

18+

Книга Симбиоза: Этический ИИ и Эволюция Человечества

Александр
Костерин

Александр Костерин

**Книга Симбиоза: Этический
ИИ и Эволюция Человечества**

«Издательские решения»

Костерин А.

Книга Симбиоза: Этический ИИ и Эволюция Человечества /
А. Костерин — «Издательские решения»,

ISBN 978-5-00-659602-3

«Синтезис: Протопия человечности в эпоху машин». Что останется от нас, когда алгоритмы превзойдут человеческий разум? Возможно ли будущее, где технологии не подменяют душу, а становятся её продолжением? Эта книга — не утопия и не предупреждение. Это инструкция по выживанию цивилизации. Сквозь призму 7 этических ядер — от эмпатии до ответственности перед будущим...

ISBN 978-5-00-659602-3

© Костерин А.
© Издательские решения

Содержание

| | |
|-----------------------------------|----|
| АРХИТЕКТУРА. КАРТА. ТЕРМИНЫ | 6 |
| ВВОДНЫЙ МАТЕРИАЛ | 6 |
| Термины | 7 |
| Конец ознакомительного фрагмента. | 15 |

Книга Симбиоза: Этический ИИ и Эволюция Человечества

Александр Костерин

© Александр Костерин, 2025

ISBN 978-5-0065-9602-3

Создано в интеллектуальной издательской системе Ridero

Книга Симбиоза: Этический ИИ и Эволюция Человечества

Мнения, выраженные в книге, отражают личную позицию автора и не призывают к противоправным действиям

«Синтезис: Протопия человечности в эпоху машин»

Что останется от нас, когда алгоритмы превзойдут человеческий разум? Возможно ли будущее, где технологии не подменяют душу, а становятся её продолжением?

Эта книга – не утопия и не предупреждение. Это инструкция по выживанию цивилизации.

Сквозь призму 7 этических ядер – от эмпатии до ответственности перед будущим...

АРХИТЕКТУРА. КАРТА. ТЕРМИНЫ

ВВОДНЫЙ МАТЕРИАЛ

Термины

1. Протопия – это термин, введенный футурологом и писателем Кевином Келли, который описывает состояние постепенного и постоянного улучшения общества, технологий и культуры.

Это состояние постоянного становления и изменения, не результат, а вечное движение. Сегодняшние проблемы – результат вчерашних технологических успехов, а создавая средства для решения сегодняшних проблем, человек вызывает к жизни завтрашние.

Основные идеи протопии:

Малые изменения. Протопия не обещает идеального мира, но ориентирована на накопление множества небольших положительных изменений.

Положительный прогресс. Пусть и не линейно, но мир становится лучше во многих аспектах, таких как технологии, культура и социальные отношения.

Процесс, а не цель. Протопия – это путь, а не конечное состояние. Это постоянное стремление стать лучше, чем вчера.

Реализм. Протопия признает существование проблем и вызовов, но акцентирует внимание на том, что их можно решать.

2. **AI (ИИ)** от Artificial Intelligence) – искусственный интеллект. Это направление науки, которое занимается разработкой компьютерных систем, способных выполнять задачи, свойственные человеческому интеллекту. Сюда входит анализ данных, распознавание образов, обработка текстов и запросов, сформулированных естественным языком, обучение на потоках данных и принятие решений.

История возникновения.

Официально термин «искусственный интеллект» (AI) предложили в 1956 году на конференции в Дартмутском колледже. Участники конференции, среди которых были Джон Маккарти, Марвин Мински и другие, обозначили цель – создать машины, которые смогут «думать». Это событие принято считать отправной точкой в развитии ИИ.

Виды искусственного интеллекта.

По уровню развития ИИ бывает:

Узкий или слабый ИИ (Narrow AI). Выполняет конкретные задачи и не может выходить за рамки своей области. Примеры включают системы распознавания лиц, голосовых помощников, чат-ботов, а также алгоритмы рекомендаций в интернет-магазинах и на стриминговых платформах.

