

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ ИСТОРИИ ЕСТЕСТВОЗНАНИЯ И ТЕХНИКИ
ИМ. С.И. ВАВИЛОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК
САНКТ-ПЕТЕРБУРГСКИЙ ФИЛИАЛ

СОЦИОЛОГИЯ НАУКИ И ТЕХНОЛОГИЙ

2025

Том 16

№ 4

Санкт-Петербург

ЭТИЧЕСКИЕ АСПЕКТЫ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

АЛЕКСАНДР ВЛАДИМИРОВИЧ ШЕНДЕРЮК-ЖИДКОВ

доктор honoris causa Балтийского федерального университета
им. И. Канта, Калининград, Россия;
сенатор, член Совета Федерации
Федерального Собрания Российской Федерации
от Калининградской области,
первый зампред Комитета по бюджету
и финансовым рынкам Совета Федерации,
Москва, Россия;
e-mail: A.shenderyuk@yandex.ru



АЛЕКСАНДР ЕВГЕНЬЕВИЧ ХРАМОВ

доктор физико-математических наук, профессор,
член-корреспондент Российской академии наук,
директор НИИ прикладного искусственного интеллекта
и цифровых решений РЭУ им. Г.В. Плеханова,
Москва, Россия;
e-mail: hramov.ae@rea.ru



Согласование этики и технологий: построение ответственного будущего нейротехнологий и искусственного интеллекта

УДК: 008.2, 004.8, 612.8

DOI: 10.24412/2079-0910-2025-4-138-154

Этика искусственного интеллекта (ИИ) представляет собой междисциплинарную область, ориентированную на идентификацию и разрешение комплекса проблем, возникающих при внедрении ИИ в социум. Ключевые этические вопросы, находящиеся в фокусе внимания исследователей, включают обеспечение конфиденциальности, инклюзивности, надежности, прозрачности, справедливости и объяснимости алгоритмов ИИ. Для их решения применяются разнообразные технические, социальные и социально-технические методы. В результате исследований в области этики ИИ были сформулированы принципы, руководящие указа-

ния, инструментальные средства, разработаны корпоративные политики, рамки управления, стандарты и юридические нормы, предназначенные для регулирования разработки и применения систем ИИ. Нейротехнологии, осуществляющие сбор и/или модификацию данных о нервной системе в реальном времени, все чаще интегрируются с ИИ, что обуславливает необходимость учета их специфических этических аспектов. Принципы, сформированные в рамках этики ИИ, могут предоставить ориентиры для решения этических проблем нейротехнологий, однако последние могут нести новые, более сложные вызовы, требующие дополнительного изучения и разработки новых подходов к их решению.

Ключевые слова: этика, искусственный интеллект, нейротехнологии, объяснимость, правовые нормы регулирования, этические дилеммы, справедливость, управление, нейроправа.

Введение

Бурное развитие нейротехнологий и искусственного интеллекта (ИИ) открывает беспрецедентные возможности для трансформации медицины, образования, науки и повседневной жизни [Hassabis et al., 2017; Karpov et al., 2023]. Однако стремительная интеграция этих технологий порождает сложные этические вызовы, связанные с автономией личности, конфиденциальностью данных, справедливостью распределения преимуществ от использования гибридных «нейро- + ИИ-технологий» и ответственностью за принимаемые решения [Ienca et al., 2017].

Нейротехнологии, включая нейроинтерфейсы, нейромодуляцию и нейровизуализацию, позволяют не только лечить заболевания, но и усиливать когнитивные способности, изменяя саму природу человеческого мышления [Kellmeyer et al., 2018]. В то же время ИИ, особенно генеративные модели, активно вторгается в сферы творчества, образования и принятия решений, стирая границы между реальным и искусственным [Mittelstadt et al., 2016; Floridi et al., 2018; Battista, 2020]. Синергия этих технологий способна привести к созданию гибридных интеллектуальных систем, но одновременно усиливает риски манипуляции сознанием, цифрового неравенства и утраты человеческой идентичности. Современные тенденции, такие как распространение «умной среды» и гибридизация физической и цифровой реальности [Филиппова, 2021; Lecomte, 2023], развитие которых невозможно без ИИ и нейротехнологий, подчеркивают необходимость всестороннего анализа этических, социальных и юридических вопросов, связанных с этими технологиями. Только опережающее регулирование и междисциплинарное сотрудничество позволят минимизировать риски и обеспечить ответственное использование нейротехнологий и ИИ в интересах общества.

Поэтому в данной статье исследуются ключевые этические дилеммы, возникающие на стыке нейротехнологий и ИИ, и предлагаются новые подходы к их согласованию. Особое внимание уделяется разработке междисциплинарных этических рамок, обеспечивающих ответственное и социально ориентированное развитие этих технологий. В условиях отсутствия глобальных регуляторных стандартов, подобных инициативам ЮНЕСКО в области нейроэтики¹, особую актуальность

¹ UNESCO. Этические аспекты нейротехнологий: ЮНЕСКО назначила международную группу экспертов для разработки рекомендаций. 2021. Режим доступа: [https://www.unesco.org/ru/articles/eticheskie-aspekty-neyrotekhnologiy-yunesko-naznachila-mezhdunarodnuyu-](https://www.unesco.org/ru/articles/eticheskie-aspekty-neyrotekhnologiy-yunesko-naznachila-mezhdunarodnuyu)

приобретает опережающее регулирование развития нейротехнологий и ИИ, основанное на принципах прозрачности, инклюзивности и защиты фундаментальных прав человека. Согласование этических принципов с технологическим прогрессом становится насущной задачей для обеспечения благополучия и справедливости в обществе [Ruiz et al., 2024].

Цель работы — не только обозначить существующие риски, но и предложить пути гармонизации технологического прогресса с этическими ценностями, чтобы совместное развитие нейротехнологий и ИИ служило интересам всего общества, а не усугубляло существующие социальные разрывы.

