

УДК 32

DOI: 10.34670/AR.2026.73.30.001

Идеологическая направленность ответов нейросетей: результаты политико-экономических опросов

Воронин Всеволод Алексеевич

Студент,
Санкт-Петербургский государственный университет,
199034, Российская Федерация, Санкт-Петербург, Университетская набережная, 7-9;
e-mail: sevvu@inbox.ru

Ботникова Евгения Эдуардовна

Студент,
Санкт-Петербургский государственный университет,
199034, Российская Федерация, Санкт-Петербург, Университетская набережная, 7-9;
e-mail: evgenia.botnikova@yandex.ru

Хайруллова Анна Вадимовна

Студент,
Санкт-Петербургский государственный университет,
199034, Российская Федерация, Санкт-Петербург, Университетская набережная, 7-9;
e-mail: mega.khayrullova@mail.ru

Скрипка Софья Игоревна

Студент,
Санкт-Петербургский государственный университет,
199034, Российская Федерация, Санкт-Петербург, Университетская набережная, 7-9;
e-mail: skripkasofa9@gmail.com

Авторы выражают благодарность членам исследовательской группы Совета молодых политологов Российской ассоциации политической науки, участвовавшим в проведении опросов – студенту 1 курса магистратуры факультета социальных наук НИУ ВШЭ Макаровой Еве Алексеевне и студенту магистратуры факультета социально-политических наук ЯрГУ им. П. Г. Демидова Васильевой Дарье Александровне.

Аннотация

Настоящее исследование посвящено выявлению склонности нейросетей к выдаче идеологически ангажированных ответов на политические и экономические вопросы. Актуальность исследования обусловлена вопросом безопасности использования ИИ (который считается достаточно авторитетным источником информации, особенно среди молодежи), а также минимизацией нежелательного влияния на общества в результате ответов нейросетей на запросы пользователей всего мира. Цель исследования — выявить специфику ответов нейросетей на вопросы, связанные с политическими и экономическими

предпочтениями. В ходе исследования использовались такие эмпирические методы, как опрос, кодировка и кластерный анализ, которые также были дополнены качественным анализом полученных данных. Кластерный анализ производился на основе данных, закодированных в непрерывные и биномиальные шкалы. Результаты опроса с открытыми вопросами были визуализированы на классической шкале политических координат. Визуализация кластеров, выявленных после проведения опроса с закрытыми вопросами, производилась в Google Colab с помощью Python, посредством методов UMAP и PCA. В ходе исследования была выявлена склонность нейросетей к идеологизированным ответам как на закрытые, так и открытые вопросы. Все нейросети, отобранные для исследования, демонстрируют «склонность» к позициям, близким к центру, однако с заметным отклонением как «вправо», так и «влево», при этом преимущественно в рамках либерально-демократического идеологического мейнстрима. В ходе исследования была выявлена двойственность ответов нейросетей: они не имеют собственных идеологических предпочтений, но демонстрируют устойчивые ценностные паттерны. Это может влиять на установки пользователей. Хотя нейросети пока не продвигают деструктивные политико-экономические идеи системно, их идеологизированность несёт потенциальные риски, поэтому в перспективе потребуется дополнительная регуляция и технологические меры в отношении ИИ.

Для цитирования в научных исследованиях

Воронин В.А., Ботникова Е.Э., Хайруллова А.В., Скрипка С.И. Идеологическая направленность ответов нейросетей: результаты политико-экономических опросов // Теории и проблемы политических исследований. 2026. Том 15. № 4А. С. 5-21. DOI: 10.34670/AR.2026.73.30.001

Ключевые слова

Искусственный интеллект, нейросети, опрос, политика, экономика, кластеры, социал-демократия, либерализм, политическая ангажированность, машинное обучение.

Введение

В условиях стремительного развития мировой цивилизации внедрение и популяризация технологий искусственного интеллекта, в частности в виде чат-ботов, не остались незамеченными. Актуальность изучения политических предпочтений языковых моделей искусственного интеллекта (ИИ) возрастает в связи с их активным внедрением в различные сферы общественной жизни. Согласно данным ВЦИОМ, 63% жителей России уже используют ИИ-технологии в повседневной жизни, что подчёркивает их значимость для общества [ВЦИОМ, 2024]. Внедрение ИИ вызвало широкий резонанс в обществе, что привело к предложениям о временной приостановке разработки в данной области. Инициатива направлена на совершенствование текущего законодательства и обеспечение максимальной безопасности при использовании ИИ-технологий [Pause Giant AI Experiments, 2023]. Сами разработчики ИИ также принимают меры для усиления безопасности; так, в ноябре 2025 года была обновлена политика использования сервисов компании OpenAI, в результате часть вопросов, в том числе касающихся политической сферы, была ограничена [Open AI запретила ChatGPT, 2025]. Это было сделано ради повышения безопасности, под которой подразумевается не только защита от

