирование современного информационного общества: когнитивные, экономические, политические и правовые аспекты»)

© Алексеева Е.А., 2024



This work is licensed under a Creative Commons Attribution 4.0 International License <https://creativecommons.org/licenses/by-nc/4.0/legalcode>

**История статьи:**

Статья поступила 21.06.2024

Статья принята к публикации 19.09.2024

**Для цитирования:** *Алексеева Е.А.* Методологическое взаимодействие экспериментальной и компьютерной философии // Вестник Российского университета дружбы народов. Серия: Философия. 2024. Т. 28. № 4. С. 1050–1066. <https://doi.org/10.22363/2313-2302-2024-28-4-1050-1066>

## Methodological Interaction between Experimental and Computer Philosophy

Ekaterina A. Alekseeva  

*State Academic University for the Humanities, Moscow, Russia*

 [eaalekseeva@gaugn.ru](mailto:eaalekseeva@gaugn.ru)

**Abstract.** The study is devoted to the problem of the relationship between the methods of two contemporary philosophical trends: experimental and computer philosophy. Both of these trends are distinguished by the fact that they depart from the principles of “armchair” philosophy, although they do not break with it completely, turn to empirical data, etc. The key principles of experimental philosophy are considered, which allow integrating the methodological structure of experimental philosophical research with approaches in the field of computer philosophy. It is also demonstrated that with the help of computer philosophy, such classical philosophical methods as conceptual analysis and thought experiments can be transformed. This, among other things, allows integrating them into experimental philosophical research. Four groups of computer technologies are distinguished that can be used in conjunction with philosophical methods. These are data analysis technologies, computer modeling (multi-agent systems, etc.), generative technologies, computer games. In the context of philosophical research, the results obtained with the help of these technological tools receive a semantic interpretation. This paper proposes the problematic of a potential experimental philosophical study devoted to the problem of confirmation/proof in the field of evidence-based medicine and medical activity. Options are offered for how the methods of computational philosophy could be used to conduct such a study.

**Keywords:** metaphilosophy, philosophical methods, conceptual analysis, conceptual engineering, thought experiment, philosophical experiment, data analysis, multi-agent systems, generative technologies, computer games

**Funding of Sources.** The research was prepared at the State Academic University for the Humanities within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (topic No. FZNF-2023-0004 “Digitalization and methods of the modern information society: cognitive, physical, political and legal aspects”).

**Article history:**

The article was submitted on 21.06.2024

The article was accepted on 19.09.2024

**For citation:** Alekseeva EA. Methodological Interaction between Experimental and Computer Philosophy. *RUDN Journal of Philosophy*. 2024;28(4):1050–1066. (In Russian). <https://doi.org/10.22363/2313-2302-2024-28-4-1050-1066>

### Ключевые методологические принципы экспериментальной философии

Экспериментальная и компьютерная философия являются относительно новыми подходами в философских исследованиях, и статус этих подходов для философского сообщества, по всей видимости, до сих пор не вполне ясен, несмотря на то, что определенная их институционализация уже присутствует. Признаками такой институционализации можно считать книжную серию *Advances in Experimental Philosophy*, в которой отдельные книги посвящены достижениям в области экспериментальной философии в различных областях (эпистемологии, метафизики, философии языка и т.д.). В области компьютерной философии так же есть как отдельные публикации, так и журналы, ориентированные на публикации статей по данному направлению. Тем не менее оба этих подхода еще не стали философским мейнстримом, а специальных методологических исследований, которые анализировали бы их взаимосвязь, практически не существует. Единственная работа, которую можно здесь отметить, это статья 2007 г. Wiegel V., van den Berg J. *Experimental Computational Philosophy: shedding new lights on old philosophical debates* [1], однако в ней речь идет, скорее, о вычислительных моделях, применимых в экспериментальной философии, а не о современных компьютерных методах философского исследования. Соответственно, вопрос о том, каким образом возможная взаимосвязь методов экспериментальной и компьютерной философии, все еще требует изучения.

Одно из достаточно широких определений экспериментальной философии звучит следующим образом: «Экспериментальная философия – это название недавно появившегося философского направления, представители которого используют методы экспериментальной психологии для исследования того, как люди выносят суждения, которые влияют на философские дебаты. Хотя у движения есть название, оно включает в себя множество проектов, направляемых различными интересами, предположениями и целями» [2. Р. 123]. Будем отталкиваться от этого определения, тем более что и другие исследователи так или иначе рассматривают экспериментальную философию в близком ключе: «При этом изучается обычное понимание людьми морали, свободы воли, счастья и других ключевых философских понятий. Цель состоит в том, чтобы погрузиться непосредственно в реальный мир и использовать психологические эксперименты, чтобы добраться до истоков философских проблем» [3. С. 29].

Ключевой задачей экспериментального философского исследования становится проверка философских интуиций относительно того, каким образом в обыденном сознании и повседневном опыте интерпретируются некоторые

понятия, для которых существуют и философские концепты, и связанные с ними концепции. Фактически это изучение мировоззренческих пресуппозиций, которыми руководствуются различные субъекты в своей деятельности. Подтверждение или опровержение подвергающихся проверке философских интуиций в свою очередь включается в структуру философского исследования, то есть рефлексия относительно определенных концептов или категорий проходит по второму кругу. Философ-исследователь здесь рекурсивно включается в изучаемую область.

