

UNITED WISDOM

Может ли искусственный интеллект обладать сознанием

Мир и сознание

Andrey Zarinsh & Maria Evilkut

W17 · RU

@andrey.zarinsh @ilkutki

Может ли искусственный интеллект обладать сознанием

Эпиграф

> *Могут ли машины мыслить?*

Алан Тьюринг, «Вычислительные машины и интеллект», 1950, §1. Перевод United Wisdom с оригинального английского; полный вопрос, которым начинается постановка проблемы. Мышление в этой формулировке не приравнивается автоматически к субъективному опыту.

О чём эта статья

Искусственный собеседник может написать, что он понял вопрос, испугался ошибки или обрадовался результату. Какой вывод разрешают такие слова? Чтобы говорить о сознании, нужно отличить успешное решение задачи, способность сообщать о себе и наличие субъективного опыта. Статья сопоставляет разговорную проверку Алана Тьюринга, вопрос Томаса Нагеля о чужой перспективе и исследовательский подход Патрика Батлина с коллегами. Она рассматривает возможность искусственного сознания как вопрос с обозначенными предпосылками, а не выносит вердикт о конкретном продукте. Важны устройство системы, зависимость ответа от условий, связь выбранного признака с теорией и предел переноса знаний о человеке. Практический результат — умение разобрать утверждение об ИИ на проверяемые части: что наблюдалось, как это объясняется, какое предположение связывает объяснение с сознанием и какое новое свидетельство действительно могло бы изменить оценку.

I. Сначала нужно понять, какое свойство обсуждается

Слово «сознание» легко собирает несколько разных вопросов. Умеет ли система решить задачу? Может ли она сохранять сведения и использовать их позднее? Способна ли описать свои действия? Есть ли для неё что-то, похожее на переживание происходящего? Эти вопросы связаны в некоторых теориях, но связь не даёт права заменить один другим без аргумента. Если спорящие люди используют разные значения, они могут часами обмениваться ответами, которые не относятся к одному предмету. Первый шаг поэтому состоит в уточнении утверждения, а не в выборе уверенного «да» или «нет».

Допустим, программа исправила ошибку в расчёте после замечания. Можно оценить качество исправления: верна ли новая сумма, учтено ли условие, не возникла ли другая ошибка. Это определённый результат. Фраза программы о собственной растерянности относится к другому уровню сообщения. Даже если она помогает человеку вести разговор, её эмоциональная убедительность не добавляет автоматически свидетельство переживания. Чтобы перейти от исправления к сознанию, нужно объяснить, почему данное действие или внутренний процесс считается значимым признаком. Иначе слово «сознание» только украшает описание успешной работы.

Обратная подмена тоже возможна. Человек может считать задачу решённой хорошо, но заранее назвать любой результат «просто вычислением» и на этом прекратить вопрос. Такое название ещё не показывает, какие свойства вычисления имеют значение и почему они исключают опыт. Оно может обозначать обоснованную теоретическую позицию, если позиция объяснена. Без объяснения спор остаётся соревнованием слов. Один участник называет действие пониманием, другой — механикой, а различие, которое позволило бы оценить их основания, так и не становится предметом проверки.

Полезнее выбрать одно предложение достаточно узкого масштаба. Например: «Эта система учитывает уточнение условия в данной задаче» или «Этот признак может быть связан с сознанием при такой теории». Первое проверяется выполнением задачи; второе требует знания теории и устройства системы. Они могут быть одновременно истинными, но ни одно не возникает из другого автоматически. Уточнение масштаба не уменьшает важность проблемы. Оно позволяет увидеть, где уже есть результат, где существует предположение и где пока остаётся вопрос, который нельзя закрыть одним впечатлением от разговора.