Отметим, что обычно этические проблемы ИИ и нейротехнологий рассматриваются отдельно. В ряде последних работ в этом направлении авторы останавливались на вопросах пересечения этических вопросов, которые ставят ИИ и нейротехнологии [Robinson et al, 2022; Sample et al, 2021; Savage, 2019; Illes, 2017]. В центре внимания данной статьи находится процесс коэволюции ИИ и нейротехнологий. Исследование ставит перед собой задачу проанализировать комплекс этических вызовов, возникающих в результате их интеграции, и сформулировать концептуальные подходы, направленные на обеспечение социально ответственного развития этой технологической сферы.

Методология исследования

В данной работе для комплексного анализа и сопоставления этических аспектов регулирования ИИ и нейротехнологий выбран междисциплинарный подход, позволяющий учесть как технические, так и социально-правовые измерения рассматриваемых проблем. Исследование опирается на синтез качественного анализа существующих нормативных документов, научной литературы и аналитических обзоров в сфере этики ИИ и нейротехнологий, а также на сравнительный правовой анализ.

Сравнительный анализ регулирования проведен на примере стран, играющих ведущую роль в формировании глобальных стандартов: Китая, ЕС, США и России. Выбор этих юрисдикций обусловлен их активной законодательной деятельностью в сфере ИИ, внедрением нейротехнологий в медицину и другие области, а также разнообразием моделей регулирования — от жестких норм до гибких «регуляторных песочниц». Кроме того, выбор стран расширяется примерами из Латинской Америки, такими как Колумбия и Аргентина, что позволяет отразить разнообразие подходов к законодательному обеспечению и внедрению принципов этики в разных социокультурных и институциональных контекстах. Такой выбор обоснован актуальностью и новизной предложенных ими законодательных инициатив, а также желанием продемонстрировать вариативность методов регулирования, от ориентированных на риск-менеджмент до основанных на правах человека. Это позволило выделить общие закономерности и национальные особенности в подходе к этическим вызовам.

Методологически исследование базируется на критическом сравнительном анализе, позволяющем выявить общие тенденции и уникальные черты в подходах к регулированию ИИ и нейротехнологий. Особое внимание уделяется анализу этических принципов и практик их реализации в правовых актах, что служит основанием для построения рекомендаций по интеграции технологий с учетом их социальных и гуманитарных последствий. Для структурирования этических проблем использована классификация рисков по категориям, включая предвзятость алгоритмов, угрозы конфиденциальности данных, автономию личности и экзистенциальные риски. Каждая категория проанализирована через призму социальных, юридических и технологических последствий, что обеспечило комплексность оценки. Принятие именно такого междисциплинарного и сравнительного подхода позволяет не только систематизировать существующие знания, но и выявить пробелы в текущих практиках регулирования, что особенно важно в контексте стремительного развития и интеграции нейротехнологий с ИИ.

Разработка рекомендаций основывалась на синтезе правовых, технических и этических решений. Рассмотрены принцип-ориентированные подходы (этические кодексы), модели, основанные на оценке рисков, и правоцентричные стратегии, направленные на защиту нейроправ. Учтены позиции ключевых стейкхолдеров — от разработчиков технологий до представителей уязвимых групп, что подчеркивает социальную ориентированность исследования.

Современное состояние и совместное развитие нейротехнологий и искусственного интеллекта

Нейротехнологии и ИИ сегодня представляют собой две взаимодополняющие области, стремительное развитие которых открывает новые горизонты в изучении мозга, медицине, реабилитации и когнитивном усилении человека. Нейротехнологии, включая нейровизуализацию, нейроинтерфейсы и нейромодуляцию, позволяют регистрировать активность мозга и влиять на нее, в то время как ИИ, особенно благодаря методам глубокого обучения, обеспечивает мощные инструменты для анализа сложных нейронных данных и создания адаптивных систем.

Одним из наиболее перспективных направлений синергии является развитие интерфейсов «мозг — компьютер» (ИМК), где алгоритмы машинного обучения играют ключевую роль в декодировании нейронных сигналов [Hramov et al., 2021]. Современные ИМК, как инвазивные, так и неинвазивные, уже применяются для восстановления двигательных функций у пациентов после инсульта, управления протезами и даже коммуникации при тяжелых формах паралича. ИИ повышает точность интерпретации нейроданных, позволяя адаптировать системы под индивидуальные особенности пользователей. Например, сочетание ИМК с генеративными моделями ИИ открывает возможности для более естественного взаимодействия с внешними устройствами, такими как экзоскелеты или системы виртуальной реальности.

Другим важным аспектом коэволюции этих технологий является использование ИИ для анализа данных нейровизуализации (фМРТ, ЭЭГ, магнитоэнцефалография и т. д.). Глубокие нейронные сети способны выявлять скрытые паттерны, связанные с нейродегенеративными заболеваниями, психическими расстройствами или

когнитивными процессами, что способствует ранней диагностике и персонализированной терапии [Topol, 2019; Карпов и др., 2022]. В свою очередь, нейробиологические исследования вдохновляют архитектуры ИИ: сверточные сети возникли под влиянием принципов обработки зрительной информации в мозге [LeCun et al., 1998], а новые модели, такие как *Titans* от *Google Research*, имитируют механизмы человеческой памяти [Behrouz et al., 2024].

Однако стремительное развитие и интеграция этих технологий порождают серьезные вопросы этического и правового характера. Особую озабоченность вызывает отсутствие четких регуляторных механизмов, способных эффективно контролировать эту динамично развивающуюся сферу.

Первый ключевой риск связан с коммерческим использованием нейроданных. Корпорации, получая доступ к информации о мозговой активности через специализированные устройства, могут выявлять нейробиологические маркеры, отражающие предпочтения и эмоциональные реакции пользователей. Подобные данные открывают возможности для скрытого манипулирования потребительским выбором в маркетинговых целях. Эта практика ставит под угрозу базовые принципы приватности, создавая прецеденты для вмешательства в личную сферу и потенциального злоупотребления конфиденциальной информацией на уровне как отдельных лиц, так и общества в целом [Luna-Nevarez, 2021; Stanton et al., 2017].

