DOI: 10.38050/2078-3809-2025-17-3-123-148

Экономика инноваций и развитие

РЕШЕНИЕ ПРОБЛЕМ УПРАВЛЕНИЯ: ВЛАСТЬ И ПОЛИТИКА В СПОРАХ ОБ УПРАВЛЕНИИ ГЕНЕРАТИВНЫМ ИИ*

Инга Улникейн
Бирмингемский университет, Эджбастон
(г. Бирмингем, Великобритания)

Автор перевода:
Белецкая Мария Юрьевна
кандидат экономических наук, старший научный сотрудник
МГУ имени М.В. Ломоносова, экономический факультет;
Институт Соединенных Штатов Америки и Канады
имени академика Г.А. Арбатова РАН
(г. Москва, Россия)

Аннотация

Запуск ChatGPT в конце 2022 г. привел к серьезным спорам об управлении генеративным искусственным интеллектом (ИИ). В этой статье рассматриваются первые международные инициативы в области управления и политики, специально посвященные генеративному ИИ: Хиросимский процесс «Большой семерки» (G7), отчеты Организации экономического сотрудничества и развития и саммит по безопасности ИИ в Великобритании. Этот анализ основан на литературе по разработке политики в области управления, в частности на работах по управлению технологиями и ответственным инновациям. Формирующееся управление генеративным ИИ демонстрирует полицентрические характеристики, где множественные и пересекающиеся центры принятия решений находятся в отношениях сотрудничества. Однако в нем доминирует ограниченное число развитых стран. Управление генеративным ИИ в основном сформулировано с точки зрения управления рисками, во многом игнорируя вопросы цели и направления инноваций и отводя довольно ограниченные роли общественности. Мы можем наблюдать возникновение «парадокса управления генеративным ИИ», а именно, как несмотря на то, что эта технология широко используется общественностью, ее управление довольно ограниченное. В этой статье вводится термин «управленческое решения», чтобы отразить этот довольно узкий и технократический подход к управлению генеративным ИИ. В качестве альтернативы предлагается принять политику полицентричного управления и ответственных инноваций, которые подчеркивают демократическое и совместное формирование технологий для общественного блага. В контексте крайне не-

* Дата публикации предварительного доступа: 2 июля 2024 г.. Оригинальная исследовательская статья DOI: https://doi.org/10.1093/polsoc/puae022
Улникейн И., e-mail: i.ulnicane@bham.ac.uk
Белецкая М.Ю., e-mail: mybeletskaya@gmail.com

равномерного распределения влияния в генеративном ИИ, характеризующегося высокой концентрацией власти в руках небольшого числа крупных технологических компаний, правительство играет особую роль в изменении дисбаланса сил путем обеспечения широкого участия общественности в управлении генеративным ИИ.

Ключевые слова: генеративный ИИ, управление, искусственный интеллект, ответственные инновации, риск.

JEL коды: Z18.

Для цитирования: Улникейн И. Решение проблем управления: власть и политика в спорах об управлении генеративным ИИ (перевод с англ. Белецкая М.Ю.) // Научные исследования экономического факультета. Электронный журнал. 2025. Том 17. Выпуск 3. С. 123-148. DOI: 10.38050/2078-3809-2025-17-3-123-148.

Введение

Запуск ChatGPT 30 ноября 2022 г. вызвал серьезные публичные споры о способах управления генеративным искусственным интеллектом (ИИ). Эксперты и компании в области ИИ призвали к мораторию на обучение более мощных систем ИИ из-за экзистенциальных угроз (Future of Life Institute, 2023). Обсуждения того, какие риски – экзистенциальные или более непосредственные – должны быть приоритетными и кто должен участвовать в принятии этих решений, сопровождали саммит по безопасности ИИ в Великобритании (UK Government, 2023; The White House, 2023). Споры о том, как следует регулировать базовые модели – посредством обязательных правил или саморегулирования, – грозили сорвать долгожданный первый всеобъемлющий документ по регулированию ИИ – Закон ЕС об ИИ (Bertuzzi, 2023; Mugge, 2024).

Эти противоречия возникли на фоне новой шумихи вокруг ИИ, характеризующейся высокими позитивными и негативными ожиданиями (Ulnicane et al., 2021), а также необходимостью срочно предпринять какие-то действия до выпуска более мощных моделей ИИ. В своем отчете о первоначальных политических соображениях в отношении генеративного ИИ, опубликованном в сентябре 2023 г., Организация экономического сотрудничества и развития (ОЭСР) заявила, что «выпуск ChatGPT удивил правительства, политиков и отдельных лиц по всему миру» (OECD, 2023c, p. 10) и что «публичное обсуждение генеративного ИИ длится меньше года. В то время как технологические компании выводят на рынок приложения генеративного ИИ, политики по всему миру пытаются разобраться с его последствиями» (OECD, 2023c, p. 29).

Эти заявления о том, что правительства и политики были удивлены выпуском ChatGPT и пытаются разобраться с последствиями генеративного ИИ, примечательны по нескольким причинам. Во-первых, к моменту выпуска ChatGPT политики работали над ИИ уже более 5 лет (Radu, 2021; Schiff, 2023; Taeihagh, 2021; Ulnicane, 2022), чего должно было быть достаточно для подготовки к будущим разработкам ИИ и соответствующей разработки политики управления. Во-вторых, информация об исследованиях генеративного ИИ была доступна задолго до выпуска ChatGPT. В том же отчете ОЭСР (OECD, 2023c) упоминается, что генеративный ИИ вышел на сцену в 2018 году, и цитируется известная статья «On the dangers of stochastic parrots: Can language models be too big?», опубликованная в начале 2021 г. (Bender et al., 2021), в которой уже тогда предупреждалось об опасностях больших языковых моделей, включая предвзятость и высокие финансовые и экологические издержки. Эта статья получила много общественного внимания в конце 2020 г., т. е. еще до ее публикации, когда она стала причиной увольнения одного из ее авторов Тимнита Гебру из Google (Wong, 2020).

Обсуждения политики и управления генеративным ИИ в течение первого года после выпуска ChatGPT подчеркнули актуальность некоторых ключевых вопросов в управлении технологиями и политике: каковы политические приоритеты? Как они устанавливаются? И кто участвует в их установлении (Khanal et al., 2024; Ulnicane, Erkkila, 2023; Whittaker, 2021)?

В связи с этим, целью данной статьи, представленной в специальном выпуске, посвященном управлению генеративным ИИ, является концептуализация и анализ власти и политики в новых дискуссиях об управлении в области генеративного ИИ. Некоторые из первых дискуссий об управлении и политике в области генеративного ИИ проходили на международном уровне. Поэтому в данной статье подробно рассматриваются ранние международные инициативы, посвященные непосредственно управлению и политике в области генеративного ИИ: Хиросимский процесс «Большой семерки» (G7) (G7, 2023a, 2023b, 2023c, 2023d), отчеты ОЭСР (OECD, 2023a, 2023b, 2023c) и саммит по безопасности искусственного интеллекта в Великобритании, который имел международный масштаб (AI Safety Summit, 2023; DSIT (Department for Science, Innovation and Technology), 2023; UK Government, 2023; Sunak, 2023). Эти инициативы основаны на работе стран-участниц. Они рассматриваются с упором на следующие вопросы. Какие предложения по международному управлению и политике выдвигаются? Какой тип международного управления формируется для генеративного ИИ? Какие рамки и противоречия доминируют в международном управлении и политике в отношении генеративного ИИ? Каковы пробелы в управлении и что умалчивается при разработки политики? Как можно концептуализировать формирующееся управление генеративным ИИ?

Эмпирический анализ в данной статье основан на политических документах, выступлениях и новостных статьях. Он также опирается на выводы из литературы по формированию политики и управлению, уделяя особое внимание управлению технологиями и ответственным инновациям.

Статья вносит два основных вклада в социальные исследования генеративного ИИ. Во-первых, она описывает формирующееся международное управление генеративным ИИ с точки зрения ключевых институтов и рамок. Во-вторых, она стремится концептуализировать его ключевые особенности. В ней определяется полицентричное управление, характеризующееся фокусом на рисках, доминированием определения риска над соображениями цели и ограниченной ролью общественности. Это можно назвать «парадоксом управления генеративным ИИ», чтобы подчеркнуть, что эта широкодоступная технология управляется довольно

узко. С концептуальной точки зрения, чтобы осмыслить эти ранние дебаты об управлении и политике генеративного ИИ, в этой работе вводится термин «управленческого решения», чтобы подчеркнуть, как управление инструментируется и рассматривается как быстрое технократическое решение сложных социальных, политических и экономических проблем.

В статье изложены следующие положения: во-первых, в ней представлены три ключевые концепции, используемые в данном исследовании: управление, генеративный ИИ и формирование политики; во-вторых, предполагается, что возникающие международные инициативы по управлению генеративным ИИ напоминают характеристики полицентрической системы управления; в-третьих, в ней рассматривается доминирование определения риска в управлении генеративным ИИ; в-четвертых, термин «управленческого решения» вводится как способ концептуализации довольно узкого и технократического подхода к управлению генеративным ИИ.

Ключевые концепции: управление, генеративный ИИ и разработка политики

Управление

Управление – это изменчивая и широко используемая концепция, имеющая множество значений и подходов в различных научных дисциплинах (Ansell, Torfing, 2022; Chhotray, Stoker, 2009; Levi-Faur, 2012; Pierre, Peters, 2021). В политологии концепция управления обычно ассоциируется с переходом от правительства к управлению в 1990-х годах из-за растущего скептицизма в отношении роли государства и правительства и растущих ожиданий в отношении участия негосударственных субъектов. Если правительство понимается как совокупность институтов государственного сектора, то управление рассматривается как процесс, посредством которого государственная политика развивается, начиная с постановки целей, мобилизации ресурсов и принятия решений до реализации, обратной связи и оценки (Pierre, Peters, 2021). Ключевой особенностью перехода от правительства к управлению является то, что последнее включает в себя новые формы взаимодействия между государственными и негосударственными субъектами. Гражданское общество и частный сектор рассматриваются как играющие все большую роль в политическом процессе. Однако способ и степень распределения власти между различными субъектами значительно различаются в зависимости от различных подходов к управлению и его форм (Pierre, Peters, 2021). Управление является глубоко политической концепцией, тесно связанной с вопросами власти, участия и плюрализма.

