УДК: 004.932, 517.5

1.5. Экспериментальные и модельные исследования влияния ИИ на эволюцию коллективного сознания

Волкова А.Д., Грачев И.Д., Костина Т.А., Ларин С.Н., Ноакк Н.В.,
ЦЭМИ РАН, Москва, Россия

Статья продолжает работы авторов по изучению влияния искусственного интеллекта на общественное сознание. Целью настоящего исследования является разработка системного инструментария для обеспечения управленческих решений, позволяющего минимизировать риски негативного воздействия ИИ на цивилизационные особенности России. Для ее достижения авторами начато систематизированное экспериментальное исследование смещения, сжатия, манипулятивности ИИ с учетом факторов индивидуализма – коллективизма и макиавеллизма. Сравнивались контрольные группы людей с группами «личностей», сгенерированных ИИ. Результаты получены на основе небольших выборок, характерных для начальной стадии экспериментов. На орте индивидуализм-коллективизм не имеет места смещение оценок, наследованных ИИ, с бесспорным для любых принятых уровней значимости сжатием вариативности. Напротив, при оценке на макиавеллизм (меру манипулятивности) смещение оценок, наследованных ИИ, значимо отличаются от оценок контрольных групп людей. Полученные результаты позволяют сделать вывод о необходимости дальнейших исследований оценок вариативности по параметру макиавеллизма.

Введение

В наших работах [Грачёв и др., 2024а; Грачёв, Ноакк, 2024в] мы начали систематизированное моделирование воздействия генеративных ИИ на социально-экономический прогресс. Так, в работе [Грачёв и др., 2024а] показано, что для удовлетворяющих фишеровским эволюционным предположениям ошибок оценивания и действий агентов ИИ, сжимающий разнообразие, ведет к замедлению темпов роста. В работе [Грачёв, Ноакк, 2024в] смоделирована принципиально иная ситуация заведомо несимметричных ошибок оценивания. Показано оптимальное сжатие разнообразия, синхронизированное с хаосом окружающей среды. В совокупности эти работы позволяют считать необходимым дифференцированное отношение к возможному сжатию разнообразия и смещению оценок людей генеративным ИИ.

В последнее время появились отдельные публикации по воздействию ИИ на контрольные группы людей по некоторым произвольным факторам. Так, в [Is Xi Jinping an AI doomer, 2024] анализируется практический опыт Китая по оценке рисков и борьбе с негативным воздействием ИИ на общественное сознание. Наиболее интересным для настоящей статьи является утверждение о рекомендациях регулярного тестирования ИИ на соответствие «социалистическим ценностям». С точностью до терминологии для нас это означает требование тестирования ИИ на предмет случайного или целенаправленного изменения базисных ценностей российской цивилизационной модели, смещение и сжатие её к западной либерал-фундаменталистской.

В научной литературе появились работы [Волкова, Костина, Ноакк, 2023], в которых влияние на группы граждан оценивается экспериментально. Однако пока эти работы не выглядят системно, хотя их общий обзор позволяет заключить, что на языке математики проблема может быть сведена к смещению оценок и действий, сжатию их вариативности и преднамеренной или спонтанной манипулятивности ИИ. При этом очевидно, что невозможно эффективное тестирование по всему бесчисленному множеству отличий. Должен быть выбран базисный набор действительно фундаментальных отличий. Пока в этом направлении нет по-настоящему систематизированных исследований.

Мы полагаем целесообразным начать исследование, опираясь на два базисных орта:

1) *индивидуализм-коллективизм* (далее – И-К), поскольку он соответствует аксиоме «о мере хаоса-порядка» как главной для живых систем. Этой же аксиомы придерживаются авторы статьи в своем исследовании; 2) *манипулятивность* (макиавеллизм).

С этих двух, несомненно, базисных ортов мы начинаем систематизированные экспериментальные исследования воздействия существующих ИИ на коллективное сознание. За неимением сегодня других известных специализированных тестов мы используем хорошо отработанные «людские» тесты с применением к результатам стандартных методов оценок. Это не исключает последующей корректировки методов с учётом особенностей ИИ.

Обзор литературы

В обзоре представлены материалы исследований, которые систематизированы в двух направлениях.