Примером исследования в области экспериментальной философии, в ходе которого эксплицируются его методологические основания, является работа *Experimental Philosophy and Statistical Learning* [4], где среди прочих проверяется «старая идея в эпистемологии, восходящая к Платону, о том, что знание требует неопровержимости» [4. Р. 18]. В работе предложен эксперимент, который должен выявить, каким образом в ходе обучения учащиеся приобретают идею того, что такое знание. Исследование состоит из нескольких этапов, где исходная проблема обозначается как *Acquirendum*, затем формулируются гипотеза относительно того, что учащиеся в ходе обучения приобретают одно из трех возможных представлений о знании: «(1) возможность того, что знание является истинным убеждением, (2) возможность того, что знание является истинным убеждением с обоснованием какого-либо рода, но без требования неопровержимости и (3) возможность того, что знание должно быть неопровержимым» [4. Р. 19]. Затем предлагается когнитивно-психологический эксперимент, который должен показать, какая из этих интерпретаций знания формируется у учащихся. Все этапы и методологические основания проведения экспериментального исследования подробно описаны в статье. Есть и другие примеры подобных исследований [5; 6].

Общее описание методологии экспериментально-философского исследования может выглядеть следующим образом: «1. Разложение выбранной философской теории на предположения, предпосылки, механизмы, предсказания; 2. они в свою очередь должны быть преобразованы в экспериментальные установки, то есть в практические примеры в противоположность абстрактным терминам; 3. затем происходит их трансформация в элементы промежуточной концептуальной структуры; 4. конкретные экспериментальные установки интерпретируются в категориях промежуточной концептуальной структуры; 5. теория реализуется в лаборатории на основе требований промежуточной концептуальной основы; 6. рефлексиируются конкретные экспериментальные установки; 7. проводится эксперимент; 8. результаты переводятся обратно в термины исходной теории; 9. и затем они используются, чтобы подтвердить, уточнить или отвергнуть теорию» [1. Р. 57].

Нужно отметить, что в настоящее время компьютерные методы уже периодически используются в экспериментальной философии. Чаще всего это методы, пришедшие из других областей цифровой гуманитаристики, например, корпусной лингвистики, использующей компьютерные технологии для

анализа корпусов текстов. Применение таких технологий описано в книге *Methodological Advances in Experimental Philosophy* [7]. Представляется, однако, что дальнейшее развитие экспериментальной философии сделает ее методологическое взаимодействие с компьютерной философией практически неизбежным. Связано это прежде всего с тем, что оба направления являются универсальными философскими методологиями, применимыми практически в любой области исследования, а также меняют сам стиль философского мышления в соответствии с трансформациями современной культуры, фундированной компьютерными технологиями.

### **Что такое компьютерная философия?**

Существует как минимум три англоязычных понятия, которые обозначают методологические изменения философии в сторону цифровизации философских исследований. Это *Digital Philosophy*, *Computational Philosophy* (встречается наиболее часто) и *Computer Philosophy* (встречается редко). Первое понятие выглядит образованным по аналогии с понятием *Digital Humanities*, но на самом деле обозначает прежде всего определенную метафизическую систему, утверждающую, что вселенная имеет цифровую основу. Второе понятие можно интерпретировать как вычислительную философию, но здесь смешиваются два аспекта: применение вычислительных моделей в философии и применение компьютерных технологий в философских исследованиях. Строго говоря, вычислительные модели не обязательно предполагают реализацию в компьютерной среде, но предшествуют ей, поэтому для того чтобы подчеркнуть, что особая философская методология предполагает применение специализированных компьютерных инструментов для философских исследований, предлагается использовать именно понятие компьютерной философии (*computer philosophy*) в его совпадении с одним из аспектов вычислительной философии (*computational philosophy*).

Понятие вычислительной (или компьютерной философии) в Стэнфордской энциклопедии определяется следующим образом: «использование вычислительных методов для усиления методологического аппарата философских исследований. Вычислительная философия – это не философия компьютерных технологий или вычислительных методов; это скорее философия с использованием компьютеров и вычислительных методов» [8]. Как видим, это в основном методологическое определение компьютерной философии, но его можно дополнить более широким пониманием компьютерной философии как фундаментального изменения самого способа философского мышления, при котором стирается онтологическая граница между человеческим и артефактным (в данном случае компьютерным). Это возможно за счет того, что современные компьютерные технологии способны воспроизводить такие интеллектуальные действия, которые ранее были принадлежностью лишь человеческого интеллекта, однако совершают они эти действия особым

нечеловеческим способом. Некоторые авторы связывают подобные эффекты с развитием GPT-подобных технологий [9], другие – с развитием интеллектуальных компьютерных систем в целом [10].

Если компьютерную философию рассматривать со стороны ее методологического потенциала, то можно обозначить несколько традиционных философских методов, которые могут быть компьютеризированы. Прежде всего, это концептуальный анализ, и хотя относительно него в современной философии нет устоявшегося консенсуса [11], но его можно в целом определить как прояснение некоторых понятий, выявляющее их разнообразные смысловые аспекты и контексты употребления. При этом концептуальный анализ иногда обозначается и в качестве метода экспериментальной философии [12]. Другой компьютеризируемый философский метод – это концептуальный инжиниринг в широком смысле слова, представляющий собой как переинтерпретацию существующих концептов с целью их «улучшения» или большей релевантности относительно некоторого контекста, так и формулировку новых концептов, в том числе в тех областях, которые остаются дискурсивно невидимыми. Наконец, компьютерная философия способна преобразовать философский метод мысленных экспериментов с помощью современных компьютерных технологий.