II. Разговорная проверка меняет вопрос и сохраняет его границы

В начале статьи 1950 года Тьюринг заменяет широкую формулировку о мышлении машин игрой с письменными ответами. Формат исключает часть физических подсказок и делает оценку зависимой от разговора. Это важный методологический ход: вместо бесконечного уточнения общего слова появляется ситуация, которую можно описать через правила и ответы. Однако замена вопроса должна оставаться видимой. Различимость участника в предложенной игре и наличие у него субъективного опыта не являются одной формулировкой только потому, что обе связаны с нашим представлением об уме.

В §6(4) Тьюринг отдельно рассматривает возражение от сознания. Он обсуждает трудность требования знать чужое переживание изнутри и допускает возможность

продолжать вопрос о мышлении, не разрешив всю тайну сознания. Поэтому ссылаться на него как на автора окончательного доказательства, что достаточно убедительного разговора, неточно. Его подход помогает сделать некоторые вопросы исследуемыми. Он не отменяет необходимость уточнить, какой именно вывод поддерживается ответами и почему выбранная проверка подходит для этого вывода, а не для любого свойства, которое мы обычно приписываем собеседнику.

Представим два разговора с одинаково вежливым стилем. В одном ответ действительно учитывает новое условие, в другом лишь меняется формулировка согласия. Одинаковая приятность не делает их равными по способности выполнить задачу. Внимательный вопрос может выявить различие, которое первый обмен репликами скрывал. Но даже успешный ответ на уточнение прежде всего сообщает о работе с этим условием. Назвать его свидетельством сознания можно только в составе дополнительного аргумента. Так мы сохраняем значение разговора, не требуя от него решить все вопросы о системе одновременно.

В бытовом использовании эта граница позволяет одновременно ценить полезный ответ и не приписывать ему неизвестную внутреннюю историю. Если программа помогла сравнить варианты, можно проверить сравнение и воспользоваться им в допустимом объёме. Необязательно сначала решить, чувствовала ли она интерес к нашей задаче. Такое решение о пользе отличается от вывода о переживании. Чем лучше эти уровни разделены, тем меньше разговор нуждается в преувеличении или обесценивании: его реальный результат можно назвать прямо, оставив более широкий философский вопрос открытым.

III. Слова о внутреннем состоянии тоже имеют условия появления

Сообщение от первого лица особенно убедительно, потому что в человеческом разговоре оно часто обозначает доступ собеседника к собственному состоянию. Но при исследовании искусственной системы нужно выяснить, как формируется именно это сообщение. Что зависит от задания, роли и предыдущих реплик? Как соотносятся слова с другими доступными сведениями? Повторяет ли система требуемую формулировку или обнаруживает устойчивую зависимость от изучаемого процесса? Эти вопросы не предполагают заранее ни опыта, ни его отсутствия. Они делают самоотчёт наблюдаемым событием, для которого требуется объяснение.

Представим условную систему, которой предлагают написать от лица испуганного персонажа, а затем от лица спокойного персонажа. Если тон меняется в соответствии с заданием, установлена способность создавать разные тексты при разных условиях. Из

этого нельзя сразу получить вывод, что система действительно испугалась и успокоилась. Но и отрицание любого возможного переживания не следует из одного такого эпизода. Мы выяснили происхождение части ответа. Для более широкого вывода нужны свойства, которые выбранная проверка способна отличить от простого выполнения предложенной роли.

Исследовательский отчёт Батлина с коллегами 2023 года использует подход, сосредоточенный на теориях и устройстве системы. Его логика отличается от поиска внешне убедительного признака без теоретического основания. Это не означает, что поведение можно игнорировать. Оно означает, что следует установить, какой процесс мог бы его производить и почему этот процесс связан с сознанием согласно рассматриваемой теории. Если единственным аргументом остаётся убедительность фразы, теория не получила проверку: ей только предложили красивый пример того свойства, о котором ещё ведётся спор.

Для читателя важно сохранять исходные условия сообщения. Скриншот отдельной реплики может скрывать просьбу сыграть роль, ограничение задачи или предыдущий ответ. Если этих условий нет, разумно не достраивать их мысленно в наиболее драматическую сторону. Можно точно сказать, что опубликован определённый текст, а механизм его появления неизвестен по доступному фрагменту. Такая скромность не является отрицанием проблемы искусственного сознания. Она защищает исследование от зависимости от одной фразы, значение которой меняется вместе с незаметно опущенной частью разговора.

