УДК 165.12:004.89
https://doi.org/10.17072/2078-7898/2024-3-317-327
EDN: OHSVIP

Поступила: 26.02.2024
Принята: 29.06.2024
Опубликована: 03.10.2024

ИСКУССТВЕННЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ПРОБЛЕМА ОПЫТА

Биричева Екатерина Вячеславовна

Институт философии и права УрО РАН (Екатеринбург)

Разработка, внедрение и дальнейшее совершенствование систем искусственного интеллекта (ИИ) тесно связаны с проблемой опыта. Такие системы, в отличие от программ как замкнутых алгоритмов, взаимодействуют с внешней по отношению к ним средой и могут вносить в нее изменения на практике. В связи с этим современный дискурс приписывает искусственным агентам «способности», «обучаемость», «принятие решений» и т.п. Однако насколько правомерно экстраполировать на искусственные интеллектуальные системы (ИИС) смысл феноменов, характерных для живых существ? Способна ли машина действительно усваивать опыт, научаться и принимать решения? Поиск ответов на подобные вопросы побуждает к раскрытию понятия опыта, его структуры и специфики его получения живыми существами. Ввиду неоднозначности самого этого понятия продуктивным оказывается применение феноменологического подхода, позволяющего не только прояснить сущностные черты опыта, но и исследовать его многомерные связи с практикой, памятью, воображением, волей, постановкой и достижением целей. Разбор конкретных примеров также помогает оценить аналоги данных компонентов для искусственных агентов и систематизировать проблемы, возникающие при дальнейшем совершенствовании ИИС. Представленные результаты показывают, что понятие опыта в строгом смысле слова неприменимо к ныне функционирующим «слабым/узким» ИИ, тем не менее возможность моделирования данного феномена открыта в рамках будущих разработок «сильного/общего» ИИ. В заключении приводятся выводы о том, какие факторы необходимо учесть и воплотить в ходе создания ИИС, которые были бы способны к переживанию опыта и осознанной практической деятельности.

Ключевые слова: искусственный интеллект, интеллектуальные системы, понятие опыта, априорные формы, практика, квалиа, память, воля.

Для цитирования:

Биричева Е.В. Искусственные интеллектуальные системы и проблема опыта // Вестник Пермского университета. Философия. Психология. Социология. 2024. Вып. 3. С. 317–327. https://doi.org/10.17072/2078-7898/2024-3-317-327. EDN: OHSVIP

https://doi.org/10.17072/2078-7898/2024-3-317-327

Received: 26.02.2024
Accepted: 29.06.2024
Published: 03.10.2024

ARTIFICIAL INTELLIGENT SYSTEMS AND THE PROBLEM OF EXPERIENCE

Ekaterina V. Biricheva

Institute for Philosophy and Law, Ural Branch, Russian Academy of Sciences (Ekaterinburg)

The development, implementation, and further improvement of artificial intelligence (AI) systems tend to be related to the problem of experience. Unlike programs as closed algorithms, such systems interact with

the environment and are able to change it in practice. Thus, contemporary discourse appears to ascribe «abilities», «learning skills», «decision-making», etc. to artificial agents. However, is it correct to extrapolate to artificial intelligent systems (AIS) the sense of phenomena characteristic of living beings? Is a machine truly able to gain experience, learn, and make decisions? The search for answers to such questions encourages investigation into the category of experience, its structure, and its gaining by living beings. Due to the ambiguity of this category itself, the use of the phenomenological approach seems to be the most productive. It allows us not only to clarify the essential features of experience but also to explore its multidimensional connections with practice, memory, imagination, will, setting and achieving goals. An analysis of specific examples also helps in evaluating analogues of these components for artificial agents and in systematizing problems that arise with further improvement of AIS. The presented results show that the category of experience, in the strict sense of the word, is not applicable to the currently functioning «weak/narrow» AI. However, the possibility of modeling this phenomenon seems to be open within the framework of future developments of «strong/general» AI. In conclusion, the reader may find factors that should be taken into account and implemented when creating AIS that would be able to gain experience and to carry out conscious practical activity.
Keywords: artificial intelligence, intelligent systems, category of experience, a priori forms, practice, qualia, memory, will.

To cite:

Biricheva E.V. [Artificial intelligent systems and the problem of experience]. *Vestnik Permskogo universiteta. Filosofia. Psihologia. Sociologia* [Perm University Herald. Philosophy. Psychology. Sociology], 2024, issue 3, pp. 317–327 (in Russian), https://doi.org/10.17072/2078-7898/2024-3-317-327, EDN: OHSVIP

Введение

Искусственные интеллектуальные системы (ИИС) все активнее внедряются в практику во всех смыслах этого слова. Речь не только об изменениях, вносимых в эмпирическую реальность, будь то генерация новых «произведений», «умная» автоматизация производства или беспилотное управление. По меткому замечанию П. Вирно, «праксис» как выбор/поступок в сфере публичного (или политического) действия сегодня неразделимо переплетается с «общим интеллектом» [Вирно П., 2013, с. 73–74]. На этом основании при работе с большими данными «принятие решений» начинают делегировать искусственным агентам, что, однако, вызывает беспокойство (экзистенциальное [Арендт Х., 2000, с. 10] и этическое [Лунь-ков А.С., 2020]), порождая также потребность в определении юридического статуса ИИС, зон ответственности и т.д. [Лаптев В.А., 2019]. Алармистские настроения указывают на непроясненность беспокоящих тенденций, поэтому подобные опасения необходимо преодолевать, опираясь на разбор лежащих в основании феноменов — в данном случае, практики и опыта. Философский анализ этих понятий также подготовит надежную базу для разработки конкретных мер этического и юридического регулирования ИИС. Так, прежде всего, необходимо исследовать структуру практического действия (в разных смыслах), приобретения опыта и его сущностные черты, характерные для живых систем, сделав также выводы о применимости данных понятий к системам искусственного интеллекта (ИИ).

