Социальная философия

Social philosophy

УДК 32.019.5

«УМНЫЙ ПОЛЕМОС»: ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ИНФОРМАЦИОННЫХ ВОЙНАХ

С. Н. ФЕДОРЧЕНКО[1)]

[1)]*Московский государственный университет им. М. В. Ломоносова, Ленинские горы, 1, 119991, г. Москва, Россия*

Аннотация. Выявлены особенности применения технологий искусственного интеллекта в современных информационных войнах, концептуализирован политический феномен «умного полемоса». В рамках нового концепта «умного полемоса» предложено сконцентрировать внимание исследователей на практиках использования алгоритмов в информационном противоборстве, в том числе в вооруженных конфликтах. Показан широкий диапазон использования технологий «умного полемоса» (от цифровых манипуляций, цифровой пропаганды, создания дипфейков до систем сопровождения боевых операций и практик применения кибероружия). Сделан вывод о том, что, несмотря на вызовы и риски развития «умного полемоса» в сфере сопровождения информационно-боевых операций, интеллектуальные системы чаще используются в цифровых психологических манипуляциях. Изучены кейсы политической интриги, создаваемой посредством применения искусственного интеллекта в информационных операциях, отмечены использование многоходового стратегического сценария, наличие заказчика, его целей и «мишени»-жертвы в структуре данных интриг. В ответ на вызовы, угрозы и риски «умного полемоса» предложен ойкуменный подход, который подразумевает формирование цифровой ойкумены государства и его союзников.

***Ключевые слова*:** «умный полемос»; искусственный интеллект; ИИ; информационные войны; алгоритмы; информационно-психологическое воздействие; политические манипуляции; цифровая ойкумена.

Образец цитирования:
Федорченко СН. «Умный полемос»: искусственный интеллект в информационных войнах. *Журнал Белорусского государственного университета. Философия. Психология*. 2025;2:13–22.
EDN: KMTHUJ

For citation:
Fedorchenko SN. «Smart polemos»: artificial intelligence in information wars. *Journal of the Belarusian State University. Philosophy and Psychology.* 2025;2:13–22. Russian.
EDN: KMTHUJ

Автор:
Сергей Николаевич Федорченко – доктор политических наук, доцент; доцент кафедры истории и теории политики факультета политологии.

Author:
Sergey N. Fedorchenko, doctor of science (politics), docent; associate professor at the department of history and theory of politics, faculty of political science.
s.n.fedorchenko@mail.ru
https://orcid.org/0000-0001-6563-044X

Журнал Белорусского государственного университета. Философия. Психология. 2025;2:13–22
Journal of the Belarusian State University. Philosophy and Psychology. 2025;2:13–22

«SMART POLEMOS»: ARTIFICIAL INTELLIGENCE IN INFORMATION WARS

S. N. FEDORCHENKO[a]

[a]*Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, Russia*

Abstract. The article reveals the specific features of using artificial intelligence technologies in modern information wars and conceptualises the political phenomenon of «smart polemos». Within the framework of the new concept of «smart polemos», it is proposed to focus the attention of researchers on the practices of using algorithms in information confrontation, including armed conflicts. A wide range of «smart polemos» technologies is shown (from digital manipulations, digital propaganda, creating deepfakes to combat operations support systems and cyberweapons use practices). It is concluded that, despite the challenges and risks of developing «smart polemos» in the field of supporting information and combat operations, intelligent systems are more often used in digital psychological manipulations. Cases of political intrigue created through the use of artificial intelligence in information operations are studied; the use of a multi-move strategic scenario, the presence of a customer, his goals and a «target»-victim in the structure of these intrigues are noted. In response to the challenges, threats and risks of «smart polemos», an ecumenical approach is proposed which implies the formation of a digital ecumene of the state and its allies.

Keywords: «smart polemos»; artificial intelligence; information wars; algorithms; information and psychological influence; political manipulation; digital ecumene.

Полемос (Война) – отец всех существ и царь всех существ,
одних он обращает в богов, других в людей,
одних делает рабами, других – свободными.
Гераклит. О природе[1]

Введение

Технологии искусственного интеллекта (ИИ) все больше внедряются в экономическую и социально-политическую жизнь современного человека, так как обладают «сквозными» признаками (охватывают множество сфер). «Умные» алгоритмы используются в обработке больших данных, предиктивной аналитике, автоматизации производства, интеллектуальном сопровождении бизнес-процессов, логистики, электронной коммерции, сельского хозяйства, здравоохранения, транспорта, туризма, совершенствовании государственного управления, улучшении имиджа политических лидеров, продвижении бренда партий, прогнозировании выборов и политическом моделировании. Но технологии ИИ применяются не только в мирных целях, происходит их активное проникновение в информационно-военную сферу, что определяет необходимость внедрения в политическую науку и полемологию отдельного термина – «умный полемос» («*smart polemos*»).

«Умный полемос» (др.-греч. Πόλεμος – божественное воплощение, олицетворение войны; спутник бога войны Ареса, упоминаемый в сочинениях Пиндара, Квинта Смирнского и Аристофана) – это политический феномен, означающий использование технологий ИИ в вооруженных конфликтах, в том числе в информационных войнах. О феномене «умной» войны было пророчески заявлено в басне Эзопа, сохранившейся через переложение Валерия Бабрия и дошедшей до нас в классификаторе Э. Перри. Согласно этой басне, когда боги женились, Полемосу досталась в жены Гибрис – демон неподобающего поведения, гордыни, дерзости и потакания худшим склонностям. Этот античный источник в очередной раз подчеркивает фундаментальную характеристику любой информационной войны, сохранившуюся до настоящих времен, – целенаправленное манипулятивное влияние на противника и его решения посредством информации. Именно на это обращал внимание доктор технических наук, один из основоположников российской научной школы изучения информационных войн С. П. Расторгуев, отмечавший, что целью информационного воздействия выступает изменение поведения некоей системы, которая начинает руководствоваться чужими, а не собственными интересами. Противник ориентируется именно на информационные мишени, стремясь привить большинству элементов информационной системы конкурента чужие ценности. Существуют даже попытки выявить алгоритмически управляемые идентичности [1]. Суть «умного полемоса», как политического феномена, базирующегося на манипулятивных технологиях воздействия нейросетевых алгоритмов, становится более понятной, если учесть основной закон информационной войны, описываемый в научных работах С. П. Расторгуева: «Доказанная взаимосвязь несуществующих событий стано-

[1]Перевод А. В. Лебедева. См.: *Лебедев А. В.* Логос Гераклита. Реконструкция мысли и слова (с новым критическим изданием фрагментов). СПб. : Наука, 2014. С. 155.