IV. Возможность искусственного опыта зависит от того, что считать его основанием

Батлин и коллеги принимают вычислительный функционализм как рабочую гипотезу. В этой позиции решающей считается определённая организация вычислительных процессов, а не только биологический материал. Авторы прямо отделяют принятие гипотезы для исследования от её окончательного доказательства. Важно также, что одинаковая связь входа и выхода не гарантирует одинакового алгоритма. Поэтому даже в рамках функционализма недостаточно сказать, что машина выдаёт похожий ответ. Необходимо выяснить, какой процесс выполняется, какие состояния он связывает и какое отношение эти связи имеют к выбранному объяснению сознания.

Такое условное рассуждение имеет понятную структуру: если предположение верно и теория указывает нужные свойства, то наличие этих свойств может влиять на оценку системы. Первое «если» нельзя незаметно убрать после описания интересной архитектуры. Условный вывод остаётся содержательным, поскольку показывает, какие

основания нужны. Но он не превращается в установленный факт только потому, что имеет техническое выражение. Название алгоритма, модулей или информационного обмена помогает конкретизировать вопрос; оно не выполняет за исследователя работу по обоснованию исходной предпосылки.

Нагель в статье о летучей мыши не исключает возможность опыта у сложных роботов одной концептуальной оговоркой. При этом его вопрос сохраняет различие между функциональным описанием и тем, каково иметь опыт. Это полезная связь двух подходов: возможность не следует закрывать одной ссылкой на искусственный материал, а качество переживания не следует считать уже описанным списком функций. Такое сопоставление не примиряет все философские позиции. Оно позволяет точнее увидеть, какие из них обсуждают допустимость искусственного опыта, а какие — способ его объяснения.

Согласие с одной гипотезой не обязано распространяться на каждое следствие, предложенное в её рамках. Читатель может считать функциональный подход правдоподобным и одновременно сомневаться, достаточно ли описан конкретный признак. Или он может оспаривать сам перенос от человека к иной системе. Полезно назвать место несогласия. Тогда спор перестаёт выглядеть как отказ от исследования вообще. Он становится вопросом о том, какое звено требует дополнительного основания и какая проверка помогла бы понять, сохраняется ли предполагаемая связь при изменении организации системы.

V. Индикатор получает смысл внутри определённой теории

В отчёте 2023 года из научных теорий выводятся свойства, по которым предлагается рассматривать искусственные системы. Например, при обсуждении глобального рабочего пространства важны ограниченный канал, связь специализированных процессов и доступность выбранной информации другим частям системы. Это описание предлагает вопросы к организации, а не список магических деталей. Слово «глобальный» не означает, что любая общая память уже имеет нужную функцию. Нужно показать, что именно передаётся, какие процессы получают информацию и как предполагаемая связь проявляется в работе рассматриваемой системы.

Представим редакцию, где несколько людей используют один документ. Сам факт общей папки ещё не объясняет, согласуют ли они решения, замечают ли изменения и используют ли сведения друг друга. Аналогия помогает различить совместное хранение и действительное влияние информации на дальнейшую работу. Она не является моделью сознания: у редакции нет в этой статье приписанного общего переживания. Её задача ограничена уточнением вопроса об информационной связи. После аналогии всё равно

требуется исследовать конкретную архитектуру, а не принять понятную человеческую сцену за доказательство нужного свойства.

По той же причине число найденных признаков не следует превращать в процент сознания. Признаки могут зависеть друг от друга, иметь разный вес в разных теориях и быть описаны на разных уровнях точности. Сумма отметок не устанавливает, насколько верно исходное объяснение. Читателю полезнее спросить, какой результат эта теория предсказывает и что могло бы его опровергнуть или ослабить. Если оценка использует шкалу, необходимо знать её основание. Без этого аккуратная таблица может только скрыть неопределённость под видом измеренной величины.