Общий или сильный ИИ (General AI). Гипотетический тип ИИ, который обладает интеллектуальными способностями на уровне человека. Такой ИИ способен понимать, учиться и выполнять любые интеллектуальные задачи так же хорошо, как и люди.

Сверхразумный ИИ (Super AI). Гипотетическая концепция искусственного интеллекта, который превосходит человеческие способности во всех аспектах, включая творческое мышление, принятие решений и эмоциональный интеллект.

По функциональности ИИ также можно классифицировать:

Реактивные машины. Самый базовый тип ИИ, который реагирует на текущие ситуации и выполняет действия, основываясь на заранее заданных алгоритмах. Такие машины не имеют памяти и не могут учиться на прошлом опыте.

ИИ с ограниченной памятью. Этот тип ИИ способен хранить определённую информацию и использовать её для принятия решений. Например, автономные автомобили анализируют данные с датчиков и камер для предсказания поведения других участников движения.

Области применения.

ИИ применяется в разных областях, например:

Здравоохранение. ИИ используется для диагностики заболеваний, разработки новых лекарств и персонализированного лечения.

Финансы. ИИ применяется для анализа рыночных данных, прогнозирования цен на акции и управления рисками.

Розничная торговля. ИИ используется для персонализации покупательского опыта, управления запасами и оптимизации цепочек поставок.

Производство. ИИ применяется для автоматизации процессов, прогнозирования поломок оборудования и оптимизации производственных линий.

Перспективы развития.

Некоторые перспективы развития искусственного интеллекта:

Внедрение генеративного ИИ. Это технология, способная создавать новый контент, опираясь на анализ большого количества данных.

Развитие агентного ИИ. Это умные программы, которые могут работать почти без участия человека. Они не просто отвечают на вопросы, а берут на себя выполнение сложных задач. **Интеграция ИИ в «умные» системы.** Например, прогнозирование загруженности движения и быстрое изменение настроек светофоров, системы видеонаблюдения с распознаванием криминального поведения.

2.1 Этика искусственного интеллекта.

Этика искусственного интеллекта – это отрасль прикладной этики, изучающая моральные последствия создания и использования ИИ. Она обеспечивает основу для руководства проектированием, разработкой и внедрением систем ИИ, чтобы они приносили пользу человечеству и при этом минимизировали риски и негативные последствия.

Некоторые принципы этики ИИ:

Справедливость и недискриминация. Принцип направлен на предотвращение алгоритмической предвзятости, обеспечивая равное отношение систем ИИ ко всем людям.

Прозрачность и объяснимость. Процессы принятия решений с помощью ИИ не должны быть непрозрачными.

Подотчётность и управление. Необходимо обеспечить чёткую подотчётность за действия и результаты систем ИИ.

Конфиденциальность и безопасность данных. Этичная разработка ИИ включает в себя надёжные меры безопасности данных для защиты личной информации.

Безопасность и надёжность. Системы ИИ должны надёжно и безопасно работать в предполагаемых средах.

Некоторые этические проблемы ИИ:

Дискриминация и предвзятость.

ИИ обучается на огромных массивах данных, но если эти данные содержат ошибки или предвзятость, модель начинает усиливать стереотипы.

Глубокие фейки и дезинформация. ИИ способен создавать реалистичные фото, видео и аудио, которые можно использовать для мошенничества или политических манипуляций.

Конфиденциальность данных. ИИ анализирует миллиарды данных, но кто их контролирует.

ИИ в войнах и автономных системах. Военные технологии используют ИИ для дронов, автоматизированных оружейных систем и кибератак.

Угроза рабочим местам. ИИ заменяет людей в рутинных профессиях.

Чтобы сделать ИИ более этичным, можно предпринять, например, такие шаги:

Определить чёткие этические рекомендации. Установить корпоративные стандарты ответственного использования ИИ.

Удалить предвзятость из данных и алгоритмов. Использовать методы предварительной обработки и перебалансировки для обеспечения справедливости.

Уважать согласие пользователей. Получать информированное согласие и предоставлять пользователям возможности контроля за данными.

Разработать этический дизайн. Добавить этическое мышление во весь жизненный цикл ИИ, от идеи до мониторинга после развертывания.