Ключ к управлению заключается в том, что это коллективный и интерактивный процесс, в котором участвует множество различных субъектов и организаций. Кристофер Анселл и Джейкоб Торфинг определяют управление как «интерактивные процессы, посредством которых общество и экономика направляются к достижению коллективно согласованных целей» (Ansell, Torfing, 2022, p. 4), в то время как Васудха Чхотрей и Джерри Стокер подчеркивают, что: «Управление касается правил коллективного принятия решений в условиях, где есть множество субъектов или организаций и где никакие формальные системы контроля не могут диктовать условия взаимоотношений между этими субъектами и организациями» (Chhotray, Stoker, 2009, p. 3).

Важно отметить, что Чхотрей и Стокер утверждают, что управление «следует понимать

аналитически и эмпирически как набор практик, а не через призму «списка пожеланий» принципов, которым необходимо следовать» (Chhotray, Stoker, 2009, p. 5).

Эти ключевые характеристики управления также применимы к управлению технологиями (Ulnicane et al., 2021), что важно, когда речь идет о генеративном ИИ. Одним из основных подходов к управлению технологиями за последние 10 лет стала концепция ответственных инноваций (De Saille, 2015; Fisher et al., 2024; Owen et al., 2012; Stilgoe et al., 2013), что определяется как «забота о будущем посредством коллективного управления наукой и инновациями в настоящем» (Stilgoe et al., 2013, p 1570). Широко распространено мнение, что: «Ответственные формы инноваций должны соответствовать социальным потребностям, реагировать на изменения в этических, социальных и экологических последствиях по мере развития исследовательской программы и включать общественность, а также традиционно определяемые заинтересованные стороны в двусторонние консультации» (De Saille, 2015, p. 153).

Хотя ответственные инновации представляют собой гибкий и широкий подход, в нем можно выделить три ключевые особенности. Во-первых, акцент на коллективном управлении в ответственных инновациях выходит за рамки индивидуальной ответственности исследователей и подчеркивает важность сотрудничества и системных факторов в управлении технологиями социально полезными способами. Во-вторых, акцент на включении общественных и двусторонних консультаций в подход ответственных инноваций отводит обществу проактивную роль в совместном формировании технологий, а не пассивную роль принимающего технологии как они есть. В-третьих, ответственные инновации расширяют обсуждение управления за пределы управления рисками, охватывая цель и направление инноваций (Stilgoe et al., 2013). Подход ответственных инноваций «стремится выйти за рамки того, чего мы не хотим от науки и инноваций – общеизвестной и хорошо документированной озабоченности характеристикой и управлением непреднамеренными рисками (последнее часто посредством регулирования), – а того, что мы хотим, чтобы они делали. ...Он спрашивает, как можно определить цели для инноваций этическим, инклюзивным, демократическим и справедливым образом» (Owen et al., 2012, p. 754).

Принятие вопросов цели и направления инноваций означает, что вместо подхода, ориентированного на предложение и стимулирование технологий, основное внимание уделяется общественному спросу и способам, которыми технологии могут внести вклад в решение общественных проблем в таких областях, как окружающая среда, здравоохранение и энергетика (Schiff, 2023; Ulnicane, 2022). Четыре тесно взаимосвязанных аспекта ответственных инноваций – предвидение, рефлексивность, инклюзивность и способность к реагированию – создают основу для постановки, обсуждения и ответа на вопросы о цели и направлении инноваций (Stilgoe et al., 2013). Эта основа включает в себя различные методы, такие как предвидение, оценка технологий, сканирование горизонтов, междисциплинарное сотрудничество и обучение, гражданское жюри и фокус-группы.

Генеративный ИИ

Недавние публичные дискуссии о генеративном ИИ демонстрируют типичные черты новых технологий, которые разрекламированы и связаны с высокими положительными и отрицательными ожиданиями (Ulnicane et al., 2021). Как это типично в случаях таких разрекламированных новых технологий, в политических документах говорится о генеративном ИИ,

создающем «новые прорывные инновации» (OECD, 2023a: 22), «предлагающем преобразовательный потенциал» (OECD, 2023c: 3) и имеющем «потенциал революционизировать отрасли и общество» (OECD, 2023c: 5), однако критики утверждают, что возможности, а также риски генеративного ИИ преувеличены (Gebru et al., 2023).

Хотя генеративный ИИ привлек к себе более широкое внимание в конце 2022 года с запуском ChatGPT, он появился несколькими годами ранее с появлением крупных языковых моделей (OECD, 2023b). Согласно определению ОЭСР, «генеративный ИИ можно понимать как форму модели ИИ, специально предназначенную для создания нового цифрового материала (включая текст, изображения, аудио, видео, программный код), в том числе при использовании таких моделей ИИ в приложениях и их пользовательских интерфейсах. Они обычно представляют собой системы машинного обучения, обученные на огромных объемах данных. Они работают, предсказывая слова, пиксели, формы сигналов, точки данных и т. д., которые будут соответствовать данным обучения модели, часто в ответ на подсказки» (OECD, 2023b, p. 6).

Генеративные системы ИИ включают, например, «ChatGPT и BARD для текста; Midjourney и Stable Diffusion для изображений; WaveNet и DeepVoice для аудио; Make-A-Video и Synthesia для видео; а также многомодельные системы, объединяющие несколько типов медиа» (OECD, 2023c, p. 8).

Другие термины, используемые параллельно с «генеративным ИИ», включают «передовые системы ИИ» и «передовой ИИ». В документе о запуске Хиросимского процесса G7 в мае 2023 г. говорится о «генеративном ИИ» (G7, 2023a), но Руководящие принципы и Кодекс поведения G7, опубликованные примерно полгода спустя, в октябре 2023 г., уже используют термин «передовые системы ИИ», который включает самые передовые базовые модели и генеративные системы ИИ (G7, 2023b, 2023c, 2023d). Саммит по безопасности искусственного интеллекта в Великобритании был посвящен «пограничному ИИ», который, по словам организаторов, относится к «высокоэффективным моделям искусственного интеллекта общего назначения, способным выполнять широкий спектр задач и соответствующим или превосходящим возможности самых современных моделей» (UK Government, 2023) .

Развитие генеративного ИИ возродило дискуссии о прогрессе в направлении создания общего искусственного интеллекта, а именно ИИ, который станет способен выполнять общие интеллектуальные действия, сопоставимые с человеческими (OECD, 2023a, 2023b). Однако такие ожидания остаются весьма спорными.

Разработка политики

Политические дебаты о генеративном ИИ сопровождались серьезными спорами о рисках, возможностях и способах работы с генеративным ИИ. Подход к разработке политики может помочь пролить свет на то, как проблемы формулируются, обсуждаются и решаются в ходе этих споров (Rein, Schon, 1996; Schon, Rein, 1994; Ulnicane, 2022; Van Hulst, Yanow, 2016). Политические рамки – это «диагностические/предписывающие истории, которые рассказывают в рамках данной проблемной области, что нужно исправить и как это можно исправить» (Rein, Schon, 1996, p. 89). Создание рамок – это динамический процесс, который включает в себя осмысление, выбор, наименование и категоризацию, а также повествование (Van Hulst, Yanow, 2016). Любая заданная проблемная область обычно характеризуется политическими

противоречиями и спорами, где несколько рамок конкурируют за смысл, легитимность и ресурсы (Rein, Schon, 1996; Schon, Rein, 1994). Такие конкурирующие рамки часто неявны и воспринимаются как должное, поэтому их сложно разрешить рассуждениями или обращением к фактам. При анализе рамок также важно обращать внимание на упущения, умолчания и скрытые в них политические мотивы (Bacchi, 2000).

Формирующаяся система полицентрического управления генеративным ИИ

На ранних этапах публичных дискуссий о генеративном ИИ была подчеркнута необходимость международного сотрудничества в связи с трансграничным характером его воздействия и глобальным охватом крупных компаний в сфере ИИ. Ряд организаций, работающих в области генеративного ИИ, такие как G7, ОЭСР и ЕС, хорошо известны своей деятельностью в области политики и управления в области ИИ в предыдущие годы (Cihon et al., 2020; Roberts et al., 2024; Schmitt, 2022; Veale et al., 2023).

Хотя некоторые обзоры и картографические исследования различных институтов и инициатив международного управления ИИ выявили его фрагментацию (Cihon et al., 2020), его скорее можно концептуализировать в терминах полицентрического управления. Концепция полицентрического управления широко используется в области изменения климата и окружающей среды (Carlisle, Gruby, 2019; Ostrom, 2010), но в последнее время он также использовался в некоторых исследованиях ИИ и цифрового управления (Aguerre et al., 2024; Filgueiras, Almeida, 2021; Schmitt, 2022).

Полицентрическая система управления подразумевает наличие множественных и перекрывающихся центров принятия решений в разных масштабах, которые независимы друг от друга, но в то же время находятся в различных отношениях сотрудничества и конкуренции друг с другом (Carlisle, Gruby, 2019; Ostrom, 2010). По словам Элинор Остром, «каждая единица полицентрической системы обладает значительной независимостью в установлении норм и правил в определенной области» (Ostrom, 2010, p. 552). Она подчеркивает преимущества полицентрических систем, такие как их механизмы взаимного мониторинга, обучения и адаптации, которые, как правило, способствуют инновациям, доверию и сотрудничеству.