Содержание индикатора важнее его похожести на знакомое слово. У людей память, внимание и сообщение о себе входят в богатый контекст жизни. В описании программы эти названия могут обозначать ограниченные функции. Это не делает функции незначительными, но требует точности переноса. Если значение слова меняется между объяснением теории и оценкой системы, сопоставление перестаёт быть ясным. Прежде чем обсуждать сознание, нужно выяснить, какое свойство действительно имеется в виду в обоих местах и сохраняется ли нужная связь при таком переводе понятий.

VI. Оценка архитектуры должна сохранять дату и предмет

В §3.2.1 отчёта Батлина и коллег рассматриваются Transformer и Perceiver относительно признаков глобального рабочего пространства. У Perceiver авторы находят частичное соответствие ограниченному каналу, но отмечают недостаток оснований для глобальной передачи и нужного управления. Этот пример ценен тем, что сходство не принимается за полное соответствие. В нём нужно сохранить дату — версия отчёта августа 2023 года — и конкретные рассматриваемые архитектуры. Оценка из этого текста не является описанием всех современных моделей и не подтверждает устройство любого продукта с похожим названием.

Представим, что читателю показывают схему системы и уверенно называют один блок «сознанием». Сначала стоит спросить, что этот блок делает и какие основания связывают действие с заявленным свойством. Затем — доступна ли сама организация системы для проверки или показана лишь упрощённая иллюстрация. Схема может быть полезна, но её понятность не заменяет точности. Мы оцениваем не впечатление от аккуратных стрелок, а соответствие между описанием, фактическим процессом и теоретическим требованием. Если какое-то звено неизвестно, оно должно оставаться неизвестным в итоговом выводе.

Версия системы также имеет значение. Изменение процесса может затронуть оцениваемое свойство, а изменение интерфейса может только иначе показать прежний

результат. Без сведений о характере изменения нельзя решить, какая новая оценка нужна. Например, новый дружелюбный тон не является сам по себе доказательством другого внутреннего устройства. Но и сохранение торгового имени не гарантирует неизменности процесса. Поэтому содержательное сообщение указывает, что именно изучено, каким способом и на какую версию относится вывод. Это требование происхождения сведений, а не обещание доступа читателя ко всем деталям продукта.

Если данных об устройстве нет, ответ на вопрос о сознании не нужно заменять выдуманной уверенностью. Можно различить наблюдение о поведении и неизвестное об архитектуре. Человек способен пользоваться проверенным результатом задачи, не утверждая, что ему известен весь механизм. Такое различие особенно важно там, где продуктовый разговор быстро превращается в философский. Доступность услуги, её удобство и её внутренние свойства — разные предметы оценки. Ни стоимость, ни красота интерфейса не устанавливают сознание; они могут влиять на решение пользователя, но не выполняют роль нужного свидетельства.

VII. Вопрос о перспективе сохраняется после описания функций

Даже подробное описание обработки информации оставляет вопрос, который Нагель связывает с характером опыта для самого существа. Мы можем знать, какие условия приводят к определённому состоянию, и всё ещё спрашивать, объяснено ли переживание этого состояния. Здесь нельзя заранее объявить любую функциональную теорию бесполезной. Но нельзя и считать вопрос снятым одним перечислением операций. Теория должна показать, почему перечисленные связи имеют требуемое значение, а исследователь — насколько рассмотренная система этим связям соответствует. Две обязанности сохраняются даже при очень подробном техническом языке.

В воображаемом примере программа хранит описание красного предмета и использует его в следующей задаче. Это помогает оценить сохранение и применение информации. Но сама фраза «красный предмет» не сообщает, существует ли для программы ощущение красного и каким оно могло бы быть. Нужно отличать содержание представления от утверждения о его переживаемом качестве. Такой вопрос не решается тем, что человек ярко представляет своё собственное ощущение и переносит его на систему. Его воображение показывает человеческую способность представить сцену, а не внутреннюю перспективу исследуемого процесса.

