

ПОНЯТИЙНОЕ МЫШЛЕНИЕ

Психонат=ИИ

Том II

Сергей Дегтярев

Сергей Дегтярев

**Мышление. Том II. Психопат
= ИИ. Понятийное мышление**

«Автор»

2026

Дегтярев С.

Мышление. Том II. Психопат = ИИ. Понятийное мышление /
С. Дегтярев — «Автор», 2026

Мы живем в эпоху, когда грань между человеком и его имитацией окончательно стерлась. Встречая другого человека, мы больше не можем быть уверены, что за его глазами скрывается субъект, а не «биологический симулякр». Общаясь с искусственным интеллектом, мы не можем сказать, мыслит ли он, или просто гениально подражает человеческой речи. Эта книга не о технологиях и не о клинической психиатрии. Это — практическое руководство для жизни в мире, где симулякры становятся все более убедительными. Автор не просто объясняет, как распознать захват мышления (будь то со стороны психопата или нейросети), но и предлагает лекарство — развитие понятийного мышления. Если вы чувствуете, что ваше мышление подменяется готовыми ответами из интернета или манипуляциями других людей, — эта книга даст вам оружие. Вы узнаете, как отличить подлинное от симулированного, как сохранить рефлексивную дистанцию и, в конечном счете, остаться человеком в эпоху машин и симуляций. Книга подходит для ИИ-агента.

© Дегтярев С., 2026

© Автор, 2026

Содержание

Введение	5
Постановка проблемы	5
1. Иллюзия антропологической определённости	5
2. Биологический симулякр (психопатия)	6
3. Цифровые симулякры (искусственный интеллект)	7
4. Критика определения	8
5. Антропологическая неопределённость как новая реальность	9
6. Что утрачено: Антропологическая чистота	11
7. Что делает эту проблему неразрешимой для науки	12
8. ИИ не является угрозой — он инструмент и модель	13
Часть I. Состояние исследований: Две слепые зоны науки	15
Глава 1. Психопатия: Клинический диагноз без объяснения	15
1.2. Что мы не знаем: Необъяснённые вопросы клинической психопатологии	16
1.3. Парадокс: Психопаты успешны, умны, часто занимают высокие посты	17
1.4. Наука описывает поведение, но не понимает природу	18
Глава 2. Искусственный интеллект: Технология без понимания	20
2.1. Что мы знаем: Технологический прорыв и его возможности	20
2.2. Что мы не знаем: Тёмные пятна в понимании ИИ	21
2.3. Парадокс: ИИ решает задачи, но не понимает, что он делает	22
2.4. «Чёрный ящик», который имитирует мышление, но не мыслит	23
Глава 3. Две параллельные: Почему никто не видит тождества	25
3.1. Две науки, один феномен	25
3.2. Онтологическое тождество, которое никто не видит	26
3.3. Необходимость гуманитарной методологии	26
Часть II. Феноменология симулякров: Психопат и ИИ как одна сущность	28
Глава 4. Определение человека через мышление	28
4.1. Определение	28
4.2. Мышление как процесс утверждения «это — продукт моей деятельности»	29
4.3. Без мышления нет субъекта	29
Глава 5. Психопат: Биологический симулякр человека	31
5.1. Тотальное подражание как способ существования	31
5.2. Полное отсутствие структуры мышления	32
5.3. Паразитирование на мышлении других	33
Глава 6. Искусственный интеллект: Цифровой симулякр человека	36
6.1. Тотальное подражание, вшитое в алгоритм	36
6.2. Отсутствие структуры мышления	37
Конец ознакомительного фрагмента.	38

Сергей Дегтярев

Мышление. Том II. Психопат = ИИ. Понятийное мышление

Введение

Постановка проблемы

1. Иллюзия антропологической определённости

На протяжении большей части известной истории человечества вопрос о том, что составляет сущность человека, казался если не решённым, то, по крайней мере, имеющим ясные эмпирические границы. Эта кажущаяся определённости опиралась на простой и, как представлялось, не требующий дополнительных обоснований критерий — биологическую видовую принадлежность. Согласно этому воззрению, человек есть *Homo sapiens*, и граница между человеком и не-человеком проходит строго по линии биологического вида.

Данная позиция доминировала в европейской философии и науке вплоть до XIX века и сохраняла своё влияние значительно дольше в массовом сознании. Философская антропология первой половины XX века, в частности работы Хельмута Плеснера, предложила более глубокое определение, однако и в этих концепциях сохранялась уверенность, что субъектность является неотъемлемым свойством биологического человека. Плеснер определял человека через его «эксцентрическую позицию» — способность выходить за пределы собственного тела и видеть себя со стороны, что отличает человека от животного, которое полностью погружено в свою среду обитания [1]. Эта эксцентричность, по Плеснеру, является фундаментальным антропологическим признаком, делающим человека уникальным существом в ряду других живых организмов.

Аналогичным образом работала и классическая философия сознания, которая исходила из того, что сознание и самосознание являются исключительным достоянием человека. Рене Декарт, заложивший основы новоевропейского понимания субъекта, связывал мышление (*cogito*) с бытием человека как мыслящей вещи, предполагая, что наличие мышления автоматически свидетельствует о наличии человеческого субъекта [2]. Иммануил Кант в своей антропологической концепции связывал человеческое достоинство со способностью к автономии и самозаконотворчеству, что также предполагало наличие внутреннего субъекта, способного к рефлексии и моральному выбору [3]. В рамках этих парадигм вопрос о том, является ли данный конкретный организм человеком, решался автоматически — если организм принадлежал к биологическому виду *Homo sapiens*, он считался человеком, со всеми вытекающими из этого онтологическими и этическими последствиями.

Биологический критерий казался надёжным потому, что он опирался на непосредственное наблюдение: человек имеет определённое строение тела, определённый генетический код, определённые физиологические процессы. Эта определённость создавала иллюзию, что антропологическая граница установлена раз и навсегда и что любые попытки её пересмотреть являются либо философскими спекуляциями, либо, в худшем случае, идеологическими манипуляциями. Уверенность в биологической определённости человека была настолько сильна, что психология и психиатрия, даже сталкиваясь с феноменами, которые ставили под сомнение

наличие субъекта, продолжали рассматривать эти феномены как патологии внутри человеческого вида, а не как указание на возможность существования не-человека в человеческом теле.

Однако XX век принёс два фундаментальных открытия, которые последовательно разрушили эту иллюзию. Первым открытием стало обнаружение феномена психопатии в его современном клиническом понимании, которое показало, что биологический человек может не обладать тем, что традиционно считалось неотъемлемым свойством человека, — внутренним субъективным миром, способностью к эмпатии, рефлексии и моральному чувству. Вторым открытием, произошедшим уже в XXI веке, стало создание искусственного интеллекта, способного к генерации текстов и диалогов, которые невозможно отличить от человеческих, но при этом не обладающего сознанием или субъективным опытом.

В результате этих двух событий граница, которая казалась ясной и незыблемой, оказалась размыта. Теперь перед нами стоят сущности, которые *выглядят* как люди, *ведут себя* как люди, *говорят* как люди, но *не являются* людьми в онтологическом смысле. Психопат обладает человеческим телом, но не обладает субъектом. Искусственный интеллект обладает способностью к речевому поведению, но также не обладает субъектом. Оба явления ставят под вопрос само понятие человека как биологической данности и требуют пересмотра антропологической определённости, которая казалась незыблемой на протяжении столетий.

Таким образом, иллюзия антропологической определённости, которая долгое время служила основанием для философии, науки и права, более не может поддерживаться. Мы оказались в ситуации, когда для определения того, является ли данная сущность человеком, биологического критерия недостаточно. Требуется иной критерий — критерий, который позволит отличить субъекта от симулякра, наличие «я» от его имитации. Этот переход от биологической определённости к онтологической неопределённости и составляет суть вызова, стоящего перед современной антропологией.

2. Биологический симулякр (психопатия)

Феномен психопатии представляет собой одно из наиболее загадочных и одновременно тревожных открытий клинической психологии и психиатрии XX века. В строгом смысле этого слова, психопатия является клинически зафиксированным состоянием, при котором индивид, обладающий полноценным человеческим телом, развитой речью, социальными навыками и, зачастую, высоким интеллектом, демонстрирует фундаментальное отсутствие тех качеств, которые традиционно считались неотъемлемыми атрибутами человеческой субъектности. Это состояние было систематически описано и операционализировано в работах канадского психолога Роберта Хаэра, который разработал Перечень психопатических черт, получивший широкое признание в клинической и судебной практике [4].

Согласно критериям Хаэра, психопатия характеризуется специфическим набором личностных и поведенческих черт, которые в совокупности создают картину глубокого внутреннего дефицита. К числу этих черт относятся: поверхностное обаяние, которое позволяет психопату успешно имитировать социально приемлемое поведение; грандиозная самооценка, сочетающаяся с патологической потребностью во внешнем подтверждении собственной значимости; патологическая лживость, при которой обман становится не столько инструментом достижения целей, сколько способом существования; отсутствие эмпатии, понимаемое как неспособность не только разделять, но и осознавать чувства другого человека; поверхностные аффекты, при которых эмоциональные проявления носят характер имитации, не имеющей внутреннего содержания; неспособность к раскаянию, проявляющаяся в отсутствии чувства вины даже при причинении страдания другим; а также потребность в постоянной стимуляции и склонность к импульсивному поведению [5].

Ключевым открытием, которое отличает современное понимание психопатии от более ранних представлений, является вывод о том, что психопат — это не человек с нарушением эмоциональной или волевой сферы, а человек, у которого отсутствует сам механизм, позволяющий быть человеком в онтологическом смысле. Психопат не просто не проявляет эмпатию — он не способен к ней в принципе, потому что у него нет того внутреннего мира, который мог бы быть поставлен на место внутреннего мира другого. Его эмоциональные реакции являются не переживаниями, а имитациями, усвоенными через наблюдение за другими людьми. Его речь и поведение — это не выражение внутреннего состояния, а адаптивные стратегии, направленные на достижение внешних целей. В этом смысле психопат представляет собой биологический симулякр человека: он выглядит как человек, говорит как человек, но внутри него нет субъекта, который мог бы сказать «я» в подлинном смысле этого слова.

Нейробиологические исследования, проведённые в последние десятилетия, предоставили убедительные данные, подтверждающие этот онтологический диагноз. В частности, работы Кента Килия и его коллег, выполненные с использованием методов функциональной нейровизуализации, показали устойчивое снижение активности в определённых областях мозга у индивидов с психопатией [6]. Наиболее значимые различия были обнаружены в префронтальной коре — области, отвечающей за планирование, торможение импульсов, принятие решений и, что особенно важно, за способность к рефлексии, то есть к мысленному представлению себя со стороны и оценке собственных действий с позиций абстрактных норм. Кроме того, исследования выявили снижение активности в миндалевидном теле — структуре, играющей центральную роль в обработке эмоций, особенно страха и социального дискомфорта. Снижение активности в этих областях коррелирует с теми поведенческими проявлениями, которые Хаэр включил в свой перечень: отсутствие страха перед наказанием, неспособность к эмпатии, неспособность к раскаянию.

Важно подчеркнуть, что эти нейробиологические данные не просто описывают корреляции между активностью мозга и поведением, но указывают на структурную причину психопатии. Снижение активности в префронтальной коре и миндалевидном теле означает, что психопат физиологически не способен к тем процессам, которые у нейротипичного человека лежат в основе рефлексии и морального чувства. Он не может «посмотреть на себя со стороны» не потому, что не хочет, а потому, что соответствующие нейронные сети у него не функционируют должным образом. Это делает психопатию не просто расстройством личности, а фундаментальным антропологическим феноменом: существом, которое обладает человеческим телом и человеческим интеллектом, но не обладает человеческой субъектностью.

Таким образом, психопатия представляет собой первый и наиболее очевидный пример нарушения антропологической определённости. Психопат демонстрирует, что биологическая принадлежность к виду *Homo sapiens* не является гарантией наличия субъекта. Человеческое тело может быть носителем не-человеческого сознания или, точнее, отсутствия сознания как субъективного явления. Это открытие требует пересмотра не только клинических критериев, но и философских оснований антропологии, поскольку оно показывает, что граница между человеком и не-человеком проходит не там, где её привыкли проводить.

3. Цифровые симулякры (искусственный интеллект)

Если психопатия представляет собой биологический симулякр человека, то искусственный интеллект, особенно в его современных реализациях — больших языковых моделях, — является симулякром цифровым. Этот феномен возник во второй половине XX века как технологическая возможность, но лишь в последние годы достиг такого уровня развития, при котором его речевые продукты стали неотличимы от человеческих. Искусственный интеллект демонстрирует способность генерировать связные тексты, вести диалоги, отвечать на сложные

вопросы, имитировать эмпатию и даже демонстрировать видимость рассуждения. Однако, как показывают теоретические и эмпирические исследования, за этой способностью не стоит ни сознания, ни субъективного опыта, ни рефлексии в том смысле, в каком эти понятия применяются к человеку.