Все три указанных философских метода не только могут быть компьютеризированы, как будет показано в дальнейшем, но и являются в той или иной степени связанными с экспериментальной философией. Экспериментальные данные могут использоваться для построения и валидации компьютерных моделей. С помощью компьютерных моделей, в свою очередь, можно генерировать гипотезы для экспериментальной проверки. Далее мы рассмотрим несколько типов компьютерных технологий, применимых в философских исследованиях, интегрирующих методы компьютерной и экспериментальной философии. К их числу относятся компьютерный анализ данных, компьютерное моделирование на основе многоагентных систем, GPT-технологии, компьютерные игры. Для того, чтобы показать потенциал этих технологий в философских исследованиях, будет предложена идея потенциально возможного философского исследования.

### **Потенциально возможное исследование на основе синтеза экспериментальной и компьютерной философии**

Чтобы продемонстрировать, каким образом возможно методологическое взаимодействие экспериментальной и компьютерной философии, будет предложен пример потенциально возможного исследования, выдержанного в рамках обоих этих направлений. Такое исследование не проводилось, но его проведение потенциально возможно, поэтому демонстрируется его предварительная структура. На основе этой структуры будет показано, каким способом возможно применение четырех типов обозначенных выше компьютерных технологий для проведения философских исследований.

В качестве исследовательской проблемы обозначим актуальную для эпистемологии тему доказательства/подтверждения, но интерпретируемую в контексте социально значимой проблемы доверия к доказательной медицине (evidence-based medicine). Доказательная медицина – это общее наименование подхода в медицине, при котором медицинская деятельность основывается на результатах исследований эффективности и безопасности лекарственных препаратов, методов лечения и т.п., осуществлявшихся согласно принципам научной рациональности и с использованием релевантной научной методологии. Существует ряд исследовательских способов, соответствующих принципам доказательной медицины (например, плацебо-контролируемое исследование), но в данный момент нет необходимости подробно на них останавливаться.

Эволюция в медицине, которая привела к появлению доказательной медицины, происходила как смена подходов: «Экспертное мнение и личный опыт долгое время оставались основой для принятия решений в области здравоохранения. Постепенно между 1950-ми и 1990-ми возросла важность научных методов, тестирования гипотез в контролируемых условиях и статистического анализа, при этом стало очевидно, что использование только личного мнения приводит к различным вредным последствиям» [13]. В то же время до сих пор внедрение методов доказательной медицины сталкивается с определенными сложностями, в частности существует так называемое явление дениализма, когда отвергаются протоколы лечения, сформулированные на основе научных результатов, а также возникают течения, подобные антивакцинаторству. Все это заставляет поставить вопрос о причинах того, почему как со стороны врачей, так и со стороны пациентов нет однозначного доверия к доказательной медицине. Далеко не все врачи в своей клинической практике целиком придерживаются принципов доказательной медицины, также далеко не все пациенты готовы предпочесть доказательную медицину разнообразным практикам из области «альтернативной» медицины. Есть различные объяснения этому феномену: социальные, экономические, психологические и т.п. Также существует ряд внутренних проблем самой доказательной медицины, которые имеют в том числе эпистемический характер [14]. Но в данном случае нас интересует смысловая интерпретация самого понятия «доказательства» в рамках доказательной медицины.

Предположим, что одной из причин того, что доказательная медицина сталкивается с рядом сложностей для реализации ее принципов в медицинской практике, является неоднозначность в понимании того, какой способ лечения, лекарственный препарат и т.п. можно считать доказавшим свою эффективность. То есть это вопрос о том, что подразумевается под доказательством. Как известно, понятие «evidence» можно перевести на русский язык по крайней мере тремя способами: доказательство, подтверждение, свидетельство. Соответственно, здесь мы сталкиваемся с проблемой подтверждения как некоторого эмпирического доказательства, а также его соотношения с доказательством в логико-дедуктивном смысле.

Существует несколько эпистемологических теорий подтверждения. В первую очередь в данной области большое значение имеют байесовские подходы: «Данные особенности находятся в соответствии с предметной спецификой байесовских подходов в эпистемологии: основным объектом исследования является рациональная уверенность и ее изменения, ввиду чего стержневой темой байесовского осмысления науки является именно операция подтверждения и ее составляющие» [15. С. 24]. Согласно байесианской теории подтверждения опытные данные  $e$  подтверждают теорию  $h$ , если и только если вероятность  $h$  возрастает после того, как  $e$  становится известно, то есть если  $\Pr(h | e) > \Pr(h)$ . Иными словами, «согласно байесианскому подходу разница между апостериорной и априорной вероятностью гипотезы  $h$  дает нам количественную оценку того, в какой степени опытные данные  $e$  поддерживают  $h$ » [16. С. 89].

