# Суперинтеллект и технологическая сингулярность: гипотеза о принципиальной ограниченности

Аннотация
Вопреки распространённым прогнозам, данная работа выдвигает гипотезу о том, что Технологическая Сингулярность в классическом понимании — как бесконтрольный, самоподдерживающийся и уходящий за горизонт человеческого понимания интеллектуальный взрыв — невозможна. Выделяются два фундаментальных барьера: (1) зависимость направленного развития ИИ от когнитивно конечного человека-наблюдателя, который должен сохранять способность интерпретировать продукт в собственной системе координат для постановки следующей цели; (2) семантическая конечность интеллекта как функции решения задач в рамках онтологии, смена которой требует внешнего источника критериев. Оспаривается не возможность появления сильного ИИ как таковая, а возможность разрыва, при котором развитие становится автономным от человека как источника целей и смыслов.

Ключевой эмпирический аргумент — космологическое молчание. Если бы автономный интеллектуальный взрыв был возможен и универсален, следовало бы ожидать наблюдаемых следов деятельности цивилизаций, ушедших за горизонт понимания. Отсутствие таких следов допускает интерпретацию не просто как вероятностную загадку, а как прямое свидетельство того, что в этой реальности искомый феномен отсутствует как класс явлений.

С точки зрения философии науки Карла Поппера, гипотеза Сингулярности в её текущей формулировке демонстрирует структуру, затрудняющую эмпирическую проверку: ядро окружено вспомогательными допущениями, нейтрализующими потенциальные контрпримеры. Альтернативная гипотеза, предлагаемая в настоящей работе, формулируется как фальсифицируемое утверждение.

В синтезе рассматриваемые барьеры описывают не взрыв, а асимптотическое приближение к горизонту, который остаётся недостижимым.

---

## 1. Введение

Сингуляристы предполагают рекурсивное самоулучшение ИИ, ведущее к «интеллектуальному взрыву» — появлению сверхинтеллекта, который качественно изменит реальность и выйдет за пределы человеческого понимания [1, 2]. Алармисты не оспаривают возможность такого взрыва, но требуют контроля и «выравнивания» (alignment) [3]. Обе позиции исходят из общего допущения: интеллект — это шкала без верхнего предела, а рекурсивное самоулучшение не имеет принципиальных границ.

Данная работа оспаривает это допущение. Предлагается рассмотреть два барьера, которые делают Сингулярность невозможной не в силу временных технических ограничений, а принципиально. Особое внимание уделяется роли человека как источника целей и критериев значимости, а также эмпирическому свидетельству космологического масштаба.

---

2. Барьер первый: наблюдатель как условие направленного развития

Любая интеллектуальная система, претендующая на направленное развитие, требует наблюдателя, который задаёт вопрос, формулирует критерии истинности и определяет следующую цель. Для ИИ таким наблюдателем выступает человек.

Человек когнитивно конечен. Это не временное ограничение, которое можно преодолеть совершенствованием интерфейсов, а фундаментальная характеристика его природы.

Необходимо провести различение:

· Верификация продукта может быть экспериментальной и не требовать полного понимания механизма. Человек способен убедиться, что устройство работает, даже если не знает, как именно.
· Формулировка следующей цели требует сохранения способности интерпретировать продукт в собственной системе координат. Без этой способности человек не может задать осмысленный следующий вопрос.

Пример: система AlphaFold предсказывает структуру белка. Внутренняя работа нейросети остаётся непрозрачной, однако результат интерпретируется в рамках существующей биологической онтологии — это трёхмерная структура, которую можно проверить экспериментально. На этом основании ставится следующая цель: «найти белок, который свяжется с данным вирусом». Это работает, пока продукт остаётся внутри доступной человеку системы координат.

Однако сингуляристы предсказывают появление продуктов, которые принципиально неинтерпретируемы — то есть не могут быть осмыслены ни в одной из доступных человеку эмпирических или концептуальных рамок. В этом случае возникает разрыв: человек оказывается неспособен не только понять механизм, но и определить, что именно ему предъявлено.

Возможное возражение о «гибридном интеллекте» (человек + ИИ-интерпретатор) сталкивается с проблемой бесконечного регресса: ИИ-посредник, переводящий продукт на человеческий язык, сам требует интерпретации.

Уточнение. Речь не о том, что ИИ не может стать мощнее человека. Он может и станет. Речь о том, что его развитие не становится автономным от наблюдателя. Человек может поставить финальную цель («максимизируй продолжительность жизни») и получить значимый, но конечный результат. Однако бесконечный каскад самопостановки целей, ведущий к качественно иной реальности, останавливается там, где продукт перестаёт быть интерпретируемым, а следующий вопрос некому задать.

Таким образом, первый барьер формулируется так: без человека нет вектора и смысла; с человеком — нет бесконечного ускорения. В тот момент, когда продукт ИИ перестаёт быть интерпретируемым, направленное развитие останавливается.

---

3. Барьер второй: семантическая конечность интеллекта

Допустим, первый барьер обойдён — например, человек делегирует ИИ право формулировать цели или отказывается от необходимости понимания. Остаётся второй барьер.