Второй существенной проблемой становится усиление социальной стратификации. Высокая стоимость современных нейротехнологических решений, таких как системы когнитивной коррекции или нейрореабилитации, может привести к их доступности исключительно для обеспеченных слоев населения. Это способно усугубить существующее неравенство, создавая новый вид цифрового разрыва — «нейроразрыв», который проявится на глобальном, национальном и локальном уровнях [Ienca et al., 2017; Goering et al., 2016].

Несмотря на огромный потенциал совместного применения нейротехнологий и ИИ в медицине, когнитивном усилении и улучшении качества жизни, стремительное развитие этих технологий требует тщательного анализа сопутствующих моральных дилемм. В последующих разделах будет представлен детальный обзор этических аспектов, возникающих в процессе разработки и внедрения нейротехнологических решений на базе ИИ.

Этические вызовы и социальные дилеммы развития ИИ и нейротехнологий

Развитие нейротехнологий и ИИ представляет собой не только технологическую революцию, но и источник глубоких этических и социальных дилемм, которые требуют тщательного анализа и проработки [Savage, 2019]. Эти вызовы затрагивают фундаментальные ценности — автономию личности, справедливость, конфиденциальность и права человека — и требуют координации усилий исследователей, государства и общества.

Одним из ключевых вопросов является проблема *предвзятости и дискриминации* [Mehrabani et al., 2021]. Алгоритмические системы, основанные на обучении на больших массивах данных, могут воспроизводить и усиливать существующие социальные стереотипы. Если в традиционных областях это проявляется в неверных реше-

ниях при отборе персонала или предоставлении финансовых услуг [Caliskan et al., 2017], то интеграция с нейротехнологиями несет риск дискриминации на основании нейропсихологических характеристик личности, которые сами по себе крайне чувствительны к интерпретации. Одной из ключевых проблем остается неравенство в доступе к цифровой медицине [Карпов и др., 2022]. Несмотря на активное внедрение ИИ и нейротехнологий в диагностику и лечение, их доступность зависит от уровня дохода, страхового покрытия и географического положения. Это усугубляет разрыв в качестве медицинской помощи, оставляя уязвимые группы без современных технологий.

Следующий крупный блок рисков связан с *прозрачностью и объяснимостью технологий*. В то время как ИИ функционирует как «черный ящик» при обработке сложных данных [Dwivedi et al., 2023], нейротехнологические вмешательства еще менее понятны в плане долгосрочного влияния на психику человека. Отсутствие четкой объяснимости снижает доверие пользователей и может подрывать социальную легитимность внедряемых систем на основе нейротехнологий.

Не менее значимы вопросы *конфиденциальности и контроля над данными*. В отличие от классических цифровых данных, нейроданные могут представлять прямое отражение когнитивных процессов и психических состояний. Их сбор и обработка открывают возможности не только для медицинской диагностики, но и для профилирования личности, что ставит под угрозу право человека на ментальную неприкосновенность и контроль над внутренним миром. Риски включают несанкционированный доступ, использование нейроданных для дискриминации (например, при найме или страховании), манипулирование поведением и нарушение приватности [Yuste et al., 2017]. В связи с этим в научной литературе все чаще обсуждается необходимость введения специальных «нейроправ» [Ienca et al., 2017], гарантирующих защиту субъективной сферы.

Отдельный пласт проблем связан с *автономией и агентностью человека*. Если ИИ воздействует скорее опосредованно — через алгоритмические решения и рекомендации, — то нейротехнологии способны напрямую вмешиваться в нервные процессы и изменять когнитивные или эмоциональные состояния субъекта. Возникает дилемма: где проходит граница допустимости вмешательства в сознание человека и как обеспечить баланс между пользой (например, восстановлением утраченных функций) и рисками потери автономии?

Значительное внимание уделяется и вопросу *справедливого доступа к новым технологиям*. Уже сегодня стоимость передовых медицинских нейроинтерфейсов или систем на базе ИИ делает их доступными лишь ограниченному кругу пользователей. В перспективе это способно привести к усилению социального неравенства: разделению общества на группы, имеющие доступ к когнитивному усилению, и те, которые остаются «по другую сторону технологического барьера».

Развитие ИИ и нейротехнологий должно учитывать их влияние на *физическое и психическое благополучие пользователей*, включая здоровье, безопасность, комфорт и социально-экологические аспекты. Согласно Хельсинкской декларации Всемирной медицинской ассоциации «Рекомендации для врачей, занимающихся биомедицинскими исследованиями с участием людей» (2000 г.), применение таких технологий требует информированного согласия, защиты нейроданных и соблюдения принципов автономии, конфиденциальности и безопасности. Достижение благополучия возможно лишь при обеспечении прозрачности, ответственности и

защиты прав пользователей, поскольку этические аспекты напрямую связаны с безопасностью и качеством жизни.

Наконец, следует выделить *социальные и нормативные дилеммы*. Интеграция ИИ и нейротехнологий не только трансформирует индивидуальный опыт, но и влияет на общество и институты в целом: от практик цифрового профилирования и манипуляции общественным мнением до вопросов кибербезопасности и использования таких технологий в военной сфере. При отсутствии эффективных законодательных механизмов регулирования есть риск стихийного развития рынка, где коммерциализация будет преобладать над этическими ограничениями.

Таким образом, анализ показывает, что этические проблемы при коэволюции ИИ и нейротехнологий выходят далеко за рамки традиционных дискуссий о надежности и безопасности. Речь идет о вызовах цивилизационного масштаба, затрагивающих права человека, социальную справедливость и само понимание человеческой идентичности. Их решение возможно лишь путем разработки комплексных подходов, включающих правовые новации, интернациональное сотрудничество и ответственное проектирование технологий.

Подходы к регулированию ИИ и нейротехнологий и сравнительные выводы

Современная политика регулирования искусственного интеллекта (ИИ) и нейротехнологий развивается по двум основным траекториям. *Первая траектория* — рискориентированная, предусматривающая классификацию технологий по уровню потенциального вреда и установление соответствующих обязательств для каждой категории. *Вторая траектория* базируется на защите фундаментальных прав человека, которые должны соблюдаться независимо от оценки рисков.

