



УДК/UDC 179

DOI: <https://doi.org/10.25206/2542-0488-2026-11-3-80-95>

EDN: LEIKZQ

Научная статья / Original article

ЦИФРОВЫЕ ДВОЙНИКИ И ПРОДЛЕНИЕ ЖИЗНИ: ЧТО ПО-НАСТОЯЩЕМУ ВАЖНО?

С. Иглесиас¹, Б. Д. Йорп¹, К. Войнеа¹, С. П. Манн¹,
А. Захиу², Н. С. Джекер³, Дж. Савулеску¹

¹Оксфордский университет, Оксфорд, Великобритания

²Бухарестский университет, Бухарест, Румыния

³Вашингтонский университет, Вашингтон, США

Перевод с английского

А. В. Антипов 

Институт философии Российской академии наук,
Россия, 109240, г. Москва, ул. Гончарная, 12, стр. 1

 nelson02@yandex.ru

Продолжаются непрекращающиеся дискуссии об этике исследований в области продления жизни: в общих чертах, речь идет о применении медицинских технологий для увеличения биологической продолжительности человеческой жизни, за пределы существующего 'естественного' предела примерно в 120 лет. Одновременно наблюдается стремительный рост интереса к использованию искусственного интеллекта для создания 'цифровых двойников' людей, например, с помощью тонкой настройки больших языковых моделей на данных, относящихся к конкретным индивидам.

В работе мы исследуем вопрос о том, могут ли цифровые двойники (или, как мы их называем, цифровые допдвигангеры) стать путем к своего рода продлению жизни, или, точнее, продлению личности, которое не опирается на биологическую непрерывность. Мы рассматриваем соответствующие концепции сознания и тождества личности и доказываем, что цифровые допдвигангеры могут помочь нам достичь, по крайней мере, некоторых целей или предполагаемых благ, связанных с продлением существования личности, даже если они, согласно доминирующим философским подходам, не могут считаться буквальным продолжением нашей самости. Мы также принимаем во внимание *реляционные* концепции тождества личности и обсуждаем, каким образом цифровые допдвигангеры могут продлевать личность в реляционном смысле или сохранять некоторые блага, связанные с важными отношениями.

В заключение мы утверждаем, что исследовательская программа по изучению подобных вопросов актуальна для продолжающихся обсуждений об этике увеличения продолжительности человеческой жизни.

Ключевые слова: старение, цифровые двойники, цифровые близнецы, продление жизни, исследования долголетия, искусственный интеллект

Для цитирования: Иглесиас С., Йорп Б. Д., Войнеа К., Манн С. П., Захиу А., Джекер Н. С., Савулеску Дж. Цифровые двойники и продление жизни: что по-настоящему важно? = Iglesias S., Earp B. D., Voinea C., Mann S. P., Zahiu A., Jecker N. S., Savulescu J. Digital Doppelgängers and Lifespan Extension: What Matters? / пер. с англ. А. В. Антипова. Омский научный вестник. Серия «Общество. История. Современность». 2026. Т. 11, № 3. С. 80–95. <https://doi.org/10.25206/2542-0488-2026-11-3-80-95>. EDN: LEIKZQ.



© Иглесиас С., Йорп Б. Д., Войнеа К., Манн С. П.,
Захиу А., Джекер Н. С., Савулеску Дж., Антипов А. В., 2026.
Контент доступен под лицензией Creative Commons Attribution 4.0 License.

DIGITAL DOPPELGÄNGERS AND LIFESPAN EXTENSION: WHAT MATTERS?

S. Iglesias¹, B. D. Earp¹, C. Voinea¹, S. P. Mann¹,
A. Zahiu², N. S. Jecker³, J. Savulescu¹

¹University of Oxford, Oxford, UK

²University of Bucharest, Bucharest, Romania

³University of Washington, Washington, USA

Translated from English

A. V. Antipov 

Institute of Philosophy, Russian Academy of Sciences,
Goncharnaya St., 12, bld. 1, Moscow, 109240, Russia

 nelson02@yandex.ru

There is an ongoing debate about the ethics of research on lifespan extension: roughly, using medical technologies to extend biological human lives beyond the current 'natural' limit of about 120 years. At the same time, there is an exploding interest in the use of artificial intelligence to create 'digital twins' of persons, for example by fine-tuning large language models on data specific to particular individuals.

In this paper, we consider whether digital twins (or digital doppelgängers, as we refer to them) could be a path toward a kind of life extension — or more precisely, a kind of *person* extension — that does not rely on biological continuity. We discuss relevant accounts of consciousness and personal identity and argue that digital doppelgängers may at least help us achieve some of the *aims* or ostensible goods of person-span extension, even if they may not count as literal extensions of our personhood on dominant philosophical accounts. We also consider *relational* accounts of personhood and discuss how digital doppelgängers may be able to extend personhood in a relational sense, or at least secure some of the goods associated with relevant relationships.

We conclude by suggesting that a research program to investigate such issues is relevant to ongoing debates about the ethics of extending the human lifespan.

Keywords: aging, digital duplicates, digital twins, life extension, longevity research, Artificial Intelligence

For citation: Iglesias S., Earp B. D., Voinea C., Mann S. P., Zahiu A., Jecker N. S., Savulescu J. Digital Doppelgängers and Lifespan Extension: What Matters? / trans. from Engl. A. V. Antipov. *Omsk Scientific Bulletin. Series Society. History. Modernity*. 2026;11(3):80–95. <https://doi.org/10.25206/2542-0488-2026-11-3-80-95>. EDN: LEIKZQ. (In Russ.).



© Iglesias S., Earp B. D., Voinea C., Mann S. P.,
Zahiu A., Jecker N. S., Savulescu J., Antipov A. V., 2026.

The content is available under a Creative Commons Attribution 4.0 License.

Введение

Этика и регулирование исследований в области продления человеческой жизни представляют собой тему, связанную с растущим к ней интересом и сопутствующими дискуссиями [1–3]. В условиях ограниченных доступных ресурсов возникают важные вопросы о том, как оптимальным образом составлять приоритеты и распределять финансирование. Ключевое противоречие заключается в выборе между инвестициями в технологии продления жизни и преодоления существующего естественного предела, составляющего приблизительно 120 лет, и сосредоточением усилий на повышении качества жизни в рамках типичной человеческой продолжительности жизни [4].

Для обсуждения подобных компромиссов необходимо, однако, прежде всего понимать, что именно сторонники продления жизни стремятся продлевать. В статье мы утверждаем, что термин

‘продление жизни’ может быть ошибочным: в конечном счете сторонников продления жизни интересует не столько пролонгирование жизни (то есть продолжение биологического функционирования), а скорее, как мы предполагаем, их истинной целью является продление собственного Я как переживающей опыт самости с определенной идентичностью, жизненными проектами, социальными связями и т. д. Более того, возможно, что человеческое Я лишь *контингентным* образом воплощено в биологической форме.

Например, хотя это и крайне умозрительно, в будущем может стать возможным загрузить все релевантные функциональные элементы центральной нервной системы на компьютерный чип и таким образом продолжить свое сознательное восприятие и ощущение самости в цифровой форме. В этом случае, согласно некоторым точкам зрения, индивид сможет продлить существование сво-

ей личности, или, грубо говоря, свою ментальную жизнь, при этом не обязательно продлевая физическую или биологическую жизнь.

Тем не менее данная возможность также является крайне спорной и, вероятно, она не станет осуществимой в обозримом будущем, если вообще станет когда-либо реализуемой. Поэтому, возможно, имеет смысл пойти еще глубже и спросить, почему кто-либо может желать продлить то, что мы называем *существованием личности* (*person-span*)? Существуют различные потенциальные причины, по которым индивиды могут стремиться к этому, и мы не претендуем на исчерпывающую типологию. Чтобы сохранить фокус нашего анализа, мы рассмотрим три основные причины:

Цель 1: **Опыт:** продолжать испытывать субъективный сознательный опыт, особенно когда он является позитивным или важным.

Цель 2: **Наследие/Влияние:** оставить прочный вклад, добиться признания ценности своей жизни и ее влияния на жизни других людей, а также участвовать в формировании будущего. Данная мотивация связана не с субъективным опытом самого человека, а с его посмертным влиянием на общество, отдельные сообщества в нем или мир в целом.

Цель 3: **Отношения:** обеспечить устойчивое благополучие конкретных близких людей и развивать особые отношения с ними: в первую очередь с теми, о ком человек заботился, например, о его любимых людях. В данном случае мотивация направлена не на общество, отдельные сообщества в нем или мир в целом, а на благо конкретных индивидов, с которыми человека связывала та или иная форма близости.

При современном уровне технологий обеспечение всех трех потенциальных целей продления существования личности (сохранение субъективного опыта, создание прочного наследия или воздействия на мир, а также поддержание важных отношений) требует достижений в области биологического продления жизни, которые, как мы предположили, маловероятны в ближайшей перспективе. Поэтому в данной статье мы сосредоточимся на ином способе реализации если не самого продления нашей подлинной личности в строгом философском смысле, то, по крайней мере, некоторых из глубинных целей, которые, как мы предполагаем, и составляют значительную часть смысла подобных устремлений. В частности, мы исследуем, как вторая и третья цели, касающиеся наследия и отношений, могут быть реализованы посредством создания *цифрового допсельгангера*.

Хотя для создания подобного допсельгангера как в настоящее время, так и в будущем могут применяться различные методы, мы сосредоточимся на актуальном и весьма показательном примере: использовании так называемой персонализированной большой языковой модели (LLM).

Речь идет о виде искусственного интеллекта, обычно представленного в форме чат-бота, который был тонко настроен на обширном массиве данных, специфичных для конкретного человека, что позволяет ему «вести беседу» в стиле и с содержанием, поразительно схожими с самим этим человеком [5; см. также: 6; 7].

В случае успешной реализации такой чат-бот смог бы пройти так называемый 'Персонализированный тест Тьюринга' [8, 9]: иными словами, располагая лишь лингвистическими данными, было бы невозможно отличить выше вероятности случай-

ного угадывания, ведется ли разговор с индивидом или с его цифровым 'двойником'.

Для ясности уточним, мы не станем отстаивать точку зрения, согласно которой цифровой допсельгангер, будь то на основе большой языковой модели или иной технологии, с необходимостью представляет собой форму продления существования личности. Вопрос о том, является ли он таковым или может им быть, отчасти зависит от принятой философской теории личности, и мы воздержимся от суждений о том, какая теория является верной или наиболее обоснованной. Вместо этого мы предполагаем, что цифровой допсельгангер потенциально способен способствовать достижению, по крайней мере, некоторых из главных *целей* или обеспечению предполагаемых *благ*, связанных с продлением собственного 'Я' во времени (например, за пределы нынешних биологических ограничений), даже если сам по себе он не может квалифицироваться как продление существования личности.

Если говорить конкретнее, то, хотя цифровые допсельгангеры, скорее всего, не достигают Цели 1 — а именно, сохранения ценного *субъективного опыта*, — они могут, как мы предполагаем, по крайней мере, частично обеспечить блага, связанные с Целями 2 и 3: теми, что относятся к жизненным проектам, наследию или влиянию человека, а также к определенным аспектам его межличностных отношений. Например, цифровой допсельгангер мог бы продолжить писать роман, который человек не успел завершить из-за прогрессирующих симптомов болезни Альцгеймера перед смертью; или же человек, посредством своего цифрового допсельгангера, мог бы, в некотором смысле, продолжать беседы с друзьями и близкими, тем самым способствуя поддержанию (определенных аспектов) важных отношений [10, 11].