Оставляя за рамками данной работы споры об определении самого ИИ, возьмем наиболее широкий смысл любой технически созданной системы, способной выполнять мыслительные операции. Распознание и обобщение (визуальной, звуковой, символьной и др. информации) выполняются т.н. «узкими» (англ. ***narrow***) или «слабыми» ИИС, которые уже повсеместно внедрены в виде чат-ботов, голосовых помощников, сканеров лиц, генераторов изображений и т.п. [Георгиу Т.С., 2022, с. 35]. «Общий» (англ. ***general***) или «сильный» ИИ, способный к самостоятельной постановке и выполнению разноплановых задач, подлинному творчеству и самосознанию специалисты надеются создать в будущем [Витяев Е.Е. и др., 2020, с. 7]. В области разработки «общего/сильного» ИИ дискуссионными остаются вопросы связи «сознание–тело», моделирования творческих функций, возможности наделения машин волей и т.д.

[Биричев Е.В., 2020а; Биричева Е.В., Стерхов Е.В., 2019; Георгиу Т.С., 2022]. Ключевым моментом, их объединяющим, также представляется проблема опыта, без которого невозможны креативные способности, обучение, достижение сложных целей на практике [Витяев Е.Е. и др., 2020], самость и субъектность искусственных агентов [Турко Д.С., 2021].

В современном дискурсе как «слабый», так и «сильный» ИИ наделяются возможностями получения опыта, совершения выбора, обучения и т.п. Это происходит, поскольку, в отличие от программы как замкнутого алгоритма, ИИС (например, нейросеть из перцептронов) является открытой, сообщаясь с внешней для нее средой [Новиков Н.Б., 2023, с. 116]. Так возникает интенция говорить о некотором *опыте* имения дела с этой средой, который может быть эвристичен, поскольку сама перцептронная математическая модель не содержит образцов результата, критериев его правильности/неправильности [Витяев Е.Е. и др., 2020, с. 8]. Однако в этом случае не избежать ряда вопросов. Можно ли применительно к имеющимся ИИС говорить о феномене опыта? Каков генезис критериев его успешности (усвоенности)? Чем отличается опыт живых организмов и человека? В каком смысле машина может «принимать решение» и совершать «практическое действие»? Какие проблемы следует решить на пути создания универсальных ИИС, способных выполнять разноплановые задачи и «дообучаться» новому? В статье предпринимается попытка систематизировать данные проблемы, уточнив границы феномена опыта и обратив внимание на его связь с практикой, волей, вниманием, целеполаганием, воображением и памятью.

1. Два смысла понятия практики

В обыденном словоупотреблении «практика» и «опыт» используются как синонимы, хотя первое наделено более объективным значением, в то время как второе характеризует скорее деятельность субъекта «изнутри» его индивидуальной ситуации. В таком смысле, с одной стороны, безусловно, транспортное средство с автопилотом на ИИ объективно осуществляет действия «на практике» (движется). Однако правомерно ли по внешним параметрам приписывать искусственному агенту формирование «опыта» вождения? Если такая система анализирует данные о своем перемещении, включает в выборку новые ситуации и на основе этого оптимизирует свою работу, то, согласно известным аргументам «китайской комнаты» и «философского зомби» [Георгиу Т.С., 2022, с. 37–38, 40–41], это вовсе не означает, что машина понимает и переживает, т.е. имеет полноценный опыт. С другой стороны, практика в широком смысле противопоставляется теории, опыт же может затрагивать не только физическую реальность, но и пространство воображения. Например, говорят об опыте эстетического переживания или о получении навыков решения математических задач, которые могут иметь место целиком «в голове», ничего не преобразуя «на практике». В связи с этим некоторые исследователи полагают, что «машина может обладать умениями, но не может обладать знанием, и это хорошая причина считать, что именно здесь и кроется отличительное метафизическое свойство человеческого разума» [Дёмин Т.С., Фролов К.Г., 2023, с. 106]. Если же сопоставлять искусственный интеллект с естественным не только человеческим, то можно заметить, что животные тоже не обладают знанием (по крайней мере, в смысле отрефлексированной информации), но очевидно получают опыт и научаются, в отличие от машин, нечто при этом *испытывая*. Таким образом, на наш взгляд, переживание одушевленного существа свидетельствует об опыте (хотя может не проявляться во внешней реальности), в то время как практическим в широком смысле можно назвать лишь действие, проявления которого можно объективно зарегистрировать по материальным коррелятам.

В узком (аристотелевском) значении «практическое действие» как *поступок* или *социально-политический выбор* характерно только для взрослого человека, этического существа. В этом смысле понятие «праксиса» неприменимо ни к жизни других живых существ, ни к действиям искусственных агентов, если у них не развиты самосознание и совесть. На данный момент системы «узкого» ИИ не способны совершать экзистенциальный и этический выбор (в т.ч. по причине отсутствия такого иррационального компонента, как воля), однако логический выбор они могут осуществлять вполне эффективно (хотя и при условии заданности кри-

териев отбора) [Биричева Е.В., 2020b, с. 93]. Скажем, навигационная система просчитывает различные возможные маршруты, предлагая один в качестве предпочтительного, поскольку ей задан критерий минимального времени, затрачиваемого на дорогу. В случае человекоразмерных задач «узкий» ИИ способен предложить набор вероятных решений, дальнейший выбор из которых (или альтернативный выбор вопреки логически просчитанному) должен осуществляться на уровне *полноты экзистеции*. Когда молодой человек использует «умный поиск» для подбора невесты на основании информации из профилей девушек, экзистенциальный выбор он совершает все равно «на свой страх и риск» по итогам реального знакомства с претендентками, т.е. на основании опыта переживания личного контакта. Поступок, в отличие от просто действия, осуществляется *перед лицом Другого*: во-первых, себя как другого, мыслящего, смотрящего «со стороны» на свое «я» (т.е. имеющего самосознание); во-вторых, в свете понимания самостоятельности других и наличия у каждого человека «внутреннего» мира, неприступного для «я» (т.е. в свете корреляции с другими посредством совести). Этический выбор также нагружен конкретикой обстоятельств, которая к тому же парадоксальным образом должна быть осмыслена в горизонте мира в целом со всем «возможным» и «невероятным». Поэтому его невозможно перевести в плоскость логических вычислений, всегда ограниченных конечной выборкой и находящихся «по ту сторону» универсальных критериев. Ниже еще вернемся к данным аспектам и подробнее их поясним на конкретных примерах, здесь же подытожим: «узкий/слабый» ИИ в отсутствие воли, самосознания и совести не выходит на уровень этического и экзистенциального выбора (т.е. «практического действия» в смысле «поступка»); его использование возможно лишь в качестве инструмента, применение же во благо или во зло оказывается целиком в области ответственности людей (создателей, обладателей и пользователей таких систем) [Лаптев В.А., 2019, с. 92, 99].