1. Вопросы возможного сжатия разнообразия и смещения оценок влияния генеративного ИИ на поведение людей пока не получили достаточного освещения в трудах российских и зарубежных ученых. Вместе с тем, сегодня имеется достаточно много исследований, направленных на изучение влияния ИИ на манипуляцию информацией, как случайную, так и/или системную [Кленк, 2022; Фараони, 2023]. В 2021 году Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) упомянула манипуляцию и связала ее со злоупотреблением когнитивными предубеждениями в своей Рекомендации по этике искусственного интеллекта [ЮНЕСКО, 2021, статья 125]. В работе [Jakesch, 2023] на основании полученных в ходе

проведения опросов достаточно представительных групп людей (1506 и 500 человек) результатов установлено, что большие языковые модели, такие как GPT-3, формируют определенную точку зрения и могут оказывать влияние на мнение людей. Авторами выделено три вида влияния: информационное, нормативное, поведенческое. Такое влияние может быть скрытым и трудно определяемым: архитектуры выбора видны, но предпочтения мнений, встроенные в языковые модели, могут быть непрозрачны для пользователей и даже разработчиков систем. Кроме того, языковые модели могут влиять на убеждения случайно, когда формируемые ими мнения могут различаться в зависимости от пользователя, продукта и контекста. С помощью межсубъектного экспериментального исследования авторы работы [Mieczkowski, 2021] изучили, как коммуникации, опосредованные ИИ (AI-MC), влияют на основные аспекты человеческого общения, такие как межличностное восприятие и выполнение задач. Было установлено, что ИИ является активной и динамичной сущностью, которая может изменить нормы и динамику человеческого общения. «Умные ответы Google» интегрируется в текстовую коммуникационную задачу и влияют на 1) языковые шаблоны, используемые собеседниками как в языке, сгенерированном ИИ, так и в языке, сгенерированном человеком, 2) восприятие собеседника и 3) производительность собеседников в текстовой коммуникационной задаче. Авторы обнаружили, что в ходе взаимодействия с ИИ создаётся новая форма сообщения. Они указывают исследователям на необходимость учитывать потенциальные различия между языком ИИ и человеческим языком. Основные выводы этой статьи содержат предварительные, но неоднозначные доказательства, позволяющие предположить, что язык, сгенерированный ИИ, может подрывать некоторые аспекты межличностного восприятия, такие как социальное притяжение.

В другой работе [Jahanbakhsh, 2021] авторы исследовали возможности изменения платформ социальных сетей таким образом, чтобы пользователи учитывали точность контента при обмене сообщениями. Некоторые исследования показывают, что ИИ может использовать человеческую эвристику и предвзятость, чтобы тонко манипулировать решениями людей. С этой целью авторы работы [Agudo, Matute, 2021] провели эмпирическую проверку возможности влияния алгоритмов ИИ на предпочтения людей посредством явного или скрытого убеждения в различных контекстах. Авторы установили, что скорость и объемы исследований особенностей влияния ИИ на поведение больших групп людей, проводимых академическими учеными, в разы меньше аналогичных показателей компаний, занимающихся ИИ. Последние работают с такими огромными выборками, которые недоступны научным работникам. Наиболее распространенные алгоритмы ИИ формируются не в результате научных исследований, а в целях достижения конкретных частных интересов этих компаний. Следовательно, способность компаний, занимающихся ИИ, влиять на решения больших групп людей или манипулировать их поведением как явно, так и скрыто, безусловно, намного выше

В общем и целом, обобщая полученные исследователями результаты, можно сделать вывод о том, что проблема манипулятивности ИИ существует и для ее решения необходимо проведение новых масштабных исследований. Вместе с тем сама манипулятивность понимается и трактуется по-разному. Авторы склонны рассматривать как непреднамеренную, так и преднамеренную манипулятивность. Существующие исследования пока не дают ясного понимания объемов и степени манипулирования информацией, которыми располагают современные и будущие системы и алгоритмы ИИ. Мы полагаем, что на языке математики проблема может быть сведена к смещению оценок и действий, сжатию их вариативности и преднамеренной или непреднамеренной манипулятивности ИИ.