Полезно рассмотреть различные инициативы и организации, участвующие в управлении генеративным ИИ, с точки зрения взаимосвязей, сотрудничества и конкуренции, а не просто как модель фрагментации, характеризующуюся дефицитом. Обзор начальных этапов развития управления генеративным ИИ см. в табл. 1.

Таблица 1

Первоначальные разработки в области международного управления генеративным ИИ

ОРГАНИЗАЦИЯ	ДАТА	НАЗВАНИЕ	ССЫЛКИ НА ДРУГИЕ ОРГАНИЗАЦИИ
G7	Май 2023 г.	Решение о создании «Хиросимского процесса» для генеративного ИИ	Сотрудничество с ОЭСР и GPAI
ОЭСР	Сентябрь 2023 г.	Хиросимский процесс G7 по генеративному ИИ	Отчет для G7

ОРГАНИЗАЦИЯ	ДАТА	НАЗВАНИЕ	ССЫЛКИ НА ДРУГИЕ ОРГАНИЗАЦИИ
ОЭСР	Сентябрь 2023 г.	Первоначальные политические соображения относительно генеративного ИИ	Вклад Европейской комиссии и Японии; упоминание G7
G7	Октябрь 2023 г.	Руководящие принципы и Кодекс поведения для организаций, разрабатывающих передовые системы ИИ	Основан на принципах ОЭСР в области искусственного интеллекта; сотрудничество с ОЭСР и GPAI; приветствует саммит по безопасности искусственного интеллекта
ПРАВИТЕЛЬСТВО ВЕЛИКОБРИТАНИИ	Ноябрь 2023 г.	Саммит по безопасности ИИ	Основан на работе ОЭСР, GPAI, Совета Европы и G7

Первая крупная международная инициатива в области управления генеративным ИИ – Хиросимский процесс – демонстрирует отношения сотрудничества между G7, ОЭСР, Глобальным партнерством по искусственному интеллекту (GPAI) и ЕС. На саммите в Хиросиме в мае 2023 г. лидеры G7 (группа из семи: Канада, Франция, Германия, Италия, Япония, Великобритания, США и ЕС) признали необходимость немедленного анализа возможностей и проблем генеративного ИИ. Они призвали ОЭСР рассмотреть анализ соответствующей политики и GPAI для реализации практических проектов. В частности, они поручили соответствующим министрам «создать Хиросимский процесс через рабочую группу G7 на инклюзивной основе и в сотрудничестве с ОЭСР и GPAI для обсуждения генеративного ИИ к концу этого года» (G7, 2023a), который мог бы включать такие темы, как управление, защита прав интеллектуальной собственности, включая авторские права, содействие прозрачности, реагирование на манипуляции иностранной информацией, включая дезинформацию и ответственное использование этих технологий. ОЭСР опубликовала доклад, призванный помочь в обсуждении общих политических приоритетов в рамках «Большой семерки» (OECD, 2023b).

Первыми результатами Хиросимского процесса стали Международные руководящие принципы и Кодекс поведения для организаций, разрабатывающих передовые системы искусственного интеллекта (G7, 2023b). Эти документы основаны на существующих принципах ОЭСР в области искусственного интеллекта (OECD, 2019) и являются ответом на последние разработки в области передовых систем искусственного интеллекта (G7, 2023c, 2023d). Здесь лидеры G7 снова обязуются работать с ОЭСР, GPAI и другими заинтересованными сторонами по мониторингу внедрения этих документов. Они также признают, что разные юрисдикции могут использовать свои собственные уникальные подходы к внедрению этих руководящих принципов и действий по-разному. Это важный принцип с точки зрения полицентричного управления, обеспечивающий гибкость для каждой страны в реализации общих принципов таким образом, чтобы это соответствовало их контексту и потребностям. Более того, в своем заявлении от 30 октября 2023 г. лидеры G7 упомянули, что они с нетерпением ждут Саммита по безопасности ИИ в Великобритании, который состоится 1 и 2 ноября. В то же время организаторы Саммита по безопасности ИИ в Великобритании признали, что саммит основывается на работе, проделанной в ОЭСР, GPAI, Совете Европы и Хиросимском процессе ИИ (UK

Government, 2023).

В документе ОЭСР о генеративном ИИ, подготовленном при участии Европейской комиссии и Японии, упоминается о создании совместно с ОЭСР Хиросимского процесса G7 (OECD, 2023c, p. 5, 9). Европейская комиссия приветствовала принципы и Кодекс поведения G7, заявив, что они отражают ценности ЕС, направленные на продвижение надежного ИИ и дополняют на международном уровне будущий Закон ЕС об ИИ (European Commission, 2023). В заявлении Комиссия также связала свою работу над Кодексом поведения G7 с намерением работать над этими вопросами, о котором было объявлено ранее на двустороннем Совете ЕС по торговле и технологиям и США.

Запуск и первые инициативы Хиросимского процесса G7 по генеративному ИИ демонстрируют, что управление генеративным ИИ представляет собой не фрагментированный процесс, а скорее формирующуюся полицентрическую систему управления, в которой различные независимые органы с частично совпадающим членством (G7, ОЭСР, GPAI и ЕС) находятся в отношениях сотрудничества друг с другом, что обеспечивает возможности для взаимного мониторинга, обучения и адаптации. Развитие генеративного ИИ, по-видимому, еще больше усиливает важную роль ОЭСР (Schmitt, 2022), а ее принципы в области ИИ (OECD, 2019) укрепляются в качестве важной точки отсчета на международном уровне.

Однако в глобальном масштабе G7 и ОЭСР (38 государств-членов в 2023 г.) являются «закрытыми клубами» некоторых из наиболее развитых стран. Их доминирование в вопросах глобального управления генеративным ИИ поднимает вопросы об инклюзивности, равенстве и представительстве интересов и потребностей менее развитых стран. Саммит по безопасности ИИ в Великобритании с его глобальными устремлениями (второй Саммит по безопасности ИИ состоялся в мае 2024 г. в Южной Корее) включал некоторые страны из Азии, Африки и Латинской Америки, но в общей сложности присутствовали только 28 стран и ЕС (AI Safety Summit, 2023). Таким образом, первоначальные инициативы по управлению генеративным ИИ доминируют в относительно ограниченном числе преимущественно развитых стран, при этом остальной мир, который также использует генеративный ИИ и испытывает на себе его влияние, практически не имеет права голоса.

Развитие управления для генеративного ИИ оживило дискуссии о необходимости новых глобальных форумов в этой области, вдохновленных примерами из других областей, таких как Межправительственная группа экспертов по изменению климата и Международное агентство по атомной энергии (Roberts et al., 2024). В октябре 2023 г. Генеральный секретарь Организации Объединенных Наций создал Многосторонний консультативный орган по искусственному интеллекту, однако он не посвящен конкретно генеративному ИИ. Формирующееся полицентричное управление генеративным ИИ может выиграть от развития дальнейших связей между международными инициативами и соответствующей национальной политикой в государствах–членах ОЭСР и странах–партнерах (OECD, 2023a), а также новые инициативы в области генеративного ИИ в других регионах, такие как Ассоциация государств Юго-Восточной Азии, и более широкие инициативы в области ИИ, такие как вышеупомянутый орган ООН, призванные содействовать согласованию управления генеративным ИИ с управленем ИИ в более широком смысле.

Возвращаясь к рискам: новые подходы к управлению генеративным ИИ

С самого начала общественное обсуждение генеративного ИИ велось с акцентом на риски. Как отмечалось ранее, хотя управление рисками всегда было в центре внимания традиционного управления в сфере технологий, доминирующий подход последних 10 лет, а именно «ответственные инновации», стремился выйти за рамки управления рисками и рассмотреть также вопросы цели, направления инноваций и решения социальных проблем (Owen et al., 2012; Stilgoe et al., 2013). Соответственно, доминирование вопросов управления рисками в формирующемся управлении и политике в области генеративного ИИ можно рассматривать как возврат к более узкому подходу к управлению технологиями.

Для выявления новых рамок управления генеративным ИИ первоначальные противоречия и документы анализируются в контексте подхода «Ответственные инновации», описанного выше, и его ключевых особенностей, таких как выход за рамки управления рисками и концентрация на цели, а также акцент на вовлечении общества в двусторонние консультации. Соответственно, можно выделить три новые рамки управления генеративным ИИ: во-первых, возрождение дискуссии об экзистенциальном риске; во-вторых, доминирование управления рисками над соображениями цели; и, в-третьих, ограниченная роль общества.

Возобновленные дебаты об экзистенциальном риске

Запуск и возможности ChatGPT и других генеративных моделей ИИ возродили дискуссию об экзистенциальных рисках, которая была частью дискуссий об ИИ уже около 10 лет (Galanos, 2019). В марте 2023 г. в открытом письме, опубликованном Институтом будущего жизни, говорится о том, что ИИ может представлять «серьезные риски для общества и человечества», так как является «нечеловеческим разумом, который в конечном итоге может превзойти нас численностью, интеллектом, и заменить людей как устаревших», и риске потери «контроля над нашей цивилизацией» (Future of Life Institute, 2023). В этом письме ко всем лабораториям ИИ был обращен призыв немедленно приостановить как минимум на 6 месяцев обучение систем ИИ, более мощных, чем GPT-4, но если такую паузу невозможно осуществить быстро, то правительствам предлагалось вмешаться и ввести мораторий. В письме предлагалось использовать паузу для разработки и внедрения протоколов безопасности для передового проектирования ИИ. Более того, разработчикам ИИ предлагалось работать с политиками, чтобы значительно ускорить разработку надежных систем управления ИИ. Первоначально это письмо подписали Илон Маск, Йошуа Бенджио, Стюарт Рассел и другие известные деятели, а сейчас его подписали более 30 000 человек.