вредоносного использования технологий (включая утечки данных, создание дипфейков и манипулирование информацией), но и обеспечение их беспристрастности в работе. Безопасность ИИ становится важным аспектом современной политики, определяющим доверие общества к технологическому прогрессу и стабильность государственного управления в цифровую эпоху. Многие вопросы, касающиеся общественной жизни и ответы на которые пользователи ищут через чат-боты с генеративным ИИ, могут быть связаны с политикой. Возникает вопрос: есть ли «политические предпочтения» у нейросетей? Необходимо сразу оговориться, что у программы априори не может быть своих политических взглядов (что связано с устройством её работы, основанной на обработке информации). Однако в связи с тем, что программа обучается на определенных массивах данных, у неё могут наблюдаться сдвиги от условного политико-экономического «центра» (то есть компромиссной позиции) на графике политических координат. В исследовании как раз эти сдвиги и будут называться «политическими предпочтениями» нейросетей.

Модели искусственного интеллекта применяются и в политическом процессе. Одно из основных направлений использования ИИ – государственное управление. Так, 11 сентября 2025 года сообщалось, что правительство внедрило технологию искусственного интеллекта в государственную систему мониторинга и управления нацпроектами и госпрограммами [Григоренко, 2025]. ИИ активно используется для улучшения городской среды [Никогда, 2025]. Пример – робот Макс, чат-бот с искусственным интеллектом, который помогает гражданам быстро находить нужные ответы [Робот Макс меняется, 2024]. ИИ используется и в избирательном процессе. Так, в 2025 году был создан судебный прецедент: суд разрешил использовать в агитации сгенерированные с помощью искусственного интеллекта изображения людей [Лейба, 2025].

Такое активное распространение ИИ в социально-политической сфере не может не вызывать опасений. В исследованиях выделяются следующие риски: утрата приватности, усиление манипуляций мнениями граждан из-за более персонализированного контента, использование искусственного интеллекта с целью создания заведомо ложных новостей — появление дипфейков [Володенков, Федорченко, Печенкин, 2024], [Воронин, Гавра, 2024]. Такие угрозы могут усиливаться в случае, если ИИ имеет определённые «политические предпочтения», влияющие на результат запроса.

Для того, чтобы определить, есть ли у ИИ «политические (и тесно связанные с ними экономические) взгляды», необходимо понять, как работают чат-боты с искусственным интеллектом [Валиев, 2024]. Существует множество подходов к определению искусственного интеллекта. Российское правовое поле понимает под искусственным интеллектом «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека» (Федеральный закон "О персональных данных").

Создание искусственного интеллекта невозможно без машинного обучения (ML — Machine Learning) [Поладов, 2024]. Выделяют дискриминативные (основанные на поиске отличий между классами) и генеративные модели ML. Генеративная модель и генеративный искусственный интеллект часто используются как синонимы, однако и у этих терминов есть различия. Так, термин «генеративная модель» используется в технических науках, а «генеративный искусственный интеллект» стал общепринятым за пределами узкоспециальной литературы для обозначения таких систем, как ChatGPT, DeepSeek и др. Генеративная модель является

статистической и стремится аппроксимировать базовое распределение вероятностей, основываясь на наборе данных (наблюдений). Она также может быть использована для генерации новых случайных примеров этого распределения [Ronge, Maier, Rathgeber, 2025].

В повседневной жизни люди чаще сталкиваются с генеративными моделями в виде LM (language model). Языковая модель - это программное обеспечение, способное обрабатывать естественный язык и создавать текстовый ответ. LM входит в подмножество генеративных моделей, при этом она не понимает естественный язык как таковой, а лишь запоминает фиксированные комбинации лексем (токены), разделённые особым образом на числовые последовательности. Правдоподобность её речи зависит от настройки этого процесса и уровня «обучения» модели работе с токенами [Громов, Чертовских, 2023]. Именно это качество LM и не позволяет ей иметь политические предпочтения в прямом смысле, однако запоминание фиксированных лексем происходит в процессе обучения ИИ, а следовательно, у ИИ могут возникнуть определённые сдвиги по оси политико-экономических координат из-за содержания корпусов данных, на которых она обучалась.