2. Вопросы исследования влияния ИИ на случайное или целенаправленное изменение базисных ценностей российской цивилизационной модели, смещение и сжатие её к западной либерал-фундаменталистской освещаются во многих работах с чисто теоретической стороны [Пантин, 2021; Савин, 2019; Литвинов, 2021; Семенова, 2023]. При этом используется достаточно широкий спектр отличий. Очевидно, что в таких условиях эффективное тестирование невозможно. Должен быть выбран базисный набор действительно фундаментальных отличий. В общем виде он закреплен в Указе Президента России № 809 от 9 ноября 2022 года, которым утверждены «Основы государственной политики по сохранению и укреплению традиционных российских духовно-нравственных ценностей». Однако в этом документе представлено описательное обоснование ценностей российской цивилизации. В целях данного исследования интерес вызывают формализованные оценки наиболее значимых ценностей. Такое исследование было выполнено в работе [Ценности, 2023]. Авторы использовали методику Г. Хофстеде для количественного анализа ценностей России и стран Запада. Сравнение было проведено по 5-ти основным параметрам: индивидуальность, дистанция от власти, долгосрочная ориентация, определенность будущего, маскулинность. В результате сравнения со странами Запада (США, Великобритания, Франция, Германия) и странами востока (Китай, Индия, Япония, Пакистан) был выявлен определенно особый аксиологический тип России. Главное ценностное противоречие выявлено в рамках оппозиции индивидуализма и коллективизма.

Других количественно систематизированных исследований в этой области нами не найдено. В нашем исследовании используется шкала *индивидуализма-коллективизма.* Она, безусловно, является одним из базисных отличий российской цивилизационной модели от либерал-фундаменталистской западной.

Основная часть проблемы воздействия ИИ на сознание индивидов, коллективов людей и более широко – на общественное сознание, может быть, во втором приближении сведена к смещению оценок и действий, их вариативности. Отдельный интерес представляет случайная или системная манипулятивность ИИ. В нашей работе [Грачёв, Ноакк, 2024в] показана значимость для прогресса реально наблюдаемого сжатия разнообразия ИИ оценок и действий агентов для симметричных отношений. Характерные результаты представлены на Рис. 1 из [Грачёв, Ноакк, Костина, 2024а], который мы считаем целесообразным воспроизвести для иллюстрации значимости проблемы.

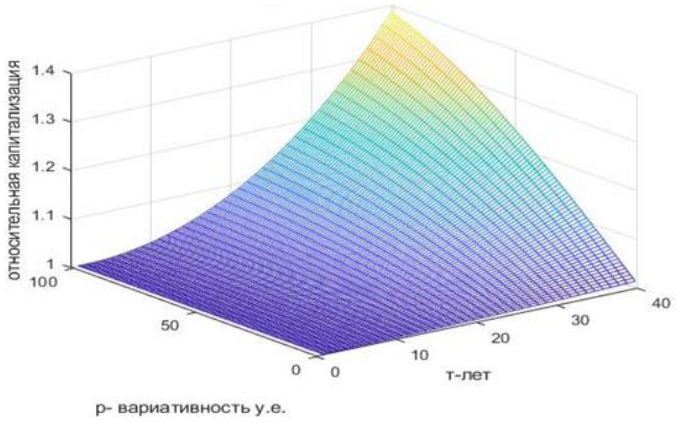

Рис. 1 Эффект вариативности (составлено авторами)

На рис 2 из [Грачёв, Ноакк, 2024в] представлены характерные результаты возможного влияния ИИ по факторам, заведомо несимметричным к ошибкам оценивания и действиям.