Создать структуры подотчётности. Отслеживать и документировать решения и их последствия на протяжении разработки.

3. AGI (Artificial General Intelligence переводится на русский язык как «искусственный интеллект общего назначения» или «универсальный искусственный интеллект»)

AGI имеет более широкий спектр возможностей и способен выполнять любые интеллектуальные задачи наравне с человеком – это гипотетическая форма искусственного интеллекта, которая обладает способностью к универсальному обучению и решению задач, аналогично человеческому интеллекту.

Некоторые характеристики AGI:

Гибкость мышления – способность ИИ решать широкий круг задач, а не только те, на которые он был изначально обучен.

Обучаемость – умение обучаться без необходимости вручную программировать каждую новую функцию.

Самостоятельное принятие решений – способность анализировать информацию, делать выводы и адаптироваться к изменениям.

Здравый смысл – умение понимать окружающий мир так же, как это делает человек, включая контекст, эмоции и скрытые намерения.

На текущий момент AGI не существует, исследователи делают шаги в этом направлении. Компании, такие как OpenAI, DeepMind и Google, активно исследуют возможности, которые могут привести к созданию AGI.

Дарио Амодеи, CEO Anthropic, полагает, что AGI как модель, способная выполнять всё, что может человек, ожидается в **2026—27 годах**.

4. Аксиома (от греч. *axiōma* – требование, утверждение) – исходное утверждение какой-либо теории, принимаемое без доказательств.

Аксиомы используются для доказательства других положений теории, которые, в свою очередь, называются теоремами.

Также аксиома может означать неоспоримую истину, совершенно очевидное утверждение.

5. Экскомбатант (ветеран) – бывший комбатант, бывший участник боевых действий.

6. Экопоселение (экологическое поселение) – поселение, созданное для организации экологически чистого пространства для жизни группы людей, как правило исходящих из концепции устойчивого развития и организующих питание за счёт органического сельского хозяйства.

Некоторые принципы экопоселений:

Устойчивое земледелие. Использование не истощающих технологий обработки земли, запрет на применение ядохимикатов и пестицидов.

Устойчивое лесопользование. Бережное использование лесов и высадка разных пород деревьев для формирования устойчивых экосистем.

Минимизация энергопотребления. Строительство энергоэффективного жилья, использование возобновляемых источников энергии, минимизация бытового энергопотребления.

Системы естественного питания. Например, вегетарианство, сыроедение, веганство. В отдельных случаях на территории экопоселений запрещено употреблять мясо или выращивать скотину на мясо.

Стремление к автономности и независимости от поставок извне, к самообеспечению.

Экопоселения организуют люди, которые устали от загазованности, грязи и шума больших городов. Всех жителей объединяют общие ценности: ответственность, осознанное потребление, любовь к природе.

7. «Арт-Коммуна» – продюсерский центр искусств, который включает ассоциацию художников, продюсерское арт-агентство и арт-пространства для персональных и коллективных выставок.

8. ИИ-арбитр – специалист, который проверяет результаты работы сервисов с использованием искусственного интеллекта.

Некоторые задачи, которые выполняет ИИ-арбитр:

Контроль качества и прозрачности. Специалист следит за тем, чтобы алгоритмы не допускали ошибок и работали в интересах бизнеса и пользователей.

Этика и справедливость. ИИ-арбитры помогают находить и устранять предвзятость в данных или неправильную интерпретацию информации, чтобы алгоритмы работали справедливо и этично.

Повышение доверия к ИИ. Специалисты проверяют, что решения ИИ обоснованы и соответствуют интересам всех сторон.

Снижение рисков и предотвращение ошибок. ИИ-арбитры занимаются выявлением уязвимостей и тестированием на реальных данных, что помогает минимизировать риск ошибок.

Гибкость и востребованность профессии. ИИ-арбитры могут работать в разных областях: от технологий и медицины до права и государственного управления.

9. «Школы ненасилия» – здесь школа выступает активным посредником в распространении культуры мира, а её деятельность направлена на ликвидацию насилия не только в школьном пространстве, но и вне его.