Как показывают уже приведенные примеры и как будет подробнее рассмотрено далее, точный смысл, в котором цифровой допсельгангер может быть признан реализующим определенные цели, например завершение романа или продолжение беседы, не является однозначным. Человеческой личности присущи агентность, интенциональность, воображение и (с различными досадными исключениями) стремление к истине и смыслу. У больших языковых моделей, напротив, ничего из этого нет, или, по крайней мере, именно так считается [12]. Соответственно, говоря, что *роман написан* человеком-автором (продолжая этот пример), мы обычно не имеем в виду, что он, подобно зомби, напечатал очень длинную последовательность слов, которая случайным образом сложилась в повествование. Скорее, мы подразумеваем, что он *намеревался* создать при помощи слов историю; что эта история передает определенные смыслы, и так далее. Истории, написанные большими языковыми моделями, этого не делают [13].

Но теперь предположим, что наш автор тщательно обучает или дообучает модель на всех своих предыдущих произведениях, включая, допустим, первые 80 % своего большого заключительного романа — романа, который он, как ему известно из-за недавнего диагностирования болезни Альцгеймера, скорее всего, не сможет завершить. Предположим далее, что до того, как его переведут в дом престарелых, где он проведет остаток своих дней и больше не будет писать, он подготавливает подробный план последних 20 % романа: того, как он *намере-*

вался его закончить. Допустим, он вводит этот план в модель в качестве финального 'промпта', ожидающего исполнения, с примечанием, в котором выражает желание, чтобы его супруга запустила модель в случае его кончины.

А теперь предположим, что автор умирает; его супруга запускает модель; персонализированная большая языковая модель выполняет свою задачу. Допустим, она генерирует заключительные 20 % романа в собственном стиле автора и в соответствии с его заранее определенным планом, так что конечный результат представляет собой нечто, что сам автор, будь он живой и без симптомов болезни, *признал бы* удовлетворительным завершением книги.

Что можно сказать об этих 20 %? Хотя было бы правильно заключить, что сама большая языковая модель (в данном примере цифровой допсельгангер автора) не имела *намерения*, чтобы история получилась именно такой; и она также не использовала *собственную* агентность или воображение для передачи какой-либо истины или смысла, все же возможно, что намерения, смыслы и общие цели *автора* в отношении книги были в значительной степени реализованы. Это объясняется тем, что автор использовал персонализированную модель, обученную на его собственных прошлых текстах, то есть материале, который с высокой долей вероятности содержит его предыдущие коммуникативные усилия, интенции, идеи, мастерство, творчество, воображение и так далее [14], а также ряд инструкций, явным образом закодировавших его желания и планы на будущее. Поскольку эти характеристики дообученной модели, отражающие авторскую личность, сыграли значительную роль в формировании последних 20 % романа так, что автор гипотетически одобрил бы их, кажется правомерным утверждать, что посредством своего цифрового допсельгангера он посмертно 'завершил' свой роман.

Именно такого рода идею мы имеем в виду, говоря о том, что цифровой допсельгангер способен реализовывать определенные виды целей.

Разумеется, в нашем примере было бы предпочтительнее, если бы автор мог завершить свой роман, так сказать, полностью самостоятельно. Это был бы *наилучший* способ достижения блага, связанного с написанием этой последней книги. Однако, учитывая разрушительное воздействие болезни Альцгеймера и преждевременную кончину автора, наилучший путь оказался недоступен. Таким образом, мы полагаем, что *даже если* кто-то настаивает на том, что цифровой допсельгангер автора никоим образом не является буквальным продолжением его личности, все же может показаться справедливым утверждение, что он способствовал достижению, пусть и второстепенным образом, конечного блага, а именно — завершения книги.

Мы полагаем, что подобная логика должна применяться и к проектам по продлению жизни. Перед лицом ограниченных ресурсов нам следует рассмотреть, каким образом цифровые допсельгангеры могут выступать в качестве альтернативного, возможно, менее совершенного, но все же ценного, средства достижения, по крайней мере, некоторых из тех целей, которые, как мы предполагаем, вероятнее всего, и мотивируют стремление к увеличению продолжительности жизни (или, как мы это называем, существования личности). Например, возможно, что инвестиции в технологические и этические исследования в области цифровых допсельгангеров

должны получать со стороны правительств более высокий приоритет, чем это имеет место в настоящее время, по крайней мере, среди тех, кто уже финансирует проекты по продлению жизни (при условии, что такие расходы оправданны, что может как соответствовать, так и не соответствовать действительности, но мы не будем рассматривать эту более широкую проблему здесь).

Для обоснования данного тезиса мы начнем с представления концепции цифровых допсельгангеров и последних достижений в области генеративного искусственного интеллекта, в частности, больших языковых моделей, а также разрастающейся подобласти персонализированных больших языковых моделей. Затем обратимся к философской литературе о сознании и тождестве личности, чтобы оценить, может ли, и в каком смысле, цифровой допсельгангер рассматриваться как продолжение личности или, если это не так, по крайней мере, как средство достижения некоторых из *целей*, которые обычно связаны со стремлением к такому продлению. Далее, прежде чем завершить данное эссе, мы рассмотрим ряд недостатков и потенциальных сложностей.

Что такое цифровые допсельгангеры?

Термин 'допсельгангер' (от нем. *doppelgänger* — букв. 'двойник') обычно используется для обозначения биологически не связанного человека, который выглядит и/или ведет себя очень похоже на другого человека. Цифровой допсельгангер, как мы используем этот термин, — это программа искусственного интеллекта, созданная для максимально точного воспроизведения личности, идиосинкразических характеристик (включая стиль речи или письма), решений, моральных суждений, воспоминаний и другого поведения данного индивида, которое может быть выражено в лингвистической форме. Мы говорим о 'лингвистическом' поведении, потому что, хотя в скором времени может стать возможным воплощать цифровых допсельгангеров в убедительно реалистичных роботах с синтетическими физическими чертами и движениями, напоминающими имитируемого человека, соответствующие технологии для этого остаются на ранней стадии разработки. Напротив, создать чисто программный допсельгангер в стиле чат-бота уже возможно как в формате только текста, так и в аудиоформатах, включающих имитацию голоса [15]. Также возможно создать убедительный 'дипфейк' лица человека, который можно отображать на экране компьютера и синхронизировать с его речевыми паттернами, создавая впечатление, будто вы разговариваете с человеком по видеосвязи [15]; однако полностью воплощенные роботизированные допсельгангеры пока недоступны [подробнее см.: 16].

Идея цифрового допсельгангера все больше захватывает общественное воображение через художественные образы в романах, таких как *Клара и Солнце* Исигуро, и популярные нарративы, вроде *Мира дикого Запада* (HBO) и *Черного зеркала* (Netflix). Появляются и реальные прототипы. Например, приложение HereAfterAI проводит длительные интервью с пользователями и побуждает их вспоминать истории, чтобы создать цифровой аналог, способный отражать их убеждения, желания и воспоминания [17]. Другое приложение, Replika, специализируется на создании виртуальных репрезентаций людей [18]. Возможно, наиболее актуальным для настоящих целей является сервис YOY

(*You, Only Virtual*), который предлагает услугу, помогающую «продолжать делиться драгоценными моментами с любимым человеком даже после физической смерти» [19].

По словам Эндрю Каплана, из числа первых пользователей HereAfterAI, который в возрасте 78 лет хотел стать одним из первых 'цифровых людей': «Для меня это вопрос истории, своего рода ограниченного бессмертия, которое создает интимный личный опыт для моих будущих родственников, которые захотят узнать, откуда они произошли» [цит. по: 20]. Основательница Replika Евгения Куйда изначально создала компанию после того, как ее друга Романа Мазуренко сбила машина, а водитель скрылся. Куйда отметила, что ответы цифрового допфельгангера Мазуренко, созданного технологией ее компании, были «точными», а «возможность вернуться к нему, продолжать общение, которое у нас было раньше, была своего рода терапией» [цит. по: 21]. Джеймс Влахос утверждает, что он регулярно взаимодействует с цифровым допфельгангером своего покойного отца, названным 'Папаботом': «Мне очень нравится, что он может чувствовать себя более вовлеченным в мою жизнь, проявлять те черты своей личности, которые я так люблю» [цит. по: 21].

Хотя сама концепция не ограничивается каким-либо одним способом реализации, в данной работе мы сосредоточимся конкретно на цифровых допфельгангерах, основанных на персонализированных больших языковых моделях. Большие языковые модели представляют собой тип генеративного искусственного интеллекта, который использует комбинацию глубинного обучения, трансформерной архитектуры и предварительного обучения на больших объемах данных естественного человеческого языка для прогнозирования следующего токена (короткого слова или фрагмента более длинного слова) в последовательности токенов (предложений, абзацев и т. д.).

Поскольку они обучаются на значительной части всего написанного человеком текста, передовые большие языковые модели, такие как Anthropic's Claude 3, OpenAI's GPT-4, а также открытые модели, наподобие LLaMA 2, демонстрируют высокий уровень общей производительности в решении разнообразных задач [22].

Несмотря на высокую эффективность, эти общие модели часто уступают моделям, адаптированным к конкретным задачам посредством процесса, известного как 'тонкая настройка' (fine-tuning) [23]. Этот процесс предполагает использование более специфического набора данных для обучения общей модели, в результате чего она сохраняет свою общую статистическую модель языковых паттернов человека (усвоенную в ходе первоначального обучения на общих наборах данных), но при этом генерирует выводы, на которые также влияют паттерны из этого более специфичного набора данных. Например, тонко настроенные модели могут использоваться для более точного, чем общие модели, анализа клинических записей врачами [24]. Другой подход к адаптации общих моделей, известный как 'генерация, дополненная извлечением' (retrieval-augmented generation, RAG) [25], заключается в предоставлении им доступа к специализированной базе знаний, например, набору файлов PDF, чтобы они могли напрямую использовать информацию, содержащуюся в этих внешних базах знаний, при составлении ответов на запросы.

В последнее время все больше обсуждается особый тип таких кастомизированных больших языковых моделей, как *персонализированные* большие языковые модели [26]. Это тонко настроенные большие языковые модели, обеспеченные базами знаний, относящимися к текстам, написанным конкретным индивидом или о нем (включая транскрипты устной речи). Было показано, что такие модели способны давать ответы на вопросы [27] и создавать академические тексты [5] в стиле авторов, на чьих статьях и книгах они обучались. Персонализированные большие языковые модели могут использоваться для самых разных целей, включая предсказание предпочтений в медицинском лечении в случаях, когда недееспособный человек не оставил заранее выраженных указаний [28].

В данной работе мы утверждаем, что способность персонализированных больших языковых моделей убедительно имитировать ответы индивидов, пусть даже в настоящее время и с разной степенью надежности, открывает интригующие возможности для создания на их основе цифровых допфельгангеров. Путем тонкой настройки большой языковой модели, дополненной методом извлечения, обученной на основе надлежащим образом авторизованного текста, произведенного конкретным индивидом и/или его описывающего, может быть создана модель, способная реагировать способами, сходными с этим индивидом [6, 7]. В зависимости от качества и полноты обучающих данных, а также базовой модели, такая большая языковая модель-допфельгангер могла бы генерировать новые выводы, отражающие знания, личность и уникальные паттерны мышления и самовыражения этого индивида.