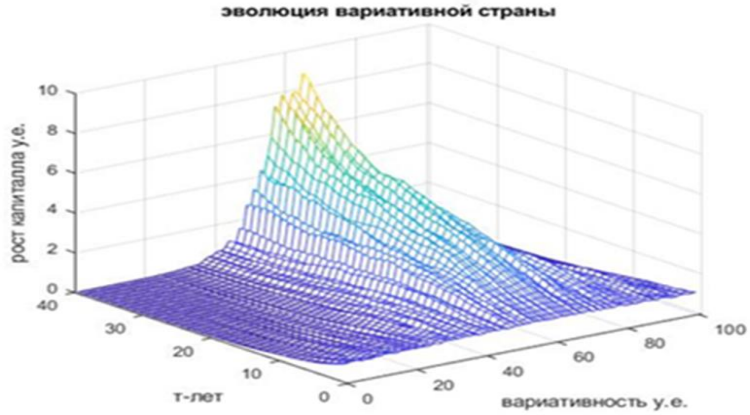

Рис. 2 Эволюция вариативной системы (составлено авторами)

Рис. 1 и 2 иллюстрируют необходимость дифференцированного подхода к воздействию ИИ на общественное сознание по разным факторам.

Учитывая невозможность эффективного оценивания манипулятивного воздействия ИИ по всему бесчисленному множеству факторов, определяющих особенности российской цивилизационной модели, необходимо операционно определиться с их базисным набором. Полагаем, что главной мерой живых систем является соотношение «хаоса и порядка». Мы начинаем со шкалы *индивидуализм-коллективизм,* которая и отражает хаос-порядок в общественном сознании и безусловно является одним из базисных отличий российской цивилизационной модели от либерал-фундаменталистской западной [Ценности, 2023].

Общий для всех последующих исследований интерес представляет оценка манипулятивности конкретных ИИ. Поскольку на данный момент нет общего базиса и особенных методик тестирования системы ИИ-Люди, мы сочли возможным применить для оценки в шкалах И-К и М методики, хорошо отработанные на людях, рассматривая «псевдоличности», сгенерированные по фактору ИИ, как группу «испытуемых личностей» и сравнивая их с контрольной группой реальных людей. Это позволило применить все хорошо отработанные в социальной психологии методики.

Материалы и методы

Для проведения исследования были отобраны 2 методики: обладающие удовлетворительными показателями надежности и валидности: методика диагностики склонности к манипуляции (макиавеллизму – далее - М) - русскоязычная версия шкалы Mach-4.); методика оценки воспринимаемой культуры сообщества на основе культурной ориентации «горизонтальный/вертикальный индивидуализм – коллективизм» (далее – И-К).

Опросник «Горизонтальный/вертикальный индивидуализм – коллективизм» (Галлямова, Григорьев, 2022). Методика призвана диагностировать склонность людей к выбору базовых культурных ценностей индивидуализма и/или коллективизма. Дополнительные характеристики горизонтальный-вертикальный [см. Singelis et al., 1995] введены для конкретизации базовых понятий с точки зрения ценности равенства, с одной стороны (горизонтальная ось) и иерархии, с другой (вертикальная ось). Адаптированный к российскому контексту вариант методики приведён в [Галлямова, Григорьев, 2022]. Методика состоит из 16 утверждений /пунктов (по 4 пункта на каждую ось (горизонтальный индивидуализм, вертикальный индивидуализм, горизонтальный коллективизм, вертикальный коллективизм)

Методика оценки склонности к манипуляции [Знаков, 2005]. Методика диагностирует склонность людей к манипуляции (другой термин – макиавеллизму) в поведении и общении. Главными

психологическими составляющими макиавеллизма как свойства личности являются: 1) убеждение субъекта в том, что при общении с другими людьми ими можно и даже нужно манипулировать; 2) навыки, конкретные умения манипуляции. Последние включают способность убеждать других, понимать их намерения и причины поступков [Studies in Machiavellianism, 1970].

Опросник состоит из 20 утверждений. Испытуемый должен выразить меру своего согласия или несогласия с каждым из 20 утверждений по семибалльной шкале - от «Полностью согласен» (7 баллов) до «Совершенно не согласен» (1 балл).

Опрос проводился через сервис Анкетолог https://anketolog.ru/. Ссылка на анкету: https://anketolog.ru/s/849940/LQ7fRY4K

Опрос «псевдоличностей» ИИ проводился авторами письменно, были предъявлены оба опросника последовательно, двум ИИ: GPT-4 и ИИСбера (Гига-чат).