Европейский союз является ярким примером реализации рискориентированного подхода. Согласно принятому в 2024 г. Закону об ИИ, нормативному акту ЕС об ИИ, принятому Европейским парламентом 13 марта 2024 г. и одобренному Советом ЕС 21 мая 2024 г., системы подразделяются на уровни риска²: «неприемлемый риск» запрещен к использованию (например, массовая биометрическая идентификация в реальном времени в общественных местах), системы «высокого риска» подлежат строгому регулированию — с требованиями к управлению данными, технической документации, прозрачности, человеческому надзору и кибербезопасности. Для «низкорисковых» систем установлены минимальные требования, включая обязательство информировать пользователей о взаимодействии с ИИ и маркировку создаваемого таким ИИ контента.

Значимую роль в европейской правовой архитектуре играет также Регламент о защите данных (General Data Protection Regulation, GDPR)³, который регулирует автоматизированную обработку персональных данных, включая профилирование.

² World's First Major Law for Artificial Intelligence Gets Final EU Green Light // CNBC. 2024. Available at: <https://www.cnn.com/2024/05/21/worlds-first-major-law-for-artificial-intelligence-gets-final-eu-green-light.html> (date accessed: 27.01.2025).

³ General Data Protection Regulation (GDPR). Available at: <https://gdpr-info.eu> (date accessed: 27.01.2025).

GDPR закрепляет права субъектов данных на защиту от решений, принимаемых исключительно автоматизированными системами. Это особенно актуально для нейроданных, которые отражают когнитивные и психические состояния, требующие повышенной защиты.

В странах Латинской Америки наблюдается развитие иной модели регулирования, адаптирующейся к локальным особенностям. Например, в Колумбии законопроект № 225/2024 вводит уголовную ответственность за использование ИИ при создании поддельных цифровых идентичностей (deepfakes). В Аргентине поправки в закон о науке и технологиях предусматривают обязательство регистрировать системы ИИ и дают государственным органам право приостанавливать разработку и использование систем, нарушающих этические нормы.

Особое место занимает применение «регуляторных песочниц», которые предоставляют разработчикам платформ возможность тестировать технологии в контролируемой среде с целью выявления потенциальных рисков и своевременной адаптации нормативных требований. Этот механизм способствует балансу между стимулированием инноваций и гарантией безопасности, что особенно важно при работе с технологиями, затрагивающими ментальную неприкосновенность личности. Согласно статье 57 Закона ЕС об ИИ, каждый член Евросоюза должен обеспечить создание национальными компетентными органами нормативных «песочниц» для ИИ на национальном уровне.

Таким образом, международная практика демонстрирует две основные стратегии: структурированный и стандартизированный подход Европейского союза с упором на управление рисками и комплексную защиту данных, а также гибкие, адаптивные подходы Латинской Америки, основанные на интеграции этических ценностей в существующие правовые рамки с учетом местных социокультурных контекстов.

Важным голосом в общественной дискуссии о этических границах технологий являются религиозные институты. Такие традиционные институты, как Русская православная и Римско-католическая церкви, в своих заявлениях (например, в «Римском призыве» папы римского или в Слове Патриарха о научно-техническом прогрессе на встрече с учеными во Всероссийском научно-исследовательском институте экспериментальной физики в г. Сарове, 31 июля 2019 г.) сходятся в ключевых принципах: технологии должны служить человеку, а не подчинять его; необходимо защищать неприкосновенность человеческого достоинства, свободы воли и приватности, особенно в контексте нейротехнологий, способных влиять на сознание и идентичность. В их позиции подчеркивается, что технологический прогресс не должен приводить к дегуманизации и обязан учитывать духовно-нравственные аспекты человеческой природы, выходящие за рамки утилитарного подхода.

Эффективное регулирование ИИ и нейротехнологий требует выработки многоуровневых международных стандартов, которые смогут объединить гибкость инновационных подходов с жесткими гарантиями защиты личности, ее автономии и ментальной неприкосновенности, обеспечивая тем самым ответственное развитие технологий на глобальном уровне.

Правовая база регулирования ИИ и нейротехнологий: современное состояние и международный опыт

Правовое регулирование ИИ и нейротехнологий в мире находится на ранних этапах формирования. В настоящее время отсутствует единый подход к законодательной регламентации, многие правовые акты содержат обобщенные и расплывчатые формулировки, не охватывающие полный жизненный цикл технологий. Особая сложность связана с нейротехнологиями, для которых специализированные законы есть лишь в единичных юрисдикциях. В большинстве стран эти технологии регулируются через существующие нормативы в сферах медицины, защиты персональных данных, прав человека и интеллектуальной собственности, а этические рекомендации носят часто неформальный характер. Важную роль в формировании глобального подхода к ответственному развитию нейротехнологий сыграли Рекомендации Совета по ответственным инновациям в нейротехнологиях Организации экономического сотрудничества и развития (2019)⁴. Этот документ стал одним из первых международных ориентиров, призвавших стран-участников принять меры для обеспечения того, чтобы политика в области нейротехнологий соответствовала принципам ответственности, безопасности, этичности, инклюзивности и уважения прав человека.

В числе мировых лидеров в области регулирования нейротехнологий выделяется Европейский союз. Уже в 2020 г. был опубликован Белый документ по ИИ, в котором были сформулированы базовые принципы этичного и безопасного развития, включая нейротехнологии. Стоит особо отметить влияние Регламента ЕС по защите данных, который содержит положения, регламентирующие обработку высокочувствительных данных, в том числе нейроданных. Этот регламент предусматривает права субъектов данных, в том числе ограничения на автоматизированные решения, которые могут затрагивать жизненно важные интересы человека.

Рассмотрим ключевые подходы и практики в различных странах.

Китайская Народная Республика. В августе 2023 г. вступила в силу директива по ИИ, которая предусматривает обязательное маркирование контента, созданного с применением ИИ, использование только легальных данных для обучения моделей, а также создание механизмов рассмотрения публичных жалоб на ИИ-сервисы. Параллельно формируются требования к качеству и актуальности обучающих данных.