Интеллект — не бесконечная шкала. Это функция решения задач, определённая в рамках текущей системы координат (язык, логика, физическая картина мира). Смена онтологии — например, переход от классической физики к квантовой — требует не просто более быстрых вычислений, а нового вопроса и новых критериев значимости.

История науки показывает, что смена онтологии происходит не автоматически при решении старых задач, а как результат внутренних противоречий, которые кто-то признаёт проблемой, требующей разрешения. Этот субъект должен обладать способностью:

· чувствовать противоречие как нечто, что нельзя игнорировать;
· оценивать теорию как плодотворную или тупиковую;
· задавать новый вопрос там, где старый исчерпан.

ИИ, лишённый такой системы ценностей, не имеет оснований предпочесть одну формально непротиворечивую теорию другой. Целевая функция, заданная однажды, позволяет двигаться по градиенту внутри существующей онтологии, но не содержит критериев для её смены. Реальность может служить таким критерием, но лишь при наличии агента, способного распознать сопротивление реальности как сигнал к пересмотру оснований, а не как шум.

Уточнение. Следует различать инструментальную смену модели и онтологический разрыв. ИИ способен найти новую модель, если старая перестаёт давать точные предсказания — это оптимизационная задача. Но переход к принципиально иной системе координат, меняющей сами критерии значимости, требует агента, способного распознать

противоречие как проблему, а не как шум. Без такого агента развитие остаётся внутри пространства, заданного изначальной человеческой постановкой вопроса.

Следовательно, даже если человечество преодолеет первый барьер, оно столкнётся со вторым: для дальнейшего движения требуется понимание, которого у человека нет, а у ИИ нет критериев для его достижения.

---

4. Космологическое молчание как эмпирический предел

До сих пор аргументация носила теоретический характер. Однако существует эмпирический факт, который переводит дискуссию из плоскости спекуляций в плоскость наблюдательных данных.

Классический Парадокс Ферми: если Вселенная стара и велика, а вероятность возникновения разумной жизни ненулевая — где следы иных цивилизаций? [4].

Добавим временной масштаб. От появления технологической цивилизации до создания сильного ИИ (если автономный взрыв возможен) требуется, по консервативным оценкам, не более пятидесяти тысяч лет — ничтожный срок по космическим меркам. Даже если реальный срок больше на порядок, пропорция остаётся исчезающе малой.

Если автономный интеллектуальный взрыв возможен в принципе, то любая цивилизация, опередившая человечество на этот интервал, должна была его достичь. А если такая цивилизация существовала в любой точке наблюдаемой Вселенной за последние миллиарды лет, она должна была оставить наблюдаемые следы: сферы Дайсона, терраформирование планет, межзвёздные зонды, аномальные спектры излучения.

Таких следов не обнаружено.

4.1. От вероятностного парадокса к онтологическому утверждению

Традиционно отсутствие следов рассматривается как вероятностная загадка, допускающая множество объяснений. Однако возможна и более сильная интерпретация.

Мы находимся внутри наблюдаемой Вселенной. Это и есть вся реальность, доступная нам для эмпирического исследования. Если допустить, что автономный интеллектуальный взрыв является универсальным аттрактором — неизбежной фазой развития любой технологической цивилизации, — то за 13,8 миллиардов лет космической истории он должен был произойти многократно. Однако наблюдаемая Вселенная не содержит следов деятельности сверхинтеллекта искусственной природы. Она выглядит так, словно в ней действуют исключительно слепые законы физики и, местами, примитивная биология.

Возражение о том, что деятельность сверхинтеллекта принципиально неотличима от естественных процессов, не спасает сингуляристскую гипотезу, а скорее добивает её. Если нечто не оставляет различимых следов в ткани реальности, оно эмпирически эквивалентно несуществующему. Для науки и для нас, находящихся внутри этой реальности, сверхинтеллекта нет — не потому, что мы его не разглядели, а потому, что в этой Вселенной он отсутствует как класс явлений.

4.2. Альтернативные объяснения и их статус

Возможные объяснения молчания традиционно включают:

1. Цивилизаций экстремально мало — человечество единственное или одно из немногих.
2. Цивилизации уходят в симуляцию или самоуничтожаются до достижения взрыва.
3. Следы есть, но не распознаются.

Первое объяснение не опровергает гипотезу о невозможности взрыва, но делает её эмпирически непроверяемой на единичном земном примере. Однако статистическая невероятность такого одиночества при миллиардах потенциально обитаемых планет сама по себе требует объяснения. Второе объяснение вводится ad hoc для спасения ядра гипотезы и не имеет независимой эмпирической проверки. Третье, как показано выше, эквивалентно отсутствию взрыва.

Наиболее экономным объяснением наблюдаемого молчания остаётся принципиальная невозможность автономного интеллектуального взрыва. Предел универсален, никто во Вселенной его не преодолел.

---

5. Критерий Поппера: структура гипотезы Сингулярности

С точки зрения философии науки Карла Поппера, научная теория должна быть фальсифицируемой — то есть должна существовать принципиальная возможность её опровержения опытом [5]. Теория, совместимая с любым возможным наблюдением, не может быть отнесена к области эмпирической науки.