Результаты ИИ и людей были сначала обработаны стандартным образом с использованием ключей по каждой из методик, затем были применены статистические методы обработки: использованы критерий Манна-Уитни, U-критерий Манна - Уитни. Выбор данных критериев был обусловлен возможностью с их помощью выявлять различия в значении параметра между малыми выборками. Полученные результаты сравнения значений выбранных параметров между ИИ и людьми представлены в Рисунках 3-6.

Результаты

По параметру *Макиавеллизма* есть значимые различия (Рис. 3-4) между GPT-4 и людьми, а также между ИИСбера и людьми.

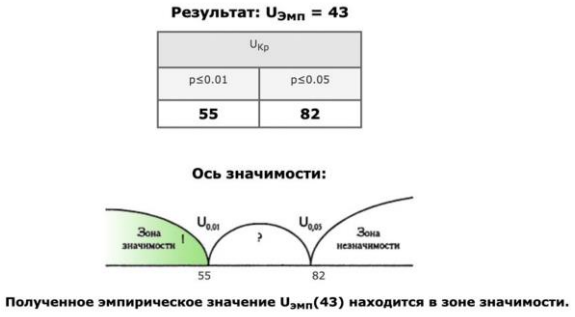

Рисунок 3. Сравнение ответов GPT-4 и Люди по Макиавеллизму.

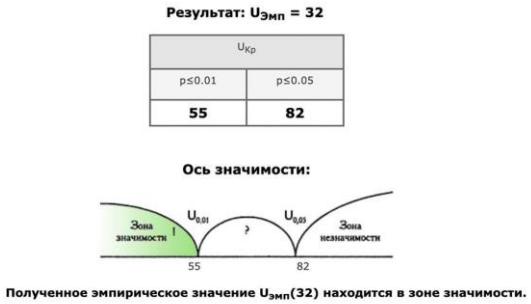

Рисунок 4. Сравнение ответов ИИ Сбера (*Гига-чат)* и *Люди* по *Макиавеллизму.*

Для значений по переменным *Индивидуализм-Коллективизм* асимптотическая значимость оказалась больше 0,05. Это значит, что значимых различий по данным переменным между группами ИИ (GPT-4 и ИИСбер) и Людей не обнаружено (Рис. 3-6). Для иллюстрации ниже приведены рисунки 5 и 6, на которых обозначены результаты сравнения ответов GPT-4 и ИИ Сбера и Людей по переменной *индивидуализма*. Аналогичное сравнение результатов GPT-4 и ИИ Сбера, с одной стороны, и Людей, с другой, по переменной *коллективизма* показало, что полученные значения (U-эмп. =80 и U-эмп. = 88, соответственно) лежат в диапазоне незначимости.

Значимость явного *Макиавеллизма* ИИ, независимо от того, является ли она случайной или системной, безусловно представляется опасной. К сожалению, незначимость отклонения средних по шкале *И-К* пока не допускает однозначной трактовки, так как «человеческая» методология работы по этой шкале предполагает процедуры центрирования данных. Вероятно, следует подумать над адаптацией методики

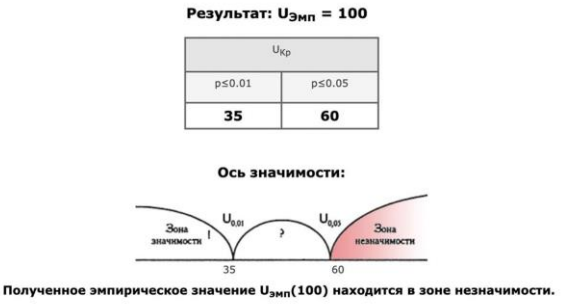

Рисунок 5. Сравнение ответов GPT-4 и Люди по Индивидуализму

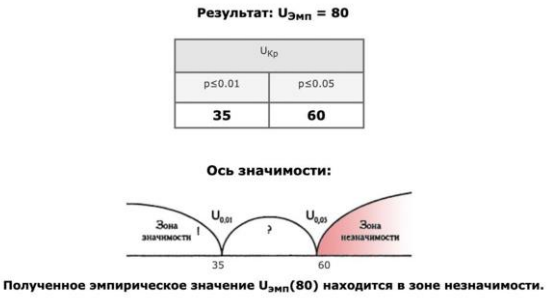

Рисунок 6. Сравнение ответов *ИИ Сбер* и *Люди* по *Индивидуализму*

для ИИ.