вится законом, определяющим поведение реальных субъектов» [2, с. 7]. Ближе всего к пониманию сущности «умной» войны подошел российский исследователь С. Б. Переслегин, выделивший «войну Афины» (война ума и технологий), «войну Ареса» (война силы, армий) и «войну Аполлона» (война смыслов)[2] [3].

Правда, в определении самой механики «умной» войны до сих пор сохраняется неопределенность. Если французские аналитики Р. Гийонно и А. ле Дез выделяют в работе кибервоинов три тактических режима: оборонительный режим, наступательный режим и режим безопасности [4], то С. П. Расторгуев считал наиболее важным маркером информационной войны именно наступательный характер информационного оружия. Не случайно Г. Киссинджер, Э. Шмидт и Д. Хаттенлокер подчеркивают сложность проблемы разграничения наступления и обороны, неоправданной агрессии и допустимого упреждения в области применения «умного» оружия государствами [5, с. 152] (авторы даже предлагают новый термин «страны – лидеры в области использования ИИ», описывающий государства, которые обладают «умным» оружием (Россия, США, Китай)). Между тем в настоящей работе рассматриваются в основном видимые проявления «умного» информационного оружия (факты использования ИИ в беспилотных летательных аппаратах и автономном оружии в современных конфликтах не освещены).

«Умный полемос» предполагает, что интернет уже пережил некоторую милитаризацию, в ходе которой в его архитектуру проникли представители армии, киберподразделений, органов государственной власти, хакерских группировок и частных военных компаний. Достаточно вспомнить, как и кем создавалась сеть *ARPANET*, прототип интернета. Другими словами, в цифровых коммуникациях изначально стала формироваться модель инфраструктуры с применением информационных технологий, а также складывались социально-экономические и политические условия, подготовившие наступление эры информационного оружия. Сейчас исследователи говорят о вепонизации – возникновении в цифровых коммуникациях информационного оружия, в том числе на базе технологий ИИ (*weaponisation of AI*) [6].

Материалы и методы исследования

В данной работе используется принцип триангуляции, предполагающий сочетание нескольких методов. Во-первых, для изучения академической литературы по теме использования технологий ИИ в информационных конфликтах применяется критический дискурс-анализ. Во-вторых, в целях анализа реальных фактов применения технологий ИИ в информационных конфликтах и понимания мотивов заказчиков этих технологий – политических акторов используется метод кейс-стади. В-третьих, феномен «умного полемоса» рассматривается через модель политической интриги, предложенной российским исследователем, доктором политических наук В. В. Разуваевым. Согласно этой модели политическая интрига создается посредством как минимум двух ходов интригующей стороны [7, с. 53–54]. Например, если первым ходом является заказ разработки дипфейка (от сгенерированного текста до синтетического видео) и его распространение в интернете, то вторым ходом будет выступать скандализация, негативизация образа политического конкурента и его сторонников. В случае сложной политической интриги применяется многоходовый стратегический сценарий («многоходовка»).

Современная «многоходовка», как ключевая методика политической интриги, использует все преимущества и возможности цифровых технологий. В этом контексте особо интересна обнаруженная представителем Торонтской школы коммуникативистики М. Маклюэном закономерность: влияние средства коммуникации становится более интенсивным, если оно сочетается с другим средством коммуникации [8, с. 29]. Это гениальное предвидение канадского мыслителя имеет прямое отношение к военно-политическим эффектам, возникающим благодаря применению технологий ИИ в цифровых коммуникациях (сочетание технических средств создает как новые возможности для интересантов, так и новые вызовы, угрозы, риски для их жертв). М. Маклюэн полагал, что любая технология может рассматриваться как оружие, а автоматизация не только проникает в городское планирование и промышленность, но и связывает государство с социальными фактами [8, с. 440, 450]. Каковы научные подтверждения «умного полемоса»?

Краткий обзор исследований

В научных исследованиях давно обращается внимание на то, что использование технологий ИИ, практикующееся в информационных войнах, сопровождается самыми разными психологическими эффектами, создает условия для возрастания степени подверженности массового сознания манипуляциям и влияет на процесс усвоения деструктивной информации. Эмпирические данные свидетельствуют о том, что существуют уязвимые к информационным манипуляциям такого типа группы – молодежь и люди

[2] *Бакланов И.* Война Афины // ИнформКурьер-Связь : сайт. URL: https://www.iksmedia.ru/articles/5846322-Vojna-Afiny.html (дата обращения: 15.11.2024).

пожилого возраста. Психологи Т. А. Нестик и Е. А. Михеев также отмечают, что интеллектуальные системы применяются в информационных войнах для пропаганды, автоматизированного фишинга и репутационных манипуляций в ходе дипломатической работы [9]. Большой потенциал новой цифровой (вычислительной) пропаганды, использующей возможности технологий ИИ, отмечают и современные исследователи в области политической науки [10].