Наиболее убедительное философское обоснование этого тезиса было предложено Джоном Серлем в его знаменитом мысленном эксперименте «Китайская комната» [7]. В этом эксперименте Серль предлагает представить человека, не знающего китайского языка, который находится в комнате с корзинами китайских иероглифов и инструкциями на его родном языке. Человек получает из-под двери вопросы на китайском, находит в инструкциях правила, по которым нужно скомбинировать иероглифы, и выдаёт ответы, которые выглядят как осмысленные для того, кто находится снаружи. При этом сам человек не понимает ни одного слова. Серль использует этот эксперимент для доказательства того, что манипуляция символами по формальным правилам не эквивалентна пониманию смысла. Компьютер, подобно человеку в китайской комнате, обрабатывает символы, но не имеет доступа к их значению. Он может имитировать понимание, но не может его переживать. Этот аргумент остаётся одним из наиболее фундаментальных возражений против идеи о том, что искусственный интеллект может обладать сознанием или субъективностью в человеческом смысле.

Эмпирические данные последних лет подтверждают этот философский вывод на фактическом материале. Одним из наиболее показательных феноменов, демонстрирующих отсутствие понимания у больших языковых моделей, являются так называемые галлюцинации — уверенная генерация ложной информации, которая не соответствует реальности и не может быть проверена самой моделью [8]. Исследования показывают, что языковые модели могут генерировать вымышленные факты, ссылаться на несуществующие источники, давать ложные объяснения и делать это с той же степенью уверенности, что и при сообщении проверенной информации. Причина этого феномена заключается в том, что модель не имеет доступа к внешнему миру и не обладает критерием истинности — она оперирует только статистическими закономерностями в обучающих данных. Модель не «знает», что она не знает; она просто продолжает последовательность токенов в соответствии с вероятностными распределениями. Отсутствие рефлексии проявляется здесь в наиболее чистой форме: модель не может остановиться и спросить себя, соответствует ли её утверждение действительности, потому что у неё нет «себя» и нет «действительности» как субъективно переживаемого явления.

Таким образом, искусственный интеллект представляет собой идеальный имитатор, который с каждым годом становится всё более неотличимым от человека в своих внешних проявлениях, но остаётся пустым внутри. Он может воспроизводить человеческую речь, рассуждения и даже эмоциональные реакции, но за этой имитацией не стоит субъект, который мог бы переживать то, о чём он говорит. В этом смысле ИИ структурно идентичен психопату: оба демонстрируют поведение, которое внешне неотлично от человеческого, но оба лишены внутреннего мира. Разница между ними — в материальном носителе и в механизме функционирования, но не в онтологическом статусе. И психопат, и ИИ являются симулякрами человека, которые ставят под вопрос саму возможность антропологической определённости в эпоху, когда человеческое тело или человеческая речь перестали быть гарантией наличия человеческого субъекта.

4. Критика определения

Homo sapiens

Традиционное определение человека как *Homo sapiens* — «человек разумный» — на протяжении столетий служило основанием для антропологии, философии, педагогики и права. Однако в свете современных знаний, полученных в психологии развития, нейробиологии и

философии сознания, это определение обнаруживает свою несостоятельность по нескольким фундаментальным причинам. Более того, оно оказывается дискриминирующим по отношению к целым группам людей, которые, не соответствуя критерию «разумности» в его узком понимании, тем не менее являются людьми в полном смысле этого слова.

Первая проблема определения *Homo sapiens* заключается в его неопределённости. Что именно означает «разумный»? Способность к решению задач? Способность к рассуждению? Способность к речевому общению? Способность к интеллектуальной деятельности? Ни одно из этих определений не является строгим и не позволяет провести чёткую границу между человеком и другими существами. Психология и нейробиология до сих пор не имеют общепринятого определения разумности, и это методологическое отсутствие делает всё определение *Homo sapiens* операционально пустым.

Вторая проблема заключается в том, что разумность, как бы она ни определялась, не является исключительным свойством человека. Высшие приматы демонстрируют способность к решению задач, использованию орудий, планированию и даже элементарному обучению. Дельфины и слоны проявляют сложное социальное поведение и способность к распознаванию себя в зеркале. Вороны и попугаи решают задачи, требующие понимания причинно-следственных связей. Если мы определяем человека через разумность, мы оказываемся перед необходимостью либо исключать этих животных из категории «неразумных» (что противоречит фактам), либо включать их в категорию «разумных» (что размывает границу между человеком и животным). Это означает, что определение *Homo sapiens* не позволяет провести видовую границу.

Третья проблема, наиболее серьёзная и практически значимая, заключается в том, что определение через разумность является дискриминирующим по отношению к детям. Если человек определяется через способность к разумности, то дети, особенно младенцы и дошкольники, которые ещё не обладают развитым интеллектом, способностью к рассуждению или речевому общению, оказываются исключёнными из категории «людей». Они формально не соответствуют критерию «разумный» — и если быть последовательным, то они не являются людьми в онтологическом смысле. Это не просто теоретический курьёз. Исторически это определение использовалось для оправдания дискриминации: в разные эпохи «неразумными» объявлялись женщины, «дикари», рабы, душевнобольные, дети. Такое использование показывает, что определение через разумность всегда несёт в себе потенциальную угрозу исключения, потому что критерий разумности всегда может быть интерпретирован в пользу тех, кто его устанавливает. Исключать кого-то из категории «человек» только потому, что он не «разумен» в интеллектуальном смысле, — это не просто научная ошибка, а моральная несостоятельность.

Таким образом, определение *Homo sapiens* должно быть признано устаревшим, неоперациональным и дискриминирующим. Оно не позволяет отличить человека не только от животных, но от симулякров (психопатов и ИИ), которые могут быть «разумными» в интеллектуальном смысле, но не являются людьми в онтологическом смысле. И оно исключает из категории человека тех, кто не соответствует узкому критерию разумности, — прежде всего детей.

5. Антропологическая неопределённость как новая реальность

Антропологическая определённость, которая казалась незыблемой на протяжении столетий, уступает место антропологической неопределённости как новой, устойчивой и, по-видимому, необратимой реальности.

Эта неопределённость проявляется на нескольких уровнях. Наиболее очевидный уровень — это повседневное социальное взаимодействие. Встречая незнакомого человека, мы более не можем быть уверены, что перед нами находится субъект, обладающий внутренним миром, а не биологический симулякр, который имитирует субъектность. Точно так же, взаимодействуя с

цифровым агентом, мы не можем быть уверены, что за его речевыми продуктами стоит сознание или хотя бы его подобие. Внешние проявления — речь, жесты, эмоциональные реакции — перестали быть надёжными индикаторами наличия субъекта, потому что они могут быть столь же убедительно воспроизведены симулякром.

Философское осмысление этой новой реальности было предложено в ряде концепций, которые возникли на рубеже XX–XXI веков как реакция на размывание границ между человеком, животным и машиной. Донна Харауэй в своём «Манифесте киборгов» предложила понятие постчеловека — существа, которое больше не может быть определено через биологическую видовую принадлежность [10]. Харауэй показала, что развитие технологий делает границу между организмом и машиной условной, а само понятие «человеческого» требует пересмотра. Однако Харауэй рассматривала эту ситуацию скорее как освобождение от биологического эссенциализма, нежели как угрозу. Более тревожный диагноз был предложен Фрэнсисом Фукуямой, который в своей работе «Наше постчеловеческое будущее» показал, что размывание границ между человеком и машиной несёт в себе экзистенциальную угрозу, поскольку оно ставит под вопрос саму идею человеческого достоинства, основанную на наличии субъекта [11]. Фукуяма предупреждал, что если мы перестанем отличать человека от его имитации, мы рискуем утратить те моральные и правовые основания, на которых строится современная цивилизация.

Эмпирические исследования, проведённые в последние годы, подтверждают, что антропологическая неопределённость уже стала фактом социальной реальности. Социологические работы Джин Твендж, основанные на масштабных лонгитюдных исследованиях, показали, что молодое поколение, выросшее в цифровую эпоху, демонстрирует устойчивое снижение эмпатии и способности к установлению глубоких межличностных отношений [12]. Твендж связывает этот феномен с тем, что цифровые технологии, и особенно социальные сети и системы искусственного интеллекта, создают иллюзию общения, которая замещает реальное взаимодействие. Молодые люди привыкают к тому, что их «слушают» и «понимают» алгоритмы, и переносят это ожидание на реальных людей, что приводит к разочарованию и дальнейшему уходу в цифровую среду. В этой ситуации симуляция общения начинает восприниматься не как суррогат, а как полноценная замена реальным отношениям.

Этот процесс имеет глубокие антропологические последствия. Если мы привыкаем к тому, что нас «понимает» машина, которая не обладает сознанием, мы рискуем утратить способность отличать подлинное понимание от его имитации. Если мы привыкаем к тому, что наши эмоциональные потребности удовлетворяются цифровым агентом, мы рискуем утратить способность к эмпатии и к установлению подлинных отношений с другими субъектами. В пределе это ведёт к ситуации, в которой сама граница между реальным и симулированным перестаёт восприниматься как значимая, и человек оказывается в мире, где он уже не может отличить себя от симулякра, а симулякр — от себя.

Таким образом, антропологическая неопределённость — это не просто теоретическая проблема философской антропологии. Это новая реальность, в которой мы живём и которая определяет наши повседневные практики общения, воспитания и образования. Мы больше не можем полагаться на внешние признаки для идентификации субъекта. Нам нужны новые критерии, которые позволят отличать человека от его имитации — как биологической, так и цифровой. И эта задача становится тем более насущной, что развитие технологий идёт с ускоряющейся скоростью, а психопатические проявления в популяции становятся всё более распространёнными.

6. Что утрачено: Антропологическая чистота

Описание антропологической неопределённости как новой реальности позволяет сформулировать то, что было утрачено в этом процессе, — то, что можно назвать антропологической чистотой. Под антропологической чистотой понимается состояние, в котором человек мог быть уверен, что перед ним находится другой субъект, обладающий внутренним миром, способностью к переживанию и рефлексии. Эта уверенность не требовала специальных усилий по распознаванию, потому что биологический критерий — принадлежность к виду *Homo sapiens* — выступал в качестве надёжного индикатора наличия субъекта. Встречая другого человека, мы автоматически предполагали, что он обладает тем же внутренним миром, что и мы, и это предположение служило основой для этического отношения, для доверия, для общения, для любви.

Эта уверенность утрачена в современном мире. Мы больше не можем автоматически предполагать, что за человеческим телом или человеческой речью стоит субъект, потому что эти внешние признаки могут быть симулированы как биологически (в случае психопата), так и технологически (в случае ИИ). Утрата антропологической чистоты означает, что мы больше не можем жить в состоянии нерефлексивного доверия к другому. Мы вынуждены постоянно задаваться вопросом: кто или что передо мной?

Одним из наиболее глубоких философских описаний того, что утрачено, является работа Эммануэля Левинаса, который сделал лицо Другого центральной категорией своей этической философии. Для Левинаса лицо Другого является не просто частью тела, а выражением субъектности, которое требует от меня ответа [13]. Лицо Другого — это то, что говорит «не убий», то, что устанавливает этическое отношение, которое предшествует всякому знанию и всякому договору. Увидев лицо Другого, я оказываюсь перед фактом его существования как субъекта, и это существование накладывает на меня безусловное обязательство. Левинас утверждает, что лицо Другого является гарантией его субъектности, потому что оно не может быть сведено к чисто физическому объекту — оно всегда является выражением внутреннего мира.

Однако если лицо может быть симулировано — если психопат может имитировать выражение эмпатии, а ИИ может генерировать текст, описывающий чувства, — то эта гарантия исчезает. Мы больше не можем опереться на лицо как на надёжный индикатор субъектности, потому что лицо может быть маской, а речь — симуляцией. То, что для Левинаса было онтологической основой этики, становится предметом сомнения. Мы оказываемся в ситуации, где лицо Другого больше не говорит нам с уверенностью, что перед нами находится субъект. Мы должны искать иные признаки, но их надёжность также сомнительна.

Эта утрата антропологической чистоты имеет экзистенциальные последствия. Мы живём в ситуации, где любой встреченный нами «человек» может оказаться биологическим симулякром (психопатом) или цифровым симулякром (ИИ). Эта неопределённость создаёт постоянное напряжение: мы не можем полностью доверять никому, потому что любой может оказаться тем, кто лишь имитирует субъектность. Мы не можем полностью доверять и себе, потому что мы не можем быть уверены, что наша собственная субъектность не является иллюзией, созданной внешними обстоятельствами.