2. Понятие опыта

Исследование проблем опыта имеет огромную философскую традицию от платоновского «припоминания» эйдосов до попыток спекулятивных реалистов выйти из «корреляционного круга». При этом нельзя сказать, что за 2500 лет значение самого этого понятия прояснилось. Как пишет Х.-Г. Гадамер, «понятие опыта относится…, — как бы парадоксально это ни звучало, — к числу наименее ясных понятий, какими мы располагаем» [Гадамер Х.-Г., 1988, с. 409]. Разработчик философской герменевтики также воздерживается от однозначной формулировки, тем не менее выделяя такие свойства опыта, как его телеологичность [Гадамер Х.-Г., 1988, с. 412], накопляемость в повторении и историчность [Гадамер Х.-Г., 1988, с. 413–415], «негативность» в плане снятия ложных представлений [Гадамер Х.-Г., 1988, с. 416–418], участность и открытость новому того, кого называют опытным [Гадамер Х.-Г., 1988, с. 418–421]. Безусловно, в литературе масса дефиниций, большинство из которых предлагают понимать под опытом совокупность знаний, умений, навыков, полученных субъектом в процессе взаимодействия с реальностью, а также в ходе внутренних переживаний, т.е. к опыту могут относить как чувственные восприятия, так и психическую деятельность человека (см., напр., [Лекторский В.А., 2010]). Однако возникают вопросы: действительно ли опытом являются «знания, умения, навыки», все ли переживания становятся опытом, остается ли опытом забытое? Кроме того, могут ли получать опыт не только взрослые люди, но и «неразумные» дети, животные, растения и, возможно, машины?

Для прояснения этих моментов обратимся к этимологии. Др.-греч. ***ἐμπειρία*** переводится как «опыт, практика, умение»; лат. ***experiri***, «пробовать, испытывать» составляется из приставки ***ex-***, «вне», и корня, означающего «изведывать» (от праиндоевроп. ****per-***, «вести, проводить»); в русском языке ***опыт*** происходит от праслав. ****pytati*** – «спрашивать, мучить, пробовать». Так, наиболее общими аспектами опыта, выражаемыми однокоренными словами, являются «попытки» (нечто опробовать, в чем-то убедиться), «испытывание» и «выпытывание» [Радеев А.Е., 2016, с. 58–60]. Эти моменты предполагают *активную позицию* и движение *воли*: решимость на пробу и моделирование способов ее осуществления; испытание (на себе) как «участное переживание» (т.е. «впускание» в себя — не пассивное состояние, а акт, требующий усилий принятия и внимания); постановка

вопроса, конструирование условий для переживания ответа на запрос при выпытывании. Так, получение любого опыта, запланированного или неожиданного (при внезапной встрече с новым), возможно только для *самостоятельных* акторов, способных хотеть, желать, стремиться, проявлять волю *сами*, без прямого принуждения извне, а также действовать вопреки принуждению.

К обозначенным трем составляющим необходимо добавить и четвертый обязательный аспект, благодаря которому попытки, испытывание и выпытывание становятся опытом, — это *запоминание* и, соответственно, актуализация в нужный момент. Если индивид получил из действительности представление о чем-то или развил какое-то умение, но забыл и не смог воспроизвести, то вряд ли в этом случае можно говорить о том, что у него есть опыт в данной сфере. К примеру, человек в юности учился играть на гитаре, но долгие годы не брал инструмент в руки. Ему дадут гитару и скажут: «Играй», — он сходу не сможет, в лучшем случае вспомнит общие моменты (что нужно определенным образом ставить пальцы одной руки, прижимая струны к ладам, а второй «бить» по струнам, извлекая нужные звуки, и т.д.). Однако он может даже не вспомнить аккорды, а для того, чтобы сыграть, придется заново натренировать пальцы, скоординировать движения и слух. Безусловно, не все запомненное можно назвать опытом: та же память о прошедшем обучении не обязательно означает возможность актуализации бывшего опыта. В целом, имение знаний не гарантирует практического освоения того, о чем получены теоретические представления. Скажем, кто-то может глубоко изучить теорию катания на сноуборде, но на деле он окажется неопытным, пока не встанет на доску и не научится телом управлять движением на ней. Возможны и знания о том, о чем принципиально не может быть опыта (например, о фантастических мирах, религиозных сущностях, метафизических основаниях).

Сама *память* живых существ является непростым и не вполне изученным феноменом. По крайней мере, реальная память далека от того, что этим словом называют применительно к носителям информации в машинах. В «Исповеди» Св. Августина есть примечательный отрывок, в котором мыслитель глубоко и многогранно исследует специфику памяти [Августин, 1991, с. 242–262], показывая, что память каждого существа кажется хранилищем содержаний лишь по видимости. Скорее она представляет собой набор зафиксированных «ходов» доступа к представлениям, о которых индивид убедился в прошлом (любым способом — от личного контакта с предметом и физического переживания до абстрактного логического доказательства и веры). Привязка содержаний памяти к материальным носителям, например, нейронам, означала бы крайне ограниченный объем памяти, которая бы терялась при отмирании клеток и в которой при нехватке места информация стиралось бы без остатка. Тем не менее мы, даже если забываем, способны восстановить представление, если нам напомнят или мы сами совершим усилие. Удивляясь способностям памяти, Августин пишет: «Велика она, эта сила памяти, Господи, слишком велика!.. И, однако, это сила моего ума, она свойственна моей природе, но я сам не могу полностью вместить себя. Ум тесен, чтобы овладеть собой же» [Августин, 1991, с. 245]. Так, уникальная память скорее не склад представлений в голове, а умение выстроить свою систему обращения к хранимому, ориентироваться и находить нужное.