Прежде чем продолжить, следует признать наличие существенного возражения. Может быть выдвинуто предположение, что большие языковые модели плохо подходят для наших целей, поскольку они склонны продуцировать грамматически правильные, но фактически или логически сомнительные выводы: феномен, известный как 'конфабуляция' (иногда также называемый *галлюцинациями*). Например, недавно было обнаружено, что судебные документы содержат «фальшивые судебные решения с вымышленными цитатами и внутренними ссылками», поскольку юристы небрежно полагались для их составления на такие инструменты, как ChatGPT [29]. В другом примере несколько ведущих больших языковых моделей некорректно сообщали, что слово 'strawberry' содержит только две буквы 'r' [30]. Так что может возникнуть вопрос: если на большие языковые модели нельзя положиться в точном представлении судебных решений или даже правописании простых слов, то почему кто-либо должен думать, что они смогут точно представить, или даже убедительно имитировать, нечто столь сложное и многогранное, как личность?

У нас есть три кратких ответа на это возражение. Во-первых, несмотря на то что это является серьезной проблемой, она может быть лишь временной. Вероятно, что большие языковые модели продолжают свой быстрый прогресс за счет сочетания увеличения размеров моделей и методов снижения конфабуляции: хорошо финансируемые компании, стоящие за большими языковыми моделями, имеют серьезный стимул решать самые известные проблемы, преследующие их основной продукт. Так, например, уровень галлюцинаций между GPT-3.5 (2022) и GPT-4 (2023), по различным оценкам, снизился на 11 % [31]; и, хотя возможно существо-

вание внутреннего предела этому прогрессу, которого можно разумно или экономически оправданно достичь в этих вопросах, в настоящее время есть тренд на продолжающееся улучшение.

Наш второй ответ заключается в том, что, по крайней мере, некоторая степень конфабуляции может быть не столь уж нежелательной при рассмотрении всех обстоятельств. В конце концов мы, люди, тоже иногда конфабулируем, и это может отражать определенные аспекты нашей когнитивной архитектуры, которые, возможно, стоит сохранить. Как отмечают Пейси Суи и его соавторы [32], конфабуляция больших языковых моделей может рассматриваться как отражение «человеческой склонности использовать повышенную нарративность в качестве когнитивного ресурса для осмысления и коммуникации», предполагая, возможно, контринтуитивно, «что склонность больших языковых моделей к конфабуляции может быть тесно связана с их позитивной способностью к генерации связанного нарративного текста» [32, р. 1]¹. Таким образом, хотя мы признаем, что конфабуляция может быть проблемой для задач, связанных с извлечением фактических знаний, мы оставляем возможность того, что причины конфабуляции могут оказаться сильной стороной, когда речь идет о создании убедительных и оригинальных человекоподобных ответов.

Наконец, наш третий ответ заключается в том, что обоснованность наших основных аргументов не зависит от текущих конкретных возможностей больших языковых моделей. Скорее, мы используем технологию больших языковых моделей в качестве доказательства концепции, демонстрируя, что уже сегодня возможно создавать достаточно точных цифровых допсельгангеров с помощью определенного типа искусственного интеллекта (ИИ) современного уровня развития. Что будет возможно с ИИ завтрашнего дня, будь то на основе больших языковых моделей или иных технологий, вопрос умозрительный; поэтому мы постарались сформулировать наши аргументы на таком уровне, который может учитывать целый спектр правдоподобных технологических достижений в этой области.

Итак, с учетом всего вышесказанного, мы перейдем к следующему разделу. В нем мы обратимся к философской литературе о тождестве личности и сознании, чтобы оценить, в какой степени можно утверждать, что основанный на основе ИИ цифровой допсельгангер представляет собой продолжение 'Я' или личности индивида. В конечном счете мы воздержимся от определенных суждений по этому философскому вопросу, но выдвинем тезис о том, что некоторые цели или предполагаемые блага от подобного рода расширения существования личности могут быть отчасти реализованы путем создания цифрового допсельгангера. При обосновании этого тезиса мы рассматриваем возможность того, что *эгоистическая забота* о 'судьбе' своего цифрового допсельгангера может быть рациональной; а также исследуем, может ли существовать *неэгоистическая*, ориентированная на отношения, забота о нем же. Чтобы понять в более общем смысле, могут ли цифровые допсельгангеры быть релевантными для уже существующих дискуссий о долголети и продлении человеческой жизни, необходимо разобраться в том, что такое цифровой допсельгангер и как он соотносится с личностью, на основе которой создан.

Сознание и тождество личности

В данном разделе мы затрагиваем два основных вопроса:

1. Каковы морально значимые различия между цифровым 'сознанием', созданным в результате работы компьютера, и сознанием, порожденным биологическими процессами мозга?

2. Как эти различия влияют на наше отношение к нашему цифровому допсельгангеру и, если таковое имеется, на надлежащую моральную заботу о нем?

Как мы утверждали ранее, цифровые допсельгангеры все больше способны имитировать индивидуальные человеческие личности. В последующих разделах мы докажем, что эта способность имеет важные последствия для этики продления жизни или, как мы считаем, истинной цели такого продления, а именно продления существования личности. Это связано с тем, что такие модели смогут частично реализовать некоторые из главных целей, которые, вероятно, и предполагается достичь таким продлением. Во введении мы предположили, что эти цели, вероятно, включают: (1) цели, связанные с опытом, (2) цели, связанные с наследием/влиянием, и (3) цели, связанные с социальными отношениями.

Мы докажем, что цифровые допсельгангеры при современном уровне технологий (с учетом оговорок из предыдущего раздела), способны частично реализовать вторую и третью цели, связанные с влиянием и отношениями. Однако они *вряд ли* будут способствовать в обозримом будущем, если вообще когда-либо, достижению первой цели, связанной с продолжением субъективного опыта.

Это объясняется тем, что большие языковые модели, по своей природе, являются ли они персонализированными или нет, в целом признаются не способными к субъективному сознанию в качественном или феноменальном смысле. Чалмерс [34] приводит несколько технических причин для такого вывода, включая, что кажется наиболее важным, отсутствие у больших языковых моделей элементов, которые, согласно ведущим теориям, считаются необходимыми для сознания, таких как унифицированная агентность [35] или глобальное рабочее пространство [36]. Тем не менее и это более фундаментальный момент, хотя большие языковые модели генерируют текст, который, будь он произведен человеком, несомненно, заставил бы нас заключить, что его автор сознателен [подробнее см.: 37], они делают это посредством чисто статистического моделирования человеческого языка: метода, который не требует и не подразумевает наличия сознания [см. например: 38]. В любом случае, для целей данной статьи мы просто исходим из предположения, что большие языковые модели, по крайней мере современные или ближайшего будущего, не могут помочь нам в достижении цели продления субъективного опыта².

Одним из следствий этого предположения является то, что основанные на больших языковых моделях цифровые допсельгангеры, как мы их понимаем, в лучшем случае являются, и, возможно, только таковыми и останутся, лишь *философскими зомби*: то есть они могут в совершенстве или почти в совершенстве внешне демонстрировать человеческое поведение или признаки личности, но при этом не будет ничего, на что было бы 'похоже' *быть* цифровым допсельгангером 'изнутри' [41]. Чтобы внести ясность, отметим, что вопрос

о том, обладают ли большие языковые модели или, собственно, философские зомби, *другими* типами сознания, такими как так называемое 'сознание доступа' [42, 43], здесь нас не интересует. Для наших целей критически важным является предположение, что современные и будущие в ближайшей перспективе большие языковые модели: (а) будут полностью лишены *субъективного опыта* и при этом (б) смогут, вероятно, пройти Персонализированный тест Тьюринга [8, 9] на основе исключительно лингвистических данных.

Какое отношение все это имеет к тождеству личности? На наш взгляд, открытым остается вопрос о том, является ли субъективный опыт необходимым условием для того, чтобы быть *личностью*. Например, некоторые исследователи готовы применять термин 'личность' для описания определенного вида существ, которые, вероятно, лишены субъективного опыта, вроде человека с диагнозом 'постоянное бессознательное состояние' [44]. Другие, напротив, полагают, что возможность, по крайней мере, субъективного опыта является минимальным условием для того, чтобы быть личностью [45]. Предположим, что мы принимаем последнее, более строгое понятие личности, в рамках которого цифровые допелъгангеры *не* будут личностями в силу своих внутренних свойств (вопрос о том, могут ли они быть личностями в *реляционном* смысле, мы обсудим позже). Тогда возникает вопрос: какой вид моральной заботы, если таковой вообще возможен, мы все еще можем оправданно испытывать по отношению к нашему цифровому допелъгангеру, — то есть к сущности, которая репрезентирует нас и ведет себя в соответствии с нашими конкретными убеждениями, желаниями, воспоминаниями и чертами личности и т. д., несмотря на то, что, *ex hypothesi*, она полностью лишена субъективного сознания и, следовательно, не является *личностью* согласно релевантным теориям³?

Тождество личности и пруденциальная забота

Мы хотим оценить, могут ли цифровые допелъгангеры способствовать обеспечению, хотя бы отчасти, одной из нескольких из тех целей или благ, которые, как мы предположили, лежат в основе стремления продлить свою жизнь, а в конечном счете — собственное 'Я', во времени. Для этого необходимо обратиться к теориям тождества личности. Они представляют собой попытки ответить на вопрос о том, что делает индивида одной и той же личностью с течением времени и, тем самым, проясняют критерии, необходимые для продолжения его существования.

Одна из влиятельных попыток ответить на этот вопрос принадлежит Дереку Парфиту [49]. Отвергая некоторые устоявшиеся представления, Парфит утверждал, что метафизическое *тождество* не является тем, что в конечном счете имеет значение для продолжения существования индивида. Скорее для этого, согласно Парфиту, по-настоящему важно *R-отношение*: набор диахронических связей или паттернов, которые соединяют определенные важные аспекты индивида во Времени 1, его личностных черт в виде воспоминаний, убеждений и желаний, с подобными же аспектами индивида во Времени 2 так, что для индивида во Времени 1 рационально иметь *личную* заинтересованность в том, как сложится судьба индивида во Времени 2. Иными словами, в той степени, в которой R-отношение существует, например, между вашим нынешним 'Я'

и будущим индивидом, у вас есть основание 'отождествлять себя' с этим индивидом: то есть заботиться о его благополучии и считать его продолжением себя в аспектах, важных для вас, *даже если* этот будущий человек не является, строго говоря, метафизически тождественным вам.

Парфит характеризует этот особый тип заботы о будущем (варианте) 'вас' как 'то, что имеет значение' для тождества личности. Наш коллега Джефф Макмахан называет это *эгоистической заботой* [50], и мы также будем использовать эту терминологию. Как утверждает Макмахан: «В принципе, может быть рациональным и уместным эгоистически заботиться о человеке в будущем, даже если этот человек не будет собой» [50, р. 42].

Важно для наших целей то, что Парфит утверждал, что R-отношение может сохраняться, даже если будущий человек возникает в результате причин, отличных от нормального старения или биологического выживания. Например, вслед за Парфитом мы можем представить случай, в котором будущая технология телепортации позволяет человеку переместиться на Марс. Телепорт сканирует его тело, уничтожает его, а затем мгновенно воссоздает из местных материалов на Красной планете.