В рамках потенциального исследования в русле экспериментальной философии предполагается выдвинуть гипотезу, согласно которой представление о том, что такое доказательство/подтверждение, какие суждения можно считать доказанными/подтвержденными, а также о том, что входит в процедуру доказательства/подтверждения, существенно расходится у трех различных групп акторов, вовлеченных в медицинскую деятельность. Первая из этих групп – исследователи, проводящие испытания методов лечения и лекарственных препаратов; вторая – врачи, которым приходится непосредственно применять разработанные методы в своей клинической практике, выбирая при этом *evidence-based* или *opinion-based* (основанный на врачебном мнении) подход; третья – пациенты, для которых ключевым становится вопрос о доверии какому-либо подходу. При этом некоторое представление о том, что такое доказательство/подтверждение, должно присутствовать в качестве фоновой пресуппозиции, в том числе при формулировке суждений с эпистемической модальностью вида «доказано, что...», на которые ориентируются представители первой группы при репрезентации результатов исследований, а представители второй и третьей группы – при принятии соответствующих решений, касающихся медицинских действий.

Структура такого потенциально возможного исследования в общем виде будет выглядеть как последовательность нескольких этапов: 1. выявление ключевых концептуальных фреймов теории доказательства/подтверждения; 2. формулировка нескольких примеров, допустим, в качестве совокупности суждений вида «доказано, что...»; 3. разработка эксперимента, в ходе которого будут выявлены предпочитаемые способы понимания подтверждения/доказательства; 4. проведение эксперимента; 5. сопоставление результатов с изначальным предположением о различии представлений о подтверждении/доказательстве у различных групп акторов, вовлеченных в медицинскую деятельность. На каждом этапе возможно применение различных компьютерных технологий и подходов из области компьютерной философии.

## Анализ данных в компьютерной и экспериментальной философии

Компьютерный анализ данных, *data-mining* и т.п. постепенно становятся частью компьютерной философии, и это вполне закономерно, поскольку философские исследования предполагают значительный объем аналитической работы. При этом компьютерный анализ текстовых данных выглядит вполне органично в рамках философских методов, ведь философия в принципе нацелена на работу с текстом, однако, как представляется, нет никаких непреодолимых препятствий для использования и нетекстовых данных в философских исследованиях, особенно в контексте экспериментальной философии. Тем не менее в настоящее время значительно проще найти примеры применения технологий анализа данных в философии, когда дело касается исследования текстов.

В области философии науки показательным примером такого философского подхода является работа *Theoretical Virtues in Scientific Practice: An Empirical Study* [17], в которой анализируется роль так называемых эпистемических добродетелей (простота, последовательность и т.п.) в практике научных исследований. Автор работы с помощью различных инструментов из области *Data Science* анализирует большие массивы научных текстов, относящиеся к различным дисциплинарным областям. Для этих текстов были определены вербальные маркеры эпистемических добродетелей (то есть слова и словосочетания, указывающие на то, что в некотором фрагменте текста речь идет об этих добродетелях), затем проанализирована частотность и контекст использования этих маркеров. Автор работы пришел к общему выводу, что «ученые ссылаются на теоретические добродетели явно, хотя и довольно редко, когда говорят о модели (менее 30 %), теории (менее 20 %) и гипотезы (менее 15 %) в своих опубликованных работах» [17]. Фактически подобное исследование обладает чертами экспериментальной философии, поскольку демонстрирует, каким образом некоторые концепты неотрефлексированно интерпретируются учеными, при этом отсутствует необходимость проводить анкетирования и опросы, которые приняты в более «классических» экспериментально-философских исследованиях.

Применение анализа данных для исследования корпуса текстов практикуется и в области экспериментальной философии, хотя преобладающим методом исследований все же остается метод опросов, анкетирование и т.п. Например, анализируются существующие в некотором сообществе интерпретации понятий, а *text-mining* применяется к корпусу соответствующих текстов. В исследовании «*Using Corpus Linguistics to Investigate Mathematical Explanation*» [18] рассматривается, каким образом ученые из различных областей интерпретируют понятие объяснения. На основании изучения корпуса математических текстов с применением методов компьютерного анализа данных делаются следующие выводы: «Наши данные проливают свет на типы объяснений, обсуждаемых математиками. Мы обнаружили, что когда

математики предлагают некоторое объяснение, они, кажется, чаще заинтересованы в объяснении того, как что-то сделать в математике, чем в объяснении того, почему все обстоит именно так, как оно есть. В физике, напротив, мы обнаружили обратное. Возможно, эта особенность была унаследована от более традиционных научных исследований, когда ученые хотят ответить на вопросы «почему» о реальном мире» [18. Р. 251].

Вернемся к нашему потенциально возможному философскому исследованию, которое совмещало бы в себе подходы в области компьютерной и экспериментальной философии. Поскольку нам необходимо было бы выявить, каким образом интерпретируется понятие «подтверждение/доказательство» в различных группах акторов, вовлеченных в медицинскую деятельность, значительную роль здесь мог бы сыграть концептуальный анализ этих понятий. С точки зрения компьютерной философии, такой концептуальный анализ возможно провести на большом объеме текстовых данных. Необходимо выделить текстовые маркеры, которые сигнализировали бы о той или иной позиции в отношении понятий «подтверждение/доказательство». При этом все тексты возможно кластеризовать по принципу наличия в них текстовых маркеров, сцепленных с понятием «доказательство/подтверждение».