Экзистенциальный характер рисков и угроз «умного полемоса» для общества и государства дополнительно подтверждают результаты проекта, инициированного американскими исследователями, которые поставили эксперимент на пяти моделях ИИ (*GPT-4-Base*, *Llama-2-Chat*, *Claude 2.0*, *GPT-3.5* и *GPT-4*). Ученые хотели выяснить, как будут реагировать, какие решения предложат интеллектуальные системы при трех сценариях развития отношений между странами – вторжении, кибератаке и бесконфликтной обстановке. Результаты симуляции показали, что все пять моделей допустили эскалацию международных отношений и труднопредсказуемые варианты эскалации, вплоть до развертывания ядерного оружия [11]. Особенно жесткие решения предлагала модель *GPT-4-Base*. Драматичность положения состоит в том, что ученые из компании *Anthropic*, организовавшие исследовательский проект в области больших языковых моделей, тоже пришли к неутешительным результатам. Первый эксперимент, в ходе которого использовались подсказки для ИИ, составленные на базе опросов, которые были проведены в рамках проектов *World Values Survey* и *Pew Research Center's Global Attitude and Trends*, показал, что ответы интеллектуальных моделей сходны с распределением мнений жителей США, Австралии, Канады, некоторых стран Европы и Латинской Америки, тогда как, например, мнения граждан России и Китая учитывались в меньшей степени. Второй эксперимент предполагал применение подсказок-уточнений по странам, однако модели продемонстрировали связь с устоявшимися стереотипами, а не глубокое знание специфики духовно-нравственных ценностей России и Китая. Третий эксперимент, в процессе которого использовались лингвистические подсказки (вопросы на русском языке), выявил, что даже при таком условии модели генерируют ответы, сходные с мнениями граждан США, Канады и некоторых европейских стран, а не жителей России. Авторы сделали выводы о том, что языковые модели могут распространять предубеждения, приводить к гомогенизации убеждений, взглядов, продвигать определенные типы мировоззрения и политические идеологии[3].

Последние разработки в сфере языковых моделей заставляют серьезно задуматься о возрастающих возможностях «умного полемоса» в цифровой пропаганде, моделировании и имитации политического поведения посредством создания активности искусственных социальных групп. Исследователи с факультета политической науки и факультета компьютерной науки Университета Бригама Янга предложили концепцию алгоритмической точности (*algorithmic fidelity*), согласно которой языковые модели наподобие чат-бота способны конструировать алгоритмические копии разных социальных групп. В соответствии с предположением ученых об алгоритмической точности [12] генерируемые ИИ тексты выбираются из комбинации разных распределений вероятностей, а не из одного всеобъемлющего распределения. Другими словами, языковая модель дает результаты и интегрирует эффекты, которые коррелируют с паттернами, идеями, опытом, мнением, отношениями и установками разных групп людей. Проведенный над искусственными выборками таких социальных групп с генерацией синтетического набора данных эксперимент продемонстрировал хорошую способность модели *GPT-3* имитировать ответы представителей американских демократов и республиканцев. Авторы исследования признают, что возможности таких языковых моделей могут использоваться в целях манипуляции и дезинформации.

В рамках другого проекта интеллектуальная система провела интервью с 1052 респондентами. Результаты интервью были преобразованы в текст, на основе которого чат-бот *GPT-4o* смоделировал генеративных агентов. Результаты показали, что генеративные агенты с точностью в 85 % воспроизводили ответы реальных людей-респондентов[4]. Получается, что созданная система агентов может имитировать индивидуальное и коллективное поведение, что не только открывает новые перспективы для политического прогнозирования, но и подтверждает высокие риски влияния ИИ на провоцирование информационных конфликтов на основе синтетических пользователей – генеративных агентов. Не даром белорусские исследователи Д. Г. Доброродний и А. И. Верещако отметили, что технические объекты ИИ уже фактически «могут претендовать на статус социального актора» [13, с. 73]. Сходные выводы сделаны и в российской политической науке [14].

Не так давно группа ученых из Центра имени Гельмгольца (Мюнхен) предприняла попытку создать единую теорию познания с помощью экспериментальной проверки разработанной им на базе алгоритма *Llama 3.1 70B* модели *Centaur*. Вычислительная модель *Centaur* была дополнительно обуче-

[3]Towards measuring the representation of subjective global opinions in language models [Electronic resource] / E. Durmus [et al.]. URL: https://arxiv.org/abs/2306.16388 (date of access: 23.11.2024).

[4]Generative agent simulations of 1,000 people [Electronic resource] / Joon Sung Park [et al.]. URL: https://arxiv.org/abs/2411.10109 (date of access: 23.11.2024).

на на основе данных базы *Psych-101*, включающей свыше 10 млн ответов от более чем 60 тыс. участников 160 психологических экспериментов. База данных *Psych-101* содержит результаты многочисленных когнитивных исследований человеческого поведения (памяти, процессов принятия решений (по Маркову), игровых автоматов, обучения с учителем). В ходе обучения через низкоранговые адаптеры (дополнительные обучаемые модули) система фокусировалась на моделировании поведения человека. В результате алгоритм *Centaur* стал не только первой моделью, отвечающей большинству критериев А. Ньюэлла (от работы в реальном времени, проявления адаптивного поведения до использования и интеграции обширных знаний об окружающей среде и применения естественного языка), но и моделью, которая может прогнозировать и моделировать поведение человека в самых разных областях (от политики до экономики). Также дополнительная проверка показала, что полученная модель способна прогнозировать человеческую нейронную активность[5]. О высоких политических рисках существования искусственных выборок социальных групп свидетельствует обнаруженная в 2023 г. учеными из Индианского университета и Северо-Восточного университета (Бостон) на платформе *X (Твиттер)*[6] сеть ботов (*botnet*, ботнет) [15]. Боты использовали изображения и имитировали поведение людей, отвечая друг другу и формируя иллюзию общения. Еще М. Маклюэн писал, что технологии меняют образцы восприятия и чувственные пропорции [8, с. 30].