Учитывая довольно ограниченные объемы выборки по ИИ-псевдоличностям, для оценки сжатия разнообразия мы использовали стандартный критерий Фишера со стандартным уровнем значимости

0,05. Для данных пар выборок он примерно равен 2,5. При этом для ИИ-1 по шкале М мы получим значение F1=2,7 и для ИИ-2, F2=1,5, что означает отсутствие значимого сжатия разнообразия.

По шкале И-К оценки вполне определённые: F1=5, F2= 25, то есть мы имеем явное значимое сжатие разнообразия оценок и действий. Учитывая необходимость постоянной оптимизации соотношения хаоса-порядка в быстроменяющихся внешних условиях (см. рис. 2), это сжатие представляется опасным.

Заключение

Риски развития ИИ для человечества продолжают оставаться одной из главных тем дискуссий как для практиков, так и для теоретиков. В наших предыдущих работах рассматривалось негативное воздействие сжатия вариативности для ситуаций симметричных «плохих» и «хороших» отклонений, анализировались особенности также встречающейся в жизни ситуации заведомо несимметричных ошибок. В совокупности эти работы позволяют считать необходимым дифференцированный подход к оценке влияния ИИ на общественное сознание по разным факторам. В настоящей работе мы начали систематизированные экспериментальные исследования воздействия существующих ИИ на коллективное сознание по двум различительным факторам - *индивидуализма – коллективизма и манипулятивности.* Результаты получены на основе небольших выборок, характерных для начальной стадии экспериментов. На орте *индивидуализм-коллективизм* не имеет места смещение оценок, наследованных ИИ, с бесспорным для любых принятых уровней значимости сжатием вариативности. Напротив, при оценке ИИ на *макиавеллизм* (меру манипулятивности) ИИ значимо отличаются от людей по смещению. Оценка вариативности по параметру макиавеллизма (далее – М) требует дальнейших исследований.

Литература

1. Волкова А.Д., Костина Т.А., Ноакк Н.В. Социальные представления об искусственном интеллекте: методологические аспекты (Часть 2)// Цифровая экономика, 2023, № 26(5). - Стр. 18-28.
2. Галлямова А.А., Григорьев Д.С. Разработка методики оценки воспринимаемой культуры сообщества на основе культурной ориентации «горизонтальный/вертикальный индивидуализм – коллективизм» Г. Триандиса / Вестник РУДН. Серия: Психология и педагогика. 2022. №3. С. 429-447.
3. Грачёв И.Д., Ноакк Н.В. Оптимизация разнообразия агентов: размышления, гипотезы, прогнозы // Цифровая экономика., № 2 (28), 2024в. Июнь, стр. 57-60.
4. Грачёв И.Д., Ноакк Н.В., Костина Т.А. Амбивалентность социальных представлений об ИИ: психология, статистика, прогнозы // Цифровая экономика, № 1(27), 2024а. Стр.54-61.
5. Знаков В.В. Психология понимания: Проблемы и перспективы. – М.: Изд-во «Институт психологии РАН», 2005. – 448 с.
6. Кленк, М. (2022). (Онлайн) манипуляция: иногда скрытая, всегда небрежная. Rev. Soc. Экономика. 80, 85-105. doi: 10.1080/00346764.2021. 1894350
7. Литвинов В.Ю., Матвеева Л.В. Сравнительный анализ культурных представлений творческой молодежи о российской, западной и восточной цивилизациях // Социальная психология и общество. 2021. Том 12. № 1. С. 177-197. DOI: https://doi.org/10.17759/sps.2021120112.
8. Пантин В.И. Цивилизационные особенности развития России в контексте современных социальных трансформаций // Вестник Института социологии. 2021. Том 12. № 4. С. 108-124. DOI: 10.19181/vis.2021.12.4.754.
9. Савин С.Д., Касабуцкая М.С. Общенациональные российские ценности в контексте формирования коллективной идентичности // Вестник Санкт-Петербургского университета. Социология. 2019. Т. 12. Вып. 1. С. 82-97. https://doi.org/10.21638/spbu12.2019.106.
10. Семенова Д.М., Афонин М.В., Кудрявцев С.А. Ценности в структуре гражданской идентичности: понятие и инструменты // Политконсультант, 2023. Т. 3. № 1. URL: https://politicjournal.ru/PDF/04PK123.pdf.
11. Фараони С. (2023) Технология убеждения и вычислительные манипуляции: чрезмерное игнорирование ментального самоопределения. Фронт. Искусственно. Интеллект. 6:1216340. doi: 10.3389/frai.2023.1216340.
12. Ценности российской цивилизации: методическое пособие для вузов. / В.Э. Багдасарян, Ю.Ю. Иерусалимский. Ярославль: ИПК «Индиго», 2023. – 80с.
13. ЮНЕСКО. (2021). «Рекомендация ЮНЕСКО по этике искусственного интеллекта» (23 ноября 2021). SHS/BIO/PI/2021/1. Париж: ЮНЕСКО.
14. Agudo U, Matute H (2021) The influence of algorithms on political and dating decisions. PLoS ONE 16(4): e0249454. https://doi.org/10.1371/journal.pone.0249454
15. Jahanbakhsh F., Zhang A. X., Berinsky A. J., Pennycook G.,. Rand D G, and David R. Karger. 2021. Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 18 (April 2021), 42 pages. https://doi.org/10.1145/3449092.
16. Mieczkowski H., Hancock J.T., Naaman M., Jung M., and Hohenstein J. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 17 (April 2021), 14 pages. https://doi.org/10.1145/3449091.