### **Многоагентное моделирование в компьютерной и экспериментальной философии**

Некоторые исследователи считают, что «Моделирование и компьютерные симуляции должны рассматриваться как ключевой философский метод» [19. Р. 2]. Действительно, в некоторых направлениях компьютерной философии именно компьютерное моделирование занимает лидирующие позиции, поскольку «Наш опыт показывает, что среда компьютерного моделирования часто заставляет нас задавать новые вопросы или задавать старые вопросы по-новому ... например, интерпретировать гоббсовские вопросы в пространстве теоретико-игровых стратегий» [20. С. 10]. А среди технологий компьютерного моделирования достаточно часто применяется многоагентное моделирование, в частности с использованием системы NetLogo. Многоагентное моделирование удобно применять в тех областях, где философская проблематика связана с анализом взаимодействий между различными акторами, и этот анализ используется в качестве базиса для исследования философских проблем. Например, «в социальной эпистемологии исследуются процессы, хорошо поддающиеся интерпретации средствами многоагентных систем: групповые убеждения, динамика обсуждения и согласия, общий вывод или подтверждение для множества акторов и т.п.» [21. С. 92].

Хороший пример социально-эпистемологического исследования с использованием многоагентного моделирования представлен в работе Yes, no, maybe so: a veritistic approach to echo chambers using a trichotomous belief model [22]. В работе проверяется гипотеза так называемой «эхо-камеры»,

предполагающая, что что определенные взгляды, убеждения, утверждения, многократно воспроизводящиеся в закрытом сообществе людей, усиливаются и становятся недоступными для критики. С помощью инструмента NetLogo моделируется взаимодействие агентов, каждый из которых может быть в одном из трех состояний убежденности в отношении некоторого утверждения: «да», «нет», «может быть так». В исследовании рассматривается вопрос о том, возможна ли беспристрастность как избегание предубеждений при таком взаимодействии между эпистемическими агентами. В результате исследования показано, каким образом работают эффекты эхо-камеры, за счет каких механизмов они поддерживаются, каким образом их можно избежать или наоборот усилить.

Если речь идет о нашем предполагаемом исследовании понятия доказательств/подтверждения в парадигме экспериментальной философии, суть которого излагалась ранее, то здесь с помощью многоагентного моделирования возможно исследовать, как взаимодействуют три группы акторов, в каждой из которой степень убежденности в доказанности/подтвержденности некоторого суждения различна. Степень доверия к некоторому суждению можно выразить промежутком вероятности от 0 до 1, тогда многоагентная модель будет представлять собой совокупность агентов, каждый из которых обладает параметром в этом промежутке. Здесь возможно использовать, например, такую формальную модель, как модель группового обсуждения Хонга и Пейджа [23]. С помощью подобной модели можно предложить решение проблемы того, как возможно формирование некоторого общего убеждения о подтвержденности/доказанности определенного суждения, даже если изначально акторы расходились в оценке того, насколько это суждение (в данном случае из области медицины, касающееся эффективности некоторого лекарственного препарата или метода лечения) является доказанным.

### **Генеративный ИИ в компьютерной и экспериментальной философии**

Произошедшая в последние годы экспансия генеративных технологий не могла не затронуть и философские исследования, которые, с одной стороны, ориентированы на рефлексию относительно генеративного ИИ, с другой – способны интегрировать GPT-подобные модели в качестве инструментария философского исследования. Однако важно понимать, что по большому счету технологии генеративного ИИ являются «стохастическим попугаем»: «Системой случайного соединения последовательностей лингвистических форм, которые присутствовали в обширных обучающих данных, в соответствии с вероятностной информацией о том, как они комбинируются, но без какой-либо ссылки на их значение» [24. Р. 126]. В то же время генеративные модели, в частности, продуцирующие текст, способны не просто стать инструментом для написания философских работ, но и эксплицировать

некоторые смысловые единицы культуры, поскольку такие системы «обучаются» на огромных объемах текстов, спродуцированных в процессе разнообразной культурной деятельности.

В настоящее время развитие генеративных технологий позволяет предложить несколько сценариев их применения в философских исследованиях. В целом, в этих сценариях можно обнаружить исходное пересечение методов компьютерной и экспериментальной философии. Первый сценарий связан с возможностью дообучать языковые модели на совокупности текстов из некоторой дисциплинарной области, например, используя работы определенного философа или нескольких философов. Дальнейший промптинг может представлять собой систему вопросов по интересующей проблематике, ответы на которые будут выдержаны в рамках системы взглядов данного философа. Подобные проекты уже разрабатывались, например, известен такой проект, как «Искусственный Деннет», но также было и несколько других попыток спроектировать искусственного философа [25]. На первый взгляд, такие проекты имеют прежде всего образовательное значение в рамках университетских курсов по изучению наследия определенных философов, однако решают они и исследовательские задачи, поскольку, например, позволяют генерировать суждения относительно различных тематик, выдержанные в русле определенной философской концепции, но затрагивающие вопросы, которые, возможно, ранее были в этой концепции недостаточно эксплицированы.

Другой аспект применения GPT-технологий в философских исследованиях связан с концептуальным инжинирингом и в целом с философской деятельностью по производству концептов. Так, например, GPT-системы способны предложить яркие метафоры, на основе которых возможна дальнейшая разработка философских концептов. Пример промптинга с запросом и ответа на него: «Придумай, как метафорически обозначить систему, которая учитывает различные взгляды на мир. GPT: Метафорически систему, которая учитывает различные взгляды на мир, можно обозначить как: „Многогранное зеркало познания“ – эта метафора показывает, как система способна отражать множество перспектив и точек зрения, подобно грани зеркала, которое преломляет свет под разными углами. „Симфония мировоззрений“ – воззрения подобны музыкальной симфонии, где каждый инструмент вносит свой уникальный голос. „Паутина взаимопонимания“ – эта метафора показывает, как система связывает и плетет воедино разные взгляды, создавая сеть взаимопроницающих и взаимодополняющих перспектив. „Калейдоскоп реальностей“ – эта метафора иллюстрирует, как система способна генерировать многообразие интерпретаций и представлений о мире, подобно калейдоскопу, который создает бесконечные узоры из отражений». Как видим, достаточно примитивный запрос привел к генерированию метафор, каждая из которых может быть либо использована для формирования концепта, либо проанализирована как некоторая экспликация общего культурного фона.