Если проводить компьютерные аналогии, то память живых существ больше похожа на поисковую систему, обращающуюся к общему серверу (или облачному хранилищу), чем на носитель информации. Этот «общий сервер» (или «облачное хранилище») — пространство Воображаемого. Многое указывает на изоморфность с воображаемыми структурами образов, к которым обращается память. Вспоминая о своих ощущениях, настроениях, эмоциях, мы не переживаем их заново, а как бы дистанцировано рассматриваем представления о них [Августин, 1991, с. 244, 249–252]. Эмпириокритицист Э. Мах замечает, что «невозможно провести абсолютно резкой границы между воспоминанием и фантазией… всякое припоминание есть “смесь действительности с фантазией”. С другой стороны, в фантастических представлениях большею частью можно доказать присутствие элементов воспоминания» [Мах Э., 2003, с. 166]. Соответственно, сами представления как виды, формы, образы содержатся в Воображаемом и являются материалом для мысленного созерцания как то-

го, с чем индивид имел дело, так и того, что можно представить независимо от его реальной жизненной истории (несуществующие вещи или связи, предположения о будущем и т.д.). Припоминание требует усилия и некоторого поиска в этом воображаемом пространстве [Августин, 1991, с. 243, 247], которое не находится в голове индивида, но в которое, если воспользоваться метафорой М.К. Мамардашвили и А.М. Пятигорского, как в «сферу», индивид может входить или «впадать».

Таким образом, на основе контакта с действительностью и переживания психических и мыслительных процессов индивиды формируют ассоциации. Устойчивые связи закрепляются благодаря либо сильному единичному впечатлению, либо многократному повторению. Если эти связи периодически актуализируются, т.е. воспроизводятся ходы мысли, представления, движения тела, то можно говорить о формировании опыта. Если определять просто, то опыт — не что иное как *постоянно обновляемое* ***освоение*** *индивида в общих пространствах*; если развернуто, то это *актуализируемые в настоящем, закрепленные и пересматриваемые (уточняемые, перезакрепляемые)* ***связи*** *реально имевших место в индивидуальном прошлом (1) проявлений активности (в т.ч. внимания), (2) отклика на них (внутреннего и/или окружающих предметов, существ, явлений) и (3) оценки состояний, при этом переживаемых*. Так, становится очевидно, что опыт каждого уникален, поэтому его не передать, однако, показав нечто (в т.ч. словами), можно приобщить к своему опыту; существа способны иметь сходный опыт — одинаковый по форме связей, — но различающийся «локацией» (индивидуальным переживанием тех же ощущений, состояний, мыслей). Тогда опыт является не знанием, умением, переживанием, но более простым: *открытым примериванием себя к миру и собственному внутримирному бытию*. Несомненно, в результате опыта могут быть получены знания, умения и т.д., однако отождествлять их с опытом представляется не совсем корректным, поскольку они являются продуктом синтеза опыта и внеопытного (критериев успешности и «условий понятности» осваиваемого).

3. Опыт и целеполагание

Интересно в этом ракурсе рассмотреть целеполагание, которым нередко наделяется феномен опыта, и по отношению к которому уже имеющийся опыт часто кажется материалом. С одной стороны, перед живыми организмами не всегда стоит задача получения опыта. Животное может быть занято каким-то своим вполне привычным делом, однако изменение окружающих обстоятельств или появление других существ могут обратить на себя внимание и подарить новый опыт. В этом случае, если уместно говорить о цели извлечения опыта, то она в качестве возможности неотъемлемо *встроена* в самостоятельных акторов, т.е. если такая цель и есть, то она безусловна (сопутствует жизни независимо от условий). Другими словами, существа живут не ради получения максимума опыта, однако возможность его получать всегда открыта, поскольку активность в реальности автоматически означает для них постоянное закрепление уже освоенных практик и освоение незнакомого. К примеру, ребенок бежал, играл и вдруг начал мотать головой или с удивлением «изучать» какой-то совершенно, казалось бы, привычный предмет. Это происходит «низачем» и «нипочему»; мы огрубим феномен, если припишем ему задачи «познания» окружающих вещей или своего тела. По крайней мере, сделать это можно лишь задним числом, сам же опыт начинается *вдруг*, а цель его получения *всегда уже* поставлена. С этим неотъемлемо связано *обращение внимания*: в одних и тех же условиях разные существа даже одного вида могут придать или не придать значение тем или иным предметам, сигналам, событиям (одного кота заинтересовала новая вещь, к которой у другого может не проявиться ни малейшего интереса). Следовательно, в зависимости от *интереса*, если обстоятельства оставляют выбор, и живые существа могут в пределе *выбирать*, получать конкретный опыт или нет. Для человека в силу развитости воображения и осознания своей конечности это особенно заметно при расстановке приоритетов для профессиональных занятий, хобби и другого времяпрепровождения, когда он отказывает себе в получении одного вида опыта в пользу другого. На самом же деле у всех существ имеет место пе-

реключение внимания и иерархизация направлений получения опыта.