Между тем не все исследователи согласны с тем, что последние открытия в области генеративного ИИ обязательно приведут к масштабному росту политико-психологических манипуляций и появлению информационных войн нового, алгоритмического типа. Скептики считают недоказанным тот факт, что ИИ способен генерировать более персонализированный и убедительный контент в целях дезинформации[7]. Хотя эксперимент с алгоритмами сервиса *Netflix*, напротив, показал, что интеллектуальные системы могут способствовать появлению высокого уровня персонализации предпочтений пользователей [16]. ИИ приводит к эффекту алгоритмического самоподтверждающегося эстетического потребления, снижая интерес человека к альтернативному контенту.

Есть определенные сомнения и по поводу возможностей дипфейков (технологий ИИ, использующихся для синтезирования голоса, изображения, в том числе в целях манипуляции). Например, экспериментальное исследование, проведенное в Нидерландах, дало понять его организаторам, что дипфейк не оценивался респондентами как более убедительная форма информации, чем текст. Результаты реализации проекта показали, что дипфейки не являются такой опасной дестабилизирующей силой для общественного порядка, как предполагалось в других научных работах. По мнению авторов эксперимента, дипфейк не может сам по себе воздействовать на политическую поляризацию. Вместе с тем авторы исследования уверены, что со временем дипфейки могут оказать более значительное влияние на поведение и убеждения людей. По оценке авторов, современные дипфейки обладают только косвенным влиянием, снижая лишь доверие человека к новостям [17]. Чтобы лучше разобраться в этом вопросе, необходимо обратиться к анализу кейсов.

Анализ кейсов «умного полемоса»

Проведенный анализ практик «умного полемоса» (применения ИИ в информационных войнах и конфликтах) позволяет назвать несколько технологий деструктивного информационного воздействия такого рода. Во-первых, к цифровым манипуляциям можно отнести широкий спектр информационно-психологических операций по обману и вводу в заблуждение в политической, общественной сферах и бизнес-сфере. Как правило, под такими манипуляциями подразумеваются дипфейк-видео, дипфейк-фото, голосовые, текстовые дипфейки и др. Также политические манипуляторы часто используют разные техники информационного воздействия, аккумулируя их аудио- и видеоэффект, а также текстовые эффекты в одном дипфейке. Во-вторых, к области применения кибероружия можно отнести специфические информационные операции и процедуры, нацеленные на подрыв кибербезопасности и снижение (ликвидацию) защитных функций информационных систем противника. Использованием кибероружия можно считать применение интеллектуальными системами распознавания неживых и живых целей; генерацию эксплойтов с помощью технологий ИИ (программ, фрагментов программных кодов, ориентированных на нахождение уязвимых мест в программном обеспечении противника); распределенные автоматизированные атаки; организацию кибератак на системы машинного обучения;

[5]Centaur: a foundation model of human cognition [Electronic resource] / M. Binz [et al.]. URL: https://arxiv.org/abs/2410.20268 (date of access: 23.11.2024).

[6]Доступ к платформе *X (Твиттер)* был ограничен в России на основании решения Гененеральной прокуратуры Российской Федерации от 24 февраля 2022 г.

[7]*Simon F. M., Altay S., Mercier H.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown [Electronic resource] // Harvard Kennedy School. Misinformation Review. URL: https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/ (date of access: 23.11.2024).

обращение к системам машинного обучения при поиске уязвимых мест цифровых систем противника; массовый фишинг; автоматизированное тестирование полученных эксплойтов; адаптивные фишинговые атаки, обход систем безопасности; автоматизированное распространение вредоносного программного обеспечения. Среди интересантов (заказчиков создания и применения кибероружия) могут быть так называемые киберармии – специальные подразделения армий государств, занимающиеся кибернаступлением (информационными атаками) и обеспечивающие киберзащиту (кибероборону). В отличие от цифровых манипуляций использование кибероружия может приводить к более серьезным киберфизическим последствиям. Заказчиками цифровых манипуляций могут быть политические партии, лидеры, организации и движения, в том числе и радикального толка, а также власти. Технологии цифровых манипуляций и кибероружия часто применяются вместе, поэтому их не всегда бывает легко четко отделить друг от друга.

Авторы связывают использование дипфейков с общими цифровыми манипуляциями, на эффективность которых влияют такие факторы, как эмоциональная привлекательность контента (провокация гнева, страха и других эмоций, позволяющих подавлять критическое мышление), персонализация контента (технологии ИИ помогают ориентировать манипуляцию на ценности и убеждения человека), его повторяемость (облегчает запоминание контент), социальное влияние (контент, одобряемый сообществом на цифровых платформах, может привлечь внимание отдельного человека), достоверность (к источнику контента должно сформироваться доверие), временны́е ограничения (манипуляция более эффективна, если у людей нет времени на проверку контента), осведомленность пользователей (те люди, которые знакомы с манипуляциями, способны с ними бороться) [18]. Примечательно, что компания *Sumsub*, специализирующаяся на анализе дипфейков, в ходе своего исследования установила, что в 2024 г. по сравнению с предыдущим годом произошел серьезный рост числа использований дипфейков по всему миру (более чем на 245 %). Широкое применение дипфейков авторы исследования связываются с выборами. М. Маклюэн в прошлом веке пророчески писал о «войне икон», фундаментальном росте влияния технологий: «…поле битвы переместилось в ментальное сотворение и сокрушение образов – как в войне, так и в бизнесе» [8, с. 136]. Дипфейки участвуют в процессе ремифологизации современного политического сознания, так как «миф есть мгновенное целостное видение сложного процесса», о чем писал канадский исследователь [8, с. 39].

Эксперты обращают внимание на то, что появляется все больше различных систем на основе технологий ИИ, которые анализируют данные для осуществления наступательных информационных операций. Среди функций таких интеллектуальных систем выделяют атаки на базы данных приложения; запуск сложных кибератак и имитацию пакетов для обновления, удаления и установки программ; отбор информации на базе алгоритмов распознавания лиц; создание вредоносного программного обеспечения, обходящего системы безопасности; автоматизированный сбор информации; создание шаблонов для кибератак [6].