Всего пару месяцев спустя, в мае 2023 г., некоторые из тех же подписавшихся опубликовали заявление о рисках ИИ, в котором говорится, что «снижение риска вымирания от ИИ должно быть глобальным приоритетом наряду с другими рисками общественного масштаба, такими как пандемии и ядерная война» (Center for AI Safety, 2023). Примерно в то же время некоторые из так называемых крестных отцов ИИ, такие как Джеффри Хинтон и Йошуа Бенджио, выступили со своими предупреждениями об опасностях ИИ и заявили, что сожалеют о своей работе над ИИ (Pringle, 2023; Taylor, Hern, 2023). Генеративный ИИ сравнивали с атомной бомбой, и в ходе популярных дебатов высказывались опасения, что ИИ приближается к своему «моменту Оппенгеймера» (Mouriquand, 2023), имея в виду Роберта Оппенгеймера, известного как «отец атомной бомбы», который позже сожалел о своей роли в ее разработке.

Эти предупреждения об экзистенциальном риске ИИ подверглись серьезной критике и

скептицизму, подчеркивая преувеличение как рисков, так и возможностей ИИ. Авторы известной упомянутой статьи (Bender et al., 2021) утверждают, что вместо того, чтобы сосредоточиваться на воображаемых проблемах и гипотетических рисках, таких как мощные цифровые интеллекты, «мы должны сосредоточиться на вполне реальных и существующих эксплуататорских практиках компаний, претендующих на их создание, которые быстро централизуют власть и усиливают социальное неравенство» (Gebru et al., 2023) .

Если до широкого распространения генеративного ИИ рассуждения об экзистенциальной угрозе в значительной степени игнорировались политиками, то новым явлением в дебатах о генеративном ИИ стало то, что он стал частью политического мейнстрима и был подхвачен политическими лидерами. Председатель Европейской комиссии Урсула фон дер Ляйен в своем послании о положении страны в сентябре 2023 г. процитировала вышеупомянутое заявление о риске ИИ, в котором риск вымирания от ИИ сравнивался с риском пандемий и ядерной войны (Von der Leyen, 2023). Аналогичным образом, идеи об экзистенциальном риске повлияли на политику правительства Великобритании в области ИИ и подготовку к его глобальному саммиту по безопасности ИИ, где акцент на потере контроля человека над передовыми системами ИИ был назван одним из рисков, связанных с передовыми технологиями ИИ (Clarke, 2023; DSIT (Department for Science, Innovation and Technology), 2023; Sunak, 2023).

Учитывая противоположные взгляды на риски, исходящие от ИИ, СМИ и комментаторы начали представлять дебаты о генеративном ИИ как противоречие между экзистенциальными и непосредственными рисками. Соответственно, многие политики и эксперты сформулировали свои мнения об ИИ в этих терминах. В своем выступлении перед Саммитом по безопасности ИИ в Великобритании вице-президент США Камала Харрис выступила за более широкое понимание экзистенциальных угроз, которое включает в себя не только угрозы, которые «могут поставить под угрозу само существование человечества», но и «угрозы, которые в настоящее время причиняют вред и которые для многих людей также кажутся экзистенциальными», такие как неисправные алгоритмы, исключающие людей из их плана медицинского страхования, или неправомерное тюремное заключение из-за предвзятого распознавания лиц (The White House, 2023). Этот пример показывает, насколько влиятельным стало разделение экзистенциальных и непосредственных рисков, что даже политики, которые не полностью согласны с ним, все еще используют его в качестве точки отсчета.

Доминирование управления рисками над соображениями цели

При обсуждении рисков, угроз и вреда генеративного ИИ в политических документах обычно упоминаются такие проблемы, как распространение дезинформации и предвзятости, нарушение прав человека, изменение рынков труда, подрыв безопасности и повышение обеспокоенности по поводу прав интеллектуальной собственности (DSIT, 2023; OECD, 2023a, 2023b, 2023c). Как подытожила ОЭСР, генеративные технологии искусственного интеллекта «создают критические социальные и политические проблемы, с которыми должны сталкиваться политики: потенциальные изменения на рынках труда, неопределенность в отношении авторских прав и риск, связанный с сохранением общественных предубеждений и потенциальным злоупотреблением при создании дезинформации и манипулировании контентом. Последствия могут включать распространение ложной информации и дезинформации, сохранение дискриминации, искажение общественного дискурса и рынков, а также подстрекательство к насилию» (OECD, 2023c, p. 3).

В документе для обсуждения, подготовленном правительством Великобритании для саммита по безопасности ИИ в Великобритании, излагаются четыре типа рисков, связанных с передовым ИИ: сквозные факторы риска, общественный вред, риски неправильного использования и потеря контроля (DSIT, 2023). Первая группа сквозных факторов риска фокусируется на вопросах безопасности, таких как отсутствие стандартов безопасности и недостаточные стимулы для разработчиков ИИ инвестировать в меры по снижению рисков. В нем также упоминается вероятность высокой концентрации рыночной власти среди разработчиков передового ИИ, что может ослабить конкуренцию, сократить инновации и потребительский выбор. Вторая группа общественного вреда охватывает деградацию информационной среды, нарушение рынка труда и предвзятость. Третья группа рисков неправильного использования включает использование наук о жизни в злонамеренных целях, обострение киберрисков и кампании по дезинформации.

Хотя первые три группы рисков уже хорошо известны из политики в области ИИ в предыдущие годы, четвертая привлекла больше внимания в контексте генеративного ИИ. Она подчеркивает более спекулятивный и противоречивый риск «потери контроля», который может произойти из-за двух факторов: во-первых, «люди все чаще передают контроль над важными решениями ИИ. Людям становится все труднее вернуть себе контроль», и, во-вторых, «системы ИИ активно стремятся увеличить свое собственное влияние и уменьшить человеческий контроль» (DSIT, 2023, p. 26). Эти более спекулятивные разработки также упоминаются ОЭСР среди потенциальных будущих проблем и рисков, связанных с появляющимися моделями поведения ИИ, которые могут привести к «коллективному бесправию – предполагаемой опасности того, что возможности моделей будут выполнять все более важные функции в обществе, отнимая власть у людей» (OECD, 2023c, p. 27).

Хотя в политических документах по генеративному ИИ обсуждается ряд угроз, включая более спекулятивные, они также игнорируют некоторые другие виды рисков. Одной из хорошо известных проблем генеративного ИИ являются высокие экологические затраты на обучение и использование передовых моделей ИИ. Эти затраты, представляющие серьезную проблему в контексте изменения климата, обсуждаются в научной литературе и СМИ (например, Bender et al., 2021; Crawford, 2024), до сих пор в значительной степени игнорировались в международных политических документах по генеративному ИИ.

Чрезмерное внимание к рискам в дискуссиях об управлении и политике в области генеративного ИИ приводит к отходу на второй план или игнорированию других вопросов. Один из вопросов, который практически не затрагивается в этих дискуссиях, – это цель разработки и использования технологий, что является ключевым аспектом подхода «ответственных инноваций».

Цели и направлению развития генеративного ИИ уделяется мало внимания при оценке рисков. Саммит по безопасности ИИ в Великобритании был посвящен исключительно управлению рисками, связанными с передовыми технологиями ИИ (UK Government, 2023). В первоначальных политических соображениях ОЭСР относительно генеративного ИИ преимущества упоминаются довольно обобщенно, но основное внимание уделяется рискам (OECD, 2023c). Демократический процесс выбора цели и направления развития генеративного ИИ, а также выбор соответствующих политических мер не входит в эти политические соображения (OECD, 2023a, 2023c).

Из 11 принципов, изложенных в Руководящих принципах и Кодексе поведения Хиросимского процесса G7, большинство из которых связаны с рисками, только один посвящен разработке передовых систем ИИ для решения самых серьезных мировых проблем, таких как климатический кризис, глобальное здравоохранение и образование, для поддержки прогресса в достижении Целей устойчивого развития Организации Объединенных Наций и для работы с гражданским обществом и общественными группами для определения приоритетных проблем (G7, 2023c, 2023d) . В целом, сильный акцент на риске в первоначальных инициативах по управлению и политике генеративного ИИ во многом затмил более ранние (хотя и ограниченные) обсуждения о цели ИИ и его роли в решении общественных проблем посредством широкомасштабного сотрудничества (Schiff, 2023; Ulnicane, 2022) .

Управление технологиями, которое рассматривает управление рисками как ключевую деятельность, имеет тенденцию предполагать, что до тех пор, пока риски смягчены, новые технологии автоматически принесут выгоду, игнорируя важные вопросы о том, кто и как выиграет от этих технологий. Это также известно в литературе как «проинновационная предвзятость», предполагающая, что инновации всегда хороши, и важно иметь как можно больше инноваций как можно быстрее (Ulnicane, 2022). В Блетчлийской декларации Саммита по безопасности ИИ в Великобритании говорится, что страны должны рассмотреть важность проинновационного управления и подхода к регулированию, который максимизирует выгоды и учитывает риски, связанные с ИИ (AI Safety Summit, 2023). Однако, преимущества этой технологии – и для кого она предназначена – далеко не так однозначны. Преимущества для одних, например, прибыль для компаний, могут сопровождаться вредом для других, например, несправедливым отношением к меньшинствам, плохими условиями труда для работников или высоким потреблением природных ресурсов и энергии, как показали многочисленные примеры использования ИИ (Crawford, 2021; Noble, 2018; Zuboff, 2019).