С другой стороны, некоторые разработчики ИИС точно подмечают *парадокс цели*: «Процесс достижения цели (как и процесс решения задачи), являясь целенаправленной деятельностью, в то же время принципиально не содержит знание о том, как, где и когда можно достичь цели» [Витяев Е.Е. и др., 2020, с. 10]. Если цель осуществляется впервые, то даже «знание желания», например, утолить жажду, не обязательно будет содержать знание о том, чем и как его удовлетворить, что для этого нужно сделать [Витяев Е.Е. и др., 2020, с. 9]. Этот пример показывает, что, во-первых, *цели не возникают из предшествующего опыта* самого по себе (хотя, безусловно, могут корректироваться и уточняться в опоре на него). Во-вторых, для того, чтобы опыт стал таковым, *до него (априори)* актор должен располагать *критерием его успешности* — чтобы, пробуя (особенно впервые), понять, получилось или нет. Таким образом, опыт тесно связан с лежащими вне его априорными формами, помогающими различать «то» и «не то»: нужное/ненужное, годное/негодное, достаточное/недостаточное и т.п. Данные критерии берутся *не из простого восприятия качеств предметов* (цветов, звуков, запахов, вкусов и т.д.), поскольку одни и те же качества способны на различных предметах или от одного объекта в разных условиях оказываться маркерами как опасного, так и жизненно важного (например, окраска ядовитых насекомых совпадает с цветом питательных ягод). Важен максимально широкий контекст других ассоциаций, в которые постоянно вписываются связи нового единичного, уточняющие предыдущие, уже сложившиеся паттерны. Хотя это может показаться противоречащим учению И. Канта, здесь очевидным становится, что априорные критерии используются всеми известными живыми существами, в т.ч. растениями, иначе они не имели бы возможности не только ориентироваться в текущих условиях, но и приспосабливаться к изменениям. Кроме того, не будь у биологических существ внеопытных ориентиров (к которым в т.ч. память привязывает встреченное в реальности), не возникало бы никакого обобщения для распознания «знакомого» и «незнакомого». В реальности происходят встречи всегда с рядами единичных уникальных предметов и явлений, и в противном случае существа реагировали бы на все каждый раз как на новое, терялись бы перед тем, с какими подобными условиями, предметами, веществами и т.д. они уже сталкивались.

4. Проблема критериев

Обозначенные моменты выводят к проблемам чисто индуктивного пути познания, квалиа и критериев истинности [Георгиу Т.С., 2022, с. 38–41; Новиков Н.Б., 2023, с. 114–116; Поппер К., 1983, с. 254–264]. В них сейчас особенно заметно упирается развитие «узкого/слабого» ИИ: на основе лишь полученных единичностей (фактов, описаний случаев, изображений и т.д.) без критериев различения можно прийти к неправильным решениям, составить ложные представления, испортить уже успешно получающееся. Это иллюстрируют, например, случаи с «предвзятостью» судебных экспертных систем [Бахтеев Д.В., Тарасова Л.В., 2020]; «нейро-инцест» генеративных ИИС, которые начинают обучаться на обильном интернет-контенте от других нейросетей (не только на оригинальных изображениях/видео от пользователей), соответственно, снижая качество новых работ [Shumailov I. et al., 2023]; недавние феномены «обучения» чат-ботов пользователями ненормативной лексике, расизму, буллингу; ложные ответы ChatGPT-4 на запросы о фактической информации (исторический событиях, личностях) и т.д. Такие проблемы связаны прежде всего с тем, что даже огромный массив данных, на которых обучается нейросеть, конечен и, следовательно, не может вместить в себя абсолютно все случаи, в т.ч. те, каких еще не было (классическая проблема индуктивного пути познания, маркируемая метафорой «черного лебедя»). Кроме того, если к достоверной информации или репрезентативной выборке примешивается по тем или иным причинам недостоверное или лишнее, то система, не обладая критерием отбора, будет ухудшать результаты своей деятельности (решения, изображения, текст и т.д.). Соответственно, если даже ИИС способна максимально качественно обобщать и систематизировать имеющееся, она не умеет отсекать ненужное и самостоятельно выходить за границы данного. Последние — фундаментальные свойства живых существ; «творчество» же современных «узких» ИИС ограничивается комбинированием известных им блоков ин-

формации без различения продуктивных и непродуктивных элементов.

Парадокс регулировки таких систем заключается в том, что, с одной стороны, им необходимы критерии отбора правильного/неправильного, однако невозможно сформулировать *универсальные критерии* из самих данных (множества которых конечны), а с другой — если такие критерии навязаны извне (запрограммированы человеком), то это однозначно захлопывает ИИС дверь к самостоятельности и подлинной эвристичности. Безусловно, существа также используют для обобщения опыта индукцию, обобщая, собственно, только то, с чем имели дело, и ни одно из них не застраховано от ошибок. Однако различие заключается в том, что примеривающийся к миру живой организм *сам* способен распознать неудачу и исправиться, учесть в т.ч. негативный опыт. Естественно, совершается ошибок достаточно много, неуспех приспособления приводит к вымиранию целых видов, исторически люди в ходе познания нередко заблуждались, платя за это здоровьем, жизнью и т.д. Как замечает Э. Мах, «познание и заблуждение вытекают из одних и тех же психических источников; только успех может разделить их. Ясно распознанное заблуждение является в качестве корректива в такой же мере элементом, содействующим познанию, как и положительное знание» [Мах Э., 2003, с. 134–135]. Спасает от ошибок целенаправленное многократное повторение и рассмотрение опыта [Мах Э., 2003, с. 142], фундаментом которого является внимание [Мах Э., 2003, с. 135]. Существа как бы априори «знают» о том, что могут ошибиться и склонны перепроверять; наконец, они наблюдают за сородичами и способны в т.ч. учитывать их опыт (как позитивный, так и негативный). У ИИС же на данном этапе их разработки внеопытных критериев нет, и сами они не расширяют себе поле возможного опыта за пределы запланированных задач. Если нейросеть плохо «обучилась», человек возвращает ее на «дообучение», изменяя или расширяя выборку (например, сам обращает внимание генеративной сети на прорисовку зубов, конечностей, пальцев и т.д., с которыми у таких ИИС были поначалу серьезные проблемы).

