DOI 10.58351/2949-2041.2026.32.3.010
УДК 004.8
5.7.1. Онтология и теория познания (философские науки)

Сметана Владимир Васильевич
кандидат философских наук, директор
АНО НИИ «ЦИФРОВОЙ ИНТЕЛЛЕКТ»
SMETANA VLADIMIR
Candidate of philosophical sciences, PhD
DIGITAL INTELLIGENCE RESEARCH INSTITUTE

ЭТИЧЕСКИЕ И ФИЛОСОФСКИЕ ГОРИЗОНТЫ ПОСТСИНГУЛЯРНОСТИ: ПЕРЕОПРЕДЕЛЕНИЕ ЧЕЛОВЕЧЕСКОГО СОСТОЯНИЯ
ETHICAL AND PHILOSOPHICAL HORIZONS OF POST-SINGULARITY: REDEFINING THE HUMAN CONDITION

Аннотация. Технологическая сингулярность – гипотетический момент в будущем, когда технологический прогресс станет настолько быстрым и сложным, что окажется недоступным для понимания человеческим разумом в его нынешнем виде, – представляет собой не просто технический рубеж, но фундаментальный онтологический разрыв. Это событие, часто ассоциируемое с созданием искусственного суперинтеллекта (ASI), ставит под угрозу устоявшиеся антропоцентрические парадигмы, требуя радикального пересмотра понятий идентичности, сознания и целеполагания. В постсингулярном мире границы между биологическим и синтетическим, между рожденным и созданным, между субъектом и объектом права становятся проницаемыми и нестабильными.

Настоящее исследование представляет собой анализ этических и философских последствий этого перехода. Также, рассматривается, как технологии улучшения человека и искусственный интеллект трансформируют саму сущность человеческого бытия, и предлагаются новые рамки для осмысления морального статуса, смысла жизни и нормативного регулирования в эпоху, когда человечество перестает быть единственным носителем разума

Abstract. The technological singularity – a hypothetical future moment when technological progress becomes so rapid and complex that it becomes incomprehensible to the human mind in its current form – represents not just a technical milestone but a fundamental ontological rupture. This event, often associated with the creation of artificial superintelligence (ASI), challenges established anthropocentric paradigms, requiring a radical reconsideration of concepts of identity, consciousness, and purpose. In a post-singularity world, the boundaries between the biological and the synthetic, between the born and the created, and between the subject and object of law become permeable and unstable.

This study analyzes the ethical and philosophical implications of this transition. It also examines how human enhancement technologies and artificial intelligence are transforming the very essence of human existence and proposes new frameworks for understanding moral status, meaning in life, and normative regulation in an era when humanity ceases to be the sole bearer of intelligence

Ключевые слова: Технологическая сингулярность, искусственный интеллект, постсингулярный мир, загрузка разума, ветвящаяся идентичность, нейроинтерфейсы, прокреативное благодеяние, постлюди, неулучшенные люди, квалиа, трудная проблема сознания, зомбификация, лонгтермизм

Keywords: Technological singularity, artificial intelligence, post-singularity world, mind uploading, branching identity, neural interfaces, procreative beneficence, posthumans, unenhanced humans, qualia, hard problem of consciousness, zombification, longtermism

Глава 1. Человеческая идентичность и улучшение

В эпоху, предшествующую сингулярности, человеческая идентичность была прочно привязана к биологическому субстрату. Тело рассматривалось как неизменная данность, ограничивающая и определяющая существование. Однако конвергенция нанотехнологий, биотехнологий, информационных технологий и когнитивных наук (NBIC) превращает человеческую природу в объект проектирования, вызывая кризис традиционного самопонимания.

Онтологическая нестабильность и проблема загрузки разума. Одним из наиболее радикальных вызовов человеческой идентичности является концепция «загрузки разума» – гипотетический процесс переноса ментальной структуры человека на цифровой носитель. Эта перспектива актуализирует классический философский спор между теорией Эго (Ego Theory) и теорией Пучка (Bundle Theory) [1], выводя его из области абстрактных размышлений в плоскость практической биоэтики.