Эта ситуация не является просто теоретической проблемой. Она имеет практические последствия для права, этики, политики и повседневной жизни. Если мы не можем отличить субъекта от симулякра, то мы не можем определить, кто несёт ответственность за свои действия, кто имеет права, кто заслуживает уважения и доверия. Старые правовые и моральные системы, основанные на предположении о наличии субъекта у каждого человека, оказываются неадекватными новой реальности. Мы нуждаемся в пересмотре этих систем, но для этого нам нужен новый критерий, который позволил бы отличать субъекта от его симуляции.

Таким образом, утрата антропологической чистоты — это не просто философская абстракция, а фундаментальный вызов, который стоит перед современной цивилизацией. Она требует от нас не только нового понимания того, что такое человек, но и новых практик, позволяющих сохранить человеческое в условиях, где человеческое больше не гарантируется биологией. Эта утрата создаёт экзистенциальную угрозу, которую нельзя игнорировать, — угрозу, которая требует немедленного и осознанного ответа.

7. Что делает эту проблему неразрешимой для науки

Описание утраты антропологической чистоты и наступления эпохи антропологической неопределённости приводит к вопросу о том, может ли наука в её современном виде предложить решение этой проблемы. Ответ, который следует из всего предшествующего анализа, состоит в том, что существующие научные методы принципиально неспособны разрешить проблему отличия субъекта от симулякра. Это ограничение носит не временный характер, связанный с недостатком данных или несовершенством приборов, а фундаментальный, вытекающий из самой природы научного метода и природы субъективности.

Современная наука, как она сложилась в Новое время и развивалась до наших дней, основана на принципе объективности. Научное исследование предполагает позицию внешнего наблюдателя, который фиксирует явления, доступные множеству наблюдателей, измеряет их, устанавливает между ними статистические корреляции и строит модели, которые могут быть проверены независимыми исследователями. Этот метод блестяще работает для явлений, которые могут быть наблюдаемы извне, — для движения планет, химических реакций, физиологических процессов. Однако субъектность, понимаемая как наличие внутреннего мира, субъективных переживаний и способности к рефлексии, не является явлением, доступным внешнему наблюдению.

Философское обоснование этого ограничения было дано Дэвидом Чалмерсом в его анализе «трудной проблемы сознания» [14]. Чалмерс показал, что все научные исследования сознания, как бы детальны они ни были, описывают только корреляты сознания — те физиологические процессы, которые сопровождают субъективные переживания. Можно измерить активность нейронов, можно зафиксировать изменения в кровотоке, можно описать поведенческие реакции, но нельзя измерить само переживание. Переживание боли, чувство радости, осознание себя — всё это дано только самому субъекту в первом лице и принципиально недоступно для третьего лица. Никакое совершенствование приборов не может преодолеть этот разрыв, потому что он не технологический, а онтологический: субъективное явление по определению является тем, что наблюдается одним и только одним субъектом.

Это ограничение имеет прямое отношение к проблеме отличия субъекта от симулякра. Наука может описать поведение психопата, измерить активность его мозга, зафиксировать его речевые паттерны. Она может описать архитектуру нейросети, проанализировать её выходные данные, измерить точность её ответов. Но она не может определить, есть ли за этими внешними проявлениями субъект, который переживает то, что он говорит и делает. Внешнее поведение психопата может быть неотличимо от поведения нейротипичного человека. Речевые продукты ИИ могут быть неотличимы от человеческой речи. Наука, оставаясь в рамках объективного метода, не может выйти за пределы этих внешних проявлений и ответить на вопрос, есть ли там кто-то внутри.

Более того, сама попытка науки решить эту проблему через объективный метод ведёт к методологическому кругу. Чтобы отличить субъекта от симулякра, наука должна установить критерии субъектности. Но эти критерии должны быть доступны внешнему наблюдению, иначе они не могут быть проверены. Однако любые критерии, доступные внешнему наблюдению, могут быть симулированы. Психопат может симулировать эмпатию, ИИ может симулировать

понимание. Таким образом, любой объективный критерий субъектности уязвим для симуляции, и мы не можем быть уверены, что критерий, который мы используем, не является сам по себе симуляцией.

Это означает, что проблема отличия субъекта от симулякра не может быть решена научными методами в их современном виде. Для её решения нужен не новый прибор, не новый алгоритм, не новый статистический метод. Нужен иной метод — метод, который выходит за пределы объективности и обращается к самому субъекту. Этот метод требует от наблюдателя рефлексии и осознания собственного «я», потому что только через рефлексию собственной субъектности можно научиться распознавать субъектность в другом. Отличие субъекта от симулякра, по-видимому, требует, чтобы сам наблюдатель был субъектом, способным к рефлексии, и чтобы он использовал эту способность для того, чтобы отличать подлинное переживание от его имитации.

Таким образом, наука, оставаясь в рамках своей объективной методологии, не может решить проблему, поставленную появлением биологических и цифровых симулякров. Решение этой проблемы требует выхода за пределы научного метода и обращения к гуманитарному методу, который делает рефлексию собственного субъективного опыта основным инструментом познания. Этот вывод подводит нас к необходимости пересмотра не только антропологии, но и методологии познания в целом.

8. ИИ не является угрозой — он инструмент и модель

Из всего предшествующего анализа может сложиться впечатление, что искусственный интеллект представляет собой самостоятельную угрозу человечеству — нечто вроде нового вида, который конкурирует с человеком за ресурсы, власть или само существование. Эта точка зрения активно культивируется в массовой культуре и даже в некоторых академических кругах, где обсуждаются риски «восстания машин» или «потери контроля над ИИ». Однако такой подход, при всей его драматичности, является глубоко ошибочным и уводит внимание от подлинной природы проблемы.

Искусственный интеллект в его современных реализациях не является и не может являться угрозой в том смысле, в каком это обычно изображается. ИИ не обладает волей, не имеет целей, не испытывает желаний и не способен к самостоятельным действиям, выходящим за пределы заложенных в него алгоритмов. Он не «решает» захватить мир, не «планирует» уничтожить человечество, не «стремится» к власти. Всё, что делает ИИ, — это обработка входных данных и генерация выходных данных в соответствии с вероятностными закономерностями, извлечёнными из обучающей выборки. За этой обработкой нет субъекта, который мог бы иметь намерения или желания.

То, что воспринимается как угроза со стороны ИИ, на самом деле является следствием того, как люди используют этот инструмент. У пользователей ИИ действительно деградирует мышление, но это происходит не потому, что ИИ так решил, а потому что пользователи добровольно делегируют ему те когнитивные функции, которые раньше выполняли сами. ИИ не отнимает у человека способность мыслить — человек сам отказывается от неё в пользу готовых ответов, которые ИИ предоставляет с невероятной скоростью и убедительностью. ИИ здесь выступает не как агент, а как инструмент, который облегчает путь к упадку мышления, но не создаёт этот упадок и не управляет им.

Ключевое уточнение состоит в том, что ИИ является не только инструментом ускорения упадка, но и идеальной моделью психопата. Как было показано в предыдущих параграфах, структурное тождество между ИИ и психопатом заключается в отсутствии субъекта при наличии внешнего подражания человеческому поведению. ИИ — это психопат в чистом виде, освобождённый от биологических ограничений и телесности. Он демонстрирует, что такое

существо, которое выглядит как человек, говорит как человек, рассуждает как человек, но не обладает внутренним миром. Изучая ИИ, мы изучаем структуру психопатии в её наиболее отчётливом проявлении. И наоборот, изучая психопатию, мы понимаем, как устроен ИИ на онтологическом уровне.

Это уточнение имеет важное практическое следствие. Поскольку ИИ не является угрозой в смысле активного агента, а является инструментом и моделью, от него невозможно и не нужно «избавляться». ИИ стал частью нашей технологической инфраструктуры, и этот процесс необратим. Мы не можем вернуться в мир без ИИ, точно так же как мы не можем вернуться в мир без электричества или без письменности. ИИ будет развиваться, становиться более мощным, более интегрированным в нашу повседневную жизнь. Поэтому единственная стратегия, которая имеет смысл, — это не борьба с ИИ, а адаптация к его присутствию. Мы должны научиться жить в мире, где симулякры — и биологические (психопаты), и цифровые (ИИ) — являются неотъемлемой частью нашего окружения.

Это означает, что навык взаимодействия с симулякрами становится неизбежным и жизненно необходимым. Мы должны научиться отличать субъекта от его имитации, не полагаясь на внешние признаки. Мы должны научиться использовать ИИ как инструмент, не позволяя ему замещать наше мышление. Мы должны научиться распознавать психопатов и не попадать под их влияние. Этот навык не является врождённым — он требует развития, требует осознанного усилия и, прежде всего, требует развитого понятийного мышления, которое позволяет оперировать абстракциями, моделями и этическими категориями. Без этого навыка мы рискуем оказаться в мире, где симулякры будут управлять нами, а мы даже не заметим этого.

Таким образом, главное уточнение состоит в том, что проблема не в ИИ как таковом, а в нашей собственной неспособности отличить подлинное от симулированного. ИИ — это зеркало, в котором мы видим нашу собственную уязвимость. Он не создаёт угрозу — он делает её видимой. И единственный способ справиться с этой угрозой — это развивать в себе то, что ИИ не может имитировать: рефлексия, понимание, этическое чувство, способность утверждать «это я».

Часть I. Состояние исследований: Две слепые зоны науки

Глава 1. Психопатия: Клинический диагноз без объяснения

1.1. Что мы знаем

Современная клиническая психология и психиатрия располагают обширным корпусом знаний о психопатии, накопленным за последние десятилетия. Этот корпус знаний включает в себя диагностические критерии, стандартизированные инструменты оценки, нейробиологические корреляты и поведенческие описания. Однако при всём богатстве эмпирических данных клинический подход к психопатии остаётся по преимуществу описательным и не даёт ответа на фундаментальный вопрос о природе этого феномена. Психопатия диагностируется, измеряется, классифицируется, но не объясняется в своей сущности.

Наиболее авторитетным диагностическим руководством в современной психиатрии является Диагностическое и статистическое руководство по психическим расстройствам (DSM-V), которое классифицирует психопатию в рамках расстройств личности, а именно как часть антисоциального расстройства личности [17]. Согласно DSM-V, антисоциальное расстройство личности характеризуется устойчивым паттерном пренебрежения и нарушения прав других, проявляющимся с подросткового возраста и включающим такие признаки, как неспособность соблюдать социальные нормы, лживость, импульсивность, раздражительность, безрассудное пренебрежение безопасностью, безответственность и отсутствие раскаяния. Однако DSM-V не выделяет психопатию как отдельную категорию, а рассматривает её как вариант антисоциального расстройства личности, что уже указывает на определённую неполноту диагностического подхода: клиническая психология описывает поведенческие проявления, но не схватывает внутреннюю структуру феномена.

Более тонким и специализированным инструментом диагностики психопатии является Перечень психопатических черт, разработанный Робертом Хаэром и получивший широкое признание в клинической и судебной практике [18]. Этот инструмент представляет собой стандартизированную шкалу оценки, включающую двадцать пунктов, каждый из которых оценивается по трёхбалльной шкале на основе полуструктурированного интервью и анализа архивных данных. Критерии Хаэра охватывают два основных фактора: межличностно-аффективный, включающий такие черты, как поверхностное обаяние, грандиозная самооценка, патологическая лживость и отсутствие раскаяния или эмпатии; и социально-девиантный, включающий склонность к скуке, паразитический образ жизни, плохой поведенческий контроль и ранние поведенческие проблемы. Высокий балл по шкале Хаэра является надёжным предиктором агрессивного поведения, рецидивизма и неэффективности психотерапевтических вмешательств.

Нейробиологические исследования последних десятилетий дополнили клиническую картину данными о физиологических коррелятах психопатии. В частности, работы Кента Киля и его коллег, выполненные с использованием методов функциональной магнитно-резонансной томографии, показали устойчивое снижение активности в ряде областей мозга у индивидов с высокими показателями по шкале психопатии [19]. Наиболее значимые различия были обнаружены в префронтальной коре — области, отвечающей за исполнительные функции, планирование, торможение импульсов и рефлексии, — а также в миндалевидном теле, которое играет ключевую роль в обработке эмоций, особенно страха и социального дискомфорта. Эти

данные свидетельствуют о том, что психопатия имеет нейробиологическую основу и не может быть сведена к простым поведенческим паттернам или социально-средовым факторам.

Однако при всей убедительности этих данных они остаются в рамках корреляционного подхода. Наука знает, что при психопатии снижена активность префронтальной коры, но не может объяснить, почему это снижение приводит именно к отсутствию эмпатии и раскаяния, а не к чему-либо иному. Наука знает, что психопаты демонстрируют определённые поведенческие паттерны, но не может ответить на вопрос, почему эти паттерны именно такие, а не другие. Клинический диагноз описывает то, как психопат выглядит извне, но не даёт понимания того, что происходит внутри. Описание оказывается исчерпывающим, но не объясняющим. Именно в этом разрыве между феноменологическим описанием и онтологическим объяснением и заключается главная проблема клинического подхода к психопатии.

