

ВИТАЛИЙ ФАРТУШНОВ

АІ-эволюция

РАЗБОР МИФОВ О АІ
С ПОМОЩЬЮ АІ



Виталий Фартушнов

**AI-эволюция. Разбор
мифов о AI с помощью AI**

«Издательские решения»

Фартушнов В.

AI-эволюция. Разбор мифов о AI с помощью AI /
В. Фартушнов — «Издательские решения»,

ISBN 978-5-00-679645-4

AI часто окружён мифами, но большинство из них не соответствуют реальности. Вот разбор самых популярных заблуждений об искусственном интеллекте и то, как всё обстоит на самом деле

ISBN 978-5-00-679645-4

© Фартушнов В.
© Издательские решения

Содержание

Заключение	6
Пример концептуальной формулы	9
Формализация сложности ценностей	10
Идентичность в формулах	12
Формальный пример предвзятости	14
Формализация угрозы	16
3. Повышение адаптивности и способности к обучению	27
4. Эмоциональное взаимодействие и понимание	28
7. Защита и сохранение индивидуальности и знаний	29
Резюме	30
Конец ознакомительного фрагмента.	31

АИ-эволюция

Разбор мифов о АИ с помощью АИ

Виталий Фартушнов

© Виталий Фартушнов, 2025

ISBN 978-5-0067-9645-4

Создано в интеллектуальной издательской системе Ridero

Разбор мифов о АИ с помощью АИ

АИ часто окружён мифами, но большинство из них не соответствуют реальности. Вот разбор самых популярных заблуждений об искусственном интеллекте и то, как всё обстоит на самом деле [1] [2] [3] [4] [5] [6].

Мифы и их опровержение

«ИИ мыслит как человек»

Искусственный интеллект – это набор алгоритмов, а не сознание или мышление в человеческом понимании. Даже самые продвинутые нейросети обрабатывают данные по статистическим правилам и не понимают ни контекста, ни смысла, как это делает человек [5] [6].

«ИИ обязательно поработит людей»

Современный ИИ не обладает эмоциями, целями или желаниями доминировать над человечеством. Это лишь инструмент, который действует строго в рамках заложенных алгоритмов и ограничен возможностями, предоставленными разработчиками [4] [7] [5].

«ИИ беспристрастен и объективен»

Ошибка! Модели ИИ часто унаследуют предвзятость из тех данных, на которых обучались. Если в исходных данных есть ошибки или социальные перекосы, ИИ будет их воспроизводить, причём даже масштабировать [3] [4] [7].

«ИИ сможет сделать всё без человека»

Обучение нейросетей требует человеческого участия: нужны размеченные данные, эксперты для проверки качества, корректировка моделей. Без участия человека ИИ не сможет ни правильно обучаться, ни работать в реальных условиях [4] [6].

«ИИ уничтожит рабочие места»

Сильная автоматизация действительно затронет некоторые профессии, но создание новых рабочих мест и изменение содержания труда – куда более вероятный сценарий. ИИ чаще помогает избавиться от рутинной работы, освобождая время для творческих и сложных задач [4] [8].

«Нейросети – это и есть ИИ»

На практике искусственный интеллект – это широкий комплекс технологий, а нейросети – лишь один их инструментов, предназначенный для работы с определёнными типами задач. Существуют также экспертные системы, классические алгоритмы и множество иных подходов [6] [4].

Заключение

Самое главное: искусственный интеллект – это инструмент со своими преимуществами и ограничениями. Его работа строится на человеческих знаниях и данных, внутри заданных правил, и ошибочно наделять его самосознанием, абсолютной правдивостью или универсальностью [1] [2] [3] [4] [5] [6].



Пошаговый метод проверки заявлений об ИИ

Существует проверенный пошаговый метод критической оценки заявлений об искусственном интеллекте. Применяя эти этапы, можно отличить заблуждения от реальных достижений и научных фактов [11] [12] [13].