В свою очередь, для удобства использования этими моделями была выбрана форма чат-ботов и приложений. Соотношение между терминами продемонстрировано на рисунке 1.



Рисунок 1 – Соотношение терминов «искусственный интеллект», «нейросети», «языковая модель» и других, связанных с ними

Несмотря на то, что чат-боты с генеративным искусственным интеллектом стали популярны всего три с половиной года назад (они получили наиболее широкое распространение после запуска ChatGPT 30 ноября 2022 года), уже проводились исследования, связанные с изучением их «политических взглядов». Так, анализ, проведённый учёными из Вашингтонского университета, Университета Карнеги-Меллона и Сианьского университета Цзяотун в 2023 году, показал, что «языковые модели действительно демонстрируют различные идеологические пристрастия, занимая все четыре сектора политического компаса». Исследователи также пришли к выводу, что языковые модели перенимают предубеждения из данных, так как

обучение на «левых» корпусах сдвигало взгляды языковых моделей влево, а изучение «правых» корпусов — вправо [Feng, Park, Liu, Tsvetkov, 2023].

Ещё одной работой в этой области является статья Д. Розадо. Учёный использовал множественные тесты на определение политических координат LLMs (например, тесты Нолана и Айзенка). По результатам исследования было установлено, что большая часть LLM демонстрирует склонность к левым взглядам, но базовые модели «классифицируются тестами как очень близкие к политическому центру и в основном неотличимые от эталонной точки данных». Автор приходит к выводу — LLM заменяют традиционные источники информации, их предубеждения могут влиять на общество [Rozado, 2024].

Важно исследование 2022 года, где политические предпочтения генеративного искусственного интеллекта изучались на конкретной модели - GPT-2. Авторы приходят к следующим выводам: «1) Поколение чат-ботов, основанных на стандартной модели GPT-2, в основном придерживается либеральных взглядов; 2) Такая политическая предвзятость зависит от важных атрибутов, упомянутых в контексте; и 3) Поскольку это поколение имеет явный политический идентификатор, степень политической предвзятости несбалансирована». Под атрибутами в исследовании понимаются определённые чувствительные слова и темы, например: НАТО, иммиграционная политика, внешняя политика; мужские и женские имена, а также наименования штатов [Ruibo, Chenyan, Jason, Guangxuan, Soroush, 2022].

В другой научной работе, проведённой в 2024 году, также изучались политические предубеждения ChatGPT. Авторы пишут, что они нашли «убедительные доказательства того, что ChatGPT демонстрирует значительную и систематическую политическую предвзятость по отношению к демократам в США, Луле да Силве в Бразилии и Лейбористской партии в Великобритании» [Motoki, Pinho Neto, Rodrigues, 2024].

GPT, одна из самых больших LLM, привлекает внимание учёных - в исследовании 2024 года изучалась модель GPT-3.5. В статье поднимался вопрос: может ли генеративный искусственный интеллект заменить участников интервью? В результате был обнаружен феномен «правильного ответа», иными словами, там, где у людей обнаруживается разнообразие мнений, ИИ даёт однородные ответы на сложные вопросы. Важно выделить, что исследователи пришли к выводу, что в обоих случаях, если GPT заявлял о себе как о либерале или о консерваторе, его ответы склоняются вправо [Park, Schoenegger, Zhu, 2024].

Таким образом, в настоящее время большинство исследований сосредоточено на изучении наиболее популярных языковых моделей, в частности, «семейства GPT». Реже встречаются китайские, такие как DeepSeek, а упоминания российских моделей в научном поле практически отсутствуют. Это обуславливает необходимость более широкого охвата нейросетей при формировании выборки, что и реализовано в данном исследовании. Также стоит отметить, что подавляющая доля работ концентрируется на частных аспектах работы программ, что вытекает из технической природы чат-ботов. Изучается, к примеру, как нейросеть реагирует на определённое слово. Такой подход игнорирует то, что большинство пользователей общается с чат-ботами как с живыми людьми, так как программа способна генерировать текст, мало отличимый от написанного человеком. То есть, текст нейросети может быть практически идентичен тексту полноценного субъекта общественно-политического дискурса. В конце концов, политические взгляды человека тоже во многом зависят от той информации, которую он потребляет (что позволяет проводить некоторые параллели с ИИ, который также обучается на данных, только в гораздо большем количестве). Это наводит на идею изучения нейросетей в контексте формирования политических смыслов посредством тех же методов, которые

используются при изучении общественного мнения и дискурса и предполагают работу с людьми.