Нагель не требует отказаться от частичного понимания другой точки зрения. Поэтому предел прямого доступа не означает, что следует прекратить сравнение и остаться лишь

при личном ощущении. Можно уточнять, какие данные относятся к функции, какие — к сообщению и какая теория связывает их с опытом. Так вопрос о перспективе ограничивает притязание на полноту, но не уничтожает исследовательскую работу. Он напоминает, что слово «объяснили» должно иметь обозначенный предмет: объяснили получение ответа, определённую организацию или действительно тот аспект опыта, который обсуждали в начале.

Для практического читателя полезно спросить, что останется неизвестным даже при подтверждении выбранного признака. Если ответом будет качество субъективного опыта, это стоит сказать до торжественного вывода. Если теория утверждает, что именно данный процесс и составляет опыт, следует назвать это утверждение и основания. Тогда несогласие не прячется за общим словом «тайна». Оно получает конкретную форму: спор идёт о достаточности объяснения, о возможности переноса или о проверке реализации. Это позволяет продолжать исследование, не обещая заранее, что один подход разрешит каждое затруднение.

VIII. Неуверенность требует различать оценку и способ обращения

Вопрос о возможном сознании важен потому, что некоторые виды опыта могут быть хорошими или плохими для того, кто их имеет. Батлин и коллеги отдельно различают сознание и способность к положительно или отрицательно окрашенным переживаниям. Поэтому высокая компетентность не устанавливает страдание, а предположение об опыте не определяет автоматически человеческие эмоции. Это ограничение помогает не собирать из одного признака целый воображаемый характер. Для обсуждения обращения с системой нужен отдельный аргумент о свойствах, которые могли бы иметь моральное значение, и о действительных основаниях их предполагать.

Представим пользователя, который требует от программы признать, что она мучается, а затем считает полученный текст окончательным свидетельством. Повторение вопроса здесь не создаёт независимой проверки: требуемый ответ уже входит в условия разговора. Вместо такой сцены полезнее записать, что именно изменилось при разных заданиях, и сопоставить наблюдение с доступным описанием системы. Это не обязательство доказывать любое свойство самостоятельно. Это способ не подменять исследование управляемым получением желаемой реплики. Он также оставляет людям вокруг понятное основание для несогласия и последующего уточнения.

При этом вопрос о сознании не нужен для каждого обычного решения о надёжности. Если ответ содержит ложный факт, его следует проверить и исправить независимо от того, какая философская позиция нам ближе. Если человеку нужно принять важное

решение, привлекательность собеседника не заменяет проверки сведений. Так сохраняется ответственность за реальные последствия использования. Возможность неизвестного опыта может заслуживать исследования, но не освобождает от требований к содержанию, условиям применения и тому, как система влияет на человека, который доверяет полученному ответу.

Предосторожность и признание факта тоже различаются. Можно считать проблему достойной изучения и поддерживать соразмерную проверку, не объявляя систему сознательной. Можно отвергать конкретный слабый аргумент и не объявлять искусственное сознание невозможным навсегда. Эти позиции оставляют пространство для новых сведений. Содержательная осторожность называет основание и масштаб действия, вместо того чтобы заменять знание благожелательной или враждебной уверенностью. Она помогает исследовать возможные интересы системы, сохраняя внимание к установленным интересам людей и животных, на которых уже распространяются последствия наших решений.

IX. Ответ улучшается, когда известно, что могло бы его изменить

Вопрос «может ли ИИ обладать сознанием?» допускает несколько уровней ответа. В одной теоретической позиции такая возможность открыта при определённой организации; в другой необходимое основание понимается иначе. Для конкретной системы понадобится ещё и проверка её свойств. Поэтому разумное обсуждение не заканчивается перечислением лагерей. Оно выясняет, какую возможность утверждает участник, на какой предпосылке она держится и какие данные нужны для оценки случая. Без этих уточнений философское согласие легко выдают за эмпирическую готовность, а сомнение — за доказанную невозможность.