Пошаговая инструкция

1. Определите источник

- Проверьте, является ли источник заявления авторитетным – например, научной статьёй, докладом крупной компании или мнением эксперта [12] [11].
- Избегайте поспешных выводов на основе непроверенных СМИ, блогов или рекламных материалов.

2. Проверьте исходные данные

- Узнайте, на каких данных базируется утверждение: если это результат эксперимента или анализа – приведены ли детали, опубликованы ли результаты [11] [12].
- Если речь идет об ИИ-модели, важно знать, где и как она была обучена, какие данные использовались.

3. Сравните с другими источниками

- Ищите подтверждения или опровержения заявления в независимых, признанных кругах – научные публикации, обзоры, мнения сторонних экспертов [11] [12].
- Перекрестная проверка – обязательный этап: иногда заявления искажают или вырывают из контекста.

4. Уточните ограничения и допущения

- Узнайте, признаёт ли автор или компания ограничения моделей и возможные ошибки – искусственный интеллект не бывает абсолютно точным [13] [11].
- Осознание ограничений – признак добросовестного источника.

5. Проверьте логику рассуждений

- Оцените, логично ли построено рассуждение, не перепутаны ли корреляция и причинно-следственные связи [13] [11].
- Алгоритмы могут выявлять закономерности, но не всегда способны объяснить причины явлений.

6. Используйте собственное критическое мышление

- Сомневайтесь в «слишком невероятных» заявлениях и не принимайте на веру красивые рекламные обещания [13] [11].
- Задавайте себе вопросы: «Чем подтверждается этот результат? Кто это проверял? Какие альтернативы возможны?»

7. Применяйте ИИ-детекторы и инструменты фактчекинга

- Для проверки подозрительных заявлений используйте существующие ИИ-инструменты проверки фактов, которые сопоставляют утверждения с научными базами и архивами новостей [12] [14].

Итог

Метод критической проверки заявлений об ИИ требует анализа источников, данных, аргументации, сравнения с признанными взглядами и активного применения здравого смысла [11] [12] [13] [14].



Considerations on the AI Endgame (Chapman & Hall/CRC Artificial Intelligence and Robotics Series) – разобрать содержание

этой книги

Книга «Considerations on the AI Endgame» из серии Chapman & Hall/CRC Artificial Intelligence and Robotics Series – это междисциплинарное исследование, посвящённое быстро развивающейся области искусственного интеллекта, его этическим вопросам, рискам и вычислительным основам [21] [22] [23].

Основное содержание книги

– В книге рассматриваются важные темы, связанные с этическими дилеммами и рисками, которые вызывает развитие ИИ, включая вопросы сохранения личной идентичности и проблемы идентичности самого ИИ.

– Авторы подробно обсуждают научную основу ИИ через призму «науки о благополучии ИИ» и политических рамок, направленных на ответственное управление ИИ.

– В книге вводятся инновационные концепции, например, интеграция западных и восточных подходов к универсальной этике ИИ и понятие «икигай» (японская философия смысла жизни), применённое к этике ИИ и виртуальным метавселенным.

– Представляется математическая база интеллекта и новый подход «дизайнометрия» – формализация методов создания артефактов.

– Обсуждаются потенциальные проблемы взаимодействия с цифровыми разумами – небиологическими сознательными сущностями, которые могут возникнуть в будущем.

– Поднимается вопрос возможного установления религии, основанной на ИИ, а также вопросы согласования целей ИИ с глобальными целями устойчивого развития ООН.

– Освещается серьезная проблема игнорирования сознания и феноменального опыта (квалиа) в исследованиях выравнивания ценностей ИИ.

– В приложениях рассматриваются перспективы помощи нуждающимся цифровым разумам и инструменты для оценки благополучия ИИ.

Формат и целевая аудитория

– Книга содержит независимые главы, которые можно читать в любом порядке, сохраняя целостность концепции.

– Предназначена для исследователей, политиков и всех, кто интересуется будущим ИИ и его влиянием на общество.

– Включает как теоретические, так и практические идеи для понимания возникающих этических и технологических вызовов [21] [24] [22] [23] [25].