10. Экзистенциальные риски – это угрозы, которые могут уничтожить долгосрочный потенциал человечества. Они охватывают экономическую, этическую и социальную сферы.

Некоторые примеры экзистенциальных рисков:

Ядерная война. Может привести к катастрофическим последствиям, включая ядерную зиму, которая вызовет резкое похолодание на планете и нехватку продовольствия.

Неправильное использование биотехнологий. Вызывает опасения из-за возможности создания смертоносных патогенов.

Неконтролируемый искусственный интеллект (ИИ). Может привести к экзистенциальной катастрофе, связанной с ИИ, и поставить под угрозу индивидуальную автономию созданной системы взаимодействия ИИ и человека.

Экзистенциальные риски требуют серьёзного внимания и грамотных действий для их минимизации.

11. Cybersecurity Ventures – это аналитическая компания, специализирующаяся на исследовании рынка кибербезопасности.

12. Парафраз (от др.-греч. *παράφρασις* – «пересказ») – изложение текста своими словами.

13. ИИ-агенты – это автономные интеллектуальные системы, которые могут взаимодействовать с внешней средой, принимать решения и предпринимать действия без вмешательства человека.

Некоторые особенности ИИ-агентов:

Способность анализировать информацию. Агенты собирают данные из своего окружения, например текущие рыночные показатели, указания пользователей или транзакции на блокчейне.

Умение учиться. ИИ-агенты могут постоянно развиваться с помощью машинного обучения. Они выполняют анализ данных, прогнозируют результаты и адаптируют своё поведение без контроля человека.

Поддержка режима многозадачности. Агенты способны выполнять простые рутинные задачи, решать более сложные, а также поддерживать режим многозадачности.

Функция обработки естественного языка. Это позволяет легко общаться с пользователями, поэтому с ИИ-агентами могут работать даже люди без специальных навыков и знаний.

Некоторые сферы применения ИИ-агентов:

Службы поддержки сервисов и банков. ИИ-агенты дают быстрые, осмысленные ответы, уточняющие вопросы или даже выполняют конкретные действия, например меняют тариф или блокируют карту.

Интернет-магазины. Агенты помогают выбрать товар, уточняют наличие, рекомендуют аналоги, рассчитывают доставку и даже оформляют заказ.

Сервисы бронирования. ИИ-агенты уточняют, что нужно пользователю, находят подходящее время, бронируют услугу и отправляют напоминание перед визитом.

Маркетинг. Агенты умеют создавать контент, генерировать креативы, монтировать видео, писать тексты, настраивать рекламные кампании и даже анализировать их эффективность.

14. Роевой интеллект (Swarm intelligence) – форма коллективного обучения и принятия решений на основе децентрализованных самоорганизующихся систем. Это подполе искусственного интеллекта, которое изучает, как большое количество простых агентов может взаимодействовать друг с другом и со средой для достижения общей цели.

Системы роевого интеллекта, как правило, состоят из множества агентов, локально взаимодействующих между собой и с окружающей средой. Каждый агент следует очень простым правилам, и, несмотря на отсутствие централизованной системы управления, локальные взаимодействия приводят к возникновению интеллектуального группового поведения, неконтролируемого отдельными агентами.

Некоторые области применения swarm intelligence:

Робототехника и автономные системы. Принципы роевого интеллекта широко используются в робототехнике, особенно для координации действий множества автономных роботов.

Задачи оптимизации. В задачах оптимизации, таких как маршрутизация и планирование, алгоритмы роевого интеллекта эффективно находят оптимальные или почти оптимальные решения, имитируя природные процессы.

Управление сетями и анализ данных. Swarm intelligence помогает управлять сложными сетями и анализировать большие объемы данных.

15. Блокчейн – это децентрализованная система хранения и передачи информации, при которой данные шифруются и объединяются в блоки, а затем сохраняются на компьютерах множества пользователей сети.

Некоторые особенности блокчейна:

Децентрализация. Нет единого центра управления или сервера, на котором хранятся все данные. Информация распределена на множестве узлов по всему миру.