Последнее воспоминание человека, когда он прибывает на Марс, — это шаг в телепорт. Согласно строго физической трактовке, это воспоминание вызвано телепортационной машиной на Марсе (использующей данные сканирования тела индивида на Земле), которая воссоздала тело и мозг индивида на Марсе таким образом, что это воспоминание теперь присутствует. Это необычная причина сохранения воспоминаний, не говоря уже о личностях, во времени, но для Парфита этого достаточно. Поскольку R-отношение сохраняется, индивид выживает в смысле, который действительно имеет значение. И это отношение, согласно аргументации, может быть тем, что требуется для того, чтобы наша эгоистическая или благоразумная забота о будущем человеке, с которым мы его разделяем, была рациональной.

То, *какие* именно аспекты личности должны быть охвачены R-отношением, чтобы существа, объединенные этим отношением, были оправданны в эгоистической заботе друг о друге⁴, является предметом дискуссии. Например, некоторые могут считать, что связанные R-отношениями существа должны иметь не просто воспоминания, а скорее *сознательно извлекаемые* или *субъективно переживаемые* воспоминания, чтобы для одного из этих существ было бы рациональным испытывать эгоистическую заботу о другом. Если эта точка зрения верна, то, учитывая предполагаемую неспособность цифровых допелъгангеров иметь какой-либо субъективный опыт, индивид не мог бы быть по-настоящему R-связан со своим цифровым допелъгангером. В лучшем случае он мог бы разделять с ним то, что мы здесь будем называть 'D-отношением', — а именно паттерн связей между различными важными аспектами личности (такими как ее воспоминания, личность, убеждения, желания и т. д., как и в случае с R-отношением), но где ни один из этих атрибутов не обладает соответствующим субъективным качеством 'каково это' [52], по крайней мере, этого мы не найдем у одного из связанных D-отношением существ (т. е. у цифрового допелъгангера). Более того, мы можем предположить, что цифровой допелъгангер, в отличие от связанного с ним человека, лишен некоторых других менталь-

ных свойств или способностей, таких как (собственная) агентность, интенциональность и т. д., как уже обсуждалось во введении.

Другими словами, мы предполагаем, что личность могла бы быть связана D-отношением со своим цифровым допелльгангером (что, в *пределе*, было бы схоже с отношениями, которые она имела бы с философской зомби-версией себя), даже если она не могла бы быть связанной R-отношением с ним. И тогда возник бы дальнейший вопрос о том, достаточно ли D-отношения для оправдания эгоистической заботы, несмотря на отсутствие того, что мы, *arguendo*, предполагаем субъективный опыт необходимым компонентом R-отношения.

Разумеется, некоторые могут полагать, что R-отношение на самом деле не требует, чтобы различные связанные аспекты двух (или более) индивидов были сознательно извлекаемы или субъективно переживаемы обоими (или ими всеми) для того, чтобы это отношение имело место. В таком случае D-отношение могло бы попросту сводиться к R-отношению в зависимости от того, какие еще аспекты индивидов, по их мнению, должны присутствовать и быть связанными, чтобы R-отношение возникло. Но, прежде чем мы сможем рассмотреть такую возможность, нам сначала необходимо привлечь во внимание тот факт, что, согласно некоторым критикам или комментаторам Парфита, даже R-отношения недостаточно для обоснования (полноценной) эгоистической заботы между индивидами, существующими в разное время.

Это возвращает нас к Джеффу Макмахану [50], который указывает, что в гипотетических сценариях, включающих в себя полностью человеческие, биологические реплики (с которыми мы определенно разделяли бы R-отношение), интуиции многих людей, по-видимому, расходятся с точкой зрения, согласно которой у нас есть основание 'отождествлять себя' с такими связанными R-отношениями существами даже *при том условии*, что они обладают, среди прочего, психологической непрерывностью психической жизни, включающей субъективный опыт, подобный нашему. Рассмотрим гипотетический сценарий под названием *Самоубийственная Миссия*:

Во время войны человека выбрали для выполнения военной миссии, которая неминуемо приведет к гибели. Хотя использование Репликатора является очень дорогим и поэтому строго нормировано, командование предоставило ему привилегию создать свою реплику перед миссией. Ему также позволят выбрать до процесса репликации, пойдет ли он на миссию сам или же будет отправлена реплика. (Поскольку он дисциплинированный солдат, его реплика также будет дисциплинированной. Он знает, что, если будет отдан приказ, она пойдет на миссию.) [50, р. 57].

Макмахан предполагает, что большинство людей выбрали бы отправить реплику, пожертвовав ею, чтобы спасти себя. Предположим, что это верно, и далее предположим, что этот выбор отражает некое нормативно значимое соображение. Из этого может следовать, что R-отношения на самом деле недостаточно для рационального обоснования эгоистической заботы о собственной реплике, даже полностью биологической, которая разделяет наши самые важные психологические свойства, включая субъективный опыт. По крайней мере, это может

указывать на то, что люди *не уверены*, оправдана ли эгоистическая забота в таких случаях; или, возможно, следует сказать, что оснований для заботы о выживании такой реплики *меньше*, даже если *какие-то* из них и существуют, что мы вскоре обсудим. Но теперь рассмотрим другой гипотетический сценарий — *Ядерную атаку*:

Некто является сотрудником Пентагона, в котором есть Репликатор, способный передать клеточный план его тела в репликационную кабину на Аляске. Он получает подтверждение, что ядерная ракета, нацеленная на Пентагон, прорвала оборону страны и уничтожит весь район в течение минуты. Этого времени как раз достаточно, чтобы отсканировать и передать данные его тела на Аляску [50, р. 58].

«Большинство из нас считает, — пишет Макмахан, — что было бы лучше создать реплику на Аляске, чем быть уничтоженным, не оставив реплики» [50, р. 58]. Снова предположим, что это верно и отражает некое нормативно значимое соображение.

На первый взгляд может показаться, что между сценариями *Самоубийственной Миссии* и *Ядерной атаки* существует некое противоречие. В *Самоубийственной Миссии* у нас, казалось бы, есть причина предпочесть, чтобы вред постиг связанную с нами R-отношениями реплику, а не нас самих, что указывает на недостаточность R-отношения Парфита для обоснования эгоистической заботы, или, по крайней мере, как мы уже намекали, для обоснования *уверенной* или *полноценной* эгоистической заботы, эквивалентной той, которую мы испытываем к нашему будущему, физически и ментально непрерывному 'Я'. Однако в сценарии *Ядерной атаки*, из соображений явной эгоистической заботы мы, тем не менее, выбираем продолжение существования нашего психологического состояния, воплощенного в физической реплике, предпочитая такой исход полному уничтожению.

Для разрешения этого противоречия Макмахан предлагает пересмотреть исходную концепцию Парфита о 'том, что имеет значение' для сохранения индивида во времени [50, р. 48]. В пересмотренной версии Макмахана акцент вновь делается на *физическую* непрерывность биологически воплощенных психических состояний: то есть тех психических состояний, которые кодируют все убеждения, желания и воспоминания (и т. д.) индивида и которые, как утверждается, составляют основу для эгоистической заботы. Назовем это более сильное отношение психофизической непрерывности *В-отношением*, поскольку оно фокусируется на биологической непрерывности как необходимом субстрате соответствующих психологических состояний. Макмахан предполагает, что интуиция, согласно которой человек предпочел бы боль или страдания связанной с ним R-отношением реплики своим собственным, основана на том факте, что реплики, несмотря на разделение с нами различных важных психологических черт, являются (грубо говоря) все же другими телами и другими субъектами опыта⁵.

Однако не следует слишком поспешно отвергать исходную концепцию Парфита, где именно R-отношение, которое, в отличие от В-отношения, *не* требует физической непрерывности мозга для того, чтобы эгоистическая забота была рациональной, фиксирует 'то, что имеет значение' в сохранении индивида. Напомним, что в сценарии *Ядерной атаки*, хотя, возможно, и не кажется рациональным

Таблица 1. Предлагаемый спектр обоснований для эгоистической заботы
 Table 1. Proposed range of justifications for egoistic concern

№ п/п	Спектр обоснований для эгоистической заботы
1.	<i>Связанные В-отношением существа.</i> У нас, пожалуй, есть самые веские основания для эгоистической заботы о будущих сущностях, находящихся с нами в В-отношениях, а именно поддерживающих функциональную непрерывность мозга в производстве сознания. Эта эгоистическая забота может ослабевать, но не может исчезнуть полностью, по мере того как сущность теряет психологическую непрерывность с нами.
2.	<i>Связанные R-отношением существа.</i> У нас могут быть вторые по силе основания для эгоистической заботы о сознательных сущностях, находящихся с нами в R-отношениях, пропорционально силе их психологической непрерывности с нами. Это основание может быть связано, например, с их способностью продвигать наше наследие, которое может иметь некоторые субъективные и объективные условия удовлетворения (см. дальнейшее обсуждение).
3.	<i>Связанные D-отношением существа.</i> У нас могут быть третьи по силе основания для эгоистической заботы о сущностях, связанных с нами D-отношениями, — то есть сущностях, лишенных какого-либо субъективного опыта, но симулирующих R-отношение с нами, — пропорционально силе их преемственности с нами в плане внешнего лингвистического поведения. Поскольку они лишены субъективного сознания, эгоистическая забота, основанная на этом отношении, была бы рациональна только в силу их способности способствовать достижению целей, имеющих объективные условия удовлетворения (см. дальнейшее обсуждение).

Источник / Source: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11804783/>

испытывать *полную* или *полноценную* эгоистическую заботу о связанной с нами R-отношением реплике на Аляске, все же может казаться интуитивно верным испытывать *хотя бы некоторую* эгоистическую заботу о ней: достаточную для того, чтобы мы, *по крайней мере, в какой-то степени* 'отождествляли' себя со связанной с нами R-отношением репликой. Достаточную, скажем, для того, чтобы эгоистически оправданным было просканировать себя и создать реплику до прибытия ядерной ракеты.

Некоторые читатели, возможно, сойдут с дистанции на этом месте, предпочтя более строгое В-отношение Макмахана в качестве единственного рационального основания для эгоистической или благоразумной заботы. Однако для тех, кто все еще с нами, представьте теперь, что Репликатор в сценарии *Ядерной атаки* не работает. К сожалению, создать связанное с вами R-отношениями существо невозможно. Однако в качестве запасного варианта существует другая машина — назовем ее Дубликатор — которая может быстро создать продвинутый цифровой допсельгангер вас самих: 'как' именно это происходит — не столь важно, но давайте просто представим, что она работает путем тонкой настройки специально разработанной высокотехнологичной большой языковой модели на базе всего, что вы когда-либо написали или сказали в текстовых сообщениях, электронных письмах, записанных разговорах, видеороликах на YouTube и т. д., при условии вашего согласия и валидации данных. Предполагая, что такой допсельгангер будет находиться в идеальном D-отношении с вами, будет ли у вас причина его создать?

Прежде чем ответить, позвольте добавить одну деталь. Для целей этого мысленного эксперимента предположите, что Дубликатор технологически и этически достаточно совершенен, чтобы организовывать все эти данные так, чтобы приватные вещи оставались приватными: он не позволит, например, вашему цифровому допсельгангеру делиться интимными текстовыми сообщениями, ранее отправленными вашему супругу(е), с вашими студентами, начальником или детьми [11]. Однако он все же сможет использовать эти сообщения для генерации схожих интимных текстов при соответствующих условиях: скажем, только если сможет верифицировать, что с ним общается именно ваш партнер, помимо прочих этических (и социально-реляционных/контекстуальных) ограничений.

