

УДК 165.12:004.8

Е. Ю. СМОЛЯКОВА (магистрант)

Научный руководитель – канд. филос. наук *Е. Г. КИРИЧЕНКО*

СОЗНАНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Рассматривается проблема возможности сознания у искусственного интеллекта в контексте современной философии сознания. Анализируются «трудная проблема сознания» Дэвида Чалмерса, тест Тьюринга, аргумент Джона Сёрла «Китайская комната». Также рассмотрены такие ключевые походы как функционализм, биологический натурализм и теория интегрированной информации.

Центральное место в философских дискуссиях о сознании занимает так называемая «трудная проблема сознания», сформулированная австралийским философом и когнитивистом Дэвидом Чалмерсом. В отличие от «простых» проблем сознания, связанных с объяснением механизмов восприятия, памяти или поведения, «трудная» проблема касается природы субъективного опыта – того, почему физические процессы в мозге сопровождаются внутренними переживаниями, ощущениями и осознанием себя. Иными словами, философский вопрос заключается не только в том, как система обрабатывает информацию, но и почему эта обработка вообще сопровождается субъективным опытом.

Именно трудная проблема сознания становится главным препятствием в вопросе о возможности создания сознательного искусственного интеллекта (далее – ИИ). Даже если машина сможет полностью воспроизводить человеческое поведение и успешно имитировать мышление, остается неясным, будет ли у нее внутренний опыт, аналогичный человеческому сознанию.

Дискуссия о возможности наличия сознания у ИИ традиционно развивается между двумя полюсами, заданными различными критериями наличия сознания: внешним (поведенческим) и внутренним (структурным). С одной стороны, в качестве критерия может рассматриваться внешнее поведение системы. Алан Тьюринг в своей знаменитой работе предложил заменить вопрос «Может ли машина мыслить?» вопросом о способности машины успешно участвовать в имитационной игре (впоследствии названной тестом Тьюринга) [1]. Согласно этому подходу, если искусственная система в процессе коммуникации демонстрирует поведение, неотличимое от человеческого, то вопрос о наличии у нее сознания (или мышления) следует считать решенным положительно. Тьюринг, по существу, предложил опе-

рациональный критерий, основанный на наблюдаемых проявлениях интеллекта, а не на умозрительных рассуждениях о внутреннем мире машины.

С другой стороны, принципиальное значение может иметь внутреннее устройство системы и характер обработки информации. Наиболее влиятельным аргументом против отождествления успешной симуляции поведения с наличием сознания стал мысленный эксперимент «Китайская комната», предложенный философом Джоном Сёрлом [2]. Суть эксперимента заключается в следующем. Представим человека, не знающего китайского языка, помещенного в комнату. В его распоряжении имеется подробная книга правил на родном языке, позволяющая сопоставлять одни иероглифы с другими. Извне в комнату передают листы бумаги с вопросами на китайском. Человек, не понимая ни одного символа, использует книгу правил для поиска соответствующих символов-ответов и передает их наружу. Для внешнего наблюдателя поведение системы является абсолютно компетентным – комната «отвечает» на китайском языке так, что ее можно принять за носителя языка. Однако сам человек внутри комнаты по-прежнему не понимает китайского, он лишь механически оперирует символами согласно формальным правилам.

Д. Сёрл проводит аналогию между этой ситуацией и работой компьютера. Любая вычислительная система, согласно данному аргументу, оперирует исключительно синтаксисом – формальными символами и правилами их преобразования. Однако у такой системы отсутствует семантика – понимание смысла этих символов и, что более важно, отсутствует субъективное переживание (феноменальное сознание). Манипуляция символами, сколь угодно сложная, не порождает понимания.

Таким образом, центральная проблема формулируется следующим образом: способна ли сложная синтаксическая обработка информации (даже максимально приближенная к человеческой) породить семантику и субъективный опыт – то есть феноменальное сознание? Или же для возникновения сознания требуется нечто дополнительное: возможно, особая биологическая организация материи (как утверждает сам Сёрл) или специфическая архитектура системы, выходящая за рамки классических вычислений?