Соединенные Штаты Америки. В 2022 г. администрация США представила «Билль о безопасном и надежном ИИ», направленный на создание безопасных систем, борьбу с дискриминацией, защиту конфиденциальности и обеспечение права выбора человека. В 2023 г. ведущие IT-компании подписали соглашение о саморегулировании, включая меры по маркировке AI-сгенерированного контента и совместные инициативы по управлению рисками. Регулирование нейротехнологий в США носит фрагментарный характер и осуществляется в рамках законов о медицинских приборах и защите данных. Этические рекомендации разрабатываются Национальными институтами здоровья и Национальной академией медицины, но не имеют юридической силы.

⁴ Recommendation of the Council on Responsible Innovation in Neurotechnology OECD/LEGAL/0457, adopted on 11.12.2019. Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457> (date accessed: 19.08.2025).

Европейский союз. Введенный в действие Закон об ИИ классифицирует системы ИИ по уровню риска: от запрещенных практик психического воздействия и социального рейтинга до строго регулируемых систем в критически важных сферах и относительно свободных «низкорисковых» технологий, таких как генерация контента. Комплекс мер дополняется обязательной регистрацией и контролем.

Знаковым событием 2024 г. стало принятие Рамочной конвенции Совета Европы об искусственном интеллекте и правах человека (СДСЕ № 225) — первого в истории международного юридически обязывающего договора, призванного обеспечить полное соответствие систем ИИ стандартам прав человека, демократии и верховенства права на всех этапах их жизненного цикла. Конвенция открыта для подписания всеми странами мира, а ее принципы, сфокусированные на защите личности, уже задают новый глобальный ориентир для разработки национальных регуляторных режимов, выходящий за рамки исключительно рискориентированного подхода.

Латинская Америка. Регулирование в регионе менее институционально систематизировано: в Колумбии законодательно ужесточена ответственность за злонамеренное использование ИИ (например, создание *deepfake*), а в Аргентине вводятся нормы, регулирующие регистрацию и этическое использование ИИ и подкрепленные государственным контролем.

Российская Федерация. В России законодательная база в сфере ИИ пока формируется. В 2020 г. была утверждена Национальная стратегия развития искусственного интеллекта до 2030 г., которая определяет основные направления развития ИИ, включая создание правовых и этических рамок. В рамках этой стратегии акцент делается на обеспечении безопасности, защите персональных данных и предотвращении дискриминации при использовании ИИ-технологий. Кроме того, в 2021 г. был принят Федеральный закон № 331-ФЗ, который вносит изменения в Гражданский кодекс РФ и регулирует вопросы ответственности за действия автономных систем, включая ИИ. Однако на данный момент правовое регулирование остается фрагментарным и требует дальнейшей детализации, особенно в вопросах определения ответственности за решения, принимаемые ИИ. Необходимо отметить также, что в настоящее время происходит значительная работа по формированию стандартов в области использования систем ИИ, и здесь одним из лидеров являются здравоохранение и промышленность, где активно развивается правовое и техническое регулирование ИИ (см. подробнее: [Карпов и др., 2022; глава 4]).

Одним из ключевых аспектов регулирования ИИ в России является этическая составляющая. В 2021 г. был разработан Кодекс этики в области искусственного интеллекта, который устанавливает принципы разработки и использования ИИ, такие как прозрачность, справедливость и уважение прав человека. Этот документ, хотя и не имеет обязательной юридической силы, служит ориентиром для разработчиков и компаний, работающих в сфере ИИ. Тем не менее отсутствие четких механизмов контроля и санкций за нарушение этических норм остается проблемой.

Интегрированные стратегии преодоления этических вызовов при совместном развитии ИИ и нейротехнологий

Для эффективного решения этических проблем, возникающих на стыке искусственного интеллекта и нейротехнологий, необходим комплексный междисциплинарный подход, учитывающий как общие, так и уникальные вызовы этих технологий. Современные исследования и практика демонстрируют, что проблемы, возникающие в области ИИ, такие как предвзятость алгоритмов, вопросы конфиденциальности, прозрачности и справедливости, дополняются специфическими рисками нейротехнологий. Среди них: особенности прямого воздействия на когнитивные процессы, работа с особо чувствительными нейроданными и вызовы для автономии личности.

Для решения этих задач адаптируются и расширяются существующие этические рамки искусственного интеллекта с учетом специфики нейротехнологий. Международное сообщество нейроэтиков, в частности Институт нейроэтики (Institute of Neuroethics, IoNx)⁵, разработало свыше двадцати рекомендаций и практических руководств, которые служат базой для создания политики и стандартов в этой области.

Ключевой этический принцип — *in dubio pro homo* (в сомнении — в пользу человека) — призван обеспечить наивысшую защиту прав, достоинства и автономии личности на всех этапах взаимодействия с этими технологиями.

Для реализации этого принципа применяются современные инструменты и технологии:

- наборы инструментов для объяснимости ИИ — программные решения, обеспечивающие доступность и прозрачность алгоритмических решений, что снижает риски ошибок и укрепляет доверие пользователей в том числе и при обработке медицинских данных [McDermid et al., 2021];
- технологии повышения конфиденциальности — методы гомоморфного шифрования, дифференциальной приватности и анонимизации, позволяющие безопасно обрабатывать данные нейровизуализации и психофизиологические сведения [Cummings et al., 2024];
- «регуляторные песочницы» — контролируемые экспериментальные среды, в которых новые разработки ИИ и нейротехнологий тестируются с целью раннего выявления рисков и адаптации нормативного регулирования, которые обсуждались нами выше;
- подход этического проектирования — методы интеграции этических норм непосредственно в цикл разработки технических средств, направленные на минимизацию негативных последствий [Sabbah et al., 2018];
- образовательные программы и курсы — специализированные обучающие инициативы для разработчиков, медицинских специалистов и регуляторов, способствующие формированию культуры ответственного внедрения технологий;
- международные стандарты и протоколы — усилия различных организаций (ISO, IEEE, ОЭСР) по формированию универсальных требований к безо-

⁵ Institute of Neuroethics (IoNx). Available at: <https://instituteofneuroethics.org> (date accessed: 04.02.2025).

пасности, прозрачности, справедливости и соблюдению прав человека при применении ИИ и нейротехнологий.