Возвращаясь к теме потенциально возможного исследования в русле экспериментальной философии, касающегося вопроса о прояснении понятия доказательство/подтверждение, можно предложить использование GPT-систем для того, чтобы выявить некоторый «общий знаменатель» того, как в целом интерпретируются эти понятия. Англиязычное понятия „evidence“ было проинтерпретировано GPT-ботом следующим образом: “Evidence is information or data that supports or refutes a claim, hypothesis, or theory. It can take many forms, including: Empirical data – Observations, measurements, and experimental results that can be objectively verified. Documentary evidence – Written records, documents, or artifacts that provide factual information. Testimonial evidence – Statements or accounts from eyewitnesses or experts. Circumstantial evidence – Indirect evidence that implies a conclusion, but does not directly prove it. Evidence is the foundation for rational, evidence-based decision making in fields like science, law, and policy. It allows us to move beyond speculation and anecdote to reach justified, well-supported conclusions about the world around us”.

Другой промпт был сформулирован следующим образом: «Какое из следующих двух суждений можно считать доказанным и почему? Суждение 1: „Пирацетам улучшает когнитивные способности“. Суждение 2: „Нет достаточных оснований считать, что пирацетам улучшает когнитивные способности“». GPT: «Рассмотрим оба суждения: Суждение 1: „Пирацетам улучшает когнитивные способности“. Для того, чтобы это суждение считалось доказанным, необходимо наличие достаточной научной доказательной базы, включающей результаты хорошо спланированных и проведенных клинических исследований. Эти исследования должны показывать статистически значимое улучшение когнитивных функций у людей, принимавших пирацетам, в сравнении с плацебо или другими контрольными группами. Только при наличии таких убедительных доказательств суждение 1 можно считать доказанным. Суждение 2: „Нет достаточных оснований считать, что пирацетам улучшает когнитивные способности“. Это суждение можно считать доказанным, если имеющиеся научные данные не подтверждают эффективность пирацетама для улучшения когнитивных функций. Отсутствие достоверных доказательств положительного эффекта пирацетама в хорошо спланированных клинических исследованиях позволяет сделать вывод, что нет достаточных оснований считать, что он улучшает когнитивные способности. Таким образом, из двух представленных суждений, второе суждение „Нет достаточных оснований считать, что пирацетам улучшает когнитивные способности“ можно считать доказанным, если имеющиеся научные данные не подтверждают эффективность пирацетама в улучшении когнитивных функций». При этом нужно учитывать, что рассуждения GPT-систем далеко не всегда являются логически корректными [26].

Как видим, данное понимание того, что такое «доказательство/подтверждение» выдержано, скорее, на уровне обыденных суждений. Поскольку у нас должно быть три исследуемых группы акторов, есть возможность

дообучить GPT-системы на трех разных группах текстов. Допустим, первая из этих групп текстов будет представлять собой совокупность научных статей, описывающих результаты клинических испытаний, вторая – тексты, продуцируемые практикующими врачами (статьи, как научные, так и популярные, реплики из обсуждений в специализированных сообществах и т.п.), третья – текстовые фрагменты, продуцируемые пациентами (реплики в околомедицинских пабликах и т.п.). Дообученным таким образом моделям можно поставить ряд общих задач, например, оценить степень доказанности суждения вида: «Доказано что X, останавливает процесс Y, приводящий к симптому Z...», а также объяснить причину сделанного вывода о доказанности/подтверждении суждения. Сопоставив полученные результаты, можно сделать вывод о том, каким образом интерпретируется «подтверждение/доказательство» различными группами акторов, а также оценить причины расхождения (или, наоборот, сходства) в интерпретации этих суждений различными группами.

### **Компьютерные игры в компьютерной и экспериментальной философии**

Взаимосвязь компьютерных игр и философии имеет достаточно интересные особенности: философские исследования компьютерных игр являются частью гейм-стадиэ, что особенно характерно для отечественной философской традиции [27]. Философскими методами исследуется как нарративный компонент (сюжет, смысловые, идейные, мировоззренческие аспекты), так и геймплейный компонент (структура игрового опыта) компьютерных игр. Данная область исследований, однако, до сих пор не стала общепризнанным философским направлением, в зарубежных научных журналах философские исследования компьютерных игр зачастую растворяются в социологических и культурологических работах, в отечественной традиции они представлены спорадическими текстами, хотя некоторые философские комьюнити, где обсуждается проблематика компьютерных игр, уже сформировались.