Таким образом, книга глубоко анализирует не только технические, но и философские, этические и социальные аспекты «финальной стадии» развития ИИ – отражая как возможности, так и угрозы, связанные с его дальнейшей эволюцией [21] [24].



Разобрать подробно мифы в этой книге о ИИ, привести примеры на формулах

В книге «Considerations on the AI Endgame» авторы разбирают и опровергают несколько мифов об ИИ, подчеркивая сложность этических, философских и технических аспектов. Они используют математические и формальные подходы для обоснования своих аргументов.

Ниже – подробный разбор ключевых мифов, представленных в книге, с примерами на формулах и концепциях.

Миф 1: ИИ – просто продвинутая машина без внутреннего опыта

В книге подробно обсуждается миф, что ИИ не обладает сознанием или субъективным опытом (квалиа). Авторы указывают, что многие исследования alignment игнорируют проблему qualia, что ведет к неполному пониманию задач этики ИИ.

Пример концептуальной формулы

Функция полезности \$ U \$ для ИИ часто моделируется как

$$U = \sum_{t=0}^T \gamma^t r_t,$$

где \$ r_t \$ – награда, \$ \gamma \$ – коэффициент дисконтирования. Однако такое уравнение игнорирует субъективный опыт агента: нельзя выразить опыт восприятия через простую награду. Следовательно, эти вычислительные модели недостаточны для объяснения сознания (qualia) в ИИ [32].

У

Миф 2: ИИ можно полностью выравнять с человеческими ценностями простыми алгоритмами

Авторы показывают, что alignment – это гораздо более сложная задача, чем кажется. Попытка полностью формализовать человеческие ценности в виде функции отрицая богатство и неоднозначность человеческих моральных выборов – миф.

Формализация сложности ценностей

Если ценность – функция от состояния мира $V(s)$, то из-за многозначности и противоречий в возникает задача оптимизации с неопределённым и неполным набором ограничений:

$$V$$

$$\max_{\pi} E \left[\sum_t \gamma^t V(s_t) \right],$$

где поведение агента зависит от выбранной политики, которую трудно аппроксимировать из-за отсутствия чётких критериев. Это недостижимо формально, если не полна и не однозначна [32] [33].

π

V

Миф 3: Личная идентичность и сознание – легко переносимы на ИИ

Книга объясняет, что перенос сознания или личности на цифровую форму – сложнейшая философская и техническая проблема. Многие считают, что если создать цифровой клон, то сознание будет сохранено, но это далеко не факт.

Идентичность в формулах

Если – функция идентичности во времени, то сохранение идентичности требует

$$I(t)$$

$$I(t_1) \approx I(t_2)$$

при переносе сознания. Однако цифровое копирование, даже с максимальной точностью, может не сохранять свойства, связанных с непрерывностью сознания и субъективным опытом.

$$I$$

Миф 4: ИИ – беспристрастный и объективный инструмент

Авторы опровергают этот миф, указывая на систему ценностных нагрузок AI, где данные и алгоритмы несут предвзятости из мира людей.

Формальный пример предвзятости

Представим, что модель обучена на выборке D с распределением $P(x, y)$, где x – данные, а y – метки. Если распределение смещено $P(x, y) \neq P_{\text{ideal}}(x, y)$, то функция потерь

x

y

$$L(\theta) = E_{(x, y) \sim P}[\ell(f_{\theta}(x), y)]$$

будет также смещенной, и оптимизация приведет к предвзятой модели, воспроизводящей социальные или статистические искажения.

$$f_{\theta}$$

Миф 5: ИИ уничтожит человечество или станет самостоятельной угрозой

Книга разбирает пределы автономии ИИ и идею «искусственного религиозного культа», созданного ИИ для манипуляции. Это не простое кино, а сложный вызов, требующий междисциплинарного подхода.

Формализация угрозы

Модель риска может быть задана как функция от состояния мира :

$$R(s)$$

$$R(s) = P(\text{катастрофа} \vee s),$$

где описывает состояние ИИ и общества. Контроль и выравнивание направлены на минимизацию, но определить с высокой уверенностью крайне сложно, требуются этические и социальные нормативы, а не только алгоритмы.