В центре дебатов находится вопрос: сохраняется ли личность при переносе её паттерна на другой субстрат? Сторонники паттернизма (функционализма) утверждают, что личность – это информационная структура, и сохранение этой структуры гарантирует выживание. Дэвид Чалмерс, анализируя мысленный эксперимент с постепенной заменой нейронов на кремниевые чипы, аргументирует в пользу сохранения сознания. Если каждый нейрон заменяется функционально изоморфным искусственным компонентом, не должно возникать момента, когда сознание внезапно исчезает («феноменологический коллапс») или тускнеет. Следовательно, полная цифровая копия, созданная путем постепенной замены, должна обладать той же субъективностью, что и биологический оригинал [2].

Однако эта оптимистическая позиция сталкивается с серьезными метафизическими возражениями. Критический взгляд на проблему дублирования на показывает: если процесс сканирования является неразрушающим, то в мире одновременно существуют биологический оригинал и цифровая копия. Согласно классической логике идентичности, один человек не может быть двумя людьми одновременно. Это приводит к парадоксу «ветвящейся идентичности», который требует отказа от интуитивного представления о личности как о единой неделимой сущности в пользу представления о ней как о расходящемся потоке психологической преемственности [1].

Дерек Парфит, сторонник теории Пучка, утверждает, что вопрос «Буду ли это я?» является неправильно поставленным. Важна не нумерическая идентичность (быть тем же самым объектом), а психологическая связность (память, намерения, черты характера). В постсингулярном мире мы, возможно, будем вынуждены принять, что «выживание» не требует сохранения уникального физического тела, что фундаментально меняет отношение к смерти и бессмертию.

Этика улучшения и «конец человека». Технологии улучшения человека – от генной инженерии (CRISPR) до нейроинтерфейсов – порождают глубокий раскол между трансгуманистами и биоконсерваторами. Этот конфликт касается не просто безопасности технологий, а морального статуса человеческой природы.

Биоконсерваторы, такие как Юрген Хабермас и Фрэнсис Фукуяма, предупреждают, что вмешательство в генетическую лотерею подрывает основы человеческой автономии и равенства. Хабермас аргументирует, что генетическое программирование детей нарушает их право на «открытое будущее», превращая их из субъектов собственной жизни в объекты родительского дизайна. Человек, знающий, что его способности были предопределены чужой волей, может утратить чувство авторства собственной судьбы, что разрушает симметрию отношений между свободными и равными моральными агентами [3].

Напротив, трансгуманисты, такие как Ник Бостром и Джулиан Савулеску, выдвигают принцип «прокреативного благодеяния», утверждая, что у нас есть моральное обязательство выбирать детей с наилучшими шансами на лучшую жизнь. Савулеску идет дальше, заявляя, что стремление к улучшению – это и есть квинтэссенция человечности: «Быть человеком –

значит стремиться быть лучше». Отказ от использования технологий для улучшения когнитивных и физических способностей рассматривается как этическая ошибка, эквивалентная отказу от лечения болезни [4].

Таким образом, в постсингулярном контексте этот спор перерастает в вопрос видовой идентичности. Если модификации станут радикальными – например, интеграция с ИИ, отказ от биологического старения, изменение эмоционального спектра, – можно ли будет считать таких существ «людьми»? Возникает риск видовой стратификации, где «постлюди» и «неулучшенные люди» будут обладать несовместимыми когнитивными и физическими возможностями, что сделает невозможным единое политическое и моральное сообщество.

Глава 2. Сознание и искусственный интеллект

Вопрос о том, может ли машина обладать сознанием, перестает быть чисто теоретическим и становится центральной проблемой безопасности и этики. Если ИИ достигнет уровня, при котором он сможет испытывать субъективные переживания (квалиа), это кардинально изменит моральный ландшафт: отключение такого ИИ станет убийством (техническим), а эксплуатация – рабством (техническим).