Однако компьютерные игры и философия могут быть связаны и другим способом. Во-первых, некоторые компьютерные игры являются буквальной визуализацией философских концепций или мысленных экспериментов, хотя их развлекательный компонент не заслоняется смысловым содержанием. К числу таких игр относится, к примеру, SOMA, где визуализированы и интегрированы в сюжет и геймплей некоторые мысленные эксперименты и концепции из области философии сознания. Во-вторых, возникает вопрос о том, могут ли компьютерные игры сами стать методологическим инструментом философского исследования, а не просто визуализировать и переводить в иммерсивный формат различные направления философской мысли. Предполагается, что такое направление развития взаимосвязи философии и компьютерных игр вполне возможно. Так, например, философ Й. Богост, которому

принадлежат такие работы по исследованию видеоигр, как *How to Talk about Videogames* [28], является еще и разработчиком компьютерных игр, с помощью которых стремится иллюстрировать собственные идеи.

Компьютерную игру в качестве философского метода можно рассматривать как некоторую иммерсивную реализацию и эмпирическую экспериментальную проверку мысленных экспериментов. Например, существует несколько не вполне серьезных онлайн-игр, которые построены на основе сюжета мысленного эксперимента «Проблема вагонетки». Эти игры собирают и предоставляют статистику о том, какой выбор совершают пользователи, проходя этот мысленный эксперимент в игровом формате. При этом можно представить исследование в рамках экспериментальной философии, которое могло бы вполне серьезно использовать такую игру, чтобы подтвердить или опровергнуть некоторые интуиции относительно того, какой выбор делают люди в предложенной в мысленном эксперименте ситуации.

Для нашего потенциально возможного экспериментального исследования проблемы «подтверждения/доказательства» можно предложить разработать онлайн-игру, в которой было бы необходимо принимать решения в области медицинской деятельности. Допустим, это была бы игра, в которой у протагониста есть задача выжить как можно дольше, выбирая предложенные NPC-врачом лекарства или методы лечения. При этом пользователю можно было предоставить несколько вариантов действия, когда он мог бы на основании предоставленной информации решать, является ли некоторое лекарственное средство или способ лечения в достаточной степени подтвержденным. Игровая статистика могла бы показать, что в повседневном мышлении воспринимается в качестве достаточного доказательства/подтверждения.

### Список литературы / References

- [1] Wiegel V, van den Berg J. Experimental Computational Philosophy: shedding new lights on old philosophical debates. In: Dastani MM, de Jong E, editors. *Proceedings of the 19th Belgian-Dutch Conference on Artificial intelligence (BNAIC'07)*. 2007. P. 291–298.
- [2] Nadelhoffer T, Nahmias E. The past and future of experimental philosophy. *Philosophical Explorations*. 2007;10(2):123–149.
- [3] Тимошенко Т.В., Скрипник К.Д. Экспериментальная философия: проблема интуиции // ЮП. 2020. № 1–2. С. 27–39.  
Timoshenko TV, Skripnik KD. Experimental philosophy: the problem of intuition. *South Pole*. 2020;(1–2):27–39. (In Russian).
- [4] Nichols Sh. Experimental Philosophy and Statistical Learning. In: Fischer E, Curtis M, editors. *Methodological Advances in Experimental Philosophy*. London: Bloomsbury Press; 2019. P. 13–41.
- [5] Griffiths PI, Machery E, Linquist S. The Vernacular Concept of Innateness. In: Knobe J, Nichols Sh, editors. *Experimental Philosophy: Volume 2*. Oxford: Oxford University Press; 2014. P. 281–306.

- [6] Andow J, Cova F. Why compatibilist intuitions are not mistaken: A reply to Feltz and Millan. *Philosophical Psychology*. 2015;29(4):550–566. <https://doi.org/10.1080/09515089.2015.1082542>
- [7] Fischer E, Curtis M, editors. *Methodological Advances in Experimental Philosophy*. London: Bloomsbury Press; 2019.
- [8] Grim P, Singer D. Computational Philosophy. In: *The Stanford Encyclopedia of Philosophy*. Available from: <https://plato.stanford.edu/archives/fall2022/entries/computational-philosophy/> (accessed: 20.06.2024).
- [9] Rees T. Non-Human Words: On GPT-3 as a Philosophical Laboratory. *Daedalus*. 2022;151(2):168–182. [https://doi.org/10.1162/daed\\_a\\_01908](https://doi.org/10.1162/daed_a_01908)
- [10] *Алексеева Е.А.* Расширенный киборг: концептуализация взаимосвязи людей и интеллектуальных систем // *Философия науки и техники*. 2024. Т. 29. № 1. С. 98–111. <https://doi.org/10.21146/2413-9084-2024-29-1-98-111>  
*Alekseeva EA.* The Extended Cyborg: Conceptualizing the Interconnection of Humans and Intelligent Systems. *Philosophy of Science and Technics*. 2024;29(1):98–111. (In Russian). <https://doi.org/10.21146/2413-9084-2024-29-1-98-111>
- [11] King JC. Philosophical and Conceptual Analysis. In: Cappelen H, Gendler T, Hawthorne J, editors. *The Oxford Handbook of Philosophical Methodology*. Oxford: Oxford University Press; 2016. P. 249–261.
- [12] Stuart MT. Philosophical Conceptual Analysis as an Experimental Method. In: Gamerschlag T, Gerland D, Osswald R, Petersen W, editors. *Meaning, Frames, and Conceptual Representation*. Düsseldorf: Düsseldorf University Press; 2015. P. 267–292.
- [13] Van Remoortel H, De Buck E, Seifried E, Vandekerckhove P. Evidence-Based Medicine: Principles and Values as Illustrated by the Case of Patient Blood Management. *Hamostaseologie*. 2023;43(1):16–21. <https://doi.org/10.1055/a-1985-7660>
- [14] Berg H. Is evidence-based practice justified? – A philosophical critique. *J Eval Clin Pract*. 2024;30:855–859. <https://doi.org/10.1111/jep.14059>
- [15] *Думов А.В.* Байесовский подход к пониманию научной рациональности // *Секулярный век: вызовы цивилизации: Материалы национальной научной конференции, посвященной Всемирному дню философии, Красноярск, 18 ноября 2021 года*. Красноярск: Красноярский государственный аграрный университет, 2021. С. 23–26.  
*Dumov AV.* Bayesian approach to understanding scientific rationality. In: *Secular age: challenges of civilization: Proceedings of the national scientific conference dedicated to World Philosophy Day, Krasnoyarsk, November 18, 2021*. Krasnoyarsk: Krasnoyarsk State Agrarian University publ.; 2021. P. 23–26. (In Russian).
- [16] *Макеева Л.Б.* Субъективная вероятность, теория подтверждения и рациональность // *Радио.ru*. 2015. № 15. С. 80–96.  
*Makeeva LB.* Subjective probability, confirmation theory and rationality. *Ratio.ru*. 2015;(15):80–96. (In Russian).
- [17] Mizrahi M. Theoretical Virtues in Scientific Practice: An Empirical Study. *The British Journal for the Philosophy of Science*. 2021;73(4):879–902. <https://doi.org/10.1086/714790>
- [18] Mejía-Ramos HP, Alcock L, Lew K, Rago P, Sangwin Ch, Inglis M. Using Corpus Linguistics to Investigate Mathematical Explanation. In: Fischer E, Curtis M, editors. *Methodological Advances in Experimental Philosophy*. London: Bloomsbury Press; 2019. P. 239–263.