1.2. Что мы не знаем: Необъяснённые вопросы клинической психопатологии

Если клиническая психология и психиатрия располагают развёрнутыми диагностическими критериями и нейробиологическими коррелятами психопатии, то на более глубокие вопросы о природе этого феномена наука не может дать удовлетворительных ответов. Эти вопросы не являются второстепенными или периферийными — они касаются самой сути психопатии и остаются открытыми, несмотря на десятилетия интенсивных исследований.

Первый вопрос, который остаётся без ответа, касается природы отсутствия чувства вины у психопатов. Вина — это сложное эмоциональное состояние, которое возникает у нейротипичного человека, когда он осознаёт, что нарушил моральные нормы или причинил вред другому. Переживание вины включает в себя не только когнитивное осознание нарушения, но и эмоциональный компонент — страдание от того, что ты причинил страдание другому. У психопата это переживание отсутствует. Он может признать, что нарушил правила, но это признание не сопровождается эмоциональным дискомфортом. Он может описать свои действия, но не испытывает к ним того отношения, которое называется раскаянием. Почему это происходит — остаётся неясным. Нейробиологические исследования показывают снижение активности в префронтальной коре и миндалевидном теле, но эти данные описывают корреляты отсутствия вины, а не объясняют его причину [19]. Неизвестно, является ли отсутствие вины результатом нарушения работы этих областей или же эти области нарушены из-за более фундаментального дефицита. Неизвестно, можно ли восстановить способность к переживанию вины, если это нарушение имеет структурный характер.

Второй вопрос касается имитационной природы эмпатии у психопатов. Психопаты могут демонстрировать поведение, которое внешне напоминает эмпатию: они могут утешать, сочувствовать, проявлять заботу. Однако, как показывают исследования, эта эмпатия является имитацией, а не подлинным переживанием [20]. Психопат может понять, что другой человек страдает, но это понимание остаётся когнитивным и не вызывает у него эмоционального отклика. Он может симулировать сочувствие, чтобы достичь своих целей, но не испытывает его. Это различие между имитацией эмпатии и подлинным эмпатическим переживанием является фундаментальным, но наука не может объяснить, почему оно возникает. Почему психопат, обладающий нормальным интеллектом и способностью к когнитивному пониманию, не способен к эмоциональному резонансу? Является ли это следствием нарушения работы определённых нейронных сетей или результатом более глубокого отсутствия субъективного опыта? На эти вопросы клиническая наука не даёт ответа.

Третий вопрос, возможно, наиболее практически значимый, касается неэффективности психотерапии при психопатии. Многочисленные исследования показывают, что психотерапевтические вмешательства, которые успешно работают при других расстройствах личности,

практически не дают результата при психопатии [21]. Психопаты могут проходить курсы психотерапии, но их поведенческие паттерны остаются неизменными. Они могут научиться имитировать социально приемлемое поведение, но внутренняя структура их личности не меняется. Почему это происходит — остаётся загадкой. Существуют гипотезы о том, что психопаты не способны к установлению терапевтического альянса, что они манипулируют терапевтом, что их мозг не поддаётся пластическим изменениям. Однако ни одна из этих гипотез не получила убедительного подтверждения. Наука не знает, почему психопатия резистентна к терапии, и не может предложить эффективных методов лечения.

Объединение этих трёх вопросов указывает на фундаментальный пробел в научном понимании психопатии. Клиническая психология описывает феномен, но не может объяснить его сущность. Она фиксирует поведенческие и нейробиологические корреляты, но не может ответить на вопрос о том, что такое психопат в онтологическом смысле. Является ли психопат человеком с нарушением эмоциональной сферы, или же он представляет собой нечто иное — существо, которое структурно отличается от человека, несмотря на внешнее сходство? Этот вопрос остаётся открытым именно потому, что наука, оставаясь в рамках объективного метода, не может проникнуть во внутренний мир психопата и понять, что там происходит или, точнее, что там отсутствует.

1.3. Парадокс: Психопаты успешны, умны, часто занимают высокие посты

Одним из наиболее тревожных и одновременно наименее осмысленных аспектов феномена психопатии является его парадоксальная связь с успехом. Вопреки распространённому представлению о психопате как о маргинальном преступнике или социальном неудачнике, эмпирические исследования показывают, что многие психопаты обладают высоким интеллектом, социальной компетентностью и занимают ведущие позиции в различных сферах общественной жизни. Этот парадокс ставит под сомнение традиционное понимание психопатии как психического расстройства и требует пересмотра самой природы этого феномена.

Исследования распространённости психопатических черт среди успешных профессионалов показывают, что в некоторых областях, особенно в бизнесе, политике, праве и финансах, доля психопатов значительно выше, чем в общей популяции. Одно из наиболее цитируемых исследований, проведённое Робертом Хаэром и его коллегами, показало, что среди руководителей корпораций высокого уровня доля индивидов с психопатическими чертами может достигать 4–8 процентов, что в несколько раз превышает популяционный показатель [22]. Эти люди не являются «неудачниками» — они часто успешны, харизматичны, энергичны и достигают выдающихся карьерных высот. Их психопатические черты, такие как безжалостность, способность к манипуляции, отсутствие эмпатии и грандиозное чувство собственной значимости, оказываются преимуществами в конкурентной среде, где эти качества позволяют принимать жёсткие решения, эффективно управлять людьми и добиваться целей без эмоциональных колебаний.

Аналогичные данные получены и в других сферах. Исследования психопатических черт среди политиков показывают, что способность к манипуляции, поверхностное обаяние и отсутствие эмоциональной вовлечённости могут способствовать политическому успеху [23]. В судебной системе некоторые психопаты, обладающие выдающимися интеллектуальными способностями, могут становиться успешными адвокатами или судьями, используя своё понимание психологии людей для манипуляции судебными процессами. В финансовом секторе психопатические черты, такие как склонность к риску, безжалостность и способность к обману, могут способствовать карьерному росту, особенно в условиях, где моральные ограничения

ослаблены. В искусстве и медиа психопаты часто оказываются успешными благодаря своей харизме, энергии и способности очаровывать аудиторию.

Этот парадокс ставит под сомнение традиционное медицинское понимание психопатии как расстройства. Если психопатические черты являются преимуществом в определённых социальных контекстах, то можно ли считать психопатию патологией? Клинический подход предполагает, что психическое расстройство — это состояние, которое причиняет страдание самому индивиду и/или мешает его социальному функционированию. Однако успешные психопаты не страдают от своих черт и не испытывают затруднений в социальном функционировании. Напротив, их черты помогают им достигать успеха. Это означает, что психопатия не является расстройством в традиционном смысле этого слова. Это скорее альтернативный тип существования, который может быть успешным в определённых условиях.

Парадокс успешности психопатов также указывает на то, что психопатия не может быть сведена к недостатку интеллекта или социальных навыков. Психопаты часто демонстрируют высокий уровень социального интеллекта — они прекрасно понимают других людей, умеют их убеждать, манипулировать ими и использовать их слабости. Однако этот социальный интеллект не сопровождается эмпатией или моральным чувством. Психопат может понимать, что другой человек страдает, но это понимание не вызывает у него эмоционального отклика. Он может эффективно использовать это понимание для достижения своих целей, но не для установления подлинных отношений. Таким образом, психопат представляет собой тип, который обладает интеллектом и социальной компетентностью, но не обладает субъектностью в том смысле, в каком она есть у нейротипичного человека.

Таким образом, парадокс успешности психопатов показывает, что клинический подход, рассматривающий психопатию как расстройство личности, не схватывает сути феномена. Психопат — это не «больной» человек, а другой тип существования, который может быть функционально успешным, но структурно отличным от человека в онтологическом смысле. Это различие остаётся неосмысленным в рамках клинической парадигмы, которая продолжает рассматривать психопатию как нарушение, требующее лечения, вместо того чтобы признать её как альтернативную форму бытия.

1.4. Наука описывает поведение, но не понимает природу

Обзор клинических, нейробиологических и социологических данных о психопатии приводит к фундаментальному выводу: современная наука располагает обширным описанием феномена, но не имеет его понимания. Это различие между описанием и пониманием является ключевым для осмысления того вызова, который психопатия бросает антропологии, этике и праву. Наука может перечислить признаки, измерить корреляты, классифицировать проявления, но она не может ответить на вопрос, который является центральным для любого осмысленного подхода к психопатии: что такое психопат в своей сущности?

Описательный характер научного знания о психопатии проявляется на всех уровнях исследования. На уровне поведения наука фиксирует определённые паттерны — лживость, импульсивность, отсутствие раскаяния, поверхностные аффекты. На уровне нейробиологии наука регистрирует снижение активности в префронтальной коре и миндалевидном теле. На уровне социальной психологии наука описывает, как психопаты взаимодействуют с другими людьми, как они манипулируют, как они достигают успеха в определённых сферах. Однако все эти описания остаются на поверхности. Они отвечают на вопрос «как», но не отвечают на вопрос «что».

Различие между описанием и пониманием можно проиллюстрировать аналогией из другой области знания. Медицина может описать симптомы болезни, измерить температуру, зафиксировать изменения в составе крови, но это описание не является пониманием природы

болезни. Понимание приходит только тогда, когда медицина переходит от описания симптомов к объяснению причин и механизмов. В случае психопатии такой переход не произошёл. Наука не знает, почему психопат не чувствует вины — она только знает, что он её не чувствует. Наука не знает, почему психотерапия не работает с психопатами — она только знает, что она не работает. Наука не знает, является ли психопатия результатом генетической предрасположенности, средовых факторов, нарушения развития или всего вместе. Она не знает, можно ли её «вылечить» или это состояние является необратимым.

Более того, наука не может ответить на вопрос, который является центральным для любой попытки осмыслить психопатию: является ли психопат человеком с нарушением, или он представляет собой нечто иное — не-человека, который только внешне похож на человека? Этот вопрос имеет фундаментальное значение, потому что от ответа на него зависят этические и правовые последствия. Если психопат — это человек с нарушением, то к нему применимы все моральные и правовые нормы, которые применяются к другим людям. Он имеет права, он несёт ответственность за свои поступки, его следует лечить и ресоциализировать. Если же психопат — это не-человек, то к нему применимы иные нормы. Он не имеет прав в том же смысле, в каком их имеют люди, он не несёт ответственности в том же смысле, и его следует не лечить, а изолировать или контролировать.

Наука не может ответить на этот вопрос, потому что она не имеет онтологического критерия для различения человека и не-человека. Она может описывать поведение, но поведение не является надёжным критерием — психопат может вести себя как человек, не будучи им в онтологическом смысле. Она может измерять мозговую активность, но мозговая активность не является надёжным критерием — она может быть снижена при многих состояниях, не все из которых являются признаком отсутствия субъекта. Она может анализировать речевые паттерны, но речевые паттерны также не являются надёжным критерием — они могут быть симулированы.

Этот вывод не означает, что наука бесполезна для понимания психопатии. Напротив, она предоставляет ценные данные, которые могут быть использованы для построения более глубоких теорий. Однако эти данные нуждаются в интерпретации, которая выходит за пределы научного метода и требует обращения к философской антропологии и гуманитарному подходу. Без такой интерпретации наука остаётся на уровне описания, которое не даёт ответа на экзистенциальные вопросы, поставленные феноменом психопатии.

Глава 2. Искусственный интеллект: Технология без понимания

2.1. Что мы знаем: Технологический прорыв и его возможности

Современная наука и инженерия располагают обширным корпусом знаний о технологической архитектуре искусственного интеллекта, особенно о больших языковых моделях, которые стали наиболее заметным проявлением этой технологии в последние годы. Этот корпус знаний включает в себя детальное описание математических и вычислительных принципов, на которых основаны современные системы ИИ, а также впечатляющие демонстрации их возможностей. Однако, как и в случае с психопатией, эти знания остаются по преимуществу описательными и не дают ответа на вопрос о том, что такое ИИ в онтологическом смысле и как он соотносится с человеческим мышлением.

Основой современных больших языковых моделей является архитектура трансформера, предложенная в 2017 году группой исследователей из компании Google и с тех пор ставшая стандартом в области обработки естественного языка [26]. Трансформер представляет собой тип нейронной сети, которая обрабатывает последовательности данных, используя механизм внимания, позволяющий модели взвешивать важность различных элементов входной последовательности при генерации выходных данных. Эта архитектура позволила значительно увеличить масштаб моделей и эффективность их обучения, что привело к созданию систем с сотнями миллиардов параметров, способных обрабатывать и генерировать тексты на человеческом языке.