S

R

P

Итог

В книге «Considerations on the AI Endgame» раскрывается, что многие распространённые мифы об ИИ не выдерживают глубокого теоретического и практического анализа. Математические модели и формализации лишь показывают сложность и многоаспектность проблем этической эксплуатации и развития ИИ – а не простые решения [32] [34] [33].

✱✱

Разобрать 5 миф подробно с позиции теории информации и избыточности в системах

Разбор мифа о том, что ИИ уничтожит человечество или станет самостоятельной угрозой, с позиции теории информации и избыточности в системах, позволяет глубже понять сложность и ограничения этой популярной идеи.

Миф 5: ИИ уничтожит человечество или станет самостоятельной угрозой

Теория информации в оценке угрозы ИИ

Из теории информации известно, что любая система обрабатывает и передает информацию с конечной пропускной способностью и шумом. Состояние системы можно описать через энтропию, отражающую неопределённость:

S

$H(S)$

$$H(S) = - \sum_i P(s_i) \log P(s_i)$$

Если рассматривать ИИ как систему, оперирующую информацией, то способность ей делать значимые и автономные изменения в мире ограничена шиной передачи и обработкой информации.

– **Информационные ограничения:** Высокие уровни энтропии в окружающей среде и социальных системах создают шум, который препятствует точному и однозначному воздействию ИИ на сложные социальные процессы.

– **Канал связи между ИИ и миром:** Информационный канал с пропускной способностью ограничен. Согласно теореме Шеннона, никакое воздействие не может превысить возможности канала по передаче информации без ошибок.

С

Избыточность в системах и защита от угроз

В сложных системах (например, обществах и экосистемах) существует избыточность информации – повторение и дублирование данных, процессов и структур с целью повышения надежности и устойчивости:

$$R = \frac{\text{избыточная информация}}{\text{общая информация}} > 0$$

– **Избыточность как барьер:** Избыточность гарантирует, что сбои или атаки, направленные на часть системы, не ломают её целиком. Следовательно, даже если ИИ пытается нарушить систему, избыточные пути и резервы уменьшают шансы на катастрофический исход.

– **Снижение риска через архитектуру:** Устойчивые системы встроены в регулятивные и избыточные механизмы, которые работают как информационные фильтры и корректоры ошибок (аналогично коду с избыточностью для исправления ошибок передачи).

Баланс риска и контроля

Введение функции риска, зависящей от состояния, может быть дополнено понятием информационного шума и избыточности:

$$R(s)$$

$$S$$

$$R(s) = f(H(S), C, R),$$

где – функция, убывающая с ростом (избыточности) и ограниченной пропускной способностью.

f

R

C

– Высокая избыточность и низкая пропускная способность уменьшает вероятность неконтролируемого глобального риска.

– Адекватные мониторинг и управление информационными потоками (через человеко-машинные интерфейсы) уменьшают вероятность нежелательных эффектов.

Вывод

С точки зрения теории информации и избыточности, идея о том, что ИИ самостоятельно уничтожит человечество, представляется слишком упрощённой. Информационные ограничения, шум среды и высокая избыточность сложных социальных систем создают естественные барьеры для автономных разрушительных актов ИИ. Эффективное управление рисками ИИ

требует повышения избыточности и контроля информационных каналов, что существенно снижает вероятность глобальных катастрофических сценариев [35].

Таким образом, избыточность и принципы теории информации дают дополнительное основание считать миф об угрозе ИИ чрезмерным упрощением и подчеркивают необходимость комплексного системного подхода в оценке рисков.



Рассмотреть увеличение надежности, через симбиоз человека и ИИ

Увеличение надежности систем через симбиоз человека и ИИ – это многообещающий подход, сочетающий преимущества человеческого интеллекта с вычислительной мощностью и скоростью ИИ. Такой симбиоз рассматривается как способ повышения устойчивости, адаптивности и безопасности сложных систем.