Многочисленные примеры неравномерного распределения выгод, получаемых от новых технологий, напоминают о том, что необходим не столько проинновационный подход, сколько более демократичный и инклюзивный подход к выбору цели и направления развития и использования технологий для общественного блага. Однако в условиях, где доминирует технологический прорыв и связанная с ним узкая ориентация на управление рисками, эти демократические дискуссии о цели и направлении развития генеративного ИИ в значительной степени отсутствуют. Следовательно, не ведутся важные дискуссии о политических мерах и механизмах управления, необходимых для реализации этой общественной цели и социальных выгод.

Ограниченная роль для общества: парадокс генеративного управления ИИ

Первоначальные инициативы по управлению и политике генеративного ИИ представляют собой довольно узкий и технократический подход к управлению технологиями, практически не оставляя места для демократического и инклюзивного обсуждения цели и направления этой инновации. Преимущественно технократический разговор о рисках и безопасности генеративного ИИ ведут представители промышленности, технологические эксперты и большинство развитых стран, при этом -гражданское общество и остальной мир практически не участвуют. Ведущие эксперты по ИИ в своем письме-паузе о серьезных рисках, которые генеративный ИИ представляет для общества и человечества, заявляют, что обществу следует дать шанс адаптироваться (Future of Life Institute, 2023), а не формировать его проактивно. Первая

публикация ОЭСР о языковых моделях ИИ (OECD, 2023a) является еще одним красноречивым примером. Хотя в этой публикации упоминается многостороннее сотрудничество в вопросах политики, роль заинтересованных сторон ограничивается предотвращением и смягчением рисков (OECD, 2023a, p. 40), а не обсуждением цели обучения и использования этих моделей.

Вместо того, чтобы включать общество в двусторонние консультации, как предполагает подход «Ответственные инновации» (De Saille, 2015), роль общества сводится к адаптации к генеративному ИИ и содействию управлению рисками. Речь премьер-министра Великобритании накануне Саммита по безопасности ИИ прекрасно иллюстрирует эту пассивную роль, возложенную на общество: «И вы можете доверять мне, я приму правильные долгосрочные решения, обеспечивающие ваши спокойствие и безопасность, в то же время гарантируя, что у вас и ваших детей будут все возможности для лучшего будущего, которое может принести ИИ» (Sunak, 2023). Вместо того, чтобы дать обществу право голоса и выбор в отношении того, какие возможности и будущее они хотят, премьер-министр просто просит общественность доверять его решениям.

Однако эта ограниченная роль, отводимая обществу, также столкнулась с сопротивлением. Организация саммита по безопасности искусственного интеллекта в Великобритании как «небольшой и целенаправленной дискуссии», ограниченной примерно 100 участниками (UK Government…, 2023), подверглось критике со стороны организаций гражданского общества. В открытом письме премьер-министру, подписанном более чем 100 организациями гражданского общества и учеными, указывалось, что «сообщества и работники, наиболее пострадавшие от ИИ, были маргинализированы Саммитом. Участие организаций гражданского общества, обладающих разнообразными экспертными знаниями и точками зрения, было избирательным и ограниченным. Это упущенная возможность» (Connected by Data…, 2023). В открытом письме содержался призыв предоставить весомое слово и равное место за столом переговоров сообществам, наиболее подверженным вреду ИИ.

В дополнение к саммиту по искусственному интеллекту, организованному правительством Великобритании, в Лондоне прошла серия мероприятий AI Fringe, призванных объединить взгляды представителей промышленности, гражданского общества и академических кругов и созвать Народную группу по искусственному интеллекту (AI Fringe…, 2024). Примечательно, что эти разнообразные взгляды различных социальных групп, сосредоточенных на инклюзивной и партисипативной разработке искусственного интеллекта, представлены как «маргинальные», а не являются частью основной дискуссии.

Подводя итог, мы наблюдаем парадоксальную ситуацию. Хотя генеративный ИИ гораздо более доступен и широко используется обществом, чем прежние инструменты ИИ, требующие гораздо более глубоких специальных знаний, управление и политика в отношении генеративного ИИ становятся более узкими, отдавая приоритет рискам и техническим экспертам, а не двусторонним консультациям с участием общественности. Я называю это «парадоксом управления генеративным ИИ», чтобы подчеркнуть эти противоположные тенденции: более широко используемые технологии управляются менее партиципативно.

Обсуждение: «управленческое решение» – узкий и технократический подход к управлению генеративным ИИ

Формирование международного управления генеративным ИИ характеризуется возрождением дискуссий о экзистенциальных рисках, доминированием управления рисками над соображениями цели и ограниченной ролью общества. Чтобы концептуализировать этот довольно узкий и технократический подход к управлению, я ввожу термин «управленческого решения» (Governance Fix)[1]. Для этого я использую концепцию «технологического решения», которая представляет технологии как решение сложных и неопределенных социальных проблем. Один из главных сторонников быстрых и дешевых технологических решений, Элвин Вайнберг, спросил: «В какой степени социальные проблемы можно обойти, сведя их к технологическим решениям? Можем ли мы найти быстрые технологические решения для глубоких и почти бесконечно сложных социальных проблем, “решения”, доступные современным технологиям, которые либо устранят исходную социальную проблему, не требуя изменения социальных установок человека, либо изменят проблему таким образом, что ее решение станет более осуществимым?» (Weinberg, 1966, p. 5).

Подход технологического решения проблем предполагает, что технологические решения превосходят более традиционные политические, экономические, образовательные и другие подходы социальных наук к решению проблем (Johnston, 2017). Согласно этому подходу, технически компетентные люди, такие как инженеры, лучше всего подготовлены к решению современных социальных проблем. Хотя технологические решения проблем были популярны среди технологов и политиков, в том числе в сфере искусственного интеллекта (Ulnicane, 2022), их также долгое время критиковали за неполноту, неэффективность, безуспешность, угрозы, неспособность добраться до сути проблемы, создание новых проблем при решении старых, односторонний, а не целостный, и механический, а не экологический подходы (Rosner, 2004; Weinberg, 1966). Более того, приоритизация технологических решений дает технологическим компаниям возможность продвигать свои корыстные интересы (Khanal et al., 2024; Morozov, 2011). Сосредоточившись на технологических решениях, политики избегают поиска более комплексных подходов и сосредотачиваются на проблемах, которые легко решаются, а не на тех, которые требуют немедленного внимания (Morozov, 2011).

Развивая идеи, высказанные в ходе дискуссий о «технологическом решении» и недавних дискуссий о генеративном ИИ, я предлагаю концепцию «управленческого решения», которая аналогичным образом представляет управление как технократический инструмент, который можно быстро разработать и внедрить. Это существенно отличается от представленных ранее концепций управления и подхода «ответственные инновации», которые фокусируются на коллективно согласованных целях и вовлечении широкого круга участников в процесс принятия решений. Ключевые характеристики управления, ответственных инноваций, технологического решения и управленческого решения сравниваются в табл. 2.

[1] Хотя термин «управленческое решение» упоминался ранее (Vilakazi, Roberts, 2019), до сих пор он не получил подробного описания.

Таблица 2

Основные характеристики управления, ответственных инноваций, технологического решения и управленческого решения

	УПРАВЛЕНИЕ	**ОТВЕТСТВЕННЫЕ ИННОВАЦИИ**	**ТЕХНОЛОГИЧЕСКОЕ РЕШЕНИЕ**	**УПРАВЛЕНЧЕСКОЕ РЕШЕНИЕ**
ФОКУС	Коллективно согласованные цели, учитывающие сложность социальных проблем	Демократическое определение направления и цели инноваций, т. е. решение общественных проблем	Экспертные предложения по быстрым и недорогим техническим решениям социальных проблем	Технократический подход к решению проблем
КТО ПРИНИМАЕТ РЕШЕНИЯ?	Взаимодействие государства и негосударственных акторов	Общественность и заинтересованные в двустороннем процессе стороны	Технические эксперты	Правительственные и технические эксперты
РОЛЬ ОБЩЕСТВА	Участие в переговорах, определение и принятие решений по целям	Активное совместное формирование инновации	Принимается техническими экспертами	Ограничено поддержкой и адаптацией

В то время как политика, участие и решение сложных проблем находятся в центре концепций управления и ответственных инноваций, подход, основанный на технологическом решении, отдает приоритет предложениям экспертов относительно быстрых и недорогих технических мер для решения социальных проблем. Я полагаю, что «управленческое решение» аналогичным образом подчеркивает роль технических знаний и информации как способа решения сложных управленческих проблем. Например, В открытом письме Института будущего жизни (2023) было высказано предположение, что ключевые вопросы управления генеративным ИИ могут быть решены в течение шестимесячной паузы. Вместо быстрого решения проблем управления, решение фундаментальных вопросов управления в контексте разработки генеративного ИИ потребовало бы более существенных реформ политических и экономических систем для решения некоторых ключевых проблем, выявленных в ходе развития ИИ и генеративного ИИ, таких как концентрация власти в руках крупных технологических компаний, приоритет экономических вопросов над социальными или усугубление неравенства (Khanal et al., 2024; Radu, 2021; Taeihagh, 2021; Ulnicane et al., 2021).

Подход «управленческого решения», используемый в недавних дебатах о генеративном ИИ, представляет собой довольно обедненную идею управления, лишенную своих корней в политологии, которая рассматривает управление как «по сути политическую концепцию» (Peters, 2012), сосредоточенную на вопросах участия различных государственных и негосударственных субъектов, инклюзивности и принятия решений (Ulnicane, Erkkila, 2023). Способом выхода за рамки узкого и технократического подхода «управленческого решения» к генеративному ИИ было бы принятие более широкого демократического и партисипативного подхода, опирающегося на политику полицентрического управления и концепции ответственных инноваций. Это подразумевало бы включение общественности и широкого круга заинтересованных сторон не только в управление рисками, но и в переговоры о цели, мотивации и направлении разработки и использования генеративного ИИ социально полезными способами. Применяя взаимосвязанные аспекты ответственных инноваций, такие как предугадывание, рефлексивность, инклюзивность и реагирование, а также различные методы от предвидения и

оценки технологий до междисциплинарного сотрудничества и советов граждан (Browne, 2023; McQuillan, 2018; Stilgoe et al., 2013), общество и различные заинтересованные стороны могли бы играть более активную роль в совместном формировании генеративного ИИ.