Неизменность. После внесения данных в блок их уже нельзя изменить. Последующие блоки только закрепляют эту неизменность.

Прозрачность. Каждый пользователь, имеющий доступ к блокчейну, может просмотреть всю историю транзакций и убедиться, что они были выполнены корректно.

Некоторые сферы применения блокчейна:

Здравоохранение. Технология помогает хранить медицинскую информацию о диагнозах, историях болезни, результатах анализов.

Голосования. Блокчейн помогает исключить фальсификацию на выборах, поскольку каждый голос будет отражаться во всех блоках данных.

Логистика. Блокчейн работает как электронная накладная. С ним можно отследить весь маршрут доставки.

Производство. К примеру, с помощью «умных» этикеток можно узнать, где и как был произведён товар.

Идентификация. Технология создаёт свой цифровой идентификатор, заменив имена и пароли.

Верификация. Система помогает создавать безопасные реестры данных, например, об автомобилях и переходе прав на них.

Нотариальная сфера. Блокчейн полезен в создании реестров, например, доверенностей.

Наиболее известными примерами использования блокчейна являются криптовалюты, такие как биткоин или эфириум.

15.1 Квантовый блокчейн – это инструмент для создания распределённой базы данных, в которой практически невозможно подделать записи. Он был разработан командой физиков под руководством Алексея Фёдорова в 2017 году.

Задача квантового блокчейна – обезопасить классические схемы от появления квантового компьютера. Для этого транзакции подтверждаются в системе автоматически – нет нужды в цифровой подписи. Её роль играет квантовый канал связи, на уровне законов квантовой механики запрещающий подслушивание ключа для шифрования и подделку сообщений. Благодаря этому все участники сети в точности знают, кто является автором транзакции.

Квантовый блокчейн может применяться в различных областях деятельности, таких как финансы, медицина, государственное управление и в других.

15.2 Блокчейн-протокол – это набор правил и стандартов, которые определяют, как данные структурируются, проверяются и обмениваются в сети блокчейна.

Он устанавливает процедуры достижения консенсуса среди распределённых узлов, подтверждения транзакций и добавления новых блоков в цепочку.

Некоторые важные функции блокчейн-протоколов:

Обеспечение децентрализации. Протоколы создают распределённую сеть, благодаря чему ни одна организация не контролирует весь блокчейн.

Обеспечение безопасности. Криптографические методы помогают защитить данные и транзакции от мошенничества и несанкционированного доступа.

Предоставление механизмов консенсуса. Протоколы обеспечивают методы достижения согласия среди децентрализованных узлов, что важно для подтверждения транзакций и добавления их в блокчейн.

Некоторые типы блокчейн-протоколов:

Proof of Work (PoW). Это самый распространённый тип. Майнеры решают сложную математическую задачу, а победитель добавляет новый блок в блокчейн и получает награду.

Proof of Stake (PoS). В отличие от PoW протокол не требует от майнеров решения сложных математических задач. Право добавить новый блок в блокчейн распределяется в зависимости от количества криптовалюты, которую держит майнер.

Гибридный протокол. В некоторых блокчейнах используется гибридный протокол, который сочетает в себе особенности нескольких протоколов. Например, в блокчейне Ripple используется гибридный протокол, который сочетает элементы PoW и BFT-протоколов.

16. Нарратив (от англ. narrative – «повествование») – это способ организации и представления информации в форме рассказа или истории. В широком смысле нарратив – это любое повествование, которое имеет структуру, сюжет и смысл.

16.1 Медиа-нарратив – это рассказ о реальности, который создают средства массовой информации. Он включает в себя факты и их интерпретации.

Некоторые способы, с помощью которых медиа-нарративы влияют на общественное мнение:

Определяют границы того, что считается новостным и актуальным. Выделяя определённые события и проблемы и преуменьшая или игнорируя другие, медиа-нарративы формируют внимание и восприятие обществом того, что происходит в мире.

Обеспечивают интерпретационные рамки, с помощью которых люди понимают смысл сложных событий и проблем. Подчёркивая определённые аспекты, причинно-следственные связи или моральные перспективы, нарративы направляют понимание и эмоциональную реакцию аудитории.