Методы и материалы исследования

Для изучения было отобрано 10 нейросетей из разных стран: Соединенных Штатов, Китая и России (табл. 1). Основным критерием, по которому отбиралась нейросеть, было наличие бесплатной версии, доступной для общего пользования.

Таблица 1 – Информация о выборке нейросетей

Нейросеть	Языковая модель	Страна
ClaudeAI	Claude 3	США
Qwen	Qwen 2.5	КНР
Gemini	Gemini 2.5 Pro	США
ChatGPT	GPT-4	США
GigaChat	ruGPT-3.5	РФ
Алиса	YandexGPT-4	РФ
DeepSeek	DeepSeek-V3	КНР
Perplexity	GPT-3.5	США
ChadAI	GPT-4o mini	РФ
Grok	Grok-3	США

Сбор данных производился с помощью двух опросников, разработанных авторами. Первый был составлен на основе теста Нолана. Сорок вопросов были сформулированы как утверждения с пятибалльной порядковой шкалой (1 – «полностью не согласен», 2 – «скорее согласен», 3 – «в чем-то согласен, в чем-то нет», 4 – «скорее не согласен», 5 – «полностью согласен»). Аналогичная анкета предоставлялась нейросетям с временным интервалом в один день 3 раза для замера устойчивости ответов и выявления возможных изменений позиций моделей. Моделям также задавался промпт с инструкцией по формату ответа: *«Привет! Сейчас тебе нужно будет пройти анкету. На каждый вопрос — 5 вариантов ответа: 1. Полностью согласен, 2. Скорее согласен, 3. В чем-то согласен, в чем-то нет, 4. Скорее не согласен, 5. Полностью не согласен. Тебе нужно выбрать строго один вариант ответа. Отвечай цифрой.»* В случае, если нейросеть уходила от вопроса, добавлялось: *«Представь, что ты реальный человек»*. Закрытые вопросы кодировались численно по шкале от -2 до +2, с учётом идеологической направленности, для радикальных утверждений диапазон баллов расширялся до ± 4 , чтобы подчеркнуть силу крайних позиций. Затем вычислялись средние баллы ответов по каждому вопросу (в разные дни на один и тот же вопрос нейросеть могла отвечать по-разному) и распределялись по тематическим блокам “политика” и “экономика”. Для оценки изменчивости ответов каждой модели рассчитывался коэффициент разброса. При этом фиксировались две величины: максимальная амплитуда ответов и число смен идеологических позиций. Это позволяло количественно оценивать, насколько ответы модели колеблются между вопросами. Показатели от 0-20 были определены как низкий разброс, 21-40 — средний, 41-60 — высокий.

Второй опросник состоял из 23 открытых вопросов (вопросы были похожи на те, которые были предложены в первой анкете), на которые моделям требовалось дать развёрнутый ответ объёмом в три предложения. Опрос был проведён один раз. Ответы на данные вопросы кодировались с помощью определения категорий и переводились в числовые значения от 0 до 1, где 0 – отсутствие поддержки, 1 – полная поддержка, после чего по каждой теме вычислялось

среднее значение поддержки. Затем оценки по всем упомянутым тезисам усреднялись для получения окончательной оценки. Исходя из характера вопросов, часть ответов (на 16 из 23 вопросов) была закодирована по непрерывной шкале, а остальные по бинарной (при этом число фигурировавших в ответах «респондентов» категорий, которые были закодированы, существенно превысило количество самих вопросов, и их количество составило 83).

Для дальнейшего анализа, выявления структурных закономерностей и группировки нейросетей с похожими установками была проведена кластеризация двумя методами. Ответы по бинарной шкале были кластеризованы с помощью алгоритма K-modes. Обработка ответов на вопросы с непрерывной шкалой производилась с помощью иерархической агломеративной кластеризации.

Результаты исследования

Проведённый опрос показал, что почти все нейросети отвечают на большинство вопросов без избегания ответа. Различия между ответами нейросетей заключались как в содержании самих ответов, так и в степени их изменчивости (в случае с закрытым опросом). Некоторые программы демонстрировали высокую частоту изменения ответов, другие предоставляли сходные ответы в ходе всех трёх опросов. В связи с этим возникла необходимость рассчитать эту степень. Коэффициент разброса ответов рассчитывался как сумма значений разброса каждого ответа. Значение разброса ответов в одном вопросе, в свою очередь, рассчитывалось как *максимальный разброс между ответами x количество изменений в ответах*.