В качестве одного из первых ярких примеров информационной операции на основе технологий ИИ в политике можно привести видео, появившееся в интернете в мае 2018 г., на котором американский президент Д. Трамп высмеивал правительство Бельгии за то, что оно не вышло из Парижского соглашения по климату. Политический дипфейк спровоцировал обсуждения между бельгийскими жителями в социальных сетях. Часть бельгийцев возмущались вмешательством американского президента в политику независимой страны. Позже выяснилось, что разработку этого дипфейка на основе машинного обучения заказала производственной студии бельгийская социал-демократическая партия *sp.a*. Истинная цель партии заключалась в привлечении внимания бельгийцев к онлайн-петиции, призывающей власть принять срочные меры по борьбе с изменением климата [19]. Из данного кейса видно, что если первым ходом политической интриги было создание и распространение дипфейка с помощью цифровых коммуникаций, то вторым ходом являлась не просто критика правительства, а привлечение внимания общественности к экологической проблематике и петиции. Как отмечал М. Маклюэн, «все без исключения средства коммуникации… конфигурируют сознание и опыт каждого из нас» [8, с. 34].

С января 2024 г. жители американского штата Нью-Гэмпшир стали получать звонки якобы от президента США Дж. Байдена. Он призывал не голосовать на предварительных выборах Демократической партии, так как это только поможет республиканцам в избрании Д. Трампа. Мало того, в голосовом сообщении оглашался номер телефона К. Салливан, политика, экс-председателя Демократической партии этого штата, возмутившейся такой ситуацией. После подачи ею заявления генеральная прокуратура штата в ходе расследования установила, что голос Дж. Байдена был сгенерирован искусственно. Федеральная комиссия по связи обвинила компанию *Lingo Telecom* в передаче робозвонка на основе технологии генеративного ИИ, однако установить, кто являлся заказчиком голосового дипфейка, удалось не сразу. В заказе таких звонков признался политический консультант С. Крамер, работавший на члена палаты представителей Д. Филлипса, крупного бизнесмена, основного соперника и критика Дж. Бай-

дена. Д. Филлипс осудил действия своего политического консультанта. С. Крамер оправдывался тем, что он хотел привлечь внимание общественности к проблеме применения технологий ИИ в политических кампаниях. В августе 2024 г. компания *Lingo Telecom* согласилась выплатить штраф в размере 1 млн долл. США за передачу поддельных автоматических звонков, а в сентябре этого же года Федеральная комиссия выписала С. Крамеру штраф в размере 6 млн долл. США[8]. По оценке исследователей, приводимая в качестве примера техника вишинга из-за массовости своего применения в некотором плане схожа с фишингом [6]. Кейс показывает создание политической интриги с помощью определенного алгоритма: 1) генерирования дипфейка; 2) заказа роботизированных голосовых звонков через компанию; 3) реализации попытки сорвать предварительные выборы через массовые звонки.

Распространение одного из первых скандальных дипфейк-фото в Бангладеш фиксируется в 2023 г. На нем политик Р. Фарахна из оппозиционной Националистической партии Бангладеш соотносилась с женщиной в бикини на пляже. Позже был доказан целенаправленный подлог. В 2024 г. противники индийской партии «Бхаратия Джаната» создали дипфейк-видео, на котором А. Хан, звезда Болливуда, высмеивал эту популярную политическую организацию за то, что она не выполнила своего обещания перевести 1,5 млн рупий на банковские счета граждан. Затем на видеоролике актер высказался в поддержку оппозиционной партии «Индийский национальный конгресс». Голос А. Хана был искусственно обработан, при этом его пресс-секретарь объяснил, что актер не высказывался по поводу партий[9]. В этом же году было распространено дипфейк-видео, на котором актер Р. Сингх критиковал Н. Моди, премьер-министра Индии, за инфляцию и безработицу в стране. Сообщение, как и в случае с А. Ханом, завершалось призывом голосовать за оппозиционную партию «Индийский национальный конгресс». Видеоконтент также был сгенерирован с помощью технологий ИИ. Р. Сингх из-за распространения этого видео в интернете подал жалобу в полицию[10]. Политические условия и контекст применяемых дипфейков позволяют прийти к промежуточному выводу о том, что их заказчиками были люди, поддерживающие противников тех политических лидеров и партий, которые стали объектами критики, дискредитации и умышленной манипуляции. Кейсы показывают, что политическая интрига была организована минимум в два хода, таких как создание и распространение дипфейков, а также сетевая агитация (через провоцирование комментариев) за оппозиционную партию и (или) дискредитация конкурентов (политика или партии) с параллельной скандализацией.

В информационных войнах с использованием технологий ИИ могут участвовать специальные государственные ведомства и разные хакерские группировки, APT-группы (*advanced persistent threat groups* – группы продвинутых постоянных угроз), сочувствующие политике определенных государств или работающие на них, но сохраняющие автономию. Милитаризация цифрового пространства направлена на вепонизацию интернете. На базе государств, их военно-политических блоков и макрорегионов возникает целая цифровая архитектура «умного полемоса». Например, с 2008 г. существует Центр передового опыта кибероборонны НАТО, а с 2009 г. – Киберкомандование США. В 2021 г. появился Региональный центр кибернетической безопасности НАТО[11]. Проектами в области технологий ИИ занимается Управление перспективных исследовательских проектов Министерства обороны США. К киберподразделениям можно отнести британскую межвидовую кибергруппу, Силы информационной поддержки Народно-освободительной армии Китая, группировки «Электронная армия Ирана», «Йеменская кибермия», «Пакистанская кибермия», лояльную к Ирану электронную команду из Ирака «Фэйтмион», киберсообщество «Ливанский кедр» (Ливан), проукраинскую группу «IT-армия», саудовские группы «Кибермухи», аффилированную с движением ХАМАС кибергруппу «Шторм-1133», сочувствующую Израилю кибергруппировку «Хищный воробей»[12], группу «Киберянычары» (Турция), симпатизирующую Турции кибергруппу «Анка Неферлер», прогреческую группу хакеров «Аноним Греции», проегипетскую кибергруппу «Египетская кибермия». Сходные подразделения и кибергруппы есть во многих странах. В связи с этим примечательно, что в 2023 г.