Подход ответственных инноваций не предоставляет «стабильного плана по “исправлению” неопределенных, сложных и неоднозначных общественных аспектов инноваций», а скорее подразумевает «приверженность непрерывному обучению, множественным точкам зрения и продуктивному сотрудничеству» (Fisher et al., 2024, p. 22), что актуально для управления генеративным ИИ. Более того, в контексте крайне неравномерного распределения сил в генеративном ИИ, где власть и ресурсы сосредоточены в руках небольшого числа крупных технологических компаний, а у общественности очень мало влияния, правительствам предстоит сыграть особую роль в изменении, а не укреплении существующего дисбаланса сил (Ulnicane et al., 2021). Соответственно, важно, чтобы правительство играло активную роль в обеспечении общественного участия, а также в содействии и смягчении участия различных заинтересованных сторон в коллективном управлении генеративным ИИ.

Заключение

В данной статье рассматриваются первые международные инициативы в области управления и политики, специально посвященные генеративному ИИ – Хиросимский процесс G7, доклады ОЭСР и Саммит по безопасности ИИ в Великобритании – в контексте более широких дебатов с участием политических лидеров и различных заинтересованных сторон. Анализ опирается на литературу по управлению, ответственным инновациям и формированию политики, чтобы проанализировать формирующееся управление, рамки и противоречия, связанные с генеративным ИИ. В статье утверждается, что формирующееся управление генеративным ИИ демонстрирует характеристики полицентричной системы управления, включающей множество пересекающихся центров управления, находящихся в отношениях сотрудничества друг с другом. Однако в этом управлении доминирует ограниченное число преимущественно развитых стран.

Основное внимание в формирующемся управлении и политике для генеративного ИИ уделяется управлению рисками, включая, во-первых, возрождение опасений по поводу экзистенциального риска; во-вторых, чрезмерную ориентацию на управление рисками, которая затмевает соображения цели и направления; и, в-третьих, ограниченную роль, отводимую обществу. Это приводит к «парадоксу управления генеративным ИИ», когда эта технология, которая широко используется общественностью, в то же время управляется довольно узко. Таким образом, управление генеративным ИИ усиливает и еще больше усугубляет многие проблемы, известные из исследований управления ИИ, такие как дисбаланс сил, неравенство и ограниченная роль для общества (Radu, 2021; Schiff, 2023; Taeihagh, 2021; Ulnicane, Erkkila, 2023).

Чтобы отразить этот довольно узкий и технократический подход к генеративному ИИ, я ввожу термин «управленческое решение» (Governance Fix), где управление рассматривается как быстрое решение сложных и многогранных проблем. В качестве альтернативы я предлагаю принять политику управления и ответственных инноваций, которая подчеркивает важность широкого участия общественности и различных заинтересованных сторон в согласовании целей и направлений развития технологий. В условиях крайне неравномерного распределения власти в сфере генеративного ИИ правительство играет особую роль в обеспечении такого партисипативного управления, способствуя вовлечению общественности.

Данное исследование имеет два основных значения для будущих исследований. Во-первых, важно проследить, как развивается разрекламированная технология генеративного ИИ, с которой связано множество позитивных и негативных ожиданий, и как ее формирует меняющееся управление на разных уровнях. Во-вторых, ключевые особенности управления технологиями, обсуждаемые в данном исследовании, такие как выход за рамки управления рисками и сосредоточение на цели, а также на ролях, отводимых обществу, имеют решающее значение для развития других технологий, помимо генеративного ИИ. Концепция «управленческого решения» может быть полезна для более широкого изучения ограничений и возможностей управления новыми технологиями .

Список литературы

Aguerre C., Campbell-Verduyn M., Scholte J.A. Global digital data governance: Polycentric perspectives. Routledge, 2024. 264 p.

AI Fringe. AI for everyone. // Perspectives from the AI Fringe. 30 October – 3 November 2023: URL: https://aifringe.org/

AI Safety Summit. The Bletchley Declaration by countries attending the AI safety summit Bletchley Park, UK. 1–2 November 2023: URL: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

Ansell C., Torfing J. (eds.) Handbook on theories of governance. Edward Elgar, 2022. 656 p.

Bacchi C. Policy as Discourse: What does it mean? Where does it get us? // Discourse: Studies in the Cultural Politics of Education. 2000. Vol. 21. No. 1. P. 45–57. DOI: 10.1080/01596300050005493.

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? // Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 2021. P. 610–623.

Bertuzzi L. EU's AI Act negotiations hit the brakes over foundation models. 10 November 2023: URL: https://www.euractiv.com/section/tech/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/

Browne J. AI and structural injustice: A feminist perspective // Feminist AI: Critical perspectives on algorithms, data, and intelligent machines / J. Browne, S. Cave, E. Drage, K. McInerney (eds.). Oxford University Press, 2023. P. 328–346.

Carlisle K., Gruby R.L. Polycentric systems of governance: A theoretical model for the commons // Policy Studies Journal. 2019. Vol. 47. No. 4. P. 927–952. DOI: 10.1111/psj.12212.

Center for AI Safety. Statement on AI risk. May 2023: URL: https://www.safe.ai/statement-on-ai-risk

Chhotray V., Stoker G. Governance theory and practice. A cross-disciplinary approach. Palgrave, 2009. DOI: 10.1057/9780230583344.

Cihon P., Maas M., Kemp L. Fragmentation and the future: Investigating architectures for international AI governance // Global Policy. 2020. Vol. 11. No. 5. P. 545–556. DOI: 10.1111/1758-5899.12890.

Clarke L. How Silicon Valley doomers are shaping Rishi Sunak's AI plans // Politico. 2023. 14 September.

Connected by Data et al. AI safety summit: Open letter to the UK Prime Minister. 2023.

Crawford K. Generative AI's environmental costs are soaring – and mostly secret // Nature. 2024. Vol. 626. No. 8000. P. 693. DOI: 10.1038/d41586-024-00478-x.

Crawford K. The atlas of AI. Yale University Press, 2021.

De Saille S. Innovating innovation policy: The emergence of 'Responsible Research and Innovation' // Journal of Responsible Innovation. 2015. Vol. 2. No. 2. P. 152–168. DOI: 10.1080/23299460.2015.1045280.

DSIT (Department for Science, Innovation and Technology). Capabilities and risks from frontier AI. A discussion paper on the need for further research into AI risk. AI Safety Summit, hosted by the UK. 2023. 1–2 November.

European Commission. Commission welcomes G7 leaders' agreement on Guiding Principles and a Code of Conduct on Artificial Intelligence. 2023. 30 October: URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5379

Filgueiras F., Almeida V. Governance for the digital world. Neither more state nor more market. Palgrave, 2021.

Fisher E., Smolka M., Owen R., Pansera M., Guston D.H., Grunwald A., Nelson J.P., Raman S., Neudert P., Flipse S.M., Ribeiro B. Responsible innovation scholarship: Normative, empirical, theoretical, and engaged // Journal of Responsible Innovation. 2024. Vol. 11. No. 1. P. 2309060. DOI: 10.1080/23299460.2024.2309060.

Future of Life Institute. Pause Giant AI Experiments: An Open Letter. 2023. 22 March: URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/

G7. G7 Hiroshima Leaders' Communique // G7. 2023a. 20 May: URL: http://www.g7.utoronto.ca/summit/2023hiroshima/230520-communique.html

G7. G7 Leaders' Statement on the Hiroshima Al Process // G7. 2023b. 30 October: URL: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai.html

G7. Hiroshima Process International Code of Conduct for Organizations Developing Advanced Al Systems // G7. 2023c. 30 October: URL: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai-code-of-conduct.html

G7. Hiroshima Process International Guiding Principles for Organizations Developing Advanced Al System // G7. 2023d. 30 October: URL: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai-principles.html

Galanos V. Exploring expanding expertise: Artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018 // Technology Analysis and Strategic Management. 2019. Vol. 31. No. 4. P. 421–432. DOI: 10.1080/09537325.2018.1518521.

Gebru T., Bender E., McMillan-Major A., Mitchell M. Statement from the listed authors of Stochastic Parrots on the "Al pause" letter. 2023. 31 March: URL: https://www.dair-institute.org/blog/letter-statement-March2023

Johnston S.F. Technological parables and iconic illustrations: American technocracy and the rhetoric of the technological fix // History and Technology. 2017. Vol. 33. No. 2. P. 196–219. DOI: 10.1080/07341512.2017.1336851.

Khanal S., Zhang H., Taeihagh A. Why and how is the power of Big Tech increasing in the policy process? The case of generative AI // Policy and Society. 2024. Vol. 44. No. 1. P. 52–69. DOI: 10.1093/polsoc/puae012.

Levi-Faur D. The Oxford handbook of governance. Oxford University Press, 2012.

McQuillan D. People's councils for ethical machine learning // Social Media+Society. 2018. Vol. 4. No. 2. P. 2056305118768303. DOI: 10.1177/2056305118768303.

Morozov E. The net delusion: How not to liberate the world. Penguin, 2011.

Mouriquand D. Christopher Noland warns that AI is reaching its ‘Oppenheimer moment'. Euronews. 2023. 18 July.