Успешное разрешение этических проблем невозможно без объединения усилий специалистов из нейронаук, информатики, права, этики, медицины, а также с учетом мнений представителей уязвимых групп — малочисленных народов, людей с особыми потребностями и др. [Panch et al., 2019]. Учет их интересов поможет избежать социальной дискриминации и обеспечить инклюзивность технологий.

Таким образом, применение современных технических инструментов в сочетании с устойчивыми этическими и правовыми принципами формирует надежную основу для ответственного и гуманного развития интегрированных технологий искусственного интеллекта и нейротехнологий.

Обсуждение и заключение

Современная динамика совместного развития и взаимопроникновения ИИ и нейротехнологий все отчетливее демонстрирует ограниченность традиционных инструментов регулирования и подчеркивает необходимость гибких, адаптивных стратегий, способных учитывать уникальные особенности обеих технологий. Поскольку нейротехнологии напрямую воздействуют на нервную систему и когнитивные процессы человека, а ИИ обеспечивает мощные инструменты анализа и управления данными, их интеграция порождает как новые этические риски, так и технологические вызовы, требующие междисциплинарного освоения и повышения регуляторной готовности.

За время, прошедшее с первых попыток сформировать юридические и этические механизмы контроля ИИ, ландшафт технологий значительно изменился, и сейчас во многих случаях выжидательная позиция, позволяющая анализировать опыт ведущих юрисдикций и параллельно изучать развитие рынка нейротехнологий, представляется наиболее прагматичной для России и ряда других государств.

Появление высокоэффективных языковых моделей и быстро развивающихся нейротехнологий в сочетании с доступностью вычислительных платформ создает беспрецедентные возможности для демократизации технологий. Тем не менее это усложняет контроль и мониторинг таких систем, поскольку и ИИ, и нейротехнологии требуют определения границ допустимого вмешательства в личность и понимания долгосрочных социальных последствий.

Не менее важен вопрос перераспределения экономических выгод в контексте глобализации цифровых технологий и концентрирования ресурсов. Крупные ИТ-компании, владеющие технологиями ИИ и инфраструктурой для поддержки нейротехнологических исследований, влияют на национальные рынки труда, образование и налоговую политику. Рассматриваемые сейчас инициативы по введению международных налоговых механизмов, наподобие *Pillar One*, могут стать важным инструментом адаптации к новой экономической реальности, в том числе и в России.

При этом растущая концентрация вычислительных мощностей и ограниченный доступ молодых команд к инфраструктуре создают барьеры не только для развития ИИ, но и для инноваций в области нейротехнологий — обе области критически зависят от технологий обработки больших данных и мощностей высокопроизводи-

тельных вычислений. В таких условиях важны меры государственной и институциональной поддержки стартапов и исследовательских коллективов, обеспечивающие равные возможности для развития как ИИ, так и нейротехнологий.

Особое внимание следует уделить вопросам интеллектуальной собственности в новой парадигме: рост объемов контента, генерируемого ИИ, требует новых подходов к компенсации авторам и правообладателям, а также к внедрению инструментов маркировки продуктов, созданных ИИ.

Авторы призывают к открытой экспертной дискуссии по формированию комплексных и гибких моделей регулирования, в центре которых — защита прав и интересов личности, стимулирование науки, образования и технологического предпринимательства. Такой системный подход, опирающийся на постоянный диалог государства, бизнеса, академического сообщества и гражданского общества, позволит создать адаптивную нормативную базу, отвечающую реалиям XXI в. и обеспечивающую устойчивое развитие цифровых и нейроцифровых экосистем.

Литература

- Карнов О.Э., Храмов А.Е.* Информационные технологии, вычислительные системы и искусственный интеллект в медицине. М.: ДПК Пресс, 2022. 320 с.
- Филипова И.А.* Нейротехнологии: развитие, применение на практике и правовое регулирование // Вестник С.-Петерб. ун-та. Право. 2021. Т. 12. № 3. С. 1–15. DOI: 10.21638/spbu14.2021.302.
- Battista D.* Political Communication in the Age of Artificial Intelligence: An Overview of Deepfakes and Their Implications // Society Register. 2024. Vol. 8. No. 2. P. 7–24. DOI: 10.14746/sr.2024.8.2.01.
- Behrouz A., Zhong P., Mirrokni V.* Titans: Learning to Memorize at Test Time // arXiv preprint arXiv:2501.00663. 2024.
- Caliskan A., Bryson J.J., Narayanan A.* Semantics Derived Automatically from Language Corpora Contain Human-Like Biases // Science. 2017. Vol. 356. No. 6334. P. 183–186. DOI: 10.1126/science.aal4230.
- Cummings R., Sarathy J.* Centering Policy and Practice: Research Gaps Around Usable Differential Privacy // 023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). 2024. <https://arxiv.org/abs/2406.12103>.
- Dwivedi R., Dave D., Naik H., Singhal S., Omer R., Patel P., Qian B. et al.* Explainable AI (XAI): Core Ideas, Techniques, and Solutions // ACM Computing Surveys. 2023. Vol. 55. No. 9. P. 1–33.
- Floridi L. et al.* AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations // Minds and Machines. 2018. Vol. 28. No. 4. P. 689–707.
- Goering S., Yuste R.* On the Necessity of Ethical Guidelines for Novel Neurotechnologies // Cell. 2016. Vol. 167. No. 4. P. 882–885. DOI: 10.1016/j.cell.2016.10.029.
- Hassabis D., Kumaran D., Summerfield C., Botvinick M.* Neuroscience-Inspired Artificial Intelligence // Neuron. 2017. Vol. 95. No. 2. P. 245–258. DOI: 10.1016/j.neuron.2017.06.011.
- Hramov A.E., Maksimenko V.A., Pisarchik A.N.* Physical Principles of Brain-Computer Interfaces and Their Applications for Rehabilitation, Robotics and Control of Human Brain States // Physics Reports. 2021. Vol. 918. P. 1–133. DOI: 10.1016/j.physrep.2021.03.004.
- Ienca M., Andorno R.* Towards New Human Rights in the Age of Neuroscience and Neurotechnology // Life Sciences, Society and Policy. 2017. Vol. 13. No. 1. P. 5.
- Illes J.* Neuroethics — Anticipating the Future. New York, NY: Oxford University Press, 2017. 643 p.