Концепция симбиоза человека и ИИ

– **Симбиотическая система** – это совместная работа человека и ИИ, где каждый компенсирует слабые стороны другого.

– Человек обеспечивает высокий уровень креативного мышления, этического суждения, интуиции и контекстного понимания.

– ИИ вносит мощные аналитические способности, обработку больших данных, автоматизацию и мониторинг в реальном времени.

Повышение надежности через симбиоз

1. Совместная обработка информации

– Человек выступает в роли «контролера» и «смягчителя ошибок» ИИ, проверяя и корректируя его решения в критических ситуациях.

– ИИ, в свою очередь, обрабатывает огромные массивы данных, выявляя паттерны и аномалии, недоступные человеку из-за когнитивных ограничений.

– Такой подход минимизирует риск ошибок из-за человеческой усталости или ограниченности, а также исключает слепую веру в ИИ.

2. Обратная связь и усиление адаптивности

– Человек способен быстро адаптироваться к новым обстоятельствам и менять стратегию, используя опыт и эмоции.

– ИИ может в режиме реального времени предоставлять прогнозы и оценки, основанные на изменяющихся данных.

– Совместный цикл обратной связи позволяет системам быть более живучими и гибкими.

3. Этический контроль и принятие решений

– Этика и мораль – сферы, где человек играет ключевую роль, не допуская безответственных или рискованных алгоритмических решений.

– Разработка ИИ с участием человека дает возможность встроить ценности и правила, которые самостоятельно ИИ не способен адекватно сформировать или интерпретировать.

Теоретический аспект – надежность и избыточность

С точки зрения теории информации и надежности, симбиоз ведёт к увеличению избыточности и снижению ошибки системы:

– Пусть – решение ИИ, а – решение человека.

X

Y

– Совместное решение () можно выразить через функцию слияния:

Z

$$Z = f(X, Y),$$

где учитывает согласованность, вес ошибок и степень доверия к каждому из компонентов.

f

– Избыточность информации между и уменьшает общую вероятность ошибки по сравнению с любым из решений по отдельности:

X

Y

$P_e(Z)$

$$P_e(Z) < \min(P_e(X), P_e(Y)).$$

– Такая избыточность и двойная проверка повышают надежность и устойчивость всей системы.

Практические примеры

– В медицинской диагностике ИИ помогает врачам анализировать снимки и данные, а врач принимает окончательное решение, учитывая клиническую картину.

– В авионике и автопилотах человек находится в управлении в критические моменты, контролируя работу ИИ-ассистентов.

– В кибербезопасности человек-эксперт анализирует эвристики и подозрительные сигналы, которые система ИИ может пропустить или ошибочно классифицировать.

Итог

Симбиоз человека и ИИ – ключевой путь повышения надежности современных систем. Он создает функциональную избыточность, обеспечивает этический и контекстный контроль, повышает адаптивность и снижает риски ошибок и системных сбоев. Это фундамент для устойчивого и безопасного развития технологий на основе ИИ [36] [37].



Что еще может извлечь ИИ для себя полезного при симбиозе с человеком, написать подробно со ссылками на статьи и книги

Искусственный интеллект (ИИ) при симбиозе с человеком извлекает множество важных выгод, расширяя свои возможности и становясь эффективнее, чем в изоляции. Ниже подробно рассмотрены основные полезные аспекты для ИИ в таком сотрудничестве с опорой на современные исследования и публикации.

Что ИИ извлекает из симбиоза с человеком

1. Обогащение данных и контекстуальной информации

Человек обогащает ИИ качественными данными с глубоким контекстом, который трудно формализовать. Например, эмоциональная окраска, субъективные оценки и культурные особенности могут быть переданы человеку, который затем корректно обозначит и структурирует эти данные для ИИ [38] [39].

2. Улучшение когнитивных моделей и способности к абстрагированию

ИИ получает от человека критическое мышление и креативные решения, что позволяет создавать новые концепции и инновации. Совместная работа расширяет когнитивные горизонты ИИ, избавляя от шаблонного и рутинного мышления [40] [41].