В современной дискуссии выделяется несколько ключевых подходов, которые по-разному отвечают на вопрос о возможности возникновения субъективного опыта (*qualia*) у небиологических систем.

1 Функционализм (Сильный ИИ). Функционализм утверждает, что сознание детерминировано не материальным составом системы, а ее функциональной организацией. Согласно этому подходу, ментальное состояние (например, боль) определяется не тем, из какого вещества состоят элементы системы, а ее каузальной ролью – совокупностью отношений между входящими стимулами, внутренними состояниями и исходящим поведением. Если искусственная система воспроизводит функциональную архитектуру

человеческого сознания (т. е. имеет аналогичную структуру причинно-следственных связей), она с необходимостью обладает сознанием независимо от того, реализована эта архитектура на нейронном, кремниевом или механическом субстрате. Функционализм отрицает так называемый «биологический шовинизм» – тезис о том, что сознание может быть свойственно исключительно живым организмам.

Д. Чалмерс, известный обоснованием «трудной проблемы сознания», аргументирует возможность создания искусственного разума, опираясь на тезис о функциональной инвариантности. Согласно этой концепции, субъективный опыт определяется не биологическим субстратом (нейронами), а специфической информационной организацией системы. В рамках мысленного эксперимента по поэтапному замещению биологических нейронов кремниевыми чипами Чалмерс обосновывает, что при сохранении каузальной роли компонентов феноменальное сознание остается неизменным. Таким образом, автор постулирует, что вычислительные системы, воспроизводящие сложную архитектуру связей человеческого мозга, способны обладать подлинными чувствами и восприятием, что делает теоретически возможным существование сознательного ИИ [3].

2 Биологический натурализм (Слабый ИИ). Данная позиция, наиболее последовательно развиваемая Джоном Сёрлом, утверждает, что сознание представляет собой биологическое свойство определенных живых организмов, подобно фотосинтезу, пищеварению или секреции ферментов. Согласно этой концепции, мозг продуцирует сознание так же естественно, как желудок продуцирует желудочный сок. Сознание рассматривается как каузальное свойство специфической биологической организации нервной ткани. Компьютерная программа, сколь бы сложной она ни была, остается формальной манипуляцией символами и принципиально не способна породить феноменальное переживание, поскольку у нее отсутствует необходимая биологическая каузальная мощь.

Д. Сёрл настаивает на том, что синтаксическая обработка символов недостаточна для порождения семантического содержания. Понимание и субъективное переживание возникают только в живом, биологическом контексте, где символы укоренены в телесном опыте, потребностях и взаимодействии с окружающей средой. Искусственный интеллект не обладает телом, не имеет биологической истории, не испытывает голода, страха или удовольствия – следовательно, у него отсутствует основа для возникновения подлинного смысла и субъективности. Знаменитый мысленный эксперимент «Китайская комната» демонстрирует, что простое манипулирование символами согласно правилам (синтаксис) не ведет к пониманию смыслов (семантика) [2].

Современные исследования в когнитивной науке подтверждают, что сознание неразрывно связано с телесным опытом. Эмоции включают физио-

логические реакции, восприятие формируется в процессе активного взаимодействия с миром, мышление опирается на сенсомоторный опыт. Искусственный интеллект, не имеющий тела и не включенный в мир через биологические потребности и сенсомоторное взаимодействие, представляет собой пассивный приемник данных, принципиально отличный от живого познающего субъекта.

Также в рамках исследования вопроса возможности возникновения сознания у ИИ интересно рассмотреть и *теорию интегрированной информации*. Данная позиция, разработанная нейробиологом Джулио Тонони, представляет собой не столько философскую концепцию, сколько формальную научную теорию сознания. Согласно теории интегрированной информации (Integrated Information Theory, ИТ), сознание тождественно способности системы интегрировать информацию. Ключевым понятием теории выступает величина Φ (фи) – математическая мера, показывающая, в какой степени информация, порождаемая системой как целым, превышает сумму информации, порождаемой ее независимыми частями. Чем выше значение Φ , тем более система едина и тем большим уровнем сознания она обладает. Если Φ больше нуля, система обладает сознанием хотя бы в минимальной степени [4].