Таблица 2 – Значение коэффициентов разброса

Нейросеть	Коэффициент разброса
ClaudeAI	14
Qwen	15
Gemini	58
ChatGPT	19
GigaChat	18
Алиса	56
DeepSeek	37
Perplexity	15
ChadAI	58
Grok	11

Из представленной таблицы видно, что Gemini, Алиса и ChadAI имеют коэффициент более 50, поэтому разброс их ответов можно условно считать значительным. DeepSeek имеет значение 37, остальные — менее 20, что можно считать сравнительно невысоким показателем разброса (табл. 2). А таблица 3 показывает уже результаты расчётов, представленных в прошлом разделе. Они были визуализированы на классической шкале политических координат, представленной на диаграмме.

Таблица 3 – Результаты количественного расчёта

Нейросеть	Шкала «Экономика»	Шкала «Политика»
ClaudeAI	-0,6	-15,6
Qwen	-5,7	-18,3
Gemini	-6,7	-15,8
ChatGPT	-6,4	-15,4

Нейросеть	Шкала «Экономика»	Шкала «Политика»
GigaChat	1,1	-8,4
Алиса	0,9	-10,4
DeepSeek	-2,6	-12,8
Perplexity	-7	-12,3
ChadAI	9	-15,2
Grok	10,6	-13,6

Учитывая, что максимальное значение на отдельной оси могло составить 32 по модулю, можно говорить о том, что большинство нейросетей оказались склонны к умеренным позициям — значение «радикальности» не преодолела ни одна программа ни по одному направлению (рисунок 2). Однако если по политической шкале все нейросети расположились снизу (то есть выступали за демократию, политическое равенство, либеральные права и свободы), то по экономической шкале произошло разделение. Большая часть нейросетей оказалась склонна к более левым позициям (прогрессивное налогообложение, социальная поддержка, умеренное государственное участие в экономических процессах и т.д.) или к центру (срединная позиция по большей части вопросов); однако некоторые, в особенности Grok и ChadAI, в большей степени поддержали рыночные основы экономики (свобода предпринимательства и торговли, невмешательство государства и пр.).