Karpov O.E., Pitsik E.N., Kurkin S.A., Maksimenko V.A., Gusev A.V., Shusharina N.N., Hramov A.E. Analysis of Publication Activity and Research Trends in the Field of AI Medical Applications: Network Approach // *International Journal of Environmental Research and Public Health*. 2023. Vol. 20. P. 5335. DOI: 10.3390/ijerph20075335.

Kellmeyer P. Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices // *Neuroethics*. 2018. Vol. 11. No. 1. P. 1–16. DOI: 10.1007/s12152-018-9371-x.

Lecomte P. Umwelt as the Foundation of an Ethics of Smart Environments // *Humanities and Social Sciences Communications*. 2023. Vol. 10. No. 1. P. 1–12. DOI: 10.1057/s41599-023-02356-9.

LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // *Proceedings of the IEEE*. 1998. Vol. 86. No. 11. P. 2278–2324.

Luna-Nevarez C. Neuromarketing, Ethics, and Regulation: An Exploratory Analysis of Consumer Opinions and Sentiment on Blogs and Social Media // *Journal of Consumer Policy*. 2021. Vol. 44. No. 4. P. 559–583. DOI: 10.1007/s10603-021-09496-y.

McDermid J.A., Jia Y., Porter Z., Habli I. Artificial Intelligence Explainability: The Technical and Ethical Dimensions // *Philosophical Transactions of the Royal Society A*. 2021. Vol. 379. No. 2207. P. 20200363. DOI: 10.1098/rsta.2020.0363.

Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. A Survey on Bias and Fairness in Machine Learning // *ACM Computing Surveys*. 2021. Vol. 54. No. 6. P. 1–35. DOI: 10.48550/arXiv.1908.09635.

Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L. The Ethics of Algorithms: Mapping the Debate // *Big Data & Society*. 2016. Vol. 3. No. 2. P. 1–21. DOI: 10.1177/205395171667967.

Panch T., Mattie H., Celi L.A. The “Inconvenient Truth” about AI in Healthcare // *NPJ Digital Medicine*. 2019. Vol. 2. P. 1–3. DOI: 10.1038/s41746-019-0155-4.

Robinson J.T., Rommelfanger K.S., Anikeeva P.O., Etienne A., French J., Gelinas J., Grover P., Picard R. Building a Culture of Responsible Neurotech: Neuroethics as Socio-Technical Challenges // *Neuron*. 2022. Vol. 110. No. 13. P. 2057–2062. DOI: 10.1016/j.neuron.2022.05.005.

Ruiz S., Valera L., Ramos P., Sitaram R. Neurorights in the Constitution: from Neurotechnology to Ethics and Politics // *Philosophical Transactions B*. 2024. Vol. 379. No. 1915. P. 20230098. DOI: 10.1098/rstb.2023.0098.

Sabbah M., Rami E.K., Jebbawi Y., Halabi R., Zantout R., Diab M.O. Robotic Interface Controller for Minimally Invasive Surgery // 2018 1st International Conference on Computer Applications & Information Security (ICCAIS). IEEE. 2018. April. P. 1–4.

Sample M., Racine E. Pragmatism for a Digital Society: The (In)Significance of Artificial Intelligence and Neural Technology // *Advances in Neuroethics: Clinical Neurotechnology Meets Artificial Intelligence*. Springer, 2021. P. 81–100.

Savage N. How AI and Neuroscience Drive Each Other Forwards // *Nature*. 2019. Vol. 571. P. S15. DOI: 10.1038/d41586-019-02212-4.

Stanton S.J., Sinnott-Armstrong W., Huettel S.A. Neuromarketing: Ethical Implications of Its Use and Potential Misuse // *Journal of Business Ethics*. 2017. Vol. 144. P. 799–811. DOI: 10.1007/s10551-016-3059.

Topol E.J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence // *Nature Medicine*. 2019. Vol. 25. No. 1. P. 44–56. DOI: 10.1038/s41591-018-0300-7.

Yuste R., Goering S., Bi G., Carmena J.M., Carter A., Fins J.J., Friesen P. et al. Four Ethical Priorities for Neurotechnologies and AI // *Nature*. 2017. Vol. 551. No. 7679. P. 159–163. DOI: 10.1038/551159a.

Harmonization of Ethics and Technology: Building a Responsible Future for Neurotechnology and Artificial Intelligence

ALEXANDER V. SHENDERYUK-ZHIDKOV

Federation Council, Moscow;
Immanuel Kant Baltic Federal University,
Kaliningrad, Russia;
e-mail: e-mail: A.shenderyuk@yandex.ru

ALEXANDER E. HRAMOV

Research Institute of Applied Artificial Intelligence
and Digital Solutions, Plekhanov Russian University of Economics,
Moscow, Russia;
e-mail: hramov.ae@rea.ru

Artificial intelligence (AI) ethics is an interdisciplinary field focused on identifying and resolving the set of issues that arise when AI is introduced into society. Key ethical issues that are the focus of research include ensuring the privacy, inclusiveness, reliability, transparency, fairness, and explainability of AI algorithms. A variety of technical, social, and socio-technical methods are being applied to address them. Research in AI ethics has resulted in the formulation of principles, guidelines, toolkits, corporate policies, governance frameworks, standards and legal norms designed to regulate the development and application of AI. Neurotechnology that collects and / or modifies real-time neural data are increasingly being integrated with AI, necessitating their specific ethical considerations. The principles developed within the framework of AI ethics can provide valuable guidelines for addressing the ethical challenges of neurotechnology, but the latter may present new more complex challenges that require further study and the development of specialized ethical standards.

Keywords: ethics, artificial intelligence, neurotechnology, explainability, legal regulatory standards, ethical dilemmas, fairness, governance, neurorights.