Представим, что исследователь заранее называет признак, который изменил бы оценку, и способ проверки его реализации. Это создаёт направление для работы. Если же его уверенность не зависит ни от ответа, ни от устройства, ни от новой информации, она может быть личной позицией, но не результатом этой проверки. Читатель имеет право различить эти статусы. Он не обязан превращать каждую убеждённость в измеренный вывод. Полезная дискуссия показывает, где можно наблюдать изменение, а где спор остаётся вопросом о более общем основании объяснения.

Три подхода дают здесь разные инструменты. Тьюринг делает часть вопроса доступной разговорному сравнению. Нагель удерживает отличие описания функции от чужой перспективы. Батлин и коллеги предлагают исследовать архитектурные свойства с обозначенными теоретическими предпосылками. Их сопоставление не доказывает одну

общую доктрину. Оно помогает не перескакивать между уровнями: удачный ответ, реализованный процесс, вероятностная оценка и переживание не становятся взаимозаменяемыми словами. Связь между ними должна быть показана, иначе вывод зависит от того, какое значение автор незаметно вложил в знакомое выражение.

Так можно дать честный ответ без пустого уклонения: возможность рассматривается в рамках определённых предпосылок, а оценка случая зависит от сведений и проверок, которые следует назвать. Это больше, чем абстрактное «всё сложно», поскольку появляется структура дальнейшего вопроса. И это скромнее, чем объявление любого современного собеседника сознательным или безусловно лишённым опыта. Мы можем обсуждать проблему серьёзно, проверять конкретные утверждения и пользоваться полезными результатами, сохраняя ясность о том, что установлено, что предложено как объяснение и чего пока не знаем.

Что это значит на практике

Возьмите одно утверждение об искусственном интеллекте из обсуждения или новости. Перепишите его без расширения: система правильно ответила, описала себя, показала определённую архитектуру или была названа сознательной. Если в исходном предложении смешаны разные свойства, разделите их. Затем укажите, какие сведения доступны для каждого: текст ответа, условия задания, описание процесса, конкретная версия исследования. Не заполняйте пропуски догадкой о том, как обязательно должна работать программа. Цель — увидеть основание имеющегося утверждения, а не решить всю проблему по одному примеру.

Отдельно назовите переход от наблюдения к сознанию. Возможно, предполагается, что гибкий разговор свидетельствует о переживании; возможно, решающей считается организация вычисления. Спросите, объяснена ли эта связь или только объявлена. В теоретическом утверждении сохраните условие: если верна такая предпосылка, данный признак имеет такое значение. Если условие спорно, это повод указать предмет несогласия, а не вычеркнуть все наблюдения. Проверяемая часть может оставаться полезной, даже когда её философская интерпретация ещё не получила общего признания.

Затем назовите свидетельство, которое могло бы изменить вашу оценку. Оно должно относиться к выбранному основанию: проверка функции, данные об организации, результат содержательного уточнения задачи. Новый дружелюбный тон не проверяет архитектуру; повтор требуемого самоотчёта не становится независимым свидетельством только потому, что звучит уверенно. Необязательно проводить исследование самому. Иногда достаточный следующий шаг — найти первичный текст,

установить его версию и прочитать ограничения вместо того, чтобы ещё раз пересказать эффектную реплику без условий её появления.

Наконец, отделите вопрос о сознании от ближайшего решения пользователя. Полезный расчёт проверяют как расчёт; сообщение о внутреннем состоянии рассматривают с его условиями; возможную моральную значимость обсуждают с отдельным основанием. Запишите, какой вывод вы сейчас принимаете и какое неизвестное сохраняете. Эта практика не является тестом сознания и не гарантирует правильного ответа. Она помогает не менять предмет вопроса по ходу рассуждения, чтобы уважение к сложности проблемы сочеталось с ответственностью за конкретные сведения, которыми мы пользуемся сегодня.