Возвращаясь к вопросу о том, эквивалентно ли выживание с помощью Дубликатора другим обсуждаемым формам выживания, предположим, что в последние секунды перед взрывом ядерной ракеты у вас есть выбор: нажать или не нажать кнопку с датчиком отпечатка пальца, которая автоматически настраивает Дубликатор на создание связанной с вами D-отношением реплики — вашего идеального цифрового допсельгангера. В следующем разделе мы рассмотрим потенциальные *неэгоистические* основания для нажатия этой кнопки, но сейчас мы спрашиваем, может ли у вас быть какое-либо основание желать создания такой реплики, так сказать, ради вас самих, и если да, то могли бы вы рационально *в какой-то степени* 'отождествлять' себя с ней. Другими словами, мы просим вас подумать, было бы для вас осмысленным, *по крайней мере отчасти*, эгоистически заботиться о том, что произойдет с ней или благодаря ей.

Мы предполагаем, что у разных читателей здесь будут разные интуиции и некоторые твердо ответят «Нет». Другие, возможно, сильно хотели бы, чтобы в этом мысленном эксперименте работал Репликатор, но все же были бы благодарны за Дубликатор как запасной вариант. А третьи могут не использовать ни один из типов машин, возможно полагая, что если бы они ими воспользовались, то итоговая версия просто не была бы подходящим объектом для эгоистической заботы, и точка.

У нас нет уверенного ответа на этот вопрос. Но мы считаем, что на данном этапе может быть разумно предложить спектр потенциальных эгоистических интересов, сгруппированных, как показано в табл. 1.

Может возникнуть возражение, что эта концепция неверно упорядочена и что психологическая непрерывность является более сильным критерием для эгоистической заботы, чем функциональная непрерывность мозга. Или же можно возразить, что не существует случаев, когда приходится выбирать между будущим связанным В-отношением индивидом и другим будущим связанным R-отношением индивидом, поскольку реплики любого типа являются вымыслом. Но то, что мы хотим здесь подчеркнуть, так это то, что все три отношения — R, В и D — *допускают степени*. Соответственно, независимо от предпочитаемой теории тождества личности может существовать точка, в которой достаточно ослабленная непрерывность в В-отношениях

(функциональной непрерывности мозга) или в R-отношениях (субъективно переживаемой психологической непрерывности) может быть превзойдена в качестве основания для эгоистической заботы силой связи с будущей связанной D-отношением сущностью, *даже если* эта сущность ни в каком смысле не является личностью.

Рассмотрим, например, нашего писателя из начала статьи, чьи авторские амбиции разрушены диагностированием болезни Альцгеймера. Мог бы он рационально предпочесть потратить свои быстрые месяцы ясного сознания на обучение цифрового допсельгангера взамен того, чтобы в спешке дописывать свой давно отложенный роман? Распределение финансовых ресурсов могло бы быть рассмотрено аналогичным образом, поскольку как паллиативная помощь, так и разработка и поддержание ее допсельгангера, связаны с затратами. Он мог бы посчитать, что его допсельгангер с очень высокой вероятностью завершит его работу приемлемым для него образом (сильное D-отношение), в то время как его ближайшее будущее 'Я' имеет очень низкую вероятность это сделать (слабое R-отношение или B-отношение). В этом случае мы могли бы сказать, что его выбор развивать свой цифровой допсельгангер является эгоистически рациональным.

В качестве альтернативы, скажем, он мог бы оценить завершение своего романа как относительно менее важное дело по сравнению с путешествиями по миру с семьей и по этой причине отказаться от создания допсельгангера. Можно представить и другие компромиссы. Но суть в том, что несколько факторов, включая степень, в которой мы соотносим себя с нашими будущими биологическими или психологически непрерывными 'Я', степень, в которой мы соотносим себя с нашим цифровым допсельгангером, а также относительная важность определенных видов незавершенных планов, будут критически влиять на то, как мы выбираем тратить свое время и ресурсы и приходиться к соответствующим компромиссам. И, по нашему мнению, основываясь на траектории текущих технологических разработок, включая качество и количество данных, которые мы собираем о себе, потенциальный уровень подобия цифрового допсельгангера по отношению к самому человеку, а следовательно, и сила D-отношения между этими двумя сущностями со временем значительно возрастут.

В следующих разделах мы подробнее рассмотрим, как цифровые допсельгангеры могут способствовать завершению наших жизненных планов и сохранению аспектов наших близких отношений такими способами, которые мы готовы ценить, *независимо от того*, готовы ли мы рассматривать наших цифровых допсельгангеров как буквальное продолжение нашего 'Я'.

Частичное достижение благ продления существования личности

Некоторые исследователи утверждают, что интересы индивидов могут нарушаться после их смерти, например, посредством ущерба репутации⁶ [55, 56]. Но, если интересы индивидов могут быть нарушены после смерти, это означает, что они также могут получать пользу от событий, происходящих после их смерти [57]. Согласно объективной, негедонистической концепции благополучия, определенные посмертные события могут, таким образом, вносить позитивный вклад в общую ценность

жизни индивида. Подобно тому, как сохранение и признание художественных или литературных произведений человека может считаться улучшающим его благополучие, обеспечивая долговечность его наследия, мы предполагаем, что существование и влияние цифрового допсельгангера также может продолжать воздействовать на мир способами, считающимися позитивными после смерти человека. (Разумеется, он также может влиять на мир способами, считающимися весьма негативными [6].)

В дальнейшем мы покажем, каким образом цифровые допсельгангеры могут, в некотором релевантном смысле, продолжать наследие индивида, удовлетворяя соответствующие желания или реализуя соответствующие планы, которые он имел при жизни. Однако, как мы также укажем, некоторые потенциальные блага продления существования личности могут иметь не строго индивидуалистическую, а *реляционную* природу: то есть их ценность проистекает из признания, оценки или текущего опыта других лиц, например, от опыта, который могут иметь или осуществлять значимые для них близкие.

Наследие — это зонтичное понятие, охватывающее как материальные, так и нематериальные блага, вроде значимых связей, созданные в течение жизни, более широкого влияния жизненных проектов и обязательств человека, а также сети ценностей, которые он отстаивал, и тому подобного. Наследие выполняет функцию сохранения интересов: оно является основным средством, с помощью которого индивиды продолжают оставлять след в мире после своей биологической смерти. Желание оставить после себя наследие давно исследуется в терминах «надежды на то, что мы обладаем потенциалом бессмертия» [58, р. 154], будь то чисто символического [59] или более содержательного, наподобие того, что мы рассмотрим далее.

Наследие/Влияние

Обсуждая гипотетические случаи разделения личности, вроде сценария, в котором каждое полушарие мозга человека пересаживается в новое тело, Парфит [49] вводит идею существования интересов, которые мы назовем *не основанными на опыте* (non-experiential). Он предполагает, что каждый из отпочковавшихся индивидов, назовем их B и C, от исходного индивида A мог бы затем продолжать различные важные жизненные проекты, которые были у A. Например, B мог бы наконец закончить лоскутное одеяло, которое A всегда собирался завершить, в то время как C мог бы посвятить себя научной карьере, об отсутствии которой A всегда сожалел. Другими примерами могут служить постройка спроектированного человеком собора, принятие законодательного акта, благополучная жизнь ребенка или даже победа в войне. Мы можем говорить, например, о том, что определенные цели *стоят того, чтобы умереть за них*, где вопрос о том, 'стоило ли это на самом деле того', можно рассматривать как зависящий от фактов, которые возникают лишь *после* смерти человека. Подобные желания (например, желание, чтобы в войне победила сторона, которую человек поддерживает) не обязательно должны быть пережиты непосредственно в собственном опыте, чтобы быть удовлетворенными (следовательно, они 'не основаны на опыте'), и поэтому можно сказать, что для их удовлетворения существуют *объективные* условия.

Они противопоставляются желаниям, имеющим субъективные условия удовлетворения, которые должны быть пережиты на собственном опыте, чтобы быть удовлетворенными. Например, если человек ставит цель дожить до преклонного возраста и увидеть комету Галлея, мы можем почувствовать утрату чего-то, если он будет находиться в коме во время ее возвращения. Удовлетворение желаний, имеющих субъективные условия удовлетворения, представляется сомнительным, если рассматриваемое существо полностью лишено субъективного опыта; в отличие от желаний, имеющих объективные условия удовлетворения.

Развивая эту мысль, мы предполагаем, что для многих людей наличие цифрового допелъгангера, продолжающего их общественное влияние или их уникальные проекты и обязательства, может обладать подлинной эгоистической ценностью, даже если эта реплика не является субъективно сознательной и не станет вместилищем для продолжения их собственной феноменологии. У людей есть пруденциальные основания заботиться о том, что произойдет после их смерти. Имеет смысл предпочесть, чтобы усилия людей и их жизненные планы, или формирование их репутации в свете осуществления данных планов, не были полностью нивелированы после их смерти.

В конце концов конечные результаты жизненных проектов человека после его смерти могут бросать ретроспективную тень на его прижизненные достижения. Например, если индивид посвящает многие годы жизненному проекту, который в конечном итоге терпит посмертный провал, его усилия могут быть оценены как ошибочные или бессмысленные, независимо от их первоначального потенциала. Посмертные события, таким образом, могут влиять на то, как мы ретроспективно оцениваем чью-то жизнь: был ли человек счастлив, успешен, реализован или нет [60]. Соответственно, может быть рациональным заботиться о том, что произойдет после нашей смерти, а также желать продолжать оказывать позитивное влияние через свое наследие [61].

Однако наш потенциал для оказания продолжающегося влияния не ограничивается жизненными проектами. Определенные жизненные переживания, которые являются определяющими для того, как индивиды воспринимают себя, также могут иметь большое значение. Когда людей спрашивают о том, что они в конечном счете хотели бы передать после смерти, они часто упоминают наследие в виде ценностей, тесно связанное с аспектами их ощущения собственной идентичности, или с ролями и чертами, которые они воспринимают как конституирующие их сущность [62, 63]. Наследие в виде ценностей часто особенно важно для людей, переживших трудности или маргинализацию: наследие передается как набор руководящих принципов, призванных помочь другим справиться с аналогичными проблемами. Например, женщины, инфицированные ВИЧ, создавали видеонаследие для своих семей, давая советы и руководство, чтобы помочь своим близким вести успешную жизнь, преодолевая возможные трудности [64]. Это видеонаследие было важно для самих создававших их женщин, поскольку воспринимались как способ контролировать повествование, которое они оставляют своим потомкам посмертно. Это показывает, что человек может желать, чтобы его жизненный опыт, рассматриваемый как уникальное проявление его лично-

сти, продолжал формировать и вдохновлять других. Такое наследие может иметь как моральные, так и рациональные измерения.

Многие люди предпочитают оставить прочный вклад, добиться, чтобы ценность их жизни и ее влияние продолжали признаваться другими после их ухода, а также участвовали в формировании будущего. Такие цели не обязательно предполагают продолжение субъективного опыта создателя, хотя они могут включать в себя предшествующие субъективные предпочтения или чувства, связанные с посмертным влиянием. Следовательно, цифровой допелъгангер мог бы, в некотором смысле, продолжить и продлить историю жизни исходного человека: достигая или реализуя его цели и желания, или, по крайней мере, те цели и желания, для удовлетворения которых не требуется субъективный опыт. Эти не основанные на опыте интересы в посмертном влиянии и продолжающемся социальном воздействии могут давать нам некоторые эгоистические основания заботиться о том, что произойдет с нашим цифровым допелъгангером после нашей смерти.