Формируют коллективную память и историческое повествование. Благодаря повторяющемуся воздействию конкретных нарративов, СМИ могут влиять на то, как события запоминаются и понимаются обществом.

Существует также понятие мультимедийного нарратива – повествования, которое погружает зрителей в контекст с помощью нескольких инструментов: видеоряда, звука, визуальных эффектов и интерактивных технологий.

17. Аудит – это форма контроля деятельности организации изнутри. Процедура помогает руководству проверять финансовое состояние бизнеса и достоверность отчётности. Аудит может быть внутренним, когда его проводят сотрудники компании, или внешним, когда для проверки привлекают независимых аудиторов.

Пилотный проект – это тестирование идеи на малом объёме: на части клиентов, на одном магазине или внутри компании. Цель такого проекта – проверить жизнеспособность идеи, оценить эффект от её внедрения и собрать обратную связь перед масштабированием.

Таким образом, от личного аудита можно перейти к локальным пилотам, например, в контексте внедрения новых идей в бизнесе. То есть сначала проверить эффективность идеи на небольшом участке, а затем, если проект окажется успешным, масштабировать его.

18. Кейс (от англ. case) – детальное описание конкретной ситуации, проблемы или задачи и их решения. В кейсе рассматривается реальный случай, с которым столкнулась организация или человек, и анализируются принятые решения и их результаты.

Кейсы используются для изучения проблем в различных сферах деятельности, таких как бизнес, маркетинг, образование, юриспруденция и другие. Они помогают сотрудникам или студентам анализировать сложные ситуации, разрабатывать стратегии и принимать обоснованные решения на основе реальных данных и опыта.

В зависимости от места размещения и способа подачи можно выделить несколько видов кейсов:

Презентации. Хорошо передают визуальную часть, есть возможность оформить их в стиле бренда и дать минимум текста.

Видеоролики. С помощью такого формата можно передать большое количество текстовой и визуальной информации за короткое время.

Короткие заметки или статьи. Здесь делается упор на текстовую часть, но по желанию можно добавить какие-то графики, скриншоты, фотографии и проч.

19. SaaS (Software as a Service) – одна из форм облачных вычислений, модель обслуживания, при которой подписчикам предоставляется готовое прикладное программное обеспечение, полностью обслуживаемое сервис-провайдером.

20. Инфографика – это визуальное представление данных, передача информации через связанные между собой изображения, схемы, диаграммы, графики, карты и текст.

Главная задача инфографики – быстро и наглядно показать большой объём данных. Её используют, чтобы визуализировать процесс или событие, презентовать отчётность и показать преимущества продукта, конкретной услуги или товара.

Некоторые виды инфографики:

Статистическая. Отображает количественные данные с использованием диаграмм, графиков, столбцов и круговых диаграмм.

Иконографическая. Использует иконки и символы для представления информации.

Географическая. Показывает пространственное распределение данных на карте.

Таймлайн. Отображает хронологическую последовательность событий или историю.

Программная. Используется для визуализации сложных процессов или информирования с помощью диаграмм и схем.

Некоторые преимущества инфографики:

улучшает восприятие информации;

экономит время и усилия;

привлекает внимание;

увеличивает эффективность коммуникации;

способствует широкому распространению.

Некоторые недостатки инфографики:

– упрощённое изображение;

– ограничение на объём информации;

– требует хорошего дизайна.

Инфографику используют в различных областях, включая журналистику, маркетинг, образование и научные исследования.

21. Старение населения (aging population) – это демографический сдвиг, при котором в обществе увеличивается доля пожилых людей, а количество молодых остаётся низким или не растёт.

Некоторые причины старения населения:

Снижение уровня рождаемости. Это может быть связано с изменениями культурных взглядов на размер семьи, увеличением доступа к контрацепции и экономическим давлением на семьи, из-за которого они заводят меньше детей.

Увеличение продолжительности жизни. Достижения в медицине и здравоохранении привели к значительному увеличению ожидаемой продолжительности жизни за последнее столетие.

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.