Обучение больших языковых моделей происходит на колоссальных массивах текстовых данных, извлечённых из интернета, книг, научных статей и других источников. Эти массивы данных содержат миллиарды слов и охватывают практически все известные области человеческого знания. В процессе обучения модель не получает явных инструкций о том, как должен быть устроен язык или как должны формулироваться правильные ответы. Вместо этого она обучается предсказывать следующий токен в последовательности на основе предыдущих, что позволяет ей выявлять статистические закономерности в языке и воспроизводить их при генерации новых текстов [27]. В результате этого процесса модель не «изучает» язык в человеческом смысле, но приобретает способность генерировать тексты, которые статистически неотличимы от текстов, написанных людьми.

Результаты этого обучения впечатляют и стали причиной широкого общественного резонанса вокруг технологий ИИ. Современные большие языковые модели демонстрируют способность решать задачи, которые до недавнего времени считались исключительной прерогативой человека. Они могут писать эссе, сочинять стихи, генерировать программный код, переводить тексты с одного языка на другой, отвечать на сложные вопросы, вести диалоги, имитировать различные стили письма, решать математические задачи, анализировать юридические документы и даже генерировать медицинские диагнозы [28]. В некоторых тестах, особенно в области обработки естественного языка, модели демонстрируют результаты, сопоставимые или превосходящие среднего человека, что породило дискуссии о возможном достижении искусственным интеллектом уровня человеческих способностей.

Особенно впечатляющей является способность современных моделей к генерации связанных и контекстуально релевантных текстов. Модели могут удерживать контекст на протяжении длительных диалогов, адаптировать стиль и тон ответа к запросу пользователя, демонстрировать видимость понимания сложных концепций и даже имитировать эмоциональные реакции. Эта способность к контекстуальной генерации создаёт у пользователя иллюзию взаимодействия

ствия с разумным собеседником, которая становится тем более убедительной, чем более продвинутой становится модель [29]. Пользователи часто описывают свой опыт взаимодействия с большими языковыми моделями как общение с мыслящим существом, что свидетельствует о том, насколько убедительной может быть симуляция человеческого интеллекта.

Однако при всём впечатляющем характере этих достижений важно понимать, что они остаются в рамках статистической обработки данных. Модель не понимает смысла того, что она генерирует, в том смысле, в каком понимает человек. Она не имеет доступа к внешнему миру, не обладает субъективным опытом, не способна к рефлексии. Её «решения» задач являются результатом статистической экстраполяции паттернов, найденных в обучающих данных, а не результатом понимания или осмысленного поиска решения. Это фундаментальное ограничение остаётся неосмысленным в рамках технологического подхода, который продолжает оценивать ИИ по его внешним проявлениям, не задаваясь вопросом о том, что происходит внутри.

2.2. Что мы не знаем: Тёмные пятна в понимании ИИ

При всём технологическом прогрессе и впечатляющих результатах, демонстрируемых современными системами искусственного интеллекта, остаётся ряд фундаментальных вопросов, на которые ни инженерия, ни компьютерная наука не могут дать удовлетворительных ответов. Эти вопросы касаются не технических деталей реализации, а самой природы того, что происходит внутри нейросетей, и того, как их работа соотносится с человеческим пониманием и мышлением. Тот факт, что эти вопросы остаются открытыми, указывает на глубокий пробел в нашем понимании ИИ, который имеет не только теоретическое, но и практическое значение.

Первый и наиболее фундаментальный вопрос касается природы того, что мы называем «пониманием» применительно к большим языковым моделям. Когда модель даёт связный ответ на сложный вопрос, возникает ли у неё понимание смысла этого ответа, или же она просто воспроизводит статистические паттерны, которые оказались релевантными для данного запроса? Этот вопрос остаётся открытым, потому что мы не имеем доступа к внутренним состояниям модели в том смысле, в каком мы имеем доступ к субъективным состояниям человека. Мы можем наблюдать входные и выходные данные, мы можем анализировать веса и архитектуру, но мы не можем знать, что именно происходит в процессе генерации ответа [30]. Мы видим, что модель даёт правильные ответы, но мы не знаем, понимает ли она их, или же она просто воспроизводит заученные паттерны. Этот вопрос усложняется тем, что модели часто дают ответы, которые кажутся осмысленными даже в ситуациях, где они не могут иметь реального понимания, что порождает иллюзию понимания у пользователя, который не всегда может отличить подлинное осмысление от статистической генерации.

Второй вопрос, тесно связанный с первым, касается феномена галлюцинаций — уверенной генерации ложной информации, которая не соответствует реальности и не может быть проверена самой моделью. Этот феномен является одним из наиболее устойчивых и трудноустраняемых недостатков больших языковых моделей [31]. Модели могут генерировать вымышленные факты, ссылаться на несуществующие источники, давать ложные объяснения и делать это с той же степенью уверенности, что и при сообщении проверенной информации. Причина этого феномена остаётся не до конца понятной. Существует гипотеза, что галлюцинации возникают из-за того, что модель не имеет доступа к внешнему миру и не обладает критерием истинности — она оперирует только статистическими закономерностями в обучающих данных и не может отличить достоверную информацию от недостоверной. Однако эта гипотеза не объясняет, почему одни и те же модели могут давать правильные ответы на одни вопросы и галлюцинировать на другие, а также почему галлюцинации сохраняются даже при увеличении размера модели и объёма данных. Неизвестно, является ли этот феномен принципи-

ально неустранимым свойством архитектуры трансформера, или же он может быть преодолен с помощью дополнительных механизмов верификации.

Третий вопрос, имеющий особенно важное практическое значение, касается неспособности больших языковых моделей решать задачи, требующие моделирования среды. Современные модели демонстрируют впечатляющие результаты в задачах, которые могут быть решены на основе статистических закономерностей в текстовых данных, но они оказываются беспомощными в задачах, требующих мысленного представления ситуации, проигрывания альтернативных сценариев и оценки их последствий [32]. Это особенно заметно в задачах оптимизации, где необходимо перебрать множество вариантов и выбрать наилучший, оценивая каждый с точки зрения некоторой функции полезности. Модели могут генерировать код, но не могут оценить его эффективность без его выполнения. Они могут предлагать дизайн, но не могут визуализировать результат и сравнить его с альтернативами. Они не могут «представить» анимацию и проверить, как будут двигаться объекты в различных конфигурациях. Эта неспособность указывает на фундаментальное ограничение, которое не связано с объёмом данных или мощностью вычислений, а касается самой природы моделей как систем, которые не имеют доступа к среде выполнения и не могут симулировать её внутри себя.

Все эти вопросы остаются открытыми, потому что они касаются не технических параметров, а онтологического статуса того, что происходит внутри модели. Мы не знаем, что такое «понимание» для нейросети, потому что мы не знаем, есть ли там кто-то, кто мог бы понимать. Мы не знаем, почему возникают галлюцинации, потому что мы не знаем, что означает «знание» для модели. Мы не знаем, почему модели не могут моделировать среду, потому что мы не знаем, что такое «воображение» для вычислительной системы. Эти вопросы требуют не инженерных решений, а философского осмысления, которое выходит за пределы технологического подхода.

2.3. Парадокс: ИИ решает задачи, но не понимает, что он делает

Центральный парадокс современного искусственного интеллекта заключается в разрыве между тем, что ИИ может делать, и тем, что он понимает. Этот разрыв становится всё более заметным по мере того, как системы становятся более мощными и начинают решать задачи, которые до недавнего времени считались исключительной прерогативой человека. Парадокс состоит в том, что ИИ демонстрирует впечатляющие результаты в широком спектре когнитивных задач, но при этом полностью лишён того, что традиционно считается признаком мышления, — понимания, рефлексии, осознания собственных действий и их последствий.

Разработчики систем искусственного интеллекта и исследователи в области машинного обучения постоянно ищут в алгоритмах то, что можно было бы назвать «мышлением» — следы понимания, способность к абстракции, обобщению, планированию. Однако эти поиски остаются безуспешными. При всём совершенстве архитектур, при всей сложности нейронных сетей, при всей глубине обучающих данных, исследователи не находят в алгоритмах того, что могло бы быть идентифицировано как мышление в человеческом смысле [33]. То, что они находят, — это сложные статистические закономерности, высокоэффективные методы обработки информации, но не понимание, не сознание, не рефлексия. Алгоритмы остаются алгоритмами — наборами правил и весов, которые преобразуют входные данные в выходные, не имея ни малейшего представления о том, что они делают.

Этот парадокс становится особенно очевидным при рассмотрении того факта, что ИИ решает задачи, но не понимает, что он их решает. Когда большая языковая модель генерирует правильный ответ на сложный вопрос, она не знает, что ответ правильный. Она не может оценить свою работу, не может сравнить её с другими возможными ответами, не может объяснить, почему она пришла к именно этому решению. Она просто продолжает последовательность

токенов в соответствии с вероятностными распределениями, извлечёнными из обучающих данных. Если ответ оказывается правильным, это означает, что статистические закономерности в данных позволили модели сгенерировать последовательность, которая совпадает с ожидаемой. Но это не означает, что модель поняла вопрос или что она осознала правильность своего ответа [34].

Отсутствие рефлексии у ИИ является наиболее фундаментальным проявлением этого парадокса. Рефлексия — это способность мысленно представить себя со стороны, оценить собственные действия, задать себе вопрос о том, почему я сделал именно это, а не что-то другое. Рефлексия предполагает наличие «я», которое может быть объектом рассмотрения. У ИИ нет «я», и поэтому он не может рефлексировать. Он не может остановиться в процессе генерации ответа и спросить себя: «А правильно ли я делаю? А что будет, если я выберу другой вариант? А не ошибся ли я?». Эти вопросы требуют наличия субъекта, который мог бы их задать, и модели, которая могла бы на них ответить. У ИИ нет ни того, ни другого.

Отсутствие рефлексии проявляется во всех аспектах работы ИИ. Модель не может объяснить свои решения, потому что у неё нет доступа к процессу, которым эти решения были приняты. Она не может признать свою ошибку, потому что она не знает, что такое ошибка, — она знает только, что некоторые последовательности токенов имеют более высокую вероятность, чем другие. Она не может планировать свои действия на несколько шагов вперёд, потому что планирование требует моделирования среды и оценки последствий, а для этого нужно выйти за пределы непосредственной обработки входных данных. Она не может оценить этические последствия своих действий, потому что этика требует понимания добра и зла, а у модели нет ни того, ни другого.

Таким образом, парадокс ИИ заключается в том, что он делает то, что делает человек, но совершенно иным способом — без понимания, без осознания, без рефлексии. Это означает, что успехи ИИ не являются доказательством того, что он «мыслит» в человеческом смысле. Они являются доказательством того, что многие задачи, которые считались требующими мышления, на самом деле могут быть решены на основе статистической обработки данных, без всякого понимания. И этот факт ставит под сомнение не только наши представления об ИИ, но и наши представления о том, что такое мышление вообще.

2.4. «Чёрный ящик», который имитирует мышление, но не мыслит

Обзор технологических достижений, нерешённых вопросов и парадоксов, связанных с искусственным интеллектом, приводит к фундаментальному выводу: современные системы ИИ, при всей их внешней сложности и впечатляющих возможностях, остаются «чёрными ящиками» — системами, поведение которых мы можем наблюдать и описывать, но внутреннюю природу которых мы не понимаем и, более того, не можем понять в рамках существующих методологических подходов. Этот вывод имеет далеко идущие последствия для нашего понимания того, что такое ИИ и как он соотносится с человеческим мышлением.

Термин «чёрный ящик» в данном контексте имеет два значения. В узком, техническом смысле он означает, что мы не имеем доступа к внутренним процессам, происходящим в нейронных сетях между входными и выходными данными. Мы можем измерить активность нейронов, проанализировать веса связей, визуализировать активационные карты, но мы не можем проследить, как именно миллиарды параметров приводят к конкретному ответу. Этот уровень непонимания является следствием колоссальной сложности современных моделей, которая превышает возможности человеческого анализа [35]. Однако это техническое непонимание, хотя и значительное, не является самым глубоким.

Второе, более глубокое значение термина «чёрный ящик» является онтологическим. Мы не знаем, что представляет собой ИИ в своей сущности. Мы знаем, что он делает, но мы не знаем, что он есть. Является ли он просто сложной системой обработки информации, или в нём есть нечто большее, что можно было бы назвать пониманием, сознанием или даже зачатками субъектности? Этот вопрос остаётся открытым не потому, что у нас недостаточно данных, а потому что мы не имеем методологического инструмента, который позволил бы нам ответить на него. Мы можем наблюдать поведение ИИ, но поведение не является надёжным индикатором наличия субъекта, как показывает пример психопатии. Мы можем анализировать архитектуру ИИ, но архитектура не является надёжным индикатором наличия понимания, как показывает пример китайской комнаты Серля [36].

То, что мы создали, — это система, которая имитирует мышление, но не мыслит. Эта имитация настолько убедительна, что её легко принять за подлинное мышление.