Препятствие в этом плане также в том, что *ИИС имеет дело не с миром в целом*, а с блоками информации, будь то набор изображений, случаев словоупотребления или количественных данных о звуковых колебаниях, температуре среды, длинах волн попадающего на сенсоры света и т.д. Классическая «проблема квалиа», представленная, например, в таких известных мысленных экспериментах, как «комната Мэри», «летучая мышь» и «философский зомби» [Георгиу Т.С., 2022, с. 38–41], в этом контексте поворачивается еще одной стороной. Дело не только в непереводимости количественного в качественное и теоретического в практическое, но и в «неумении» современных нейросетей обращать внимание. Наличие сенсоров, считывающих те или иные показания, и элементов, обрабатывающих эту информацию, самих по себе недостаточно для переживания и получения опыта. Требуется отмеченное выше волевое начало, которое, само будучи бессодержательным «переключателем», обращает внимание (выделяет) что-то в действительности благодаря уже до всякого опыта имеющимся критериям. Существо не является сплошным чувствилищем, оно использует тело и априорные формы как инструменты ориентирования в мире. Хотя тело постоянно получает сигналы с рецепторов (внешних и внутренних), если бы не было некого начала, регулирующего восприятие, мы бы сошли с ума, выражаясь словами Ж. Делеза, от «хаоса» непрерывного чувствования. На самом же деле некоторые даже интенсивные восприятия мы можем не замечать, к примеру, долго не чувствуем голода или забываем о боли в теле, если полностью захвачены каким-то интересным делом.

Заключение

Таким образом, что все это означает для разработки «общего/сильного» ИИ и совершенствования «узких/слабых»? Получать опыт означает постоянно выходить за границы данного, открыто осваиваться, соотнося встречаемое (физически воспринятое и/или психически пережитое) с априорными формами-ориентирами. Сенсоров недостаточно для получения опыта; даже если бы системы воспринимали качества, помимо практической, нужна теоретическая составляющая, а также воля, обращающая внимание и направляющая усилия к целям. В связи с этим возникла идея создавать ИИС из нескольких нейросетей, комбинирующих индуктивные и дедуктивные методы и контролирующих друг друга (например, генеративно-состязательные

модели (см., напр., [Аверченков А.В. и др., 2020])). Для решения проблемы «черного ящика» и повышения доверия к ИИС аналогичным образом разработали концепцию т.н. «объяснимого ИИ» (англ. ***explainable artificial intelligence, XAI***), в котором одна нейросеть анализирует деятельность другой и объясняет на естественном языке выбор решения первой (см., напр., [Srihari S.N., 2022]). Все подобные модели, безусловно, являются перспективными в плане совершенствования «узких/слабых» ИИС, однако простого комбинирования функций и усложнения архитектур нейросетей недостаточно для создания «общего/сильного» ИИ. Поскольку опыт — не хранилище информации, а закрепленные связи уровня информации, уровня практического действия и критериев успешности, необходим еще один момент, а именно их органичное соединение. Последнее оказывается слишком простым, чтобы его запрограммировать: в живом оно предшествует разделению на эти пространства (общих критериев-ориентиров и индивидуально испытанного), поэтому сложением отдельно смоделированных теоретической, дедуктивной способности и практических умений, переживаний его вряд ли можно получить.

При этом надежды подают модели, опирающиеся на эволюционную парадигму и идеи самоорганизации (см., напр., [Сергеев С.Ф., 2023]). В этом случае предполагается, что в своем становлении ИИ должен пройти некоторые стадии развития, подобно существам в живой природе. К примеру, чувственно-эмоциональная составляющая появляется у живых существ в качестве усовершенствования механизма приспособления к сложным, изменчивым условиям окружающей среды, помогая более эффективно в них ориентироваться. На определенном уровне развития можно было бы ожидать, что система породит в себе нечто вроде приложения, отвечающего за чувственное восприятие и эмоциональную сферу. Интересно заметить в этом контексте, что логическая способность развивается в полной мере только у человека в процессе взросления (и с этой точки зрения было странно начинать разработку ИИ «с конца» эволюционной цепочки, которую прошел естественный интеллект). Несмотря на концептуальную целостность и перспективность, в рамках этого направления неясным остается как вложить хотя бы в самую простейшую ИИС базовую способность любого живого существа — волю. Без нее система не будет самостоятельно ни к чему стремиться и вряд ли пожелает самосовершенствоваться, развивать в себе эмоции и логическое мышление. В связи с этим, пожалуй, ключевой проблемой создания «сильного» ИИ следует считать моделирование воли; решение же вопросов «квалиа», связи «сознание-тело», реализации подлинных коммуникативных, когнитивных и креативных способностей и т.д., как представляется, в этом отношении носят вторичный характер.

Список литературы

Августин. Исповедь / пер. с лат. М.К. Сергеенко. М.: Ренессанс, ИВО–СиД, 1991. 488 с.

Аверченков А.В., Андросов А.А., Малахов Ю.А. Анализ и применение генеративно-состязательных сетей для получения изображений высокого качества // Эргодизайн. 2020. № 4. С. 167–176. DOI: https://doi.org/10.30987/2658-4026-2020-4-167-176

Арендт Х. Vita activa, или О деятельной жизни / пер. с нем. и англ. В.В. Бибихина. СПб.: Алетейя, 2000. 437 с.

Бахтеев Д.В., Тарасова Л.В. Применение искусственного интеллекта в деятельности арбитражных судов РФ: перспективные направления и проблемы // Вестник Костромского государственного университета. 2020. Т. 26, № 4. С. 249–254. DOI: https://doi.org/10.34216/1998-0817-2020-26-4-249-254

Биричева Е.В. Воля в концепции сильного искусственного интеллекта // Гуманитарное знание и искусственный интеллект: стратегии и инновации: 4-й молодежный конвент УрФУ: материалы международ. конф. (Екатеринбург, 26 марта 2020 г.). Екатеринбург: Изд-во Урал. ун-та, 2020. С. 1009–1012.