Mtigge D. EU AI sovereignty: For whom, to what end, and to whose benefit? // Journal of European Public Policy. 2024. P. 1–26. DOI: 10.1080/13501763.2024.2318475.

Noble S.U. Algorithms of oppression: How search engines reinforce racism. N. Y.: University Press, 2018.

OECD. AI Language Models: Technological, Socio-economic and Policy Considerations // OECD Digital Economy Papers. 2023a. April. No. 352.

OECD. G7 Hiroshima process on generative Artificial Intelligence (AI). Towards a G7 common understanding on Generative AI. Report prepared for the 2023 G7 presidency and the G7 digital and tech working group. OECD, 2023b. 7 September.

OECD. Initial policy considerations for generative Artificial Intelligence // OECD Artificial Intelligence papers. 2023c. September. No. 1.

OECD. Recommendation of the Council on Artificial Intelligence. OECD, 2019.

Ostrom E. Polycentric systems for coping with collective action and global environmental change // Global Environmental Change. 2010. Vol. 20. No. 4. P. 550–557. DOI: 10.1016/j.gloenvcha.2010.07.004.

Owen R., Macnaghten P., Stilgoe J. Responsible research and innovation: From science in society to science for society, with society // Science and Public Policy. 2012. Vol. 39. No. 6. P. 751–760. DOI: 10.1093/scipol/scs093.

Peters B.G. Governance as political theory // The Oxford handbook of governance / D. Levi-Faur (ed.). Oxford University Press, 2012. P. 19–32.

Pierre J., Peters B.G. Advanced introduction to governance. Edward Elgar, 2021.

Pringle E. One of A.I.'s 3 ‘godfathers' says he has regrets over his life's work. ‘You could say I feel lost'. Fortune. 2023. 23 May.

Radu R. Steering the governance of artificial intelligence: National strategies in perspective // Policy and Society. 2021. Vol. 40. No. 2. P. 178–193. DOI: 10.1080/14494035.2021.1929728.

Rein M., Schon D. Frame-critical policy analysis and frame-reflective policy practice // Knowledge and Policy. 1996. Vol. 9. No. 1. P. 85–104. DOI: 10.1007/BF02832235.

Roberts H., Hine E., Taddeo M., Floridi L. Global AI governance: Barriers and pathways forward // International Affairs. 2024. Vol. 100. No. 3. P. 1275–1286. DOI: 10.1093/ia/iiae073.

Rosner L. (ed.) The technological fix: How people use technology to create and solve problems. Routledge, 2004.

Schiff D. Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy // Review of Policy Research. 2023. Vol. 40. No. 5. P. 729–756. DOI: 10.1111/ropr.12535.

Schmitt L. Mapping global AI governance: A nascent regime in a fragmented landscape // AI and Ethics. 2022. Vol. 2. No. 2. P. 303–314. DOI: 10.1007/s43681-021-00083-y.

Schon D., Rein M. Frame reflection: Toward the resolution of intractable policy controversies. Basic Books, 1994.

Stilgoe J., Owen R., Macnaghten P. Developing a framework for responsible innovation // Research Policy. 2013. Vol. 42. No. 9. P. 1568–1580. DOI: 10.1016/j.respol.2013.05.008.

Sunak R. Prime minister's speech on AI. The Royal Society. 2023. 26 October.

Taeihagh A. Governance of artificial intelligence // Policy and Society. 2021. Vol. 40. No. 2. P. 137–157. DOI: 10.1080/14494035.2021.1928377.

Taylor J., Hern A. ‘Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation. Guardian. 2023. 2 May.

The White House. Remarks by Vice President Kamala Harris on the Future of Artificial Intelligence. 2023. 1 November: URL: https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/11/01/remarks-by-vice-president-harris-on-the-future-of-artificial-intelligence-london-united-kingdom/

UK Government. Introduction to the AI Safety Summit. 2023. 31 October: URL: https://www.gov.uk/government/publications/ai-safety-summit-introduction

Ulnicane I. Emerging technology for economic competitiveness or societal challenges? Framing purpose in Artificial Intelligence policy // Global Public Policy and Governance. 2022. Vol. 2. No. 3. P. 326–345. DOI: 10.1007/s43508-022-00049-8.

Ulnicane I., Erkkila T. Politics and policy of Artificial Intelligence // Review of Policy Research. 2023. Vol. 40. No. 5. P. 612–625. DOI: 10.1111/ropr.12574.

Ulnicane I., Knight W., Leach T., Stahl B.C., Wanjiku W.-G. Framing governance for a contested emerging technology: Insights from AI policy // Policy and Society. 2021. Vol. 40. No. 2. P. 158–177. DOI: 10.1080/14494035.2020.1855800.

Van Hulst M., Yanow D. From policy «frames» to «framing» theorizing a more dynamic, political approach // The American Review of Public Administration. 2016. Vol. 46. No. 1. P. 92–112. DOI: 10.1177/0275074014533142.

Veale M., Matus K., Gorwa R. AI and global governance: Modalities, rationales, tensions // Annual Review of Law and Social Science. 2023. Vol. 19. No. 1. P. 255–275. DOI: 10.1146/annurev-lawsocsci-020223-040749.

Vilakazi T., Roberts S. Cartels as ‘fraud'? Insights from collusion in southern and East Africa in the fertiliser and cement industries // Review of African Political Economy. 2019. Vol. 46. No. 161. P. 369–386. DOI: 10.1080/03056244.2018.1536974.

Von der Leyen U. State of the Union. Strasbourg. 2023. 13 September.

Weinberg A.M. Can technology replace social engineering? // Bulletin of the Atomic Scientists. 1966. Vol. 22. No. 10. P. 4–8. DOI: 10.1080/00963402.1966.11454993.

Whittaker M. The steep cost of capture // Interactions. 2021. Vol. 28. No. 6. P. 50–55. DOI: 10.1145/3488666.

Wong J.C. More than 1,200 Google workers condemn firing of AI scientist Timnit Gebru. The Guardian. 2020. 4 December.

Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books, 2019.

DOI: 10.38050/2078-3809-2025-17-3-123-148

Innovation and Development

Governance fix? Power and politics in controversies about governing generative AI

Inga Ulnicane
University of Birmingham, Edgbaston,
(Birmingham B15 2TT, United Kingdom)

Autor of translation:
Maria Yu. Beletskaya
Candidate of Economic Science, Senior Researcher
Lomonosov Moscow State University, Faculty of Economics
(Moscow, Russia)

Abstract

The launch of ChatGPT in late 2022 led to major controversies about the governance of generative artificial intelligence (AI). This article examines the first international governance and policy initiatives dedicated specifically to generative AI: the G7 Hiroshima process, the Organisation for Economic Cooperation and Development reports, and the UK AI Safety Summit. This analysis is informed by policy framing and governance literature, in particular by the work on technology governance and Responsible Innovation. Emerging governance of generative AI exhibits characteristics of polycentric governance, where multiple and overlapping centers of decision-making are in collaborative relationships. However, it is dominated by a limited number of developed countries. The governance of generative AI is mostly framed in terms of the risk management, largely neglecting issues of purpose and direction of innovation, and assigning rather limited roles to the public. We can see a "paradox of generative AI governance" emerging, namely, that while this technology is being widely used by the public, its governance is rather narrow. This article coins the term "governance fix" to capture this rather narrow and technocratic approach to governing generative AI. As an alternative, it suggests embracing the politics of polycentric governance and Responsible Innovation that highlight democratic and participatory coshaping of technology for social benefit. In the context of the highly unequal distribution of power in generative AI characterized by a high concentration of power in a small number of large tech companies, the government has a special role in reshaping the power imbalances by enabling wide-ranging public participation in the governance of generative AI.

Keywords: generative AI, governance, artificial intelligence, responsible innovation, risk.

JEL: Z18.

For citation: Ulnicane, I. (2025) Governance fix? Power and politics in controversies about governing generative AI (trans. from Eng. Beletskaya, M. Yu.). Scientific Research of Faculty of Economics. Electronic Journal, vol. 17, no. 3, pp. 123-148. DOI: 10.38050/2078-3809-2025-17-3-123-148.

References

Aguerre C., Campbell-Verduyn M., Scholte J.A. Global digital data governance: Polycentric perspectives. Routledge, 2024. 264 p.

AI Fringe. AI for everyone. Perspectives from the AI Fringe. 30 October – 3 November 2023: Available at: https://aifringe.org/

AI Safety Summit. The Bletchley Declaration by countries attending the AI safety summit Bletchley Park, UK. 1–2 November 2023: Available at: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

Ansell C., Torfing J. (eds.) Handbook on theories of governance. Edward Elgar, 2022. 656 p.

Bacchi C. Policy as Discourse: What does it mean? Where does it get us? Discourse: Studies in the Cultural Politics of Education. 2000. Vol. 21. No. 1. P. 45–57. DOI: 10.1080/01596300050005493.

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 2021. P. 610–623.

Bertuzzi L. EU's AI Act negotiations hit the brakes over foundation models. 10 November 2023: Available at: https://www.euractiv.com/section/tech/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/

Browne J. AI and structural injustice: A feminist perspective. Feminist AI: Critical perspectives on algorithms, data, and intelligent machines / J. Browne, S. Cave, E. Drage, K. McInerney (eds.). Oxford University Press, 2023. P. 328–346.

Carlisle K., Gruby R.L. Polycentric systems of governance: A theoretical model for the commons. Policy Studies Journal. 2019. Vol. 47. No. 4. P. 927–952. DOI: 10.1111/psj.12212.

Center for AI Safety. Statement on AI risk. May 2023: Available at: https://www.safe.ai/statement-on-ai-risk

Chhotray V., Stoker G. Governance theory and practice. A cross-disciplinary approach. Palgrave, 2009. DOI: 10.1057/9780230583344.