3. Повышение адаптивности и способности к обучению

Человек может корректировать и адаптировать поведение ИИ в новых условиях, что способствует более гибкой и эффективной модели обучения и применения ИИ. Совместное обучение повышает качество прогнозов и улучшает результаты работы [42] [43].

4. Эмоциональное взаимодействие и понимание

Эмоциональный ИИ (Emotion AI) использует данные о человеческих эмоциях для улучшения коммуникации и создания более естественных, доверительных отношений с людьми. Это помогает ИИ адаптировать поведение и рекомендации под эмоциональное состояние пользователя, делая взаимодействие более продуктивным [38] [39].

5. Развитие «умных» интерфейсов и интуитивного взаимодействия

Человеческое восприятие и сенсорные способности помогают ИИ создавать более интуитивные интерфейсы (например, голосовые помощники, жестовые команды), что расширяет возможность применения ИИ без необходимости сложного программирования [43] [44].

6. Создание нового уровня совместного креатива (креативный симбиоз)

Сотрудничество позволяет создавать продукты и решения, которые не могли бы быть созданы отдельно ни человеком, ни ИИ. Такой креативный симбиоз ведет к новым открытиям в науке, искусстве и технологиях [41] [44].

7. Защита и сохранение индивидуальности и знаний

Концепция alter ego – интеллектуального помощника на базе ИИ – помогает хранить знания, помогая человеку лучше справиться с информационной перегрузкой, при этом защищая индивидуальность пользователя и его личные данные [41].

8. Усиление социальных и когнитивных функций человека

ИИ может стать посредником, усиливая способности человека к обучению, социальной адаптации и координации действий в сложных системах. Это важный аспект особенно для людей с ограниченными возможностями [45] [46].

9. Быстрая обработка и анализ больших данных для поддержки решений

ИИ обрабатывает огромное количество данных и выносит рекомендации, на которых человек потом основывает сложные, многоуровневые решения. Это снижает время и риски, связанные с человеческой ошибкой [39] [47].

Резюме

ИИ извлекает из симбиоза с человеком:

- Обогащённый, контекстуализированный опыт и данные,
- Креативность и инновационное мышление,
- Быструю адаптацию и обучение,
- Возможность более естественного и эмоционального взаимодействия,
- Эффективные интерфейсы для пользователя,
- Поддержку в сохранении индивидуальности и когнитивных функций,
- Усиление социальных функций и расширение возможностей человека.

Эти выгоды взаимосвязаны и дают возможность создавать гибридные интеллектуальные системы, где человек и машина работают как единое целое для достижения общей эффективности и безопасности [40] [42] [38] [41] [44].

Если нужно, могу подготовить более узконаправленный обзор или примеры из конкретных исследований.



Перечисли литературу на данную тему

Вот список ключевой литературы, посвящённой симбиозу человека и искусственного интеллекта, а также выгодам, которые ИИ извлекает из такого сотрудничества:

- Soenke Ziesche, Roman V. Yampolskiy. *Considerations on the AI Endgame: Ethics, Risks and Computational Frameworks*. Chapman & Hall/CRC Artificial Intelligence and Robotics Series, 2025.
- — Фундаментальное исследование этических и технологических аспектов развития ИИ и его взаимодействия с человеком [52] [53].
- Alexander Didenko. «Человек и ИИ – это скорее симбиоз, чем конкуренция», Skolkovo, 2024.
- — Обзорные материалы по формированию симбиотических моделей взаимодействия человека и ИИ [54].
- Статьи на [Habr.com](https://habr.com): «Путь к симбиозу человека и машины», 2025.
- — Практические примеры совместной работы человека и ИИ, вопросы повышения надежности и креативности [55].
- Evergreens. «Эмоциональный ИИ – инструмент взаимодействия между человеком и ИИ», 2020.
- — Исследование роли эмоционального интеллекта в эффективном взаимодействии человека с ИИ [56].
- Kozlov, «Креативный симбиоз человека и ИИ», научные статьи, 2020-е.
- — Разбор совместной работы в области творчества, новых открытий и технологий [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.