Биричева Е.В. Проблематизация информации как важнейшего понятия современности // Манускрипт. 2020. Т. 13, вып. 2. С. 90–95. DOI: https://doi.org/10.30853/manuscript.2020.2.15

Биричева Е.В., Стерхов Е.В. Парадоксальность творческой деятельности и интеллектуальные системы: поиск альтернатив информационному подходу // Российский гуманитарный журнал. 2019. Т. 8, № 6. С. 390–402. DOI: https://doi.org/10.15643/libartrus-2019.6.2

Вирно П. Грамматика множества: к анализу форм современной жизни / пер. с ит. А.Г. Петровой. М.: Ад Маргинем Пресс, 2013. 176 с.

Витяев Е.Е., Гончаров С.С., Свириденко Д.И. О задачном подходе в искусственном интеллекте и когнитивных науках // Сибирский философский

журнал. 2020. Т. 18, № 2. С. 5–29. DOI: https://doi.org/10.25205/2541-7517-2020-18-2-5-29

Гадамер Х.-Г. Истина и метод: Основы филос. герменевтики / пер. с нем. М.: Прогресс, 1988. 704 с.

Георгиу Т.С. Решение проблемы «сознание–тело» и искусственный интеллект // Вестник Тверского государственного университета. Серия: Философия. 2022. № 4(62). С. 32–45. DOI: https://doi.org/10.26456/vtphilos/2022.4.032

Демин Т.С., Фролов К.Г. Знание-что, знание-как, сознание и искусственный интеллект // Омский научный вестник. Серия: Общество. История. Современность. 2023. Т. 8, № 1. С. 102–109. DOI: https://doi.org/10.25206/2542-0488-2023-8-1-102-109

Лаптев В.А. Понятие искусственного интеллекта и юридическая ответственность за его работу // Право. Журнал Высшей школы экономики. 2019. № 2. С. 79–102. DOI: https://doi.org/10.17323/2072-8166.2019.2.79.102

Лекторский В.А. Опыт // Новая философская энциклопедия (ИФ РАН): в 4 т. / под науч. ред. В.С. Степина и др. М.: Мысль, 2010. Т. 3. С. 158–159.

Луньков А.С. Проблема стандартизации этики искусственного интеллекта // Искусственные общества. 2020. Т. 15, № 2. URL: https://artsoc.jes.su/s207751800009044-7-1/ (дата обращения: 19.02.2024). DOI: https://doi.org/10.18254/s207751800009044-7

Мах Э. Познание и заблуждение: Очерки по психологии исследования / пер. с нем. Г.А. Котляра. М.: БИНОМ. Лаборатория знаний, 2003. 456 с.

Новиков Н.Б. Искусственный интеллект должен научиться делать случайные открытия // Аллея науки. 2023. Т. 1, № 4(79). С. 108–138.

Поппер К. Логика и рост научного знания: избранные работы / пер. с англ. Л.В. Блинникова и др. М.: Прогресс, 1983. 606 с.

Радеев А.Е. Что же имеется в виду под опытом, когда мы называем его эстетическим? // Вестник Санкт-Петербургского университета. Серия 17: Философия. Конфликтология. Культурология. Религиоведение. 2016. Вып. 4. С. 53–62. DOI: https://doi.org/10.21638/11701/spbu17.2016.406

Сергеев С.Ф. Компоненты техноразума: искусственное сознание // Теоретическая и экспериментальная психология. 2023. Т. 16, № 1. С. 5–18. DOI: https://doi.org/10.11621/tep-23-01

Турко Д.С. Феноменальный минимализм в онтологии самости // Антиномии. 2021. Т. 21, вып. 4. С. 7–30. DOI: https://doi.org/10.17506/26867206_2021_21_4_7

Shumailov I., Shumaylov Z., Zhao Y., Gal Y., Papernot N., Anderson R. The Curse of Recursion: Training on Generated Data Makes Models Forget / ArXiv; Cornell University. 2023. URL: https://arxiv.org/pdf/2305.17493.pdf (accessed: 19.02.2024). DOI: https://doi.org/10.48550/arXiv.2305.17493

Srihari S.N. Explainable Artificial Intelligence: An Overview. 2022. URL: https://cedar.buffalo.edu/~srihari/papers/XAI-Overview.pdf (accessed: 20.02.2024).

References

Arendt, H. (2000). *Vita activa, ili O deyatel'noy zhizni* [The human condition]. St. Petersburg: Aleteiya Publ., 437 p.

Augustine (1991). *Ispoved'* [Confessions]. Moscow: Renessans Publ., IVO–SiD Publ., 488 p.

Averchenkov, A.V., Androsov, A.A. and Malakhov, Yu.A. (2020). [Analysis and application of generative-adversarial networks for producing high quality images]. *Ergodizayn* [Ergodesign]. No. 4, pp. 167–176. DOI: https://doi.org/10.30987/2658-4026-2020-4-167-176

Bakhteev, D.V. and Tarasova, L.V. (2020). [The application of artificial intelligence in commercial courts of the Russian Federation: perspectives and issues]. *Vestnik Kostromskogo gosudarstvennogo universiteta* [Vestnik of Kostroma State University]. Vol. 26, no. 4, pp. 249–254. DOI: https://doi.org/10.34216/1998-0817-2020-26-4-249-254

Biricheva, E.V. (2020). [Problematization of «information» as the most important contemporary notion]. *Manuskript* [Manuscript]. Vol. 13, iss. 2, pp. 90–95. DOI: https://doi.org/10.30853/manuscript.2020.2.15

Biricheva, E.V. (2020). [Will in the strong artificial intelligence concept]. *Gumanitarnoe znanie i iskusstvennyy intellekt: strategii i innovatsii*: *4-y molodezhnyy konvent UrFU: materialy mezhdunarod. konf. (Ekaterinburg, 26 marta 2020 g.)* [Humanities and Artificial Intelligence: Strategies and Innovations 4th Youth Convention: Proceedings of the International Conference (Ekaterinburg, March 26, 2020)]. Ekaterinburg: UrFU Publ., pp. 1009–1012.