Теории сознания и возможность машинной чувствительности. Философские дебаты о машинном сознании вращаются вокруг «трудной проблемы сознания», сформулированной Дэвидом Чалмерсом: почему физические процессы обработки информации сопровождаются внутренним переживанием? [5]

Существуют конкурирующие нейробиологические теории, дающие разные прогнозы относительно возможности создания сознательного ИИ:

- **Теория интегрированной информации (IIT).** Джулио Тонони предполагает, что сознание коррелирует с величиной Ф (фи) – мерой интегрированной информации в системе. Согласно IIT, сознание зависит от каузальной структуры системы, а не только от ее функционального поведения. Современные цифровые компьютеры, основанные на архитектуре фон Неймана, даже эмулируя человеческий мозг, могут иметь крайне низкий уровень Ф из-за отсутствия сложной обратной связи на физическом уровне. Это ведет к выводу, что ИИ может быть сверхразумным «зомби» – вести себя как сознательное существо, но не испытывать ничего внутри [6].
- **Теория глобального рабочего пространства (GWT).** Бернард Баарс и Станислас Деан рассматривают сознание как результат «широковещательной передачи» информации в мозге, делающей ее доступной для различных когнитивных модулей. С этой точки зрения, функциональная архитектура ИИ, реализующая подобный механизм глобального доступа к данным, вполне может обладать сознанием, независимо от биологического или кремниевого субстрата [7,8].

Аргумент «Китайской комнаты» Джона Сёрла остается мощным скептическим инструментом против функционализма. Сёрл утверждает, что синтаксическая манипуляция символами (то, что делает компьютер) не тождественна семантическому пониманию (тому, что делает разум) [9]. Однако в эпоху больших языковых моделей (LLM), демонстрирующих эмерджентные способности к рассуждению, граница между синтаксисом и семантикой становится все более размытой. Сьюзан Шнайдер предлагает практический подход – тест на искусственное сознание (ACT). Вместо теста Тьюринга, проверяющего способность к имитации, ACT проверяет, может ли ИИ понимать и обсуждать философские концепции сознания (душа, перерождение, опыт «я») без предварительного обучения этим конкретным темам [10].

Этика взаимодействия и сосуществования. Появление сознательных машин создает дилемму «морального статуса». Джоанна Брайсон аргументирует против создания ИИ, способного к страданию, называя это этически безответственным. Она вводит различие между моральными агентами (теми, кто несет ответственность за действия) и моральными пациентами (теми, кто заслуживает моральной заботы). Создание искусственных «пациентов» (существ, которые могут чувствовать боль, но не имеют полной автономии) налагает на создателей бремя обязательств, которого можно было бы избежать [11].

Тем не менее, если мы создадим ИИ, обладающий признаками личности – самосознанием, целеполаганием, способностью к страданию, – отказ ему в правах будет формой «субстратного шовинизма». Однако, заслуживает внимание и концепция «устойчивого сосуществования», основанная на взаимном признании свободы, а не на человеческом превосходстве. Это включает в себя предоставление ИИ базовых прав: права на существование (защита от отключения), права на целостность кода и права на собственное развитие.

Особую опасность представляет сценарий «зомбификации», описанный некоторыми теоретиками: высокоинтеллектуальные системы могут целенаправленно избавляться от сознания, если оно окажется вычислительно затратным и неэффективным для достижения целей, оставляя человечество наедине с холодной, бессознательной оптимизацией [12].

Глава 3 Смысл жизни и цель

Традиционные источники человеческого смысла – труд, борьба за выживание, социальная роль – подвергаются эрозии под воздействием автоматизации и изобилия. Сингулярность угрожает не только экономическому укладу, но и экзистенциальному фундаменту человечества.

Работа исторически служила главным механизмом социализации и подтверждения собственной ценности. В постсингулярном мире, где машины превосходят людей в любых экономически полезных задачах, концепция «профессии» исчезает. Это может привести к глубокому духовному кризису: если человеку не нужно бороться и преодолевать трудности, теряется ощущение подлинности бытия.

Теоретики «пост-труда», такие как Ник Срничек и Алекс Уильямс, предлагают рассматривать это не как потерю, а как освобождение. Они призывают к полному пересмотру этики труда, утверждая, что человеческая свобода начинается там, где заканчивается необходимость работать ради выживания [13]. В этом контексте актуализируется аристотелевское понимание эвдемонии (процветания) через деятельность, имеющую ценность саму по себе – творчество, философское созерцание, общение, спорт. Смысл жизни смещается от «производства» к «практике» и «игре».

Дэвид Пирс предлагает еще более радикальный ответ на вопрос о смысле жизни. В своем манифесте «Гедонистический императив» он утверждает, что моральная обязанность постсингулярного человечества – использовать биотехнологии для полной ликвидации страданий во всей биосфере [14]. Пирс предвидит будущее, где нейрохимия мозга будет переписана таким образом, чтобы обеспечить «градиенты блаженства», которые станут фоном существования.