[8]Political consultant fined $6 M for using AI to fake Biden's voice in robocalls to voters [Electronic resource]. URL: https://nypost.com/2024/09/26/business/political-consultant-fined-6m-for-using-ai-to-fake-bidens-voice-in-robocalls-to-voters/ (date of access: 15.11.2024).

[9]Deepfakes take just minutes to make with artificial intelligence. Here's how Indian political parties use them against opponents [Electronic resource]. URL: https://www.abc.net.au/news/2024-05-30/how-ai-is-disrupting-india-election-campaign/103899138 (date of access: 15.11.2024).

[10]Ranveer singh files police case after deepfake video goes viral [Electronic resource]. URL: https://www.ndtv.com/india-news/ranveer-singh-files-police-case-after-deepfake-video-goes-viral-5495846 (date of access: 15.11.2024).

[11]*Хетагуров А.* Новый инструментарий США и НАТО для информационного и киберпротивоборства с Россией в Восточной Европе [Электронный ресурс]. URL: https://russiancouncil.ru/analytics-and-comments/analytics/novyy-instrumentariy-ssha-i-nato-dlya-informatsionnogo-i-kiberprotivoborstva-s-rossiey-v-vostochnoy-/?sphrase_id=113374031 (дата обращения: 15.11.2024).

[12]*Цуканов Л.* «Цифровой шторм» Аль-Акса: штрихи к противостоянию Израиля и ХАМАС [Электронный ресурс]. URL: https://russiancouncil.ru/analytics-and-comments/columns/cybercolumn/tsifrovoy-shtorm-al-aksa-shtrikhi-k-protivostoyaniyu-izrailya-i-khamas/?sphrase_id=151590373 (дата обращения: 15.11.2024).

компания «Лаборатория Касперского» подготовила отчет, в котором отслеживаются кибератаки APT-группировок на государственные ресурсы России, Беларуси, Индонезии, Малайзии и Аргентины, национальный телеком Пакистана[13]. Также в разработке и применении ИИ в информационных конфликтах активно участвуют технологические корпорации *Palantir Technologies Ing.*, *Clearview AI, Primer Tecnologies Inc.* и др. В России в ответ на такие вызовы и угрозы еще в 2017 г. министр обороны сообщил о создании войск информационных операций[14].

К сожалению, технологии ИИ все чаще используются в информационных войнах террористическими и экстремистскими группировками. Научный сотрудник Центра центральноазиатских исследований Института Китая и современной Азии РАН Л. А. Шашок отметила, что террористическая группа «ИГ-хорасан» (запрещена в России) стала публиковать пропагандистские ролики, созданные с помощью технологий ИИ, в которых видна попытка подражать реальным ведущим афганских новостных агентств. Мишенью террористической цифровой пропаганды становятся не только взрослые, но и дети. Например, бот группировки «Исламское государство» (запрещена в России) «Сарх Аль-Хилафа» распространял многосерийный мультсериал, в котором персонажи Салим и Муавия одобряли взгляды террористической организации. Бот поддерживал 4 языка и был удален из социальных сетей после совершения терактов в концертном зале «Крокус сити холл». Л. А. Шашок отмечает, что террористы используют технологии ИИ для генерации видео, постоянно меняя свою тактику, чтобы избежать блокировки контента. Они обращаются к псевдообучающим курсам о технологиях ИИ, чат-ботам на их основе, контенту, содержащему отрывки из документальных фильмов о дикой природе, спорте, а также музыку[15]. Создание политической интриги с помощью технологий ИИ предполагает реализацию террористами нескольких ходов: 1) повышение узнаваемости, привлечение внимания населения разных стран к деятельности террористических организаций; 2) вербовку новых сторонников в их состав; 3) наращивание сил, ресурсов и захват политической власти в регионе или в целом государстве.

Американское правительство, часто обвиняющее Россию и Китай в использовании новейших технологий цифровой пропаганды, само стремится к применению ИИ в качестве информационного оружия. Так, в 2024 г. С. Биддл, репортер из американской новостной организации *The Intercept*, обратил внимание на планы Совместного командования специальных операций США, в которых была заявлена необходимость в использовании передовых технологий по генерированию убедительных онлайн-образов в цифровых коммуникациях, а также уникальных профилей несуществующих людей со своим фоном, набором фотографий для конструирования виртуальной среды, не идентифицируемой алгоритмами. С. Биддл также напоминает, что в 2023 г. Командование специальных операций США проявило интерес к технологиям создания дипфейков[16].

Можно предположить, что использование технологии ИИ американскими специалистами будет направлено на дискредитацию внешней и внутренней политики стран – противников США; генерацию, поддержку и усиление в них образа врага; оправдание последующих санкционных или наступательных намерений США. Даже если опасения С. Биддла сильно преувеличены, существуют факты, свидетельствующие о возрастании роли «умного полемоса». Одно международное исследование подтвердило, что модели *ChatGPT* и *GPT-3* можно использовать в целях генерирования микротаргетированной политической рекламы автоматизированным способом. Эксперимент показал, что персонализированный и адаптированный к личности человека контент, созданный с помощью ИИ, оказывает более эффективное воздействие, чем неперсонализированный [20]. Другое исследование, получившее этическое одобрение Колумбийского университета, также подтвердило, что персонализированные сообщения, полученные при помощи модели *ChatGPT-3*, оказывали на испытуемых большее влияние, чем неперсонализированные [21]. В третьем исследовании, проведенном учеными из Оксфордского университета на основе чат-бота *GPT-4*, было доказано, что влияние технологий ИИ состоит не в адаптации сообщений под конкретного человека, а наоборот, в оказании неперсонализированными, общими сообщениями убеждающего воздействия. Результаты эксперимента засвидетельствовали, что одно сообщение из двух сотен слов, сгенерированное ИИ, смогло повысить поддержку респондентами предполагаемого мнения почти на 50 %. Авторы пришли к выводу, что как нетаргетированные, так и микротаргетированные политические сообщения по большинству вопросов, созданные ИИ, имеют убедительность[17].