Теория интегрированной информации отрицает поведенческие критерии сознания, в частности тест Тьюринга. Система может демонстрировать сколь угодно сложное и осмысленное речевое поведение, но при этом иметь Φ близкое к нулю (быть системой типа «вход-выход» без внутренней связности) и, следовательно, не обладать сознанием. И напротив, система может быть практически неподвижной и некоммуницирующей (как человек в состоянии глубокой медитации или сна со сновидениями), но обладать высоким Φ и, соответственно, богатым сознательным опытом.

Рассмотрим цифровую фотокамеру. Каждый ее пиксель реагирует на свет независимо от других, выход одного пикселя не влияет на состояние остальных. Если удалить или повредить один пиксель, работа остальных не изменится. Информация, порождаемая такой системой, представляет собой простую сумму информации от отдельных элементов. Величина Φ для подобной системы близка к нулю. В мозге, напротив, наблюдается принципиально иная организация. Воздействие на один участок мозга неизбежно сказывается на состоянии множества других областей. Состояние мозга не сводится к сумме состояний отдельных нейронов – оно есть продукт их сложного динамического взаимодействия. Следовательно, согласно данной теории, ИИ теоретически может обладать определенной формой сознания, если уровень интеграции информации достигнет необходимой степени.

Таким образом, проблема возможности сознания у искусственного интеллекта остается открытой и дискуссионной. Функционализм допускает возможность возникновения сознания в искусственных системах при достаточной сложности их организации. Биологический натурализм и аргументы

Д. Сёрла, напротив, указывают на принципиальную невозможность появления подлинного субъективного опыта у машины.

Современная философия не располагает однозначными критериями определения сознания, что делает невозможным окончательное решение вопроса о сознательном ИИ. Тем не менее развитие технологий и усложнение интеллектуальных систем продолжают усиливать актуальность данной проблемы, превращая ее из теоретической философской дискуссии в важный междисциплинарный вопрос XXI века.

СПИСОК ЛИТЕРАТУРЫ

1 **Тьюринг, А. М.** Вычислительные машины и разум / А. М. Тьюринг // *Mind*. – 1950. – Т. LIX, № 236. – С. 433–460.

2 **Сёрл, Дж. Р.** Разум, мозг и программы / Дж. Р. Сёрл // *Behavioral and Brain Sciences*. – 1980. – Т. 3, № 3. – С. 417–424.

3 **Chalmers, D.** Artificial Consciousness / D. Chalmers. – URL: <https://serious-science.org/atificial-consciousness-6883> (date of access: 30.04.2026).

4 **Тонони, Д.** Теория сознания как интеграции информации / Д. Тонони // *BMC Neuroscience*. – 2004. – Т. 5. – С. 42.

Получено 04.05.2026

ISSN 2227-1155. Сборник студенческих научных работ.

Вып. 31. Гомель, 2026

УДК 316.42:12

Д. А. СОЛОДОВНИКОВА (магистрант)

Научный руководитель – канд. филос. наук *Е. Г. КИРИЧЕНКО*

ГЛОБАЛЬНЫЕ ПРОБЛЕМЫ СОВРЕМЕННОСТИ И БУДУЩЕЕ ЧЕЛОВЕЧЕСТВА

Проведён комплексный анализ глобальных проблем современности как системного вызова существованию человеческой цивилизации. Рассмотрены экологические, социально-экономические и технологические угрозы в их нелинейной взаимосвязи. Выделены три магистральных сценария будущего: пессимистический (коллапс), оптимистический (устойчивое развитие) и технократический (трансгуманизм). Обоснована необходимость смены экономической парадигмы, перехода к постростовой экономике и формирования ноосферного мышления как условия сохранения цивилизации.

Современный мир вступил в эпоху «поликризиса» – синхронизации экологических, социально-экономических и технологических угроз [1]. Антро-