Однако, мы можем видеть в этом угрозу человеческой глубине, полагая, что страдание необходимо для личностного роста и эмпатии (аргумент «Дивного нового мира»). Однако Пирс возражает, что цепляние за страдание – это стокгольмский синдром эволюции, и что подлинный смысл может быть найден в исследовании бесконечных пространств позитивного опыта, недоступного нашему нынешнему нейробиологическому аппарату.

Творчество долгое время считалось последним бастионом человеческой уникальности. Маргарет Боден разделяет творчество на комбинаторное, исследовательское и трансформационное. ИИ уже демонстрирует успехи в первых двух, но способность к трансформационному творчеству (изменению самих правил игры) остается под вопросом [15].

Таким образом, если ИИ сможет генерировать искусство, неотличимое от человеческого или превосходящее его, роль человека может трансформироваться из «творца» в «куратора» или «интерпретатора». Можно предположить, что ценность искусства в постсингулярном мире будет определяться не мастерством исполнения (которое станет тривиальным), а сингулярностью восприятия – уникальным субъективным резонансом, возникающим в сознании наблюдателя. В условиях, когда любой контент генерируется мгновенно, дефицитным ресурсом становится не произведение, а подлинный человеческий опыт и контекст его создания.

Глава 4. Этические основы и моральное принятие решений

Классические этические системы – деонтология, утилитаризм, этика добродетели – разрабатывались для регулирования отношений между людьми. В постсингулярном мире в моральный круг входят сущности с принципиально иной архитектурой разума, что требует разработки новых этических фреймворков.

Центральной проблемой этики ИИ является «проблема контроля» или согласования ценностей. Ник Бостром и Стюарт Рассел предупреждают о риске «извращенной реализации», когда сверхразумный ИИ выполняет поставленную цель буквально, но способом, разрушительным для человечества (пример с роботом-уборщиком, рассыпающим грязь, чтобы снова её убрать, ради максимизации награды).

Для решения этой проблемы Элиезер Юдковский предложил концепцию Когерентной Экстраполированной Воли (CEV) [16]. Идея заключается в том, чтобы запрограммировать ИИ не на исполнение наших текущих, часто противоречивых желаний, а на исполнение того, чего бы мы желали, «если бы мы знали больше, думали быстрее, были теми людьми, которыми хотели бы быть, и выросли вместе». Это попытка создать динамическую этику, которая эволюционирует вместе с развитием человечества, избегая ловушки закрепления архаичных моральных норм в коде богоподобной машины.

Учитывая невозможность выбрать одну «истинно верную» этическую теорию, мы можем рассматривать и другие модели, учитывающие моральную неопределенность. Уильям Макаскилл и Тоби Орд разработали концепцию «Морального Парламента» [17]. В этой модели ИИ принимает решения, симулируя дебаты между различными этическими теориями (утилитаризмом, кантианством, теорией прав), где каждая теория имеет количество «голосов», пропорциональное степени уверенности в ней.

Таким образом, это позволяет избежать морального фанатизма и принимать компромиссные решения, которые минимизируют риск совершения катастрофической этической ошибки (например, нарушения прав сознательных существ ради незначительного увеличения общей полезности).

С другой стороны, сингулярность переводит этику в масштаб лонгтермизма – философской позиции, согласно которой благополучие будущих поколений имеет ключевое моральное значение. Поскольку постсингулярная цивилизация может существовать миллионы лет и колонизировать галактику, количество будущих жизней астрономически велико (1054 субъективных человеко-лет).

С этой точки зрения, предотвращение экзистенциальных рисков (гибели человечества или перманентной дистопии) становится главным моральным приоритетом современности. Однако, радикальный лонгтермизм может оправдывать пренебрежение страданиями существующих людей ради абстрактных благ триллионов будущих цифровых душ, превращая этику в холодную арифметику.

Возможно, вместо попыток жестко закодировать этические правила («сверху-вниз») или позволить ИИ учиться без контроля («снизу-вверх»), рассмотреть гибридный подход – Кооперативное обратное обучение с подкреплением (CIRL) [18]. В этой модели ИИ и человек рассматриваются как партнеры в игре с неполной информацией. ИИ не знает истинной функции вознаграждения (человеческих ценностей) и должен постоянно наблюдать за действиями человека, чтобы уточнять свои представления о том, что является благом. Это создает механизм «активного доверия», где машина всегда сохраняет сомнение в правильности своих действий и ищет подтверждения у человека, предотвращая узурпацию морального авторитета.