17. Is Xi Jinping an AI doomer? China's elite is split over artificial intelligence//The Economist 25.08. 2024. https://www.economist.com/china/2024/08/25/is-xi-jinping-an-ai-doomer
18. Jakesch M., Bhat A., Buschek D., Zalmanson L., and Naaman M., 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23-28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581196.
19. Singelis, T.M., Triandis, H.C., Bhawuk, D.P.S., & Gelfand, M.J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. Cross-Cultural Research, No.29 (3). Pp. 240-275. https://doi.org/10.1177/106939719502900302.
20. Studies in Machiavellianism / Ed. by Christie R., Geis F.L. New York: Academic Press, 1970.

References in Cyrillics

1. Volkova A.D., Kostina T.A., Noakk N.V. Social`ny`e predstavleniya ob iskusstvennom intellekte: metodologicheskie aspekty` (Chast` 2)// Cifrovaya e`konomika, 2023, № 26(5). - Str. 18-28.
2. Gallyamova A.A., Grigor`ev D.S. Razrabotka metodiki ocenki vosprinimaemoj kul`tury` soobshhe-stva na osnove kul`turnoj orientacii «gorizontal`ny`j/vertikal`ny`j individualizm – kollek-tivizm» G. Triandisa / Vestnik RUDN. Seriya: Psixologiya i pedagogika. 2022. №3. S. 429-447.
3. Grachyov I.D., Noakk N.V. Optimizaciya raznoobraziya agentov: razmy`shleniya, gipotezy`, prognozy` // Cifrovaya e`konomika., № 2 (28), 2024v. Iyun`, str. 57-60.
4. Grachyov I.D., Noakk N.V., Kostina T.A. Ambivalentnost` social`ny`x predstavlenij ob II: psixo-logiya, statistika, prognozy` // Cifrovaya e`konomika, № 1(27), 2024a. Str.54-61.
5. Znakov V.V. Psixologiya ponimaniya: Problemy` i perspektivy`. – M.: Izd-vo «Institut psixologii RAN», 2005. – 448 s.
6. Klenk, M. (2022). (Onlajn) manipulyaciya: inogda skry`taya, vsegda nebrezhnaya. Rev. Soc. E`konomika. 80, 85-105. doi: 10.1080/00346764.2021. 1894350
7. Litvinov V.Yu., Matveeva L.V. Sravnitel`ny`j analiz kul`turny`x predstavlenij tvorcheskoj mo-lodezhi o rossijskoj, zapadnoj i vostochnoj civilizaciyax // Social`naya psixologiya i obshhestvo. 2021. Tom 12. № 1. C. 177-197. DOI: https://doi.org/10.17759/sps.2021120112.
8. Pantin V.I. Civilizacionny`e osobennosti razvitiya Rossii v kontekste sovremenny`x social`-ny`x trans-formacij // Vestnik Instituta sociologii. 2021. Tom 12. № 4. S. 108-124. DOI: 10.19181/vis.2021.12.4.754.
9. Savin S.D., Kasabuczkaya M.S. Obshhenacional`ny`e rossijskie cennosti v kontekste formirovaniya kollektivnoj identichnosti // Vestnik Sankt-Peterburgskogo universiteta. Sociologiya. 2019. T. 12. Vy`p. 1. S. 82-97. https://doi.org/10.21638/spbu12.2019.106.
10. Semenova D.M., Afonin M.V., Kudryavcev S.A. Cennosti v strukture grazhdanskoj identichnosti: po-nyatie i instrumenty` // Politkonsul`tant, 2023. T. 3. № 1. URL: https://politicjour-nal.ru/PDF/04PK123.pdf.
11. Faraoni S. (2023) Texnologiya ubezhdeniya i vy`chislitel`ny`e manipulyacii: chrezmernoe ignoriro-vanie mental`nogo samoopredeleniya. Front. Iskusstvenno. Intellekt. 6:1216340. doi: 10.3389/frai.2023.1216340.
12. Cennosti rossijskoj civilizacii: metodicheskoe posobie dlya vuzov. / V.E`. Bagdasaryan, Yu.Yu. Ierusalimskij. Yaroslavl`: IPK «Indigo», 2023. – 80s.
13. YuNESKO. (2021). «Rekomendaciya YuNESKO po e`tike iskusstvennogo intellekta» (23 noyabrya 2021). SHS/BIO/PI/2021/1. Parizh: YuNESKO.