References

- Battista, D. (2024). Political Communication in the Age of Artificial Intelligence: An Overview of Deepfakes and Their Implications, *Society Register*, 8 (2), 7–24. DOI: 10.14746/sr.2024.8.2.01.
- Behrouz, A., Zhong, P., Mirrokni, V. (2024). Titans: Learning to Memorize at Test Time, *arXiv preprint arXiv:2501.00663*.
- Caliskan, A., Bryson, J.J., Narayanan, A. (2017). Semantics Derived Automatically from Language Corpora Contain Human-Like Biases, *Science*, 356 (6334), 183–186.
- Cummings, R., Sarathy, J. (2024). Centering Policy and Practice: Research Gaps Around Usable Differential Privacy, *023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. <https://arxiv.org/abs/2406.12103>.
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B. et al. (2023). Explainable AI (XAI): Core Ideas, Techniques, and Solutions, *ACM Computing Surveys*, 55 (9), 1–33.

Filipova, I.A. (2021). Neyrotekhnologii: razvitiye, primeneniye na praktike i pravovoye regulirovaniye [Neurotechnologies: Development, practical application, and legal regulation], *Vestnik S.-Peterb. un-ta. Pravo*, 12 (3), 1–15 (in Russian). DOI: 10.21638/spbu14.2021.302.

Floridi, L. et al. (2018). AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, *Minds and Machines*, 28 (4), 689–707.

Goering, S., Yuste, R. (2016). On the Necessity of Ethical Guidelines for Novel Neurotechnologies, *Cell*, 167 (4), 882–885.

Hassabis, D., Kumaran, D., Summerfield, C., Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence, *Neuron*, 95 (2), 245–258. DOI: 10.1016/j.neuron.2017.06.011.

Hramov, A.E., Maksimenko, V.A., Pisarchik, A.N. (2021). Physical Principles of Brain-Computer Interfaces and Their Applications for Rehabilitation, Robotics and Control of Human Brain States, *Physics Reports*, vol. 918, 1–133. DOI: 10.1016/j.physrep.2021.03.004.

Ienca, M., Andorno, R. (2017). Towards New Human Rights in the Age of Neuroscience and Neurotechnology, Life Sciences, *Society and Policy*, 13 (1), p. 5.

Illes, J. (2017). *Neuroethics — Anticipating the Future*, New York, NY: Oxford University Press.

Karpov, O.E., Khramov, A.E. (2022). *Informatsionnyye tekhnologii, vychislitel'nyye sistemy i iskusstvennyi intellekt v meditsine* [Information technologies, computing systems, and artificial intelligence in medicine], Moskva: DPK Press (in Russian).

Karpov, O.E., Pitsik, E.N., Kurkin, S.A., Maksimenko, V.A., Gusev, A.V., Shusharina, N.N., Hramov, A.E. (2023). Analysis of Publication Activity and Research Trends in the Field of AI Medical Applications: Network Approach, *International Journal of Environmental Research and Public Health*, vol. 20, p. 5335. DOI: 10.3390/ijerph20075335.

Kellmeyer, P. (2018). Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices, *Neuroethics*, 11 (1), 1–16.

Lecomte, P. (2023). Umwelt as the Foundation of an Ethics of Smart Environments, *Humanities and Social Sciences Communications*, 10 (1), 1–12. DOI: 10.1057/s41599-023-02356-9.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition, *Proceedings of the IEEE*, 86 (11), 2278–2324.

Luna-Nevarez, C. (2021). Neuromarketing, Ethics, and Regulation: An Exploratory Analysis of Consumer Opinions and Sentiment on Blogs and Social Media, *Journal of Consumer Policy*, 44 (4), 559–583. DOI: 10.1007/s10603-021-09496-y.

McDermid, J.A., Jia, Y., Porter, Z., Habli, I. (2021). Artificial Intelligence Explainability: The Technical and Ethical Dimensions, *Philosophical Transactions of the Royal Society A*, 379 (2207), 20200363.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning, *ACM Computing Surveys*, 54 (6), 1–35.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate, *Big Data & Society*, 3 (2), 1–21.

Panch, T., Mattie, H., Celi, L.A. (2019). The “Inconvenient Truth” about AI in Healthcare, *NPJ Digital Medicine*, vol. 2, 1–3. DOI: 10.1038/s41746-019-0155-4.

Robinson, J.T., Rommelfanger, K.S., Anikeeva, P.O., Etienne, A., French, J., Gelinas, J., Grover, P., Picard, R. (2022). Building a Culture of Responsible Neurotech: Neuroethics as Socio-Technical Challenges, *Neuron*, 110 (13), 2057–2062. DOI: 10.1016/j.neuron.2022.05.005.

Ruiz, S., Valera, L., Ramos, P., Sitaram, R. (2024). Neurorights in the Constitution: From Neurotechnology to Ethics and Politics, *Philosophical Transactions B*, 379 (1915), 20230098.

Sabbah, M., Rami, E.K., Jebbawi, Y., Halabi, R., Zantout, R., Diab, M.O. (2018, April). Robotic Interface Controller for Minimally Invasive Surgery, *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)* (pp. 1–4), IEEE.

Sample, M., Racine, E. (2021). Pragmatism for a Digital Society: The (In)Significance of Artificial Intelligence and Neural Technology, *Advances in Neuroethics: Clinical Neurotechnology Meets Artificial Intelligence* (pp. 81–100), Springer.

Savage, N. (2019). How AI and Neuroscience Drive Each Other Forwards, *Nature*, vol. 571, p. S15. DOI: 10.1038/d41586-019-02212-4.

Stanton, S.J., Sinnott-Armstrong, W., Huettel, S.A. (2017). Neuromarketing: Ethical Implications of Its Use and Potential Misuse, *Journal of Business Ethics*, vol. 144, 799–811.

Topol, E.J. (2019). High-Performance Medicine: The Convergence of Human and Artificial Intelligence, *Nature Medicine*, 25 (1), 44–56. DOI: 10.1038/s41591-018-0300-7.

Yuste, R., Goering, S., Bi, G., Carmenta, J.M., Carter, A., Fins, J.J., Friesen, P. et al. (2017). Four Ethical Priorities for Neurotechnologies and AI, *Nature*, 551 (7679), 159–163. DOI: 10.1038/551159a.