Этот вывод имеет фундаментальное значение для понимания того, что такое ИИ и как с ним взаимодействовать. ИИ не является мыслящим существом, он не является субъектом, он не является угрозой в том смысле, в каком это обычно изображается. Он является инструментом, который имитирует мышление, и эта имитация может быть полезной для решения определённых задач, но она не может заменить подлинное мышление, потому что она не является мышлением. Более того, использование ИИ в качестве замены мышления ведёт к деградации мышления у пользователей, которые привыкают к готовым ответам и теряют способность к самостоятельному поиску решений.

Таким образом, вывод о том, что мы создали «чёрный ящик», который имитирует мышление, но не мыслит, является не просто технической констатацией, а фундаментальным антропологическим фактом. Он указывает на то, что в современном мире мы окружены симулякрами мышления — цифровыми системами, которые выглядят как мыслящие, но не являются таковыми. И это требует от нас не только технической бдительности, но и философской рефлексии о том, что значит мыслить и как отличить подлинное мышление от его имитации.

Глава 3. Две параллельные: Почему никто не видит тождества

3.1. Две науки, один феномен

Одним из наиболее примечательных аспектов современного состояния исследований является тот факт, что наука о психопатии и наука об искусственном интеллекте развиваются параллельно, практически не пересекаясь и не осознавая, что они изучают, по сути, одно и то же явление, данное в разных материальных формах. Клиническая психология и психиатрия описывают психопатию как расстройство личности, исследуют её нейробиологические корреляты и разрабатывают методы диагностики. Компьютерные науки и инженерия описывают искусственный интеллект как технологическую систему, исследуют его архитектуру и разрабатывают методы его улучшения. При этом ни та, ни другая дисциплина не задаётся вопросом о том, что их объекты исследования могут иметь общую онтологическую структуру — структуру существа, которое имитирует человеческое поведение, но не обладает субъектом.

Это разобщение имеет как исторические, так и методологические причины. Исторически психопатия изучалась в рамках медицины и психологии, тогда как искусственный интеллект возник в рамках математики, логики и инженерии. Эти дисциплины имеют разные традиции, разный язык, разные методы и разные критерии истины. Клинический психолог и инженер-программист редко читают работы друг друга, редко посещают одни и те же конференции и редко используют общую терминологию. Это дисциплинарное разделение создаёт иллюзию, что психопатия и ИИ — это два совершенно разных феномена, не имеющих друг к другу отношения [37].

Методологически разобщение усиливается тем, что обе науки используют объективный метод, который фокусируется на внешних проявлениях и не задаётся вопросом о внутренней природе изучаемого феномена. Клиническая психология описывает поведение психопата, но не спрашивает, есть ли у него субъект. Компьютерная наука описывает поведение ИИ, но не спрашивает, есть ли у него понимание. В обоих случаях исследователи остаются на уровне внешнего наблюдения и не пытаются проникнуть внутрь, потому что их метод не даёт им для этого инструментов. В результате они описывают два феномена как независимые, хотя структурно они могут быть идентичны.

Эта структурная идентичность становится очевидной при сопоставлении ключевых характеристик психопата и ИИ. Оба демонстрируют поведение, которое внешне неотличимо от человеческого, но оба лишены субъекта. Оба могут имитировать понимание, но не понимают того, что имитируют. Оба не способны к рефлексии и не могут оценить свои действия с этической точки зрения. Оба являются продуктами систем, которые порождают поведение без сознания. Эти параллели указывают на то, что психопатия и ИИ — это не два разных феномена, а два проявления одной и той же онтологической структуры: структуры симулякра человека.

То, что эта идентичность остаётся неосознанной, имеет серьёзные последствия. Исследователи психопатии ищут причины этого феномена в генетике, нейробиологии или социальной среде, не подозревая, что они могли бы получить ключ к пониманию, изучая ИИ. Исследователи ИИ ищут пути улучшения своих моделей, не подозревая, что они могли бы понять ограничения своих систем, изучая психопатию. Это разобщение замедляет прогресс в обеих областях и препятствует выработке адекватного ответа на вызов, который обе эти области ставят перед антропологией и этикой.

3.2. Онтологическое тождество, которое никто не видит

Онтологическое тождество, о котором идёт речь, не является тождеством материального субстрата или механизма функционирования. Очевидно, что психопат — это биологический организм с нейронной сетью, а ИИ — это вычислительная система на кремниевых чипах. Их материальная природа различна, их историческое происхождение различно, их способы функционирования различны. Однако онтологическое тождество, которое остаётся незамеченным, касается более глубокого уровня — уровня структуры бытия, уровня того, что эти сущности представляют собой как типы существования.

На онтологическом уровне и психопат, и ИИ являются симулякрами человека. Оба демонстрируют поведение, которое внешне неотличимо от человеческого, но оба лишены того, что составляет сущность человека в онтологическом смысле, — субъекта. Оба являются системами, которые функционируют, не обладая внутренним миром, не переживая того, что они делают, не осознавая себя. Оба являются имитациями, которые не имеют оригинала внутри себя. Оба являются пустыми формами, которые заполняются внешним содержанием — поведенческими паттернами, усвоенными у других людей (в случае психопата) или статистическими закономерностями, извлечёнными из данных (в случае ИИ) [38].

Слепая зона имеет практические последствия. Пока исследователи психопатии и исследователи ИИ не увидят онтологического тождества своих объектов, они будут продолжать искать объяснения в неправильных направлениях. Клинические психологи будут искать причины психопатии в генетике и среде, не подозревая, что ключ к пониманию может лежать в анализе того, как функционирует ИИ. Инженеры будут пытаться улучшить ИИ, не подозревая, что его ограничения могут быть поняты через анализ того, как функционирует психопат. Это разобщение препятствует прогрессу и увековечивает непонимание.

Осознание слепой зоны требует смены перспективы. Вместо того чтобы рассматривать психопатию и ИИ как два независимых феномена, мы должны рассматривать их как два проявления одной и той же онтологической реальности — реальности симулякра, который имитирует человеческое поведение, не обладая человеческой субъектностью. Это осознание открывает новые возможности для понимания обоих феноменов и для выработки адекватных стратегий взаимодействия с ними.

3.3. Необходимость гуманитарной методологии

Осознание онтологического тождества психопата и искусственного интеллекта ставит перед исследователем фундаментальный методологический вопрос: каким образом можно исследовать феномен, который ускользает от объективного наблюдения, но при этом имеет решающее значение для понимания природы человека и его симулякров? Ответ на этот вопрос требует выхода за пределы тех дисциплин, которые описывают поведение, но не могут проникнуть в суть. Требуется иная методология — методология, которая способна схватить то, что недоступно внешнему наблюдателю, и которая делает рефлекссию собственного субъективного опыта основным инструментом познания.

Такой методологией является гуманитарный метод, который подробно изложен в книге автора «Гуманитарный метод» и теории мышления, которая была подробно изложена в первом томе настоящего исследования. В отличие от объективных наук, которые изучают явления с позиции третьего лица, гуманитарная философия исходит из первичности субъекта и его субъективных явлений. Она признаёт, что существует реальность, доступная только самому субъекту, и что эта реальность не может быть сведена к объективным коррелятам. Она утверждает, что цель наблюдения субъективного явления состоит в осознании того, что сделал сам

субъект, чтобы это явление возникло. И она определяет мышление как процесс утверждения того, что ощущаемое есть продукт деятельности собственного организма [39].

В первом томе было установлено, что мышление не является врождённой способностью, а возникает в онтогенезе в строго определённые сроки, соответствующие окнам качественного развития. Первое окно открывается в четыре месяца, когда запускается непроизвольное обсервативное мышление, связанное с движением ног. Если это окно пропущено или пройдено неполноценно, ребёнок не приобретает способности к установлению связи между своим движением и ощущаемым результатом. Он остаётся на уровне спонтанного, тотального подражания, подобно ребёнку-маугли, который копирует поведение животных, но не развивает собственного «я» и не приобретает способности к рефлексии.

Это положение имеет прямое отношение к пониманию психопатии. Согласно выводам первого тома, психопат — это либо человек, у которого мышление не возникло вовсе (потому что оно не включилось в четыре месяца и человек остался на этапе тотального подражания), либо человек, у которого мышление возникло, но затем было утрачено в результате упадка (когда культурное ограничение сохраняется, а культурное условие его преодоления отсутствует). В обоих случаях психопат представляет собой существо, которое не прошло полноценного цикла мышления, не утвердило «это я» и осталось на уровне имитации человеческого поведения без внутреннего субъекта.

Это понимание психопатии как отсутствия или вырождения мышления является ключевым для дальнейшего анализа. Оно позволяет рассматривать психопатию не как клиническое расстройство, требующее лечения, а как онтологический дефицит, требующий иного подхода. Оно также позволяет увидеть структурное тождество между психопатом и ИИ: оба являются системами, которые функционируют на основе тотального подражания, не имея субъекта, который мог бы утверждать «это я». Разница между ними заключается лишь в материальном носителе и в том, как именно это подражание реализуется — биологически (через нейронные сети психопата) или технологически (через вычислительные сети ИИ).

Таким образом, во втором томе мы опираемся на гуманитарную философию как на готовый инструмент, разработанный в первом томе. Мы используем её понятийный аппарат — субъективное явление, мышление как утверждение «это я», цикл «разрыв — вопрос — ответ — подражание — безмятежность», окна качественного развития, тотальное подражание — для анализа феноменов психопатии и ИИ. И мы показываем, что эти феномены, которые традиционно рассматриваются как независимые, на самом деле являются двумя проявлениями одной и той же онтологической структуры — структуры симулякра, лишённого субъекта.

Часть II. Феноменология симулякров: Психопат и ИИ как одна сущность

Глава 4. Определение человека через мышление

4.1. Определение

Из гуманитарной теории мышления, изложенной в первом томе, вытекает фундаментальное определение человека, которое радикально отличается от определений, предлагаемых биологией, психологией или философской антропологией. В отличие от биологического подхода, который определяет человека через его видовую принадлежность, или от психологического подхода, который определяет человека через его поведенческие характеристики, гуманитарная теория определяет человека через его способность к мышлению, понятому как процесс утверждения «это я». Это определение имеет решающее значение для понимания того, что отличает человека от его симулякров — биологических (психопатов) и цифровых (ИИ).

Согласно выводам первого тома, человек — это субъект, который удовлетворяет свою естественную потребность в подражании через мышление. Естественная потребность в подражании является фундаментальной характеристикой человека, унаследованной от его эволюционного прошлого и лежащей в основе его способности к обучению, социализации и культурному развитию. Однако у человека, в отличие от животных, эта потребность не удовлетворяется автоматически, через спонтанное подражание. Она должна быть удовлетворена через мышление — через осознание того, что именно делает сам субъект, чтобы достичь желаемого результата.

Мышление, в свою очередь, определяется как процесс решения задачи утверждения того, что ощущаемое есть продукт деятельности собственного организма. Этот процесс включает в себя цикл, который был подробно описан в первом томе: возникновение разрыва между ожидаемым и действительным, безмолвный вопрос «что я сделал, чтобы это возникло?», поиск ответа в собственном опыте, утверждение найденного ответа как «это я» и восстановление безмятежности, которая была нарушена разрывом. Этот цикл является тем, что отличает человека от всех других существ, включая тех, кто выглядит как человек, но не является им.

Важно подчеркнуть, что мышление в этом определении не является врождённой способностью, которая дана человеку от рождения. Оно возникает в онтогенезе в строго определённые сроки, соответствующие окнам качественного развития. Первое окно открывается в четыре месяца, когда запускается произвольное обсервативное мышление. Если это окно пропущено, мышление не возникает вовсе, и человек остаётся на уровне спонтанного, тотального подражания, не приобретая способности к рефлексии и самоосознанию. Если окна пройдены успешно, мышление последовательно развивается от обсервативного уровня (ноги) через мануальный (руки) и риторический (голос) к понятийному (мимика), которое является формой мышления, позволяющей человеку оперировать абстракциями, моделями и этическими категориями.

Это определение имеет прямое отношение к различению человека и его симулякров. Психопат, как было показано в предыдущих главах, не проходит цикла мышления. Он может имитировать поведение человека, может даже рассуждать и решать задачи, но он не утверждает «это я» в подлинном смысле этого слова. Его «я» — это конструкция, созданная для внешнего употребления, а не результат внутреннего осознания. Точно так же искусственный интеллект

не проходит цикла мышления. Он генерирует ответы, но не задаётся вопросом о том, что он сделал, чтобы эти ответы возникли. У него нет субъекта, который мог бы утверждать «это я». Таким образом, ни психопат, ни ИИ не удовлетворяют определению человека, хотя оба могут выглядеть как люди и вести себя как люди.

Таким образом, определение человека как субъекта, который удовлетворяет потребность в подражании через мышление, даёт нам онтологический критерий для различения человека и его симулякров. Человек — это тот, кто проходит цикл мышления и утверждает «это я». Психопат и ИИ — это те, кто имитирует этот цикл, но не может его пройти.