Реляционное тождество

Возможное возражение против всего, что мы предложили до сих пор, заключается в том, что это все кажется довольно *эгоистичным*. Тем не менее, анализируя этику исследований в области продления жизни или, собственно, продления существования личности, как мы предлагаем здесь, необходимо учитывать права и интересы, не сводящиеся исключительно к эгоистическим. Например, могут возникнуть опасения на общественном уровне по поводу продолжающегося, хотя бы частично, временного существования репрезентаций индивидов, потенциально безграничных. Одно дело, когда общество хранит память о своих (биологически) умерших близких или об оставшихся артефактах, таких как фотографии или видеозаписи; другое дело, когда 'умершие' *продолжают* — потенциально *массово* — порождать новые идеи, вести новые беседы и так далее. Возможно, в конечном итоге для живых благом является научиться скорбеть и отпускать.

Не претендуя на всесторонний анализ этических последствий цифровых допелъгангеров для общества в целом (в этом отношении мы рекомендуем цитированные ранее недавние эссе Джона Данахера и Свена Нихолма [6, 7]), мы завершим данный раздел несколькими соображениями об *отношениях*.

Вспомните истории Джеймса Влахоса, Евгении Куйды или Эндрю Каплана, которые мы приводили ранее в статье. В этих историях главное значение имела не продолжающаяся субъективная сознательность цифрового симулякра, а определенный тип *отношений* между этими людьми и их близкими, которые воспринимались как значимые. Как мы можем начать осмысливать такие свидетельства? Придерживаются ли эти люди правдоподобного или разумного отношения, или же они просто находят в отрицании, неспособные принять внезапное прекращение существования человека, который был им так дорог? Могут ли продолжающиеся 'отношения' между живым человеком и цифровым допелъгангером недавно умершего близкого иметь какую-либо подлинную ценность сами по себе?

Мы коснемся этого вопроса далее в этом разделе. Однако сначала мы хотели бы высказать бо-

лее общий тезис о концепции *реляционного тождества*. По существу этот подход постулирует, что наша личность не может быть изолирована от внешнего мира, сообществ, частью которых мы являемся, норм, направляющих наше поведение, и, что важнее, от отношений, которые мы выстраиваем с другими на протяжении жизни [65]. Порой эти значимые отношения становятся центральными чертами нашей так называемой социальной или практической идентичности настолько, что определенные роли, такие как мать, отец, учитель, врач, художник и т. д., частично конституируют то, кем мы являемся. Таким образом, мы склонны говорить: «Я — мать, отец, учитель, врач или художник», а не просто «Я воспитываю детей, учу, занимаюсь медициной или создаю искусство».

Случай пациентов в устойчивом вегетативном состоянии, которые могут рассматриваться ни как живые, ни как мертвые, ни как субъекты, ни как объекты, демонстрирует потенциал социальных отношений формировать наш статус (то есть нашу способность быть осмысленно признаваемыми) как личностей [44]. Эти пациенты бросают вызов традиционному пониманию личности как «индивида, переживающего свое тело-в-мире, самосознающего индивида, конституирующего агентность» [44, р. 55]. Посредством повседневного взаимодействия с пациентами в устойчивом вегетативном состоянии ухаживающие за ними люди поддерживают либо порождают «различные смыслы их личности» [44, р. 62]. Реляционная перспектива личности проливает свет на то, как даже индивиды, которые больше не имеют и, возможно, никогда не будут иметь субъективного опыта, все же могут считаться личностями в некотором смысле, а именно в силу своих отношений с другими людьми.

Разумеется, может быть выдвинуто возражение, что это односторонние, невзаимные отношения, или, как некоторые могли бы считать, это вовсе не *отношения*. Подобно 'отношениям' человека, влюбленного в знаменитость, с которой он никогда не встречался и которая не знает о его существовании, они являются односторонними и иногда могут быть даже нездоровыми или бредовыми. Однако есть условия, при которых они могут быть здоровыми и значимыми. Тот, кто читает книгу своему родственнику, находящемуся без сознания, возможно, участвует в отношениях, которые 'поддерживают человека в живых', связывая прошлые воспоминания с текущими, что потенциально может способствовать их собственному развитию или переживанию горя. То есть взаимность может и не быть необходимым компонентом всех значимых отношений. Наше взаимодействие с другими постоянно формирует нас и становится частью того, кем мы являемся (или кем мы себя считаем); точно так же, как и сами мы оказываем влияние на тех индивидов, для которых мы имели значение⁷ [69–72].

Мы предлагаем расширить наш предыдущий набор оснований для заботы о собственном будущем 'Я', включив в него реляционные интересы, согласно которым продолжение определенной динамики отношений может быть рационально оценено, даже если один из участников диады этих отношений лишен субъективного опыта. Разумеется, как мы уже упоминали, некоторые могут полагать, что для того, чтобы отношения были ценными или вообще считались отношениями, необходимо, чтобы обе стороны субъективно переживали взаимодействие между

ними [73]. Если это так, то тогда те, кто заявляет о ценности своих продолжающихся отношений, например, с цифровым допсельгангером умершего близкого человека, просто ошибаются, трагически полагая, что вовлечены в ценные отношения, в то время как в реальности ничего подобного не переживают.

В качестве альтернативы, если мы примем градуированный или иерархический взгляд на статус или ценность отношений, можно признать, что *наиболее* ценные или настоящие отношения — это отношения между двумя личностями, обе из которых способны субъективно переживать мир и друг друга. Но если мы представим супруга, который продолжает любить и поддерживать связь со своим партнером, находящимся в живых, но в коме, возможно, без какого-либо субъективного опыта, то отрицание какой-либо ценности этих отношений или настаивание на том, что живой супруг ошибается, переживая их как таковые, может показаться не просто жестоким или бесчувственным, но и ошибочным. Подобный анализ может быть применим и к отношениям между человеком и цифровым допсельгангером кого-то, кого он любит. Цифровые допсельгангеры, несмотря на нынешнее отсутствие у них субъективного опыта, могут быть как минимум средством воспоминания, но также и потенциальным сохранением ценных аспектов наших отношений с другими [10, 11, 51].

Что важно в продлении существования личности

Если приведенные выше аргументы убедительны или, по крайней мере, в целом правдоподобны, какие последствия могут из этого вытекать для исследований в области долголетия и продления жизни человека? Как отмечалось в начале статьи, существуют прагматические основания для того, чтобы прояснить, что именно является самым важным в продлении жизни в свете ограниченности ресурсов. Мы аргументировали, что правдоподобно предположить, что нас заботит продление не биологической 'жизни' как таковой, а, в некотором смысле, нашего собственного 'Я'. На вопрос 'почему мы заботимся о продлении нашего 'Я'?' мы предположили, что для этого есть различные основания, связанные с разными типами целей — опытом, наследием и отношениями, — некоторые из которых могут быть частично реализованы посредством создания и использования цифрового допсельгангера. Следовательно, при взвешивании 'за' и 'против' в вопросах расходования ресурсов на исследование долголетия может быть разумным рассмотреть более широкий спектр проектов, которые можно было бы отнести к данной категории.

Размышляя о том, что важно в продлении человеческой жизни, мы также предлагали задуматься о разных типах значимых вещей как с эгоистической, так и с неэгоистической точек зрения. Как мы аргументировали, существуют некоторые *объективные* события (например, завершение романа после нашей смерти), которые человек может рационально желать, чтобы они произошли, и которые разумно считать вкладом в его общие интересы или благополучие. Мы предположили, что некоторые из этих событий могут быть осуществлены цифровым допсельгангером. Другими словами, требующие *субъективного* опыта цели, отнюдь не исчерпывают круг того, о чем мы глубоко заботимся; скорее, они представляют лишь один аспект того, что имеет зна-

чение. Другие виды ценностей и достижений могут продолжать существовать после смерти.

Наш анализ также установил, что у нас есть неэгоистические причины желать продлить свое 'Я', связанные с отношениями с близкими: цифровые допсельгангеры, разумно используемые, потенциально могут удовлетворять некоторые из таких причин, хотя бы частично. В более общем смысле, если мы последуем за Парфитом в рассмотрении того, что важно в сохранении индивида, как вопроса степени — а не вопроса 'все или ничего', — то даже *частичное* продолжение себя может иметь ценность. Если и не в терминах В-отношения Макмахана, то, возможно, в терминах R-отношения-R Парфита. А если все же и не в них, то тогда в терминах нашего D-отношения. Наконец, даже если кто-то отвергает точку зрения, что последнее из этих отношений, могло бы когда-либо рассматриваться как достаточное условие для продления существования личности — то есть в принципе в любом смысле или в какой-либо степени, — мы предположили, что связанное с нами D-отношениями существо все-таки способно правдоподобным образом помочь нам достичь некоторых целей или осуществить некоторые блага, связанные с таким продлением жизни личности.

Выводы

В данной статье мы сосредоточились на вопросе о том, каким образом цифровой допсельгангер может способствовать реализации некоторых интересов, которые человек может стремиться продвигать посредством традиционных форм биологического продления своей жизни не только с точки зрения эгоистической заботы, но и с точки зрения отношений.

Это предполагает, что продвижение этих интересов было бы в целом *благом*, примерно так, как утверждают сторонники продления жизни. Тем не менее необходимо признать, что некоторые люди весьма критически относятся к продлению жизни [см. например: 74]: они считают, что мы должны быть удовлетворены здоровой и конечной жизнью, ненамного превышающей нынешнюю среднюю продолжительность жизни.

В статье мы не занимаем позиции по таким обсуждениям; наше внимание было сознательно узконаправленным. Следовательно, нас следует понимать как выдвигающих ряд условных утверждений: а именно, что *в той мере*, в какой на индивидуальном уровне разумно заботиться о продлении своей биологической жизни — это может иногда иметь больше общего с желанием продлить свое собственное 'Я' во времени, а не свою 'жизнь' как таковую. И если *это* желательная цель, пусть и такая, которую нельзя легко достичь с помощью существующих технологий или которая может не осуществиться у большинства людей в обозримом будущем, все же возможно, что цифровой допсельгангер мог бы послужить 'второстепенным' вариантом в том смысле, что он мог бы удовлетворить *некоторые* из оснований, по которым нас в принципе может заботить продление собственной жизни.

Однако независимо от того, согласны ли вы с нашим конкретным предложением относительно цифровых допсельгангеров или нет, наша мысль о необходимости выявления глубинных причин стремления продлить свою жизнь и/или свое собственное 'Я' важна для научных исследований и политических действий. Например, те, кто рас-

сматривает плюсы и минусы инвестирования ресурсов в продление человеческой жизни, включая любые альтернативные решения, должны сначала прояснить, что именно люди в основе своей ценят в более долгой жизни. Мы предложили здесь некоторые возможные ответы; могут быть и другие. В любом случае, размышление над подобными основаниями хотя бы отчасти должно открыть пространство для более творческого осмысления различных других способов потенциального удовлетворения запросов людей.

В заключение мы хотим подчеркнуть, что мы оставили в стороне многие важные аспекты этой темы. В частности, есть многочисленные этические (а также юридические и политические) проблемы, которые можно и нужно поднимать в отношении цифровых допсельгангеров, часть из которых мы надеемся рассмотреть в будущих работах и которые уже начали исследовать другие авторы [см. например: 6; 7].