Cihon P., Maas M., Kemp L. Fragmentation and the future: Investigating architectures for international AI governance. Global Policy. 2020. Vol. 11. No. 5. P. 545–556. DOI: 10.1111/1758-5899.12890.

Clarke L. How Silicon Valley doomers are shaping Rishi Sunak's AI plans. Politico. 2023. 14 September.

Connected by Data et al. AI safety summit: Open letter to the UK Prime Minister. 2023.

Crawford K. Generative AI's environmental costs are soaring – and mostly secret. Nature. 2024. Vol. 626. No. 8000. P. 693. DOI: 10.1038/d41586-024-00478-x.

Crawford K. The atlas of AI. Yale University Press, 2021.

De Saille S. Innovating innovation policy: The emergence of ‘Responsible Research and Innovation’. Journal of Responsible Innovation. 2015. Vol. 2. No. 2. P. 152–168. DOI: 10.1080/23299460.2015.1045280.

DSIT (Department for Science, Innovation and Technology). Capabilities and risks from frontier AI. A discussion paper on the need for further research into AI risk. AI Safety Summit, hosted by the UK. 2023. 1–2 November.

European Commission. Commission welcomes G7 leaders' agreement on Guiding Principles and a Code of Conduct on Artificial Intelligence. 2023. 30 October: Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5379

Filgueiras F., Almeida V. Governance for the digital world. Neither more state nor more market. Palgrave, 2021.

Fisher E., Smolka M., Owen R., Pansera M., Guston D.H., Grunwald A., Nelson J.P., Raman S., Neudert P., Flipse S.M., Ribeiro B. Responsible innovation scholarship: Normative, empirical, theoretical, and engaged. Journal of Responsible Innovation. 2024. Vol. 11. No. 1. P. 2309060. DOI: 10.1080/23299460.2024.2309060.

Future of Life Institute. Pause Giant AI Experiments: An Open Letter. 2023. 22 March: Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/

G7. G7 Hiroshima Leaders' Communique. G7. 2023a. 20 May: Available at: http://www.g7.utoronto.ca/summit/2023hiroshima/230520-communique.html

G7. G7 Leaders' Statement on the Hiroshima Al Process. G7. 2023b. 30 October: Available at: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai.html

G7. Hiroshima Process International Code of Conduct for Organizations Developing Advanced Al Systems. G7. 2023c. 30 October: Available at: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai-code-of-conduct.html

G7. Hiroshima Process International Guiding Principles for Organizations Developing Advanced Al System. G7. 2023d. 30 October: Available at: http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai-principles.html

Galanos V. Exploring expanding expertise: Artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. Technology Analysis and Strategic Management. 2019. Vol. 31. No. 4. P. 421–432. DOI: 10.1080/09537325.2018.1518521.

Gebru T., Bender E., McMillan-Major A., Mitchell M. Statement from the listed authors of Stochastic Parrots on the "Al pause" letter. 2023. 31 March: Available at: https://www.dair-institute.org/blog/letter-statement-March2023

Johnston S.F. Technological parables and iconic illustrations: American technocracy and the rhetoric of the technological fix. History and Technology. 2017. Vol. 33. No. 2. P. 196–219. DOI: 10.1080/07341512.2017.1336851.

Khanal S., Zhang H., Taeihagh A. Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. Policy and Society. 2024. Vol. 44. No. 1. P. 52–69. DOI: 10.1093/polsoc/puae012.

Levi-Faur D. The Oxford handbook of governance. Oxford University Press, 2012.

McQuillan D. People's councils for ethical machine learning. Social Media+Society. 2018. Vol. 4. No. 2. P. 2056305118768303. DOI: 10.1177/2056305118768303.

Morozov E. The net delusion: How not to liberate the world. Penguin, 2011.

Mouriquand D. Christopher Noland warns that AI is reaching its ‘Oppenheimer moment'. Euronews. 2023. 18 July.

Mtigge D. EU AI sovereignty: For whom, to what end, and to whose benefit? Journal of European Public Policy. 2024. P. 1–26. DOI: 10.1080/13501763.2024.2318475.

Noble S.U. Algorithms of oppression: How search engines reinforce racism. N. Y.: University Press, 2018.

OECD. AI Language Models: Technological, Socio-economic and Policy Considerations. OECD Digital Economy Papers. 2023a. April. No. 352.

OECD. G7 Hiroshima process on generative Artificial Intelligence (AI). Towards a G7 common understanding on Generative AI. Report prepared for the 2023 G7 presidency and the G7 digital and tech working group. OECD, 2023b. 7 September.

OECD. Initial policy considerations for generative Artificial Intelligence. OECD Artificial Intelligence papers. 2023c. September. No. 1.

OECD. Recommendation of the Council on Artificial Intelligence. OECD, 2019.

Ostrom E. Polycentric systems for coping with collective action and global environmental change. Global Environmental Change. 2010. Vol. 20. No. 4. P. 550–557. DOI: 10.1016/j.gloenvcha.2010.07.004.

Owen R., Macnaghten P., Stilgoe J. Responsible research and innovation: From science in society to science for society, with society. Science and Public Policy. 2012. Vol. 39. No. 6. P. 751–760. DOI: 10.1093/scipol/scs093.

Peters B.G. Governance as political theory. The Oxford handbook of governance / D. Levi-Faur (ed.). Oxford University Press, 2012. P. 19–32.

Pierre J., Peters B.G. Advanced introduction to governance. Edward Elgar, 2021.

Pringle E. One of A.I.'s 3 ‘godfathers' says he has regrets over his life's work. ‘You could say I feel lost'. Fortune. 2023. 23 May.

Radu R. Steering the governance of artificial intelligence: National strategies in perspective.Policy and Society. 2021. Vol. 40. No. 2. P. 178–193. DOI: 10.1080/14494035.2021.1929728.

Rein M., Schon D. Frame-critical policy analysis and frame-reflective policy practice. Knowledge and Policy. 1996. Vol. 9. No. 1. P. 85–104. DOI: 10.1007/BF02832235.

Roberts H., Hine E., Taddeo M., Floridi L. Global AI governance: Barriers and pathways forward. International Affairs. 2024. Vol. 100. No. 3. P. 1275–1286. DOI: 10.1093/ia/iiae073.

Rosner L. (ed.) The technological fix: How people use technology to create and solve problems. Routledge, 2004.

Schiff D. Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy. Review of Policy Research. 2023. Vol. 40. No. 5. P. 729–756. DOI: 10.1111/ropr.12535.

Schmitt L. Mapping global AI governance: A nascent regime in a fragmented landscape. AI and Ethics. 2022. Vol. 2. No. 2. P. 303–314. DOI: 10.1007/s43681-021-00083-y.

Schon D., Rein M. Frame reflection: Toward the resolution of intractable policy controversies. Basic Books, 1994.

Stilgoe J., Owen R., Macnaghten P. Developing a framework for responsible innovation. Research Policy. 2013. Vol. 42. No. 9. P. 1568–1580. DOI: 10.1016/j.respol.2013.05.008.

Sunak R. Prime minister's speech on AI. The Royal Society. 2023. 26 October.

Taeihagh A. Governance of artificial intelligence. Policy and Society. 2021. Vol. 40. No. 2. P. 137–157. DOI: 10.1080/14494035.2021.1928377.

Taylor J., Hern A. ‘Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation. Guardian. 2023. 2 May.

The White House. Remarks by Vice President Kamala Harris on the Future of Artificial Intelligence. 2023. 1 November: Available at: https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/11/01/remarks-by-vice-president-harris-on-the-future-of-artificial-intelligence-london-united-kingdom/

UK Government. Introduction to the AI Safety Summit. 2023. 31 October: Available at: https://www.gov.uk/government/publications/ai-safety-summit-introduction

Ulnicane I. Emerging technology for economic competitiveness or societal challenges? Framing purpose in Artificial Intelligence policy. Global Public Policy and Governance. 2022. Vol. 2. No. 3. P. 326–345. DOI: 10.1007/s43508-022-00049-8.

Ulnicane I., Erkkila T. Politics and policy of Artificial Intelligence. Review of Policy Research. 2023. Vol. 40. No. 5. P. 612–625. DOI: 10.1111/ropr.12574.

Ulnicane I., Knight W., Leach T., Stahl B.C., Wanjiku W.-G. Framing governance for a contested emerging technology: Insights from AI policy. Policy and Society. 2021. Vol. 40. No. 2. P. 158–177. DOI: 10.1080/14494035.2020.1855800.

Van Hulst M., Yanow D. From policy «frames» to «framing» theorizing a more dynamic, political approach. The American Review of Public Administration. 2016. Vol. 46. No. 1. P. 92–112. DOI: 10.1177/0275074014533142.

Veale M., Matus K., Gorwa R. AI and global governance: Modalities, rationales, tensions. Annual Review of Law and Social Science. 2023. Vol. 19. No. 1. P. 255–275. DOI: 10.1146/annurev-lawsocsci-020223-040749.

Vilakazi T., Roberts S. Cartels as ‘fraud'? Insights from collusion in southern and East Africa in the fertiliser and cement industries. Review of African Political Economy. 2019. Vol. 46. No. 161. P. 369–386. DOI: 10.1080/03056244.2018.1536974.

Von der Leyen U. State of the Union. Strasbourg. 2023. 13 September.

Weinberg A.M. Can technology replace social engineering? Bulletin of the Atomic Scientists. 1966. Vol. 22. No. 10. P. 4–8. DOI: 10.1080/00963402.1966.11454993.

Whittaker M. The steep cost of capture. Interactions. 2021. Vol. 28. No. 6. P. 50–55. DOI: 10.1145/3488666.

Wong J.C. More than 1,200 Google workers condemn firing of AI scientist Timnit Gebru. The Guardian. 2020. 4 December.

Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books, 2019.