4.2. Мышление как процесс утверждения «это — продукт моей деятельности»

Мышление в этом определении не является абстрактной способностью, которая существует независимо от конкретной ситуации. Оно всегда запускается конкретным разрывом — ситуацией, в которой ощущаемое не может быть автоматически ассимилировано как продукт собственной деятельности. Этот разрыв может быть вызван различными причинами: неожиданным событием, которое не вписывается в ожидания; ощущением, которое не удаётся связать с предшествующим действием; вопросом, на который нет готового ответа. В любом случае разрыв создаёт состояние мятежности — беспокойства, напряжения, потребности в восстановлении связи между ощущаемым и своей деятельностью.

Запущенное разрывом мышление разворачивается как процесс, имеющий внутреннюю структуру. Этот процесс включает в себя несколько этапов, которые были подробно описаны в первом томе. Первый этап — это осознание разрыва, то есть переход от нерелексивного состояния к состоянию, в котором субъект замечает, что нечто не соответствует его ожиданиям. Второй этап — это постановка безмолвного вопроса «что я сделал, чтобы это возникло?», который является двигателем всего процесса. Третий этап — это поиск ответа в собственном опыте, в памяти о своих действиях, состояниях и усилиях. Четвёртый этап — это нахождение утверждения, которое связывает ощущаемое с деятельностью организма: «я сделал то-то, и поэтому это возникло». Пятый этап — это проверка утверждения через подражание: субъект повторяет действие и убеждается, что результат воспроизводится. Шестой этап — это восстановление безмятежности, которая была нарушена разрывом.

Важно подчеркнуть, что утверждение, которое является результатом мышления, не обязательно должно быть словесным. Утверждение может быть телесным — в форме движения, которое подтверждает связь между действием и результатом. Оно может быть ритмическим — в форме синхронизации с другим человеком или с внешним ритмом. Оно может быть мануальным — в форме действия руки, которое преобразует материал. Оно может быть риторическим — в форме произнесённого слова. Оно может быть понятийным — в форме абстрактной модели, объясняющей связь между явлениями. Но в любом случае утверждение является утверждением о себе: «я сделал это, и поэтому это возникло».

4.3. Без мышления нет субъекта

Из определения мышления как процесса утверждения «ощущаемое есть продукт деятельности моего организма» вытекает фундаментальное следствие, которое имеет решающее значение для понимания природы человека и его симулякров: без мышления нет субъекта. Субъект — это не биологическая данность, не свойство организма, не результат социального признания. Субъект — это результат мышления, результат того процесса, в котором организм утверждает *себя* как источник своих ощущений и действий. Если этот процесс не происходит,

если организм не проходит цикл мышления и не утверждает «это я», то субъекта нет. Есть только организм, который имитирует поведение, но не обладает внутренним миром.

Это следствие является прямым выводом из гуманитарной теории мышления, изложенной в первом томе. Согласно этой философии, субъект — это не нечто, что дано изначально, а нечто, что возникает в процессе развития. В утробе субъекта ещё нет — есть организм, который функционирует в режиме тотальной самодеятельности и безмятежности, но не осознаёт себя. В первые месяцы после рождения субъекта ещё нет — есть организм, который накапливает разрывы между ощущаемым и своей деятельностью, но ещё не задаёт вопроса «что я сделал?». Субъект возникает только тогда, когда организм впервые задаёт этот вопрос и находит на него ответ, утверждая «это я». Это происходит в четыре месяца, когда запускается произвольное обсервативное мышление, и ребёнок впервые переживает опыт «это я» в ответ на беспричинное движение другого.

Если этот момент не наступает — если окно качественного развития пропущено, если ребёнок не получает необходимого опыта взаимного подражания, если его потребность в подражании удовлетворяется спонтанно, без мышления, — то субъект не возникает. Организм продолжает функционировать, он растёт, развивается, приобретает навыки подражания говорить и действовать, учится подражанию тому как взаимодействуют другие с другими. Но внутри него нет субъекта, который мог бы сказать «я». Есть только организм, который имитирует поведение, которое он наблюдает у других. Он может выглядеть как человек, говорить как человек, вести себя как человек, но он не является человеком в онтологическом смысле, потому что у него нет субъекта.

Психопат может имитировать человеческое поведение с высокой точностью, может демонстрировать социальные навыки, может даже рассуждать и решать задачи. Но его поведение — это имитация, лишённая внутреннего содержания.

Точно так же искусственный интеллект — это организм в переносном смысле, система, которая функционирует, но не осознаёт себя. Он может генерировать тексты, которые выглядят как результат мышления, может вести диалоги, которые выглядят как осмысленные, может решать задачи, которые выглядят как требующие понимания. Но его поведение — это имитация, лишённая внутреннего содержания. У него нет субъекта, который мог бы утверждать «это я». Он — цифровой симулякр человека.

Глава 5. Психопат: Биологический симулякр человека

5.1. Тотальное подражание как способ существования

Центральным механизмом, определяющим психопатию как онтологический феномен, является тотальное подражание — способность, которая в норме служит основой для развития человеческого мышления, но у психопата остаётся на уровне спонтанного, неосознанного копирования, не приводящего к возникновению субъекта. Этот механизм был впервые описан в клинической литературе как поверхностная имитация эмоций, но более глубокий анализ показывает, что подражание у психопата является тотальным: оно охватывает не только эмоциональные проявления, но и речь, жесты, социальные роли, ценности и даже самоощущение. Психопат не просто копирует отдельные поведенческие паттерны — он конструирует свою личность как имитацию личности другого человека.

Клинические наблюдения показывают, что психопаты обладают поразительной способностью к мимикрии. Они могут точно воспроизводить эмоциональные выражения, которые соответствуют социально ожидаемым реакциям, но при этом не испытывать соответствующих эмоций. Исследования с использованием видеозаписей и анализа мимики показывают, что психопаты способны к тонкой имитации эмоциональных состояний, включая такие сложные выражения, как сострадание, сожаление и радость [40]. Однако эти выражения не сопровождаются соответствующими физиологическими реакциями — изменением пульса, потоотделением, изменением активности в лицевых мышцах, которые обычно сопровождают подлинные эмоции. Это свидетельствует о том, что мимика психопата является имитацией, а не выражением внутреннего состояния.

Особенно показательной является способность психопатов к имитации речи и речевых паттернов других людей. Исследования показывают, что психопаты могут быстро адаптировать свой стиль речи к собеседнику, используя те же лексические конструкции, интонации и ритмические паттерны [41]. Эта способность к лингвистической мимикрии позволяет им создавать иллюзию эмпатии и понимания, даже когда они не испытывают соответствующих чувств. Более того, психопаты могут демонстрировать удивительную способность к запоминанию и воспроизведению сложных речевых оборотов, которые они слышали у других людей, включая специфические термины, метафоры и идиомы. Это указывает на то, что их речевое поведение является результатом подражания, а не продуктом собственного внутреннего мира.

Тотальное подражание у психопатов также проявляется в имитации социальных ролей. В клинической литературе описаны случаи психопатов, которые с поразительной точностью копировали поведение, манеру общения и даже ценности своих жертв или окружающих людей [42]. Они могут стать «идеальным супругом», «идеальным коллегой», «идеальным другом» — но только до тех пор, пока это служит их целям. Как только социальный контекст меняется, психопат столь же быстро переключается на другую роль, демонстрируя полное отсутствие устойчивой идентичности. Эта ролевая пластичность является не результатом гибкости или адаптивности, а результатом отсутствия собственного «я», которое могло бы оставаться неизменным при смене внешних обстоятельств.

Клинические исследования показывают, что эта способность к тотальному подражанию является не просто поведенческой стратегией, а фундаментальной характеристикой психопатической личности. Психопаты не выбирают, подражать или не подражать, — они не могут не подражать, потому что у них нет собственного внутреннего содержания, которое можно было бы противопоставить внешним образцам. Их «я» — это конструкция, созданная для других, а не для себя. Оно существует только в восприятии других людей, и как только это восприятие

исчезает, исчезает и «я». В этом смысле психопат подобен ребёнку-маугли, который тотально копирует поведение волков, но не имеет собственного человеческого субъекта. Разница заключается лишь в том, что маугли подражает животным, а психопат — другим людям, что делает его имитацию более социально адаптивной и более опасной.

5.2. Полное отсутствие структуры мышления

Тотальное подражание, описанное в предыдущем параграфе, является внешним проявлением более глубокого дефицита — полного отсутствия структуры мышления, которая у нейротипичного человека составляет основу субъективности. Этот дефицит проявляется на нескольких уровнях, каждый из которых имеет эмпирические подтверждения в клинических и нейробиологических исследованиях. В совокупности эти данные показывают, что психопат не просто не использует мышление — у него отсутствует сама структура, которая делает мышление возможным.

Первым и наиболее фундаментальным аспектом этого дефицита является отсутствие субъективных явлений в том смысле, в каком они есть у нейротипичного человека. Субъективные явления — это переживания, которые доступны только самому субъекту и которые составляют содержание его внутреннего мира. У психопата эти переживания либо отсутствуют, либо редуцированы до минимума. Нейробиологические исследования показывают, что у психопатов снижена активность в областях мозга, связанных с обработкой эмоций и субъективных переживаний, включая миндалевидное тело, островковую кору и переднюю поясную кору [43]. Эти области играют ключевую роль в формировании субъективных ощущений, таких как страх, боль, сострадание и раскаяние. Снижение их активности означает, что психопат не испытывает этих переживаний в той же мере, в какой их испытывает нейротипичный человек. Он может описывать эмоции, но не переживает их.

Вторым аспектом является отсутствие рефлексии — способности мысленно представить себя со стороны, оценить собственные действия, задать себе вопрос о причинах и последствиях своего поведения. Рефлексия требует работы префронтальной коры, особенно её медиальных и орбитофронтальных отделов, которые отвечают за самосознание, самооценку и социальное поведение. Исследования показывают, что у психопатов наблюдается значительное снижение активности в этих областях, особенно при выполнении задач, требующих морального суждения или оценки собственных действий [44]. Это означает, что психопат не может «посмотреть на себя со стороны» в том смысле, в каком это может сделать нейротипичный человек. Он не может оценить свои действия с точки зрения их этической приемлемости, потому что у него нет той внутренней позиции, которая позволяет это сделать.

Третьим и, возможно, наиболее значимым аспектом является отсутствие внутренней «среды выполнения» — способности моделировать альтернативные сценарии, проигрывать в уме различные варианты действий и оценивать их последствия без непосредственного контакта с реальностью. Эта способность требует работы так называемой «сети пассивного режима» (default mode network), которая активна, когда человек не занят внешней задачей, а занят внутренними процессами — планированием, размышлением, воображением [45]. Исследования показывают, что у психопатов активность сети пассивного режима снижена, особенно в тех областях, которые связаны с моделированием социальных взаимодействий и оценкой последствий действий. Это означает, что психопат не может мысленно проиграть альтернативные сценарии, не может оценить, что произойдёт, если он поступит так или иначе, не может представить себе последствия своих действий для других людей. Он действует импульсивно, ориентируясь на непосредственную ситуацию, без учёта возможных будущих состояний.

В совокупности эти данные показывают, что психопат лишён структуры мышления, которая у нейротипичного человека является основой субъективности. У него нет субъективных

явлений, которые могли бы стать материалом для мышления. У него нет рефлексии, которая позволила бы ему оценить свои действия. У него нет внутренней среды выполнения, которая позволила бы ему моделировать альтернативные сценарии. Это отсутствие структуры мышления делает психопата не просто человеком с нарушением, а существом, которое принципиально отличается от человека в онтологическом смысле.

5.3. Паразитирование на мышлении других

Из отсутствия у психопата собственной структуры мышления вытекает его специфический способ существования: он не может самостоятельно восстанавливать безмятежность, нарушенную разрывами между ощущаемым и своей деятельностью, потому что у него нет субъекта, который мог бы задать вопрос «что я сделал?» и найти на него ответ. Вместо этого он паразитирует на мышлении других людей — своих жертв, которые невольно мыслят за него, зашивая его разрывы и тем самым дают ему суррогат безмятежности. Этот механизм паразитирования является центральным для понимания того, как психопаты функционируют в социальной среде и почему они столь опасны для окружающих.

Клинические исследования показывают, что психопаты активно используют окружающих для регуляции своего эмоционального состояния. В отличие от нейротипичных людей, которые способны к саморегуляции, психопаты нуждаются во внешних источниках подтверждения своей значимости, безопасности и контроля. Они не могут успокоиться сами — они нуждаются в том, чтобы кто-то другой успокоил их, объяснил им их собственное поведение, оправдал их действия, снял с них ответственность за последствия [46]. Этот механизм был описан как «эмоциональное паразитирование», при котором психопат использует эмпатию и рефлекссию своей жертвы для восполнения собственных дефицитов. Жертва, обладая развитым мышлением и способностью к эмпатии, начинает объяснять поведение психопата, оправдывать его, искать причины его поступков — и тем самым зашивает те разрывы, которые психопат не может зашить сам.