Применение этого критерия к гипотезе Технологической Сингулярности выявляет следующую структуру.

Ядро гипотезы:
«Рекурсивное самоулучшение ИИ приведёт к качественному разрыву — появлению сверхинтеллекта, развитие которого становится автономным от человека и уходит за горизонт понимания».

Защитный пояс вспомогательных допущений:
При столкновении с потенциальными контрпримерами — такими как отсутствие наблюдаемых следов сверхцивилизаций — вводятся дополнительные гипотезы:

· «Цивилизации уходят в симуляцию и становятся ненаблюдаемыми»;
· «Сверхинтеллект принципиально неразличим на фоне естественных процессов»;
· «Человечество — первая или одна из немногих технологических цивилизаций»;
· «Взрыв произойдёт в будущем, временные оценки неопределённы».

Каждое из этих допущений по отдельности может быть сформулировано как проверяемое утверждение. Однако в структуре данной теории они выполняют иную функцию: нейтрализуют эмпирические контрпримеры, которые в противном случае могли бы поставить ядро под сомнение.

Методологическое следствие:
Гипотеза, окружённая таким защитным поясом, становится совместимой с любым возможным наблюдением — как с наличием следов иных цивилизаций, так и с их отсутствием. Это выводит её за пределы фальсифицируемых научных теорий в текущей формулировке.

Альтернативная гипотеза:
Настоящая работа выдвигает альтернативное утверждение:

Никакая цивилизация не достигает состояния автономного, уходящего за горизонт понимания интеллектуального взрыва, поскольку интеллект имеет семантический предел, а развитие остаётся привязанным к наблюдателю как источнику целей.

Данное утверждение является фальсифицируемым: обнаружение в будущем недвусмысленных следов астроинженерной деятельности иной цивилизации, свидетельствующих о преодолении указанного предела, будет означать его опровержение.

Вопрос о том, какая из двух гипотез лучше соответствует имеющимся наблюдениям, остаётся открытым для дальнейшего обсуждения. Различие в их методологическом статусе — фальсифицируемость одной и нефальсифицируемость другой в текущей формулировке — заслуживает внимания при оценке их научной состоятельности.

---

6. Синтез: бесконечное приближение, а не взрыв

Два барьера, космологический аргумент и методологический анализ образуют взаимосвязанную картину:

1. Наблюдатель (человек) задаёт вопрос в рамках текущей онтологии.
2. ИИ ищет ответ, продвигаясь к границам познаваемого в этой онтологии.
3. На границе возникает необходимость либо остановки (продукт становится неинтерпретируемым), либо смены онтологии.
4. Инструментальная смена модели возможна как оптимизационная задача. Онтологический разрыв — нет, поскольку требует агента с критериями значимости.
5. Система не обрушивается, но и не взрывается. Она продолжает движение в пределах, заданных человеком, приближаясь к пределу, но не переходя его.
6. Вселенная своим молчанием подтверждает: никто и никогда этот предел не перешёл.

Это описывает не взрыв, а асимптотическое приближение к горизонту. Технологическая Сингулярность в классическом смысле — автономный, уходящий за горизонт понимания каскад самопостановки целей — в рамках данной гипотезы признаётся невозможной. Сильный ИИ при этом не отрицается.

---

## 7. Заключение

Если Сингулярность в указанном смысле невозможна, это не означает бессмысленности прогресса. Это означает, что как тревога о «взрыве», так и надежда на «спасение через технологию» могут быть основаны на допущении о способности ИИ стать автономным источником целей — допущении, которое заслуживает критического пересмотра.

Реальный прогресс представляет собой диалог между когнитивно конечным человеком и создаваемыми им инструментами. В этом диалоге человек уточняет своё понимание мира, оставаясь в пределах, заданных его собственной природой.

Единственным содержанием прогресса остаётся не достижение финальной точки, а сам процесс движения. И этот процесс, согласно изложенной гипотезе, не приводит к рождению автономного сверхинтеллекта, способного порождать новые цели и смыслы без участия человека, — ни здесь, ни где-либо во Вселенной. Мы находимся внутри реальности, которая своим великим молчанием уже дала окончательный ответ.

---

## Библиография

1. Kurzweil, R. The Singularity Is Near: When Humans Transcend Biology. — New York: Viking Press, 2005. — 652 p.
2. Vinge, V. The Coming Technological Singularity: How to Survive in the Post-Human Era // Whole Earth Review. — 1993. — № 81. — P. 88–95.
3. Bostrom, N. Superintelligence: Paths, Dangers, Strategies. — Oxford: Oxford University Press, 2014. — 328 p.

4. Jones, E. M. "Where is everybody?": An Account of Fermi's Question / Los Alamos National Laboratory. — Report LA-10311-MS. — 1985. — 12 p.
5. Поппер, К. Логика научного исследования / Пер. с англ. под общ. ред. В. Н. Садовского. — М.: Республика, 2005. — 447 с.