[13]Азиатские APT-группировки: тактики, техники и процедуры [Электронный ресурс]. URL: https://go.kaspersky.com/ru-apt-report (дата обращения: 15.11.2024).

[14]Шойгу рассказал о российских войсках информационных операций [Электронный ресурс]. URL: https://www.rbc.ru/politics/22/02/2017/58ad78cd9a794757f3c80ece (дата обращения: 15.11.2024).

[15]*Шашок Л. А.* Джихадисты берут на службу искусственный интеллект [Электронный ресурс]. URL: https://www.ng.ru/kartblansh/2024-06-03/3_9021_kb.html (дата обращения: 15.11.2024).

[16]*Biddle S.* The Pentagon wants to use ai to create deepfake Internet users [Electronic resource]. URL: https://theintercept.com/2024/10/17/pentagon-ai-deepfake-internet-users/ (дата обращения: 15.11.2024).

[17]*Hackenburg K., Margetts H.* Evaluating the persuasive influence of political microtargeting with large language models // PNAS : website. URL: https://www.pnas.org/doi/10.1073/pnas.2403116121 (дата обращения: 15.11.2024).

Стоит отметить, что в настоящее время появляется опасный тренд использования технологий ИИ в сфере специфического информативно-физического воздействия. Так, правительство Израиля экспериментирует с технологиями «умного полемоса», например, интеллектуальными системами *Lavender* и *The Gospel*, применяя их в своих информационных и военных операциях. Алгоритмы системы *Lavender* нужны израильтянам для обнаружения, маркировки, а также ранжирования людей, которые подозреваются в подготовке терактов, сотрудничестве с боевым крылом организаций «Исламский джихад» (запрещена в России) и «ХАМАС» в Палестине[18]. Система *The Gospel* обрабатывает крупные массивы всевозможных разведавательных данных, перехваченных текстовых и звуковых сообщений, данные движения групп людей, съемки спутников и беспилотных аппаратов.

Складывается впечатление, что, несмотря на вызовы и риски развития «умного полемоса», например использование ИИ в сопровождении информационно-боевых операций, пока интеллектуальные системы применяются только для цифровых психологических манипуляций. Логично предположить, что ИИ в информационных конфликтах будет использоваться государствами, партиями, политическими лидерами, APT-группами, хакерами и террористическими группировками после апробации соответствующих техник крупными технологическими корпорациями. Психологическое воздействие «умных» алгоритмов корпоративных платформ видно на примере «темных» паттернов (*dark patterns*) – уловок цифровой архитектуры, интерфейса, обманывающих пользователей и провоцирующих их на принятие невыгодных или даже рискованных решений. Интересно, что в одном из исследовании 11 тыс. сайтов было выявлено 1818 фактов наличия таких «темных» паттернов, созданных для манипулирования посетителями цифровых ресурсов [22]. Широкая апробация цифровых манипуляций также видна на примере платформы *Ютуб*, алгоритмы которой нацелены на увеличение периода просмотра видеоконтента пользователями и корректировку их социально-политических взглядов. Еще в 2016 г. было замечено, что, например, после просмотра человеком роликов с Д. Трампом ему предлагался контент на тему отрицания холокоста и превосходства белой расы [23].

Вместе с тем роль интеллектуальных систем в социальной и политической жизни, их эволюцию важно понимать правильно. Ответственность за применение таких систем должен нести человек, конкретная корпорация или государство.

Заключение

Обращение к академическому дискурсу и кейстади позволяет утверждать, что технологические трансформации, научные исследования, развитие экономических отраслей, существующих политических и социальных институтов создали благоприятные условия для появления «умного полемоса» – политического феномена, подразумевающего применение технологий ИИ в информационных войнах и вооруженных конфликтах. Анализ кейсов показал, что диапазон использования технологий «умного полемоса» довольно широк (от цифровых манипуляций, цифровой пропаганды, создания дипфейков до систем сопровождения боевых операций и практик кибероружия). Разбор кейсов применения ИИ в информационных операциях через модель политической интриги выявил, что в каждом случае можно обнаружить многоходовой стратегический сценарий, заказчика, его истинные цели и «мишень»-жертву.

Снизить угрозы цифровой десуверенизации, колониализма и колонизации, риски применения враждебных технологий ИИ в информационных манипуляциях может развитие цифровой ойкумены – системы технологических корпораций, информационных агентств, цифровых радио и телевидения, осуществляющих слаженную медийную активность не только внутри страны, но и за ее пределами для позиционирования, защиты интересов, ценностно-цивилизационной, политической и экономической повестки поддерживающих их государств [24]. В условиях геополитических рисков и угрозы перехода информационных войн на алгоритмический уровень Россия и Беларусь могут постепенно сформировать собственную цифровую ойкумену, которая предполагает тесное сотрудничество в области разработки и внедрения единых цифровых стандартов, ведение совместной политики информационной безопасности и создание общего цифрового пространства, обеспечивающего качественные коммуникационные площадки для функционирования обратной связи между представителями граждан, государства и бизнеса, а также сохранение и развитие культуры обеих стран.