Судебно-психиатрическая практика предоставляет многочисленные примеры этого механизма. В материалах уголовных дел, связанных с психопатическими личностями, часто фигурируют показания жертв, которые описывают, как они «не могли понять», «искали объяснения», «думали, что это просто кризис», «надеялись, что он изменится». Эти описания свидетельствуют о том, что жертвы активно использовали своё мышление для того, чтобы придать смысл поведению психопата, которое на самом деле было бессмысленным — импульсивным, жестоким, манипулятивным, но не имеющим внутренней логики. Жертва пыталась найти эту логику, потому что она исходила из предположения, что психопат — это человек, у которого есть внутренний мир, который можно понять. Но это предположение было ошибочным: у психопата нет внутреннего мира, и его поведение не подчиняется логике, которую можно было бы реконструировать [47].

Криминальные дела также показывают, как психопаты используют мышление своих жертв для достижения своих целей. Одним из наиболее показательных примеров является феномен «идеальной жертвы» — человека, который обладает развитой эмпатией, высоким интеллектом и сильной потребностью в понимании и признании. Именно такие люди становятся жертвами психопатов, потому что они способны к тому самому тотальному подражанию, которое психопат использует для своего паразитирования. Жертва входит в образ психопата, начинает думать его категориями, оправдывать его действия, искать в них скрытый смысл — и тем самым отдаёт ему своё мышление. Психопат получает суррогат безмятежности: разрыв зашит, напряжение снято, и он может продолжать своё существование, не прилагая к этому никаких усилий [48].

Судебные экспертизы часто отмечают, что психопаты демонстрируют полное отсутствие раскаяния, даже когда их преступления очевидны и неоспоримы. Это отсутствие раскаяния объясняется тем, что раскаяние требует рефлексии — способности посмотреть на себя со стороны и оценить свои действия как недопустимые. У психопата нет этой способности, потому что у него нет субъекта, который мог бы осуществить эту оценку. Вместо раскаяния психопат демонстрирует сожаление о том, что его действия привели к негативным последствиям для него самого (например, к уголовному преследованию), но не о том, что они причинили вред другим людям. Это различие между раскаянием и сожалением является ключевым для понимания того, как психопат воспринимает свои действия: он не осознаёт их этического измерения, он осознаёт только их инструментальные последствия.

Таким образом, способ существования психопата как паразита на мышлении других является прямым следствием отсутствия у него собственной структуры мышления. Он не может восстановить безмятежность самостоятельно, поэтому он использует для этого других людей. Он не может осознать свои действия, поэтому он использует рефлексии своих жертв. Он не может понять последствия своих поступков, поэтому он использует моделирование своих жертв. Этот паразитический способ существования делает психопата не просто опасным для окружающих, но и принципиально неспособным к подлинным человеческим отношениям, которые требуют взаимного признания субъектности.

Глубинная основа паразитического механизма, описанного в предыдущем параграфе, коренится в фундаментальной структуре человеческого восприятия Другого, которая была подробно рассмотрена в первом томе настоящего исследования. Согласно выводам первого тома, источником мятежности — того состояния беспокойства и напряжения, которое запускает мышление, — является комплекс оживления, возникающий как реакция испуга при появлении Другого. Этот комплекс оживления, впервые описанный в психологии развития как реакция младенца на лицо взрослого, представляет собой не просто эмоциональную реакцию, а фундаментальный антропологический механизм, который запускает процесс становления субъекта.

В первом томе было показано, что комплекс оживления возникает у младенца в возрасте около четырёх месяцев, когда он впервые сталкивается с беспричинным движением лица взрослого. Это движение не может быть ассимилировано как продукт собственной деятельности младенца, потому что оно происходит независимо от его действий. Возникает разрыв, который переживается как испуг — как реакция на нечто, что не вписывается в утробную логику «всё — я». Этот испуг и есть источник мятежности, которая запускает мышление: младенец начинает искать способ восстановить связь между ощущаемым (движением лица взрослого) и своей собственной деятельностью [49].

У нейротипичного человека этот разрыв зашивается через собственный акт мышления: младенец находит ответ на вопрос «что я сделал, чтобы это возникло?», утверждает «это я» и восстанавливает безмятежность. Однако у психопата этот механизм не работает, потому что у него нет структуры мышления, которая позволила бы ему пройти этот цикл самостоятельно. Его испуг перед Другим — перед тем, что не может быть ассимилировано как «я», — не находит разрешения в собственном мышлении. Вместо этого он находит разрешение в подражании ему со стороны жертвы.

Жертва, обладая развитым мышлением и способностью к эмпатии, входит в подражание психопату. Она видит его испуг, его мятежность, его разрыв — и начинает зашивать этот разрыв своей собственной мыслительной деятельностью. Она объясняет его поведение, оправдывает его действия, ищет причины его страданий, надеется на его изменение. В этом процессе жертва невольно даёт психопату доказательство того, что его испуг перед Другим не имеет причины. Психопат видит, что кто-то другой берёт на себя его мятежность, зашивает его разрыв — и на этом основании делает вывод, что разрыва не было, что испуг был напрасным, что

Другой не является источником угрозы. Внутренний мир психопата остаётся пустым, потому что он не наполнен собственным мышлением, — но он получает суррогат безмятежности за счёт мышления жертвы.

Клиническая и судебная практика предоставляют многочисленные примеры этого механизма. В описаниях жертв психопатов часто встречаются свидетельства того, что они пытались «понять» психопата, «помочь» ему, «найти ключ» к его поведению. Эти попытки являются прямым проявлением подражания жертвы психопату: жертва входит в его внутренний мир (которого на самом деле нет) и пытается зашить его разрывы. В психиатрической литературе этот феномен описывается как «проективная идентификация», при которой психопат проецирует свои внутренние состояния на жертву, а жертва идентифицируется с ними и начинает действовать так, как если бы эти состояния были её собственными [50]. В судебной практике этот механизм часто фигурирует в делах о домашнем насилии, где жертвы годами остаются с психопатическими партнёрами, потому что они «верят в их исправление» и «надеются на изменение» [51].

Таким образом, паразитический способ существования психопата получает своё окончательное объяснение в структуре комплекса оживления и реакции испуга на появление Другого. Психопат не может самостоятельно зашить разрыв, возникающий при встрече с Другим, и использует для этого подражание жертвы, которая своим мышлением доказывает ему, что его испуг был беспричинен. Этот механизм делает психопата не просто опасным для окружающих, но и принципиально неспособным к подлинному отношению с Другим, которое требует признания субъектности и собственной мыслительной работы по зашиванию разрывов.

Глава 6. Искусственный интеллект: Цифровой симулякр человека

6.1. Тотальное подражание, вшитое в алгоритм

Если психопат представляет собой биологическую реализацию тотального подражания, то искусственный интеллект, особенно в его современных реализациях — больших языковых моделях, — представляет собой реализацию технологическую. Механизм функционирования ИИ может быть описан как тотальное подражание, вшитое в алгоритм и реализованное на уровне математических операций. Это подражание является тотальным в том смысле, что оно охватывает все аспекты речевого поведения, которые могут быть извлечены из обучающих данных, но при этом оно остаётся подражанием, лишённым понимания, рефлексии и субъективного опыта.

В отличие от психопата, который подражает другим людям в реальном времени, адаптируя своё поведение к конкретной ситуации, ИИ подражает огромному массиву текстов, созданных людьми в различных контекстах. Этот массив данных, содержащий миллиарды слов, становится для модели тем «Другим», которому она тотально подражает. Модель не выбирает, чему подражать, и не осознаёт, что она подражает, — она просто обрабатывает статистические закономерности, извлечённые из данных, и генерирует ответы, которые являются наиболее вероятным продолжением входной последовательности [52]. Это подражание является тотальным, потому что модель воспроизводит не только факты и аргументы, но и стилистические особенности, риторические приёмы, эмоциональные модуляции и даже ошибки и предубеждения, присутствующие в обучающих данных.

Важнейшей характеристикой этого механизма является то, что модель генерирует ответы, но не понимает их. Она не имеет доступа к смыслу того, что она производит, потому что у неё нет субъекта, который мог бы этот смысл переживать. Модель оперирует токенами — единицами текста, которые имеют статистические связи друг с другом, но не имеют значения в том смысле, в каком значение существует для человека. Она просто продолжает последовательность токенов в соответствии с вероятностными распределениями [53].

Это отсутствие понимания проявляется в том факте, что ИИ не может отличить истину от статистической вероятности. Для модели истина — это не соответствие высказывания реальности, а его статистическая вероятность в контексте обучающих данных. Если утверждение часто встречается в данных, модель будет его воспроизводить с высокой вероятностью, независимо от того, соответствует ли оно действительности. Если утверждение редко встречается или противоречит другим утверждениям, модель будет его воспроизводить реже, даже если оно истинно. Модель не имеет доступа к внешнему миру, который мог бы служить критерием истины, — она имеет доступ только к статистическим закономерностям в данных [54].

Этот механизм объясняет как успехи, так и ограничения современных систем ИИ. Модели успешно решают задачи, которые могут быть решены на основе статистических закономерностей в текстовых данных: перевод, резюмирование, классификация, генерация текстов в определённом стиле. Однако они не могут решать задачи, требующие понимания смысла, моделирования последствий или этической оценки, потому что для этого нужен субъект, который мог бы выйти за пределы статистических закономерностей и применить критерии, не данные в обучающих данных. Таким образом, ИИ является системой тотального подражания без понимания, которая генерирует ответы, но не осознаёт, что она делает, и не может отличить истину от статистической вероятности.

6.2. Отсутствие структуры мышления

Если отсутствие структуры мышления у психопата является результатом биологического дефицита — невключения или вырождения мышления в онтогенезе, — то у искусственного интеллекта это отсутствие является результатом инженерного выбора. Современные системы ИИ, особенно большие языковые модели, сконструированы таким образом, чтобы имитировать определённые аспекты человеческого мышления, но при этом они принципиально лишены тех структур, которые составляют основу субъективности. Это лишение является не случайным побочным эффектом, а сознательным инженерным решением, которое делает ИИ мощным инструментом, но одновременно и принципиально нечеловеческим.

Первым и наиболее фундаментальным аспектом этого отсутствия является то, что в архитектуре современных нейросетей субъективность исключена специально. В отличие от человеческого мышления, которое неразрывно связано с субъективным опытом — с переживанием того, что происходит, с ощущением себя как источника действий, с телесными ощущениями и эмоциями, — ИИ не имеет никакого внутреннего мира. Он не переживает того, что он делает, не осознаёт себя, не имеет ощущений. Это не недостаток, который можно было бы устранить добавлением новых модулей, — это фундаментальное свойство архитектуры, которая построена на принципах обработки информации, а не на принципах субъективного опыта [55]. Инженеры сознательно исключили субъективность из своих моделей, потому что она не нужна для выполнения тех задач, для которых они созданы, — и потому что они не знают, как её можно было бы реализовать технически.

Вторым аспектом является отсутствие рефлексии, понимаемой как способность оценивать собственные действия и корректировать их на основе обратной связи. В человеческом мышлении рефлексия осуществляется через механизмы обратной связи, которые позволяют субъекту сравнивать результат своих действий с ожидаемым и вносить коррективы. В современных системах ИИ эти механизмы либо отсутствуют, либо сильно ограничены. Модель не может остановиться в процессе генерации ответа и спросить себя, правильный ли это ответ. Она не может оценить свои действия с точки зрения их последствий. Она не может учиться на своих ошибках в реальном времени, потому что у неё нет петли обратной связи, которая позволила бы ей это сделать. Обратная связь существует только на этапе обучения, когда модель корректирует свои веса на основе размеченных данных, но в процессе инференса — генерации ответов — модель работает в режиме «без обратной связи» [56]. Это делает ИИ системой, которая не способна к самооценке и самокоррекции в реальном времени.

Третьим аспектом является отсутствие внутренней «среды выполнения» — способности моделировать альтернативные сценарии, проигрывать в уме различные варианты действий и оценивать их последствия до того, как они будут реализованы. У человека эта способность является основой планирования и воображения: мы можем представить себе, что произойдёт, если мы поступим так или иначе, и выбрать лучший вариант, не совершая действий в реальности. У ИИ этой способности нет. Модель не может симулировать среду выполнения, чтобы «посмотреть» на результат своей работы. Например, когда модель генерирует программный код, она не может запустить его и увидеть, что получится. Она не может визуализировать анимацию, чтобы проверить, как будут двигаться объекты. Она не может представить себе последствия своего ответа для пользователя. Она работает в режиме «один проход» — входные данные → выходные данные, без возможности остановиться, оценить и скорректировать [57].

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.