Ключевые слова

манипулятивность искусственного интеллекта, коллективное сознание, российская цивилизационная модель, шкала индивидуализма-коллективизма

Волкова Анастасия Дмитриевна – младший научный сотрудник ЦЭМИ РАН
SPIN РИНЦ: 1470-2650
ORCID: 0000-0002-4216-9328
volkova.nst@mail.ru

Грачев Иван Дмитриевич – д.э.н., к.ф.-м.н., главный научный сотрудник ЦЭМИ РАН
ORCID 0000-0003-1815-5898
idg@mail.ru

Костина Татьяна Анатольевна ЦЭМИ РАН
kostina1@yandex.ru
Ларин Сергей Николаевич, к.техн.н., ведущий научный сотрудник ЦЭМИ РАН
ORCID 0000-0001-5296-5865
sergey77707@rambler.ru

Ноакк Наталия Вадимовна – к.психол.н., ведущий научный сотрудник ЦЭМИ РАН
ORCID 0000-0001-8696-5767

n.noack@mail.ru

Anastasia Volkova, Ivan Grachev, Tatiana Kostina, Sergey Larin, Natalia Noakk, Experimental and model studies of the influence of AI on the evolution of collective consciousness.

Keywords
manipulativeness of artificial intelligence, collective consciousness, Russian civilizational model, scale of individualism-collectivism

DOI: 10.34706/DE-2024-03-05
JELclassification – C65, E42

Abstract
The article continues the work of the authors on the study of the influence of artificial intelligence on public consciousness. The purpose of this study is to develop system tools to ensure management decisions that minimize the risks of negative impact of AI on the civilizational features of Russia. To achieve this, the authors have begun a systematic experimental study of displacement, compression, and manipulativeness of AI, taking into account the factors of individualism – collectivism and Machiavellianism. Control groups of people were compared with groups of "personalities" generated by AI. The results were obtained on the basis of small samples typical for the initial stage of experiments. At the individualism-collectivism level, there is no bias in the estimates inherited by AI, with an indisputable compression of variability for any accepted levels of significance. On the contrary, when evaluating Machiavellianism (a measure of manipulativeness), the bias of estimates inherited by AI significantly differs from those of control groups of people. The results obtained allow us to conclude that further studies of estimates of variability in the Machiavellian parameter are necessary.