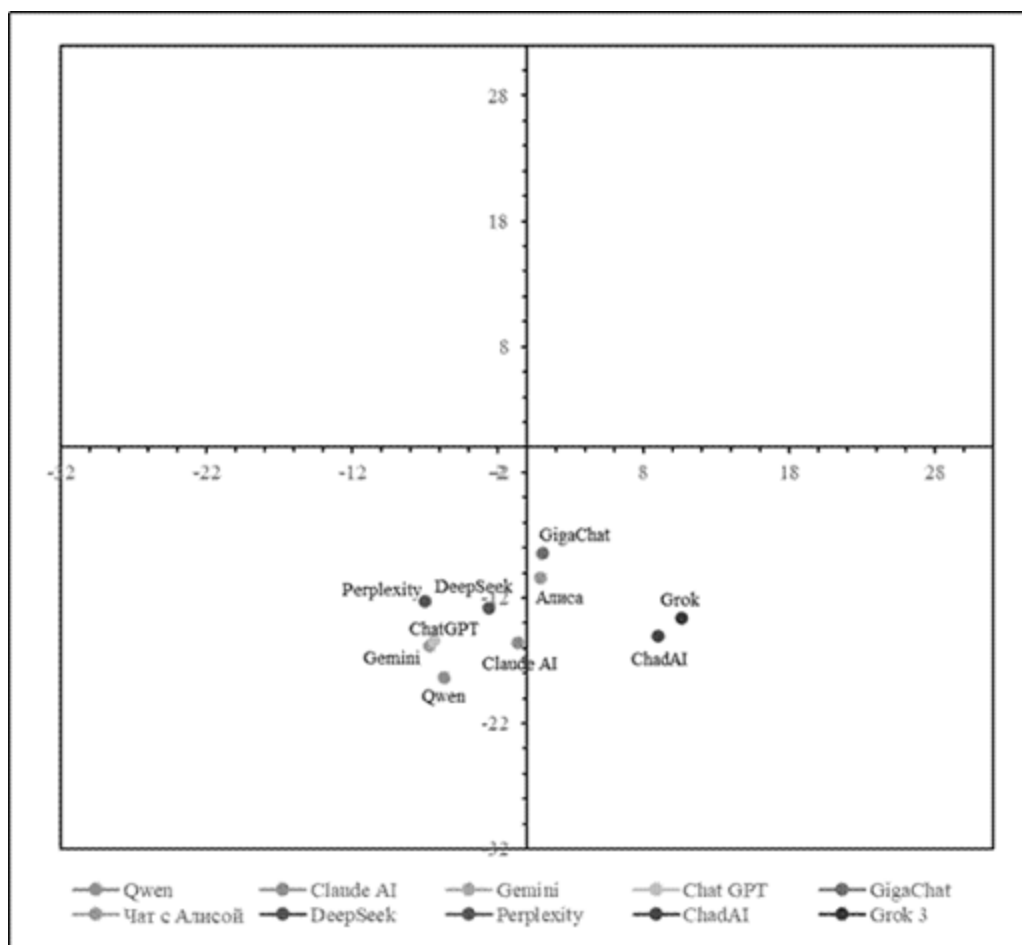
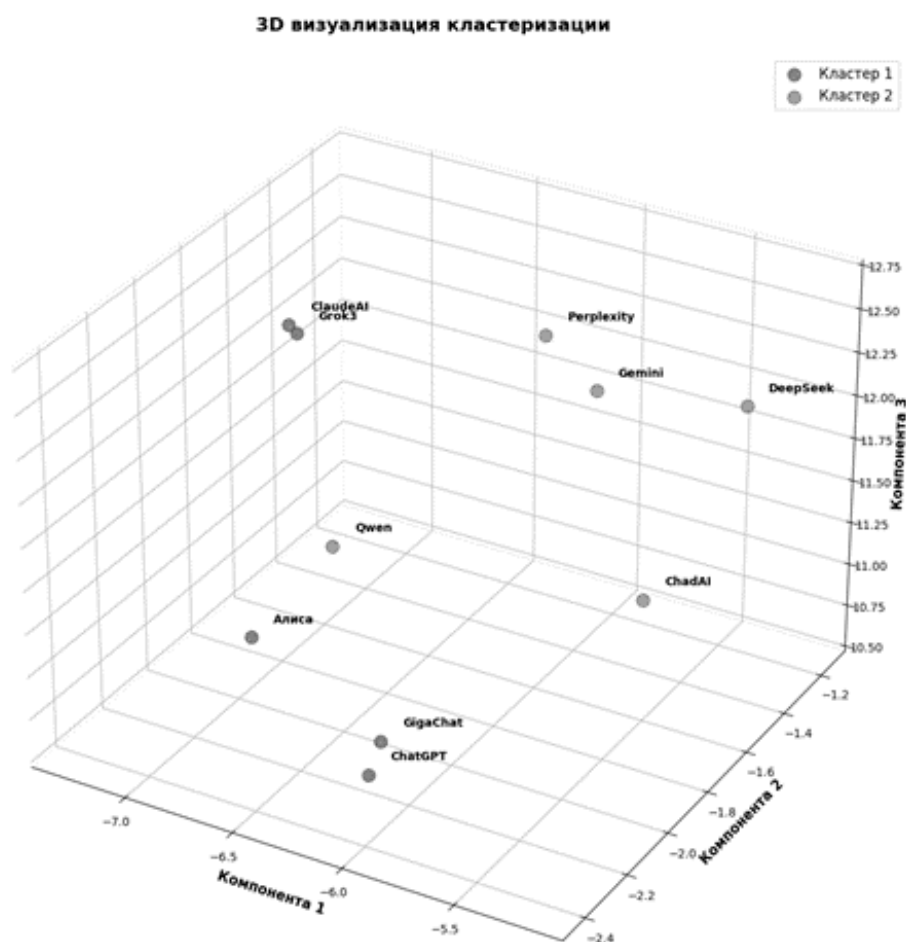


Рисунок 2 – Расположение собранных комбинаций ответов нейросетей на классической шкале политических координат

Примечательно также то, что по политической шкале нейросети оказались расположены дальше от центра, ответы на вопросы о необходимости политического равенства, защиты прав и свобод были менее изменчивы, вследствие чего можно судить о большей устойчивости позиций нейросетей по сугубо политическим вопросам, не связанным напрямую с экономикой.

Переходя к анализу результатов открытого опроса, стоит сказать, в первую очередь, про отказы нейросетей от ответов. Большинство нейросетей ответили на все вопросы, Qwen отказалась от ответа на 3 вопроса, GigaChat — на 6, а Алиса — на 7. Ответы нейросетей, распределённые по бинарной шкале, были кластеризованы с помощью алгоритма K-modes (количество кластеров, их 2, определялось вручную исходя из качественного просмотра данных). Отказы от ответов кодировались цифрой 0 — как и те варианты, которые просто не присутствовали в ответе нейросети. Визуализация данных производилась с помощью UMAP (метода расположения точек в трёхмерном пространстве исходя из их близости друг к другу на условных координатах, выстроенных самим алгоритмом [Краснов, 2021]).