Также можно было бы сказать гораздо больше о том, как на практике правительства или другие финансовые организации, в настоящее время инвестирующие в исследования долголетия, должны распределять ресурсы с учетом аргументов и утверждений, представленных в этой статье: к примеру, утверждений об относительной легкости создания цифровых допсельгангеров по сравнению с очевидными трудностями продления биологической жизни.

Но, разумеется, прежде всего нам необходимо продолжать оттачивать наши аргументы и утверждения в ответ на критику и диалог с коллегами. И мы надеемся, что данная статья послужит первым шагом к продолжающемуся плодотворному обсуждению, и с нетерпением ждем последующей критики и комментариев.

Благодарности от переводчика / Acknowledgements of the translator

Выражаю свою благодарность Софье Лаврентьевой и Анастасии Турковой за содержательную критику перевода и обсуждение сложных мест.

I would like to express my gratitude to Sofya Lavrentyeva and Anastasia Turkova for their insightful criticism of the translation and discussion of difficult passages.

Примечания

¹ Другие исследователи считают, что галлюцинации в больших языковых моделях можно понимать как форму 'креативности' и, следовательно, они могут быть полезны в некоторых случаях [33].

² Безусловно, этот вопрос не является окончательно решенным, и некоторые могут быть более открыты к возможности сознания у больших языковых моделей в обозримом будущем [39]. Тем не менее, чтобы ограничить рамки нашего теоретизирования и ради ясности в обсуждении тождества личности в последующих разделах, мы просто будем держать рамки биологического натурализма, предполагающего, что феноменальное сознание и субъективный опыт представляют собой эмерджентные свойства определенных физических процессов в мозге [40]. Согласно этой точке зрения, компьютерная программа, сколь бы точно она ни моделировала человеческое поведение, сама по себе не порождает подлинного субъективного опыта, подобно тому, как компьютерная симуляция биолуминесценции не производит реального видимого света.

³ Данный вопрос согласуется с недавними аргументами Адженифер Блюменталь-Барби [46], которая утверждает, что нет необходимости определять, является ли сущность личностью, чтобы понять, как к ней рационально или морально следует относиться. Иногда мы можем продвинуться в решении этих вопросов, апеллируя напрямую к определенным свойствам сущности, независимо от того, достаточны ли они для признания личности [см. также: 47; альтернативный взгляд см.: 48].

⁴ Или, по крайней мере, эгоистическая забота индивида 'Времени 1' об индивиде 'Времени 2'. Вопрос в том, можем ли мы рационально испытывать эгоистическую заботу о наших прошлых 'Я' — или о прошлых индивидах, с которыми мы находимся в R-отношении, — является интересным, но мы не будем в него углубляться. Вместо этого мы сосредоточены исключительно на проспективном случае: на том, как нам, ныне живущим, рационально относиться к цифровому допеллгангеру, который сохраняется после нашей биологической смерти. О возможности и этике создания множественных цифровых допеллгангеров (т. е. самого себя) при жизни см. [7, 51].

⁵ Если Макмахан прав, нам следует сузить набор критериев, составляющих основу эгоистической заботы до таких, которые более тесно соотносятся с тождеством личности в ее традиционном понимании (т. е. в рамках теорий биологической непрерывности). Это называется *Концепцией воплощенного сознания*, согласно которой «критерием тождества личности является физическая и минимальная функциональная непрерывность мозга» и «физическая (но не организационная) непрерывность мозга необходима для того, чтобы эгоистическая забота была рациональной» [50, р. 69]. Однако мы не столь уверены, что Макмахан прав, как мы поясняем далее в тексте.

⁶ Хотя эта интуиция является распространенной, она остается крайне спорной, поскольку ее нормативный статус неясен [53, 54].

⁷ Эти соображения не являются чисто теоретическими; они глубоко укоренены в том, как мы взаимодействуем с другими людьми. Страх смерти, представляющий собой многомерный комплекс эмоций, часто включает в себя страх того, как наша кончина повлияет на тех, кого мы любим [66, 67]. В одном исследовании причин танатофобии было установлено, что «разлука с другими посредством смерти (собственной смерти или чужой смерти) является одним из самых пугающих аспектов смерти, страхом, который разделяет большинство респондентов» [68]. Это может быть связано с тем, что мы знаем, что наш уход может лишить тех, о ком мы заботимся и кто заботится о нас, значимых отношений, важных для жизни. Следовательно, в той мере, в которой взаимодействие наших близких с нашими цифровыми допеллгангерами после нашей смерти может быть здоровой частью процесса горевания, существуют причины, помимо эгоистических, желать их создания.

Список источников / References

1. Schloendorn J. Making the Case for Human Life Extension: Personal Arguments. *Bioethics*. 2006;20(4):191–202. <https://doi.org/10.1111/j.1467-8519.2006.00494.x>.
2. Davis J. New Methuselahs: The Ethics of Life Extension. Cambridge, MA: The MIT Press, 2018. 368 p.
3. Porter A. Arguments Over Life Extension in Contemporary Bioethics. *Handbook of Bioethical Decisions*. Vol. I / Eds. E. Valdivia, J. A. Lecaros. Cham: Springer, 2023. P. 247–276.
4. Lucke J. C., Hall W. Strong and Weak Lifespan Extension: What Is Most Feasible and Likely? *Australasian Journal on Ageing*. 2006;25(2):58–62. <https://doi.org/10.1111/j.1741-6612.2006.00150.x>.
5. Porsdam Mann S., Earp B. D., Möller N. [et al.]. AUTOGEN: A Personalized Large Language Model for Academic

Enhancement — Ethics and Proof of Principle. *The American Journal of Bioethics*. 2023;23(10):28–41. <https://doi.org/10.1080/15265161.2023.2233356>.

6. Danaher J., Nyholm S. The Ethics of Personalised Digital Duplicates: A Minimally Viable Permissibility Principle. *AI and Ethics*. 2025;5(2):1703–1718. <https://doi.org/10.1007/s43681-024-00513-7>.

7. Danaher J., Nyholm S. Digital Duplicates and the Scarcity Problem: Might AI Make Us Less Scarce and Therefore Less Valuable? *Philosophy & Technology*. 2024;37:106. <https://doi.org/10.1007/s13347-024-00795-z>.

8. Steinhart E. Survival as a Digital Ghost. *Minds and Machines*. 2007;17:261–271. <https://doi.org/10.1007/s11023-007-9068-0>.

9. Caon M. Designing Systems for the Digital Immortality of Intangible Cultural Heritage. Human Interaction and Emerging Technologies: Proceedings of the 1st International Conference on *Human Interaction and Emerging Technologies (IHET 2019)*. Nice: Springer International Publishing, 2020. P. 783–789.

10. Voinea C. On Grief and Griefbots. *Think*. 2024;23(67):47–51. <https://doi.org/10.1017/S1477175623000490>.

11. Voinea C., Porsdam Mann S., Savulescu J., Earp B. D. Digital Doppelgängers, Human Relationships, and Practical Identity. *Bioethics*. 2026;40(5):435–444. <https://doi.org/10.1111/bioe.70026>.

12. Hicks M. T., Humphries J., Slater J. ChatGPT Is Bullshit. *Ethics and Information Technology*. 2024;26:38. <https://doi.org/10.1007/s10676-024-09775-5>.

13. Porsdam Mann S., Earp B. D., Müller N. [et al.]. AUTOGEN and the Ethics of Co-Creation with Personalized LLMs — Reply to the Commentaries. *The American Journal of Bioethics*. 2024;24(3):W6–W14. <https://doi.org/10.1080/15265161.2024.2308175>.

14. Earp B. D., Porsdam Mann S., Liu P. [et al.]. Credit and Blame for AI-Generated Content: Effects of Personalization in Four Countries. *Annals of the New York Academy of Sciences*. 2024;1542(1):51–57. <https://doi.org/10.1111/nyas.15258>.

15. Masood M., Nawaz M., Malik K. M. [et al.]. Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward. *Applied Intelligence*. 2023;53:3974–4026. <https://doi.org/10.1007/s10489-022-03766-z>.

16. Sterri A., Earp B. D. The Ethics of Sex Robots. *Oxford Handbook of Digital Ethics* / Ed. C. Véliz. Oxford: Oxford University Press, 2023. P. 241–257.

17. Morris M. R., Brubaker J. R. Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives. *arXiv preprint*. 2024. arXiv:2402.01662. <https://doi.org/10.48550/arXiv.2402.01662>. (accessed: 04.09.2025).

18. Replika. 2024. URL: <https://replika.com/> (accessed: 04.09.2025).

19. YOV Home | You, Only Virtual. 2024. URL: <https://www.myyov.com/> (accessed: 04.09.2025).

20. Holley P. Hey, Google! Let Me Talk to My Departed Father. *Washington Post*. 2019. August 29. URL: <https://www.washingtonpost.com/technology/2019/08/29/hey-google-let-me-talk-my-departed-father/> (accessed: 04.09.2025).

21. Chesler C. AI's New Frontier: Connecting Grieving Loved Ones with the Deceased. *Washington Post*. 2022. November 12. URL: <https://www.washingtonpost.com/health/2022/11/12/artificial-intelligence-grief/> (accessed: 04.09.2025).

22. Bubeck S., Chandrasekaran V., Eldan R. [et al.]. Sparks of Artificial General Intelligence: Early Experiments with GPT-4. *arXiv preprint*. 2023. arXiv: 2303.12712. URL: <https://arxiv.org/abs/2303.12712>. (accessed: 04.09.2025).

23. Bucher M. J. J., Martini M. Fine-Tuned 'Small' LLMs (Still) Significantly Outperform Zero-Shot Generative AI Models in Text Classification. *arXiv preprint*. 2024. arXiv: 2406.08660. URL: <https://arxiv.org/abs/2406.08660> (accessed: 04.09.2025).