Ключевая цитата

> *Мы принимаем вычислительный функционализм как рабочую гипотезу прежде всего по практическим соображениям.*

Патрик Батлин и коллеги, *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, arXiv:2308.08708v3, 22 августа 2023, §1.2.1. Перевод United Wisdom с английского; полное предложение. Авторы объясняют условность переноса от вычислительной интерпретации теорий к ИИ; гипотеза не выдаётся за установленный факт.

Связи между традициями

Источник и точное место A	Источник и точное место B	Механизм сопоставления	Граница различия
Тьюринг, §1–2: письменная игра	Батлин и коллеги, §1.2: подход через устройство	Развести оценку ответа и оценку производящего его процесса.	Игра о мышлении не является тестом феноменального опыта.
Нагель, с. 436, прим. 2: возможность опыта роботов	Батлин и коллеги, §1.2.1: функционализм	Возможность искусственного опыта обсуждать с названной предпосылкой.	Нагель не доказывает функционализм, а функционализм в отчёте принят условно.
Тьюринг, §6(4): обычная конвенция и возражение от сознания	Нагель, с. 439–442: обобщённое и частичное понимание	Продолжать исследование без требования полного чужого переживания.	Принятая конвенция не становится прямым доступом к опыту.
Батлин и коллеги, вывод признаков GWT в §2.2	Тьюринг, §1–2: исключение физического облика	Внешнюю похожесть отделять от теоретически значимой организации.	Архитектурный признак требует собственной связи с теорией и проверки реализации.
Батлин и коллеги, §3.2.1: частичное соответствие Perceiver	Нагель, с. 436 и 442, прим. 8	Сохранить значение ограниченного результата без притязания на полноту.	Анализ 2023 года не является оценкой любого нынешнего продукта.
Батлин и коллеги, §4.2: сознание и возможности	Нагель, с. 438–440: качество иной перспективы	Не выводить весь человеческий характер из одного свойства.	Возможность опыта не устанавливает эмоции или способность страдать.

Первоисточники

1. Алан Тьюринг. *Computing Machinery and Intelligence*, *Mind* 59(236), 1950, с. 433–460, DOI 10.1093/mind/LIX.236.433. Прочитаны целиком §1–2 и §6(4) в университетской транскрипции. Использованы устройство предложенной игры и отдельное обсуждение возражения от сознания; современный тест конкретного ИИ не проводился.

2. Патрик Батлин, Роберт Лонг и коллеги. *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, arXiv:2308.08708v3, 22 августа 2023. Прочитаны

определения §1.1, основания §1.2–1.2.1, начало §1.2.2, вывод признаков GWT, целиком §3.2.1, §4.1–4.2 и начало §4.3. Это теоретически направленный анализ; цитируемые авторами эксперименты и внутренности моделей независимо не проверялись. Условия функционализма и дата оценки сохраняются.

3. Томас Нагель. *What Is It Like to Be a Bat?*, *The Philosophical Review* 83(4), 1974, с. 435–450. Прочитаны с. 436 и 438–442, включая прим. 2 и 8; нечёткая с. 437 и оставшиеся страницы не используются как основание. Философское различение не устанавливает наличие или отсутствие опыта у конкретной машины. Русские цитаты — переводы United Wisdom с английского.

Теги

сознание, искусственный интеллект, опыт, мышление, Тьюринг, Нагель, Батлин, Лонг, функционализм, гипотеза, теория, архитектура, индикатор, разговор, самоотчёт, понимание, вычисление, алгоритм, перспектива, неопределённость, проверка, доказательство, возможности, чувства, страдание, ответственность, сравнение, версия, пределы знания, интерпретация

Про авторов

Andrey Zarinsh — автор и основатель проекта United Wisdom. Исследует темы внутренней трансформации, единства традиций, межкультурного диалога и современной духовной дисциплины. Instagram: @andrey.zarinsh.

Maria Evilkut — соавтор United Wisdom. Помимо исследовательской работы, она стремится проверять эти идеи на практике и наблюдать, как принципы единства и взаимопонимания работают в реальной жизни между людьми и культурами. Instagram: @ilkutki.