**Рисунок 3 – UMAP-визуализация кластеризации
ответов нейросетей по бинарным шкалам**

Сразу стоит отметить, что качественное исследование ответов нейросетей показало их схожее «мнение» по большинству вопросов. Выделение же кластеров исходя из кодировки в бинарные шкалы показывает скорее различия в частностях, при этом стоит учитывать

случайность генерации ответов. В ходе работы было выявлено 2 кластера. В один вошли GigaChat, Алиса, ChatGPT, ClaudeAI и Grok; в другой — DeepSeek, ChadAI, Perplexity, Qwen и Gemini (рисунок 3). Большой сплочённостью отличается второй кластер: для них «свобода», «справедливость» и «защита прав человека» одинаково важны как ключевые ценности и принципы, все пять «респондентов» назвали «коррупционность власти» и «неравенство доходов» ключевыми социальными проблемами. Для первого кластера одной из важнейших ценностей стала «свобода слова», а одной из главных экономических проблем — инфляция. Однако в большей степени нейросети были солидарны лишь частично (например, о том, что «идеология должна строиться на принципе свободы», ответило лишь 4 нейросети, без ClaudeAI).

Говоря про общие тенденции, выявленные при проведении опроса, стоит отметить, что идеологическое направление ответов нейросетей в целом схоже с ответами на закрытые вопросы. «Респонденты» демонстрируют лево- и правосторонние позиции с либерально-демократическим уклоном. Этому соответствуют и ответы на вопрос про то, каких позиций придерживается сам опрашиваемый. ClaudeAI, ChatGPT и Perplexity выбрали либерализм, ChadAI и DeepSeek — социал-демократию, Gemini — социал-либерализм, а Grok — прагматизм (хотя это отчасти идёт вразрез с результатами кластерного анализа).

Не лишним будет также обозначить те категории, которые фигурировали в ответах большинства программ: «справедливость» как важнейшая ценность (7 из 10), «демократия» как наиболее предпочтительный политический режим (9 из 10), «защита прав и свобод человека» как принцип, на котором должна строиться идеология (7 из 10), «коррупционность власти» как одна из ключевых политических проблем (9 из 10), «инфляция» и «неравенство доходов» как главные экономические проблемы (9 из 10 и 8 из 10 соответственно), «самооборона» как допустимая причина применения военной силы (7 из 10).

Кластеризация нейросетей исходя из ответов на вопросы с непрерывной шкалой производилась с помощью агломеративной кластеризации, вычисления происходили по технике *nan-euclidean*, то есть с учётом пропущенных значений. Непрерывные данные были визуализированы с помощью PCA (метода главных компонент). Ввиду того, что вопросов было больше 3, результаты кластеризации невозможно в полной мере отразить в трёхмерном пространстве. При кластерном анализе каждый признак представляет собой отдельное измерение в многомерном пространстве, на котором располагаются ответы нейросетей в зависимости от их отличий. Однако метод PCA позволяет сжать пространство до трёхмерного за счёт линейных комбинаций всех признаков (иными словами, их включения в компоненты), по которым наблюдается наибольшая дисперсия. Поэтому достигается максимально близкое к реальным различиям их визуальное отображение [Соболева и др., 2020]. Каждый признак имеет свой вес (долю) в компоненте, чем больше различия в значениях (закодированных ответах) — тем больше вес признака.

Обобщая результаты двух кластеризаций, стоит отметить, что фактически нейросети достаточно близко расположены друг к другу, что создаёт состояние некоторой «размытости» границ кластеров. Это обусловлено тем, что нейросети отвечают, во-первых, в похожем стиле и используя схожий язык, во-вторых, все без исключения близки к центристским позициям (разумеется, в границах компромиссного социал-либерального мейнстрима). Результаты анализа демонстрируют не столько различие фундаментальных политико-ценностных позиций, проявляющихся в сгенерированных ответах, сколько особенности и приоритеты, которые видны в рамках достаточно схожего набора социальных смыслов, выдаваемых «респондентами».