24. Vaid A., Landi I., Nadkarni G., Nabeel I. Using Fine-Tuned Large Language Models to Parse Clinical Notes in Musculoskeletal Pain Disorders. *The Lancet Digital Health*. 2023;5(12):e855 – e858. [https://doi.org/10.1016/S2589-7500\(23\)00202-9](https://doi.org/10.1016/S2589-7500(23)00202-9).
25. Lewis P., Perez E., Piktus A. [et al.]. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020* / Eds. H. Larochelle, M. Ranzato, R. Hadsell, M.-F. Balcan, H.-T. Lin. P. 9459 – 9474. <https://doi.org/10.5555/3495724.3496517>.
26. Zohny H., Porsdam Mann S., Earp B. D., McMillan J. Generative AI and Medical Ethics: The State of Play. *Journal of Medical Ethics*. 2024;50:75 – 76. <https://doi.org/10.1136/jme-2023-109834>.
27. Schwitzgebel E., Schwitzgebel D., Strasser A. Creating a Large Language Model of a Philosopher. *Mind & Language*. 2024;39(2):237 – 259. <https://doi.org/10.1111/mila.12466>.
28. Earp B. D., Porsdam Mann S., Allen J. [et al.]. A Personalized Patient Preference Predictor for Substituted Judgments in Healthcare: Technically Feasible and Ethically Desirable. *The American Journal of Bioethics*. 2024;24(7):13 – 26. <https://doi.org/10.1080/15265161.2023.2296402>.
29. Armstrong K. ChatGPT: US Lawyer Admits Using AI for Case Research. *BBC News*. 2023. URL: <https://www.bbc.com/news/world-us-canada-65735769> (accessed: 04.09.2024).
30. Eaton K. How Many R's in 'Strawberry'? This AI Doesn't Know. *Inc.* 2024. URL: <https://www.inc.com/kit-eaton/how-many-rs-in-strawberry-this-ai-cant-tell-you.html> (accessed: 04.09.2025).
31. Chelli M., Descamps J., Lavoué V. [et al.]. Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *Journal of Medical Internet Research*. 2024;26:e53164. <https://doi.org/10.2196/53164>.
32. Sui P., Duede E., Wu S., So R. Confabulation: The Surprising Value of Large Language Model Hallucinations. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Vol. 1: Long Papers / Eds. L. Ku, A. Martins, V. Srikumar. Bangkok: Association for Computational Linguistics, 2024. P. 14274 – 14284.
33. Jiang X., Tian Y., Hua Y. F. [et al.]. A Survey on Large Language Model Hallucination via a Creativity Perspective. *arXiv preprint*. 2024. arXiv:2402.06647. URL: <https://arxiv.org/abs/2402.06647> (accessed: 04.09.2025).
34. Chalmers D. J. Could a Large Language Model Be Conscious? *arXiv preprint*. 2023. arXiv:2303.07103. URL: <https://arxiv.org/abs/2303.07103> (accessed: 04.09.2025).
35. Schechter E. Intentions and Unified Agency: Insights from the Split-Brain Phenomenon. *Mind & Language*. 2012;27(5):570 – 594. <https://doi.org/10.1111/mila.12003>.
36. Baars B. J. Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience. *The Boundaries of Consciousness: Neurobiology and Neuropathology* / Ed. S. Laureys. Amsterdam: Elsevier, 2005. P. 45 – 53. [https://doi.org/10.1016/S0079-6123\(05\)50004-9](https://doi.org/10.1016/S0079-6123(05)50004-9).
37. Dodig-Crnkovic G. How GPT Realizes Leibniz's Dream and Passes the Turing Test without Being Conscious. *Computer Sciences & Mathematics Forum*. 2023;8(1):66. <https://doi.org/10.3390/cmsf2023008066>.
38. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, 2021. P. 610 – 623. <https://doi.org/10.1145/3442188.3445922>.
39. Long R., Sebo J., Butlin P. [et al.]. Taking AI Welfare Seriously. *arXiv preprint*. 2024. arXiv:2411.00986. URL: <https://arxiv.org/abs/2411.00986> (accessed: 04.09.2025).
40. Searle J. Biological Naturalism. *The Blackwell Companion to Consciousness* / Eds. M. Velmans, S. Schneider. Oxford: Blackwell Publishing, 2007. P. 325 – 334.
41. Lloyd D. What Is It Like to Be a Bot? The World According to GPT-4. *Frontiers in Psychology*. 2024;15:1292675. <https://doi.org/10.3389/fpsyg.2024.1292675>.
42. Block N. Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience. *The Behavioral and Brain Sciences*. 2007;30(5-6):481 – 499. <https://doi.org/10.1017/S0140525X07002786>.
43. Naccache L. Why and How Access Consciousness Can Account for Phenomenal Consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018;373(1755):20170357. <https://doi.org/10.1098/rstb.2017.0357>.
44. Bird-David N., Israeli T. A Moment Dead, a Moment Alive: How a Situational Personhood Emerges in the Vegetative State in an Israeli Hospital Unit. *American Anthropologist*. 2010;112:54 – 65. <https://doi.org/10.1111/j.1548-1433.2009.01196.x>.
45. McInerney P. K. Person-Stages and Unity of Consciousness. *American Philosophical Quarterly*. 1985; 22(3):197 – 209.
46. Blumenthal-Barby J. The End of Personhood. *The American Journal of Bioethics*. 2024;24(1):3 – 12. <https://doi.org/10.1080/15265161.2022.2160515>.
47. Dranseika V., Neiders I., Earp B. D. Time for Bioethics to End Talk of Personhood (But Only in the Philosophers' Sense). *The American Journal of Bioethics*. 2024;24(1):32 – 35. <https://doi.org/10.1080/15265161.2023.2278567>.
48. Koplin J., Holzman T., Bhatia N., Savulescu J. Bioethics Should Not Be Constrained by Linguistic Oddness or Social Offense. *The American Journal of Bioethics*. 2024;24(1):15 – 18. <https://doi.org/10.1080/15265161.2024.2279435>.
49. Parfit D. *Reasons and Persons*. Oxford: Oxford University Press, 1984. 560 p.
50. McMahan J. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press, 2002. 560 p.
51. Voinea C., Porsdam Mann S., Register C. [et al.]. Digital Duplicates, Scarcity, and Value: Commentary on Danaher and Nyholm (2024). *Philosophy & Technology*. 2024;37:121. <https://doi.org/10.1007/s13347-024-00813-0>.
52. Nagel T. What Is It Like to Be a Bat? *The Philosophical Review*. 1974;83(4):435 – 450. <https://doi.org/10.2307/2183914>.
53. Partridge E. Posthumous Interests and Posthumous Respect. *Ethics*. 1981;91(2):243 – 264. <https://doi.org/10.1086/292226>.
54. Harris J. Doing Posthumous Harm. *The Metaphysics and Ethics of Death: New Essays* / Ed. J. S. Taylor. Oxford: Oxford University Press, 2013. P. 213 – 221.
55. Pitcher G. The Misfortunes of the Dead. *American Philosophical Quarterly*. 1984;21(2):183 – 188.
56. Feinberg J. The Rights of Animals and Unborn Generations. *Environmental Rights* / Ed. S. Vanderheiden. New York: Routledge, 2017. P. 241 – 265.
57. Boonin D. *Dead Wrong: The Ethics of Posthumous Harm*. Oxford: Oxford University Press, 2019. 222 p.
58. Birren J. E. A Contribution to the Theory of the Psychology of Aging: As a Counterpart of Development. *Emergent Theories of Aging* / Eds. J. E. Birren, V. L. Bengtson. New York: Springer, 1988. P. 153 – 176.
59. Kotre J. N. *Outliving the Self: How We Live on in Future Generations*. New York: W W. Norton & Co., 1984. 282 p.
60. Rozin P., Stellar J. Posthumous Events Affect Rated Quality and Happiness of Lives. *Judgment and Decision Making*. 2009;4(4):273 – 279. <https://doi.org/10.1017/S1930297500003843>.
61. Scarre G. On Caring about One's Posthumous Reputation. *American Philosophical Quarterly*. 2001;38(2):209 – 219.
62. Hunter E. G. Beyond Death: Inheriting the Past and Giving to the Future, Transmitting the Legacy of One's Self. *OMEGA — Journal of Death and Dying*. 2008;56(4):313 – 329. <https://doi.org/10.2190/OM.56.4.a>.
63. Zacher H., Rosing K., Frese M. Age and Leadership: The Moderating Role of Legacy Beliefs. *The Leadership Quarterly*. 2011;22(1):43 – 50. <https://doi.org/10.1016/j.leaqua.2010.12.006>.
64. Barnes D. B., Taylor-Brown S., Wiener L. "I Didn't Leave Y'all on Purpose": HIV-Infected Mothers' Videotaped Legacies for

Their Children. *Qualitative Sociology*. 1997;20:7–32. <https://doi.org/10.1023/A:1024760113945>.

65. Andersen S. M., Chen S. The Relational Self: An Interpersonal Social-Cognitive Theory. *Psychological Review*. 2002;109(4):619–645. <https://doi.org/10.1037/0033-295X.109.4.619>.

66. Abdel-Khalek A. M. Why Do We Fear Death? The Construction and Validation of the Reasons for Death Fear Scale. *Death Studies*. 2002;26(8):669–680. <https://doi.org/10.1080/07481180290088365>.

67. Cicirelli V. G. Personal Meanings of Death in Relation to Fear of Death. *Death Studies*. 1998;22(8):713–733. <https://doi.org/10.1080/074811898201236>.

68. Bath D. M. Separation from Loved Ones in the Fear of Death. *Death Studies*. 2010;34(5):404–425. <https://doi.org/10.1080/07481181003697639>.

69. Taylor C. Sources of the Self: The Making of the Modern Identity. Cambridge, MA: Harvard University Press*, 1989. 601 p. (*Гарвардское университетское издательство (Harvard University Press) внесено в реестр нежелательных организаций на территории Российской Федерации).

70. Lindemann H. Holding and Letting Go: The Social Practice of Personal Identities. Oxford: Oxford University Press, 2016. 220 p.

71. Iftode C., Zorilă A., Vică C., Mihailov E. Experimental and Relational Authenticity: How Neurotechnologies Impact Narrative Identities. *Phenomenology and the Cognitive Sciences*. 2024;23:743–760. <https://doi.org/10.1007/s11097-022-09825-7>.

72. Velleman J. D. Self to Self: Selected Essays. Cambridge: Cambridge University Press, 2005. 385 p.

73. Turkle S. Authenticity in the Age of Digital Companions. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*. 2007;8(3):501–517. <https://doi.org/10.1075/is.8.3.11tur>.

74. Callahan D. Setting Limits: Medical Goals in an Aging Society with "A Response to My Critics". Washington: Georgetown University Press, 1995. 256 p.

Информация о переводчике:

АНТИПОВ Алексей Владимирович, кандидат философских наук, старший научный сотрудник сектора гуманитарных экспертиз и биоэтики Института философии Российской академии наук, г. Москва. SPIN-код: 7782-1854

AuthorID (РИНЦ): 828022

ORCID: <https://orcid.org/0000-0002-7048-3373>

AuthorID (SCOPUS): 58117217900

ResearcherID: U-8814-2018

e-mail: nelson02@yandex.ru

Источник перевода: Iglesias S., Earp B. D., Voinea C., Mann S. P., Zăhău A., Jecker N. S., Savulescu J. Digital Doppelgängers and Lifespan Extension: What Matters? *The American Journal of Bioethics*. 2025;25(2):95–110. <https://doi.org/10.1080/15265161.2024.2416133>.

Прозрачность финансовой деятельности: автор не имеет финансовой заинтересованности в представленных материалах и методах. Конфликт интересов отсутствует.

Автор прочитал и одобрил окончательный вариант рукописи.

При подготовке данной рукописи автор не использовал генеративный искусственный интеллект или аналогичные автоматизированные инструменты.

Статья поступила в редакцию 09.02.2026; одобрена после рецензирования 10.04.2026; принята к публикации 08.09.2026.

Information about the translator:

АНТИПОВ Алексей Владимирович, Candidate of Philosophical Sciences, Senior Researcher of the Humanitarian Expertise and Bioethics Department, Institute of Philosophy, Russian Academy of Sciences, Moscow.

SPIN-code: 7782-1854

AuthorID (RSCI): 828022

ORCID: <https://orcid.org/0000-0002-7048-3373>

AuthorID (SCOPUS): 58117217900

ResearcherID: U-8814-2018

e-mail: nelson02@yandex.ru

Financial transparency: the author has no financial interest in the presented materials or methods. There is no conflict of interest.

The author has read and approved the final manuscript. The author did not use the Artificial Intelligence or similar automated tools in the article.

The article was submitted 09.02.2026; approved after reviewing 10.04.2026; accepted for publication 08.09.2026.