- [19] Mayo-Wilson C, Zollman KJS. The computational philosophy: simulation as a core philosophical method. *Synthese*. 2021;(199):3647–3673. <https://doi.org/10.1007/s11229-020-02950-3>
- [20] Mar GR, Grim P, Denis PS. *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling*. US Massachusetts: MIT Press; 1998.
- [21] Алексеева Е.А. Проекты компьютерной эпистемологии // Философия науки и техники. 2023. Т. 28. № 4. С. 88–101. <https://doi.org/10.21146/2413-9084-2023-28-2-88-101>  
Aleksееva EA. Projects of computer epistemology. *Philosophy of science and technology*. 2023;28(4):88–101. (In Russian). <https://doi.org/10.21146/2413-9084-2023-28-2-88-101>
- [22] Baumgaertner B. Yes, no, maybe so: a veritistic approach to echo chambers using a trichotomous belief model. *Synthese*. 2014;(191):2549–2569. <https://doi.org/10.1007/s11229-014-0439-9>
- [23] Hong L, Page SE. Groups of diverse problem solvers can outperform groups of highability problem solvers. *Proceedings of the National Academy of Sciences*. 2004;101(46):16385–16389. <https://doi.org/10.1073/pnas.0403723101>
- [24] Bender E, Gebru T, McMillan-Major A, Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York NY United States: Association for Computing Machinery; 2021. P. 610–623.
- [25] Schwitzgebel E, Schwitzgebel D, Strasser A. Creating a large language model of a philosopher. *Mind & Language*. 2024;39(2):237–259. <https://doi.org/10.1111/mila.12466>
- [26] Зайцев Д.В. Почему большие языковые модели не (всегда) рассуждают как люди? // Вестник Московского университета. Серия 7: Философия. 2024. Т. 48. № 1. С. 76–93. <https://doi.org/10.55959/MSU0201-7385-7-2024-1-76-93>  
Zaitsev DV. Why don't large language models (always) reason like people? *Bulletin of Moscow University. Series 7: Philosophy*. 2024;48(1):76–93. (In Russian). <https://doi.org/10.55959/MSU0201-7385-7-2024-1-76-93>
- [27] Ветушинский А.С., Салин А.С. Game Studies в России: год восьмой // Социология власти. 2020. Т. 32. № 3. С. 8–13. <https://doi.org/10.22394/2074-0492-2020-3-8-13>  
Vetushinsky AS, Salin AS. Game Studies in Russia: Year Eight. *Sociology of Power*. 2020;32(3):8–13. (In Russian). <https://doi.org/10.22394/2074-0492-2020-3-8-13>
- [28] Bogost I. *How to Talk about Videogames*. Minneapolis: University of Minnesota Press; 2015.

### Сведения об авторе:

Алексеева Екатерина Алексеевна – кандидат философских наук, доцент, кафедра эпистемологии и логики, философский факультет, Государственный академический университет гуманитарных наук, Российская Федерация, 119049, Москва, Ломоносовский проспект, д. 26. ORCID: 0000-0002-0006-5942. SPIN-код: 2328-7230. E-mail: eaallekseeva@gaugn.ru

### About the author:

Alekseeva Ekaterina Alekseevna – PhD in Philosophy, Associate Professor, Department of Epistemology and Logic, Faculty of Philosophy, State Academic University for the Humanities, 26 Maronovskiy Pereulok, Moscow, 119049, Russian Federation. ORCID: 0000-0002-0006-5942. SPIN-code: 2328-7230. E-mail: eaallekseeva@gaugn.ru