Biricheva, E.V. and Sterkhov, E.V. (2019). [Intelligent systems and the paradoxicalness of creative activity: In search for alternatives to the informational approach]. *Rossiiskiy gumanitarnyy zhurnal* [Liberal Arts in Russia]. Vol. 8, no. 6, pp. 390–402. DOI: https://doi.org/10.15643/libartrus-2019.6.2

Demin, T.S. and Frolov, K.G. (2023). [Knowledge-that, knowledge-how, consciousness and artificial in-

telligence]. *Omskiy nauchnyy vestnik. Seriya: Obschestvo. Istoriya. Sovremennost'* [Omsk Scientific Bulletin. Series: Society. History. Modernity]. Vol. 8, no. 1, pp. 102–109. DOI: https://doi.org/10.25206/2542-0488-2023-8-1-102-109

Gadamer, H.-G. (1988). *Istina i metod: Osnovy filosofskoy germenevtiki* [Truth and method]. Moscow: Progress Publ., 704 p.

Georgiou, T.S. (2022). [Solution of the problem «mind–body» and artificial intelligence]. *Vestnik Tverskogo gosudarstvennogo universiteta. Seriia: Filosofiia* [Herald of Tver State University. Series: Philosophy]. No. 4(62), pp. 32–45. DOI: https://doi.org/10.26456/vtphilos/2022.4.032

Laptev, V.A. (2019). [Artificial intelligence and liability for its work]. *Pravo. Zhurnal Vysshey shkoly ekonomiki* [Law. Journal of the Higher School of Economics]. No. 2, pp. 79–102. DOI: https://doi.org/10.17323/2072-8166.2019.2.79.102

Lektorsky, V.A. (2010). [Experience]. *Novaya filosofskaya entsiklopediya (IF RAN): v 4 t., pod nauch. red. V.S. Stepina i dr.* [V.S. Stepina et al. (eds.) New philosophical encyclopedia (IF RAS): in 4 vols]. Moscow: Mysl' Publ., vol. 3, pp. 158–159.

Lun'kov, A.S. (2020). [The problem of standardizing the ethics of artificial intelligence]. *Iskusstvennye obschestva* [Artificial Societies]. Vol. 15, no. 2. Available at: https://artsoc.jes.su/s207751800009044-7-1/ (accessed 19.02.2024). DOI: https://doi.org/10.18254/s207751800009044-7

Mach, E. (2003). *Poznanie i zabluzhdenie: Ocherki po psikhologii issledovaniya* [Knowledge and error: Sketches on the psychology of enquiry]. Moscow: BINOM. Laboratoriya Znaniy Publ., 456 p.

Novikov, N.B. (2023). [Artificial intelligence has to learn to make random discoveries]. *Alleya nauki* [Science Alley]. Vol. 1, no. 4(79), pp. 108–138.

Popper, K. (1983). *Logika i rost nauchnogo znaniia: Izbrannye raboty* [The logic of scientific discovery: Selected works]. Moscow: Progress Publ., 606 p.

Radeev, A.E. (2016). [What is meant by experience when we call it aestetic?] *Vestnik Sankt-Peterburgskogo universiteta. Seriya 17: Filosofiya. Konfliktologiya. Kul'turologiya. Religiovedeniye* [Vestnik of Saint Petersburg University. Series 17: Philosophy. Conflict Studies. Culture Studies. Religious Studies]. Iss. 4, pp. 53–62. DOI: https://doi.org/10.21638/11701/spbu17.2016.406

Sergeev, S.F. (2023). [Techno-mind components: Artificial consciousness]. *Teoreticheskaya i eksperimental'naya psikhologiya* [Theoretical and Experimental Psychology]. Vol. 16, no. 1, pp. 5–18. DOI: https://doi.org/10.11621/tep-23-01

Shumaylov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N. and Anderson, R. (2023). *The curse of recursion: Training on generated data makes models forget*. ArXiv, Cornell University. Available at: https://arxiv.org/pdf/2305.17493.pdf (accessed 19.02.2024). DOI: https://doi.org/10.48550/arXiv.2305.17493

Srihari, S.N. (2022). *Explainable artificial intelligence: an overview*. Available at: https://cedar.buffalo.edu/~srihari/papers/XAI-Overview.pdf (accessed 20.02.2024).

Turko, D.S. (2021). [Phenomenal minimalist ontology of the self]. *Antinomii* [Antinomies]. Vol. 21, iss. 4, pp. 7–30. DOI: https://doi.org/10.17506/26867206_2021_21_4_7

Virno, P. (2013). *Grammatika mnozhestva: k analizu form sovremennoy zhizni* [A grammar of the multitude: for an analysis of contemporary forms of life]. Moscow: Ad Marginem Publ., 176 p.

Vitiaev, E.E., Goncharov, S.S. and Sviridenko, D.I. (2020). [On the task approach in artificial intelligence and cognitive sciences]. *Sibirskiy filosofskiy zhurnal* [Siberian Journal of Philosophy]. Vol. 18, no. 2, pp. 5–29. DOI: https://doi.org/10.25205/2541-7517-2020-18-2-5-29

Об авторе

Биричева Екатерина Вячеславовна
кандидат философских наук,
научный сотрудник

Институт философии и права УрО РАН,
620108, Екатеринбург, ул. С. Ковалевской, 16;
e-mail: ek.v.bir@gmail.com
ResearcherID: F-2980-2016

About the author

Ekaterina V. Biricheva
Candidate of Philosophy, Researcher

Institute for Philosophy and Law, Ural Branch,
Russian Academy of Sciences,
16, S. Kowalevskaya st., Ekaterinburg, 620108, Russia;
e-mail: ek.v.bir@gmail.com
ResearcherID: F-2980-2016