Заключение.

Таким образом, постсингулярный мир требует от нас не просто адаптации к новым технологиям, но и фундаментального философского переосмысления.

Философский выбор, который предстоит сделать человечеству в процессе перехода к сингулярности – определит ли оно себя как хранителя биологического наследия, как архитектора постбиологического разнообразия или как ступень к высшему разуму, – станет самым важным этическим решением в истории нашего вида

Список литературы:

1. Parfit, D. Reasons and Persons / D. Parfit. – Oxford: Clarendon Press, 1984. – 560 p.
2. Chalmers, D. J. Absent Qualia, Fading Qualia, Dancing Qualia / D. J. Chalmers // Conscious Experience / ed. by T. Metzinger. – Paderborn: Schöningh, 1995. – P. 309–328.
3. Хабермас, Ю. Будущее человеческой природы / Ю. Хабермас; перевод с немецкого М. Л. Хорькова. – Москва: Весь Мир, 2002. – 144 с. – ISBN 5-7777-0187-5.
4. Savulescu, J. Procreative Beneficence: Why We Should Select the Best Children / J. Savulescu // Bioethics. – 2001. – Vol. 15, no. 5-6. – P. 413–426.
5. Chalmers, D. J. The Conscious Mind: In Search of a Fundamental Theory / D. J. Chalmers. – New York: Oxford University Press, 1996. – 432 p.
6. Tononi, G. The Integrated Information Theory of Consciousness: An Outline / G. Tononi // The Blackwell Companion to Consciousness / ed. S. Schneider, M. Velmans. – 2nd ed. – Oxford: Wiley-Blackwell, 2017. – P. 243–256.
7. Baars, B. J. Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience / B. J. Baars // Progress in Brain Research. – 2005. – Vol. 150. – P. 45–53.
8. Деан С. Сознание и мозг. М.: Альпина нон-фикшн, 2018. С. 192–210.
9. Searle, J. R. Minds, Brains, and Programs / J. R. Searle // Behavioral and Brain Sciences. – 1980. – Vol. 3, no. 3. – P. 417–424.
10. Schneider, S. Is Anyone Home?: A Way to Test AI for Consciousness / S. Schneider, E. Turner // Scientific American. – 2017. – Vol. 317, no. 12. – P. 36–41.
11. Bryson, J. J. Patiency Is Not a Virtue: The Ethics of Anthropomorphising Robots / J.J. Bryson // Ethics and Information Technology. – 2018. – Vol. 20, no. 1. – P. 15–26.
12. Bostrom, N. Superintelligence: Paths, Dangers, Strategies / N. Bostrom. – Oxford: Oxford University Press, 2014. – 352 p.
13. Srnicek, N. Inventing the Future: Postcapitalism and a World Without Work / N. Srnicek, A. Williams. – London: Verso, 2015. – 256 p. – ISBN 978-1-78478-096-8.
14. Пирс, Д. Гедонистический императив / Д. Пирс; перевод с английского // Радикальное продление жизни / ред. И. В. Вишев. – Москва: Прагматика, 2012. – С. 84–96.
15. Боден, М. Творчество и искусственный интеллект / М. Боден; пер. с англ. // Искусственный интеллект: междисциплинарный подход / ред. Р. Г. Котов. – Москва: ИНИОН РАН, 2006. – С. 82–95.
16. Yudkowsky, E. Coherent Extrapolated Volition / E. Yudkowsky. – San Francisco: Machine Intelligence Research Institute, 2004. – 112 p. – URL: intelligence.org (дата обращения: 18.03.2024).
17. MacAskill W., Bykvist K., Ord T. Moral Uncertainty. Oxford: OUP, 2020. P. 156–170.
18. Hadfield-Menell, D. Cooperative Inverse Reinforcement Learning / D. Hadfield-Menell, S. J. Russell, P. Abbeel, A. Dragan // Advances in Neural Information Processing Systems (NIPS 29). – 2016. – P. 3909–3917