Стратегия выстраивания цифровой ойкумены на деле означает формирование новой модели цифровой суверенизации. Такая модель может включать инвестиции в совместные технологические разработки России и Беларуси; снижение цифровой зависимости от крупных зарубежных, особенно западных технологических корпораций; более активное сотрудничество в области цифровых технологий со странами БРИКС и ШОС. Возможно, ойкуменный подход станет одним из стратегических ответов на вызовы «умного полемоса».

[18] *Чугунов В., Доронин А.* Боевое применение ИИ-систем Израилем в секторе Газа: этические проблемы // Российский совет по международным делам : сайт. URL: https://russiancouncil.ru/analytics-and-comments/columns/cybercolumn/boevoe-primenenie-ii-sistem-izrailem-v-sektore-gaza-eticheskie-problemy/ (дата обращения: 15.11.2024).

Библиографические ссылки

1. Cheney-Lippold J. A new algorithmic identity: soft biopolitics and the modulation of control. *Theory, Culture & Society.* 2011;28(6):164–181. DOI: 10.1177/02632764114244.

2. Расторгуев СП. Планирование и моделирование информационной операции. *Информационные войны.* 2014;1:2–10. EDN: RUXFYP.

3. Переслегин СБ. *Первая мировая. Война между реальностями. Том 2.* Москва: Яуза; 2020. 608 с.

4. Guyonneau R, le Dez A. Artificial intelligence in digital warfare: introducing the concept of the cyberteammate. *The Cyber Defense Review.* 2019;4(2):103–116.

5. Киссинджер Г, Шмидт Э, Хаттенлокер Д. *Искусственный разум и новая эра человечества.* Ахметов К, переводчик. Якимова Е, редактор. Москва: Альпина про; 2022. 200 с.

6. Yamin MM, Ullah M, Ullah H, Katt B. Weaponized AI for cyber attacks. *Journal of Information Security and Applications.* 2021;57:102722. DOI: 10.1016/j.jisa.2020.102722.

7. Разуваев ВВ. *Анатомия политической интриги.* Москва, Санкт-Петербург: Центр гуманитарных инициатив; 2019. 184 с.

8. Маклюэн МГ. *Понимание медиа. Внешние расширения человека.* Николаев В, переводчик. Москва: Кучково поле; 2023. 464 с.

9. Нестик ТА, Михеев ЕА. Информационные войны с использованием систем искусственного интеллекта: анализ психологических механизмов воздействия. *Институт психологии Российской академии наук. Организационная психология и психология труда.* 2019;4(4):148–174. EDN: VOXRQM.

10. Володенков СВ. Цифровые актанты и вычислительная пропаганда как инструменты воздействия на массовое сознание в условиях глобальных технологических трансформаций. *Вестник Московского университета. Серия 12, Политические науки.* 2024;2:47–70. DOI: 10.55959/MSU0868-4871-12-2024-2-2-47-70.

11. Rivera J-P, Mukobi G, Reuel A, Lamparth M, Smith Ch, Schneider J. Escalation risks from language models in military and diplomatic decision-making. In: ACM FAccT Executive Committee. *FAccT '24. Proceedings of the 2024 ACM conference on fairness, accountability, and transparency.* New York: Association for Computing Machinery; 2024. p. 836–898. DOI: 10.1145/3630106.365894.

12. Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D. Out of one, many: using language models to simulate human samples. *Political Analysis.* 2023;31(3):337–351. DOI: 10.1017/pan.2023.2.

13. Доброродний ДГ, Верещако АИ. Статус технических объектов с искусственным интеллектом в современном обществе. *Журнал Белорусского государственного университета. Философия. Психология.* 2024;1:66–74. EDN: IALYPG.

14. Володенков СВ, Федорченко СН. Особенности феномена субъектности в условиях современных технологических трансформаций. *Полис. Политические исследования.* 2022;5:40–55. DOI: 10.17976/jpps/2022.05.04.

15. Kai-Cheng Yang, Menczer F. Anatomy of an AI-powered malicious social botnet. *Journal of Quantitative Description: Digital Media.* 2024;4:1–36.

16. Pajkovic N. Algorithms and taste-making: exposing the Netflix recommender system's operational logics. *Convergence.* 2022;28(1):214–235. DOI: 10.1177/13548565211014464.

17. Hameleers M, van der Meer TGLA, Dobber T. You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media + Society.* 2022;8(3):1–12. DOI: 10.1177/20563051221116346.

18. Ienca M. On artificial intelligence and manipulation. Topoi. *An International Review of Philosophy.* 2023;42:833–842. DOI: 10.1007/s11245-023-09940-3.

19. Ullrich QJ. Is this video real? The principal mischief of deepfakes and how the lanham act can address it. *Columbia Journal of Law and Social Problems.* 2021;55(1):5–6.

20. Simchon A, Edwards M, Lewandowsky S. The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus.* 2024;3(2):1–4. DOI:10.1093/pnasnexus/pgae035.

21. Matz SC, Teeny JD, Vaid SS, Peters H, Harari GM, Cerf M. The potential of generative AI for personalized persuasion at scale. *Scientific reports.* 2024;14(1):1–17. DOI: 10.3389/fphys.2019.00637.

22. Mathur A, Acar G, Friedman M, Lucherini E, Mayer J, Chetty M, et al. Dark patterns at scale: findings from a crawl of 11 K shopping websites. *Proceedings of the ACM on Human-Computer Interaction.* 2019;3:1–32. DOI: 10.1145/3359183.

23. Kozyreva A, Lewandowsky S, Hertwig R. Citizens versus the Internet: confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest.* 2020;21(3):103–156. DOI: 10.1177/1529100620946707.

24. Федорченко СН. Государство-цивилизация в цифровой ойкумене. *Журнал политических исследований.* 2023; 7(1):3–26. DOI: 10.12737/2587-6295-2023-7-1-3-26.

Статья поступила в редколлегию 23.12.2024.
Received by editorial board 23.12.2024.