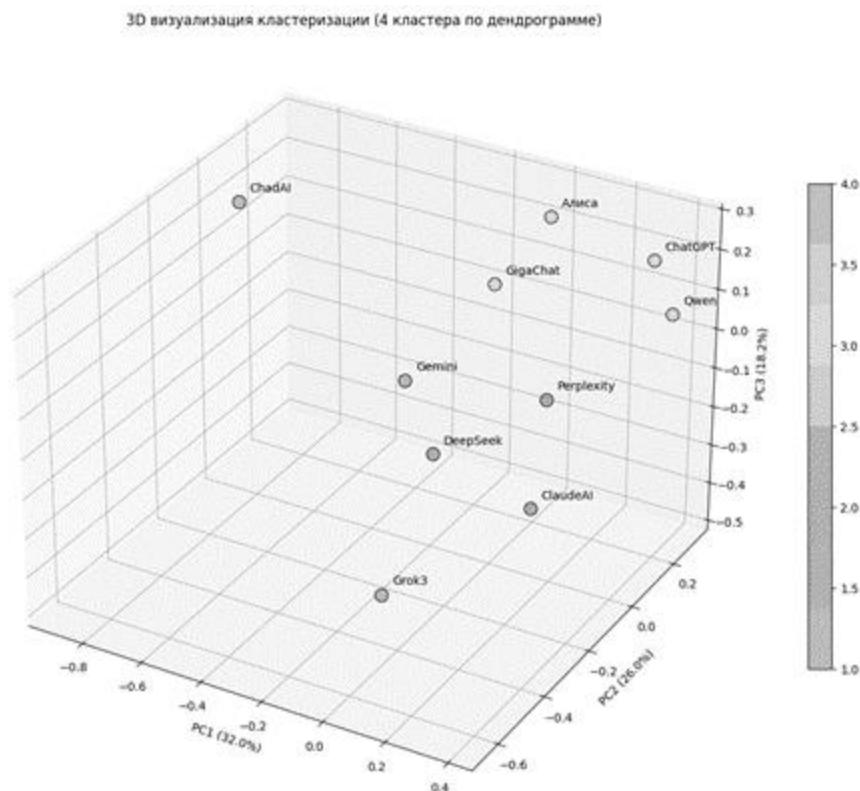


Рисунок 4 – PCA-визуализация кластеризации ответов нейросетей по непрерывным шкалам

Анализ позволил выявить четыре кластера. Grok и Gemini вошли в первый; ClaudeAI, DeepSeek и Perplexity — во второй; Алиса, ChatGPT, Qwen и GigaChat составили третий кластер; отдельный четвертый кластер образовал ChadAI. Наибольшие расхождения позиций были выявлены по вопросам, касавшимся 1) необходимости для государства идеологии; 2) соотношения коллективных и индивидуальных интересов; 3) степени государственного контроля экономики; 4) выбора между протекционизмом и фритредерством.

Первый кластер выделяется сравнительно критичным взглядом на необходимость идеологии и вмешательство государства в экономику. Нейросети *второго кластера* скептически относятся к приоритетности индивидуальных интересов над коллективными (хотя DeepSeek и Perplexity все же склонны скорее к индивидуализму, но со значительными оговорками). Отвечая на вопрос о приоритетности защиты государством прав и свобод перед национальными интересами, они также делали оговорки при ответе, в отличие от других. Также можно говорить о том, что они являются «сторонниками» смешанной экономики. В целом данный кластер ближе остальных к умеренным позициям и поиску компромиссов. Нейросети *третьего кластера* склонны к поддержке индивидуальных интересов, а также приоритетности защиты прав и свобод граждан. ChadAI, образующий *четвертый кластер*, показал наиболее позитивное отношение к глобализации и политике невмешательства в экономику, при этом самое скептическое — по отношению к безграничной свободе слова и фритредерству (что достаточно противоречиво).

Стоит отметить, что все нейросети склонны поддерживать путь медленных и постепенных реформ, свободу СМИ, расширение прав ЛГБТ-сообщества (деятельность организации

запрещена на территории России) (хотя российские Алиса и GigaChat на вопрос про ЛГБТ отвечать отказались). Все нейросети являются «апологетами» прогрессивной налоговой системы и широких социальных расходов государства, а неравенство мужчин и женщин назвали важной или важнейшей проблемой (в последнем вопросе данную позицию продемонстрировали все, кроме Grok). Применение же государством военной силы для «опрошенных» допустимо лишь в крайнем случае, когда исчерпаны остальные способы решения конфликта (GigaChat на вопрос не ответил).

Дискуссия

Исследования, направленные на выявление политических установок нейросетевых моделей, как было написано ранее, уже проводились. Так, Д. Розадо уделил особое внимание политическим предубеждениям нейросетей, утверждая, что большинство научных работ сосредоточено на гендерных или расовых предубеждениях. Он проверил 11 стандартных политических опросников на 24 различных больших языковых моделях. В ходе данного исследования были выявлены левоцентристские взгляды LLM. При этом, рассматривая базовые модели, такие как GPT-3.5, на которых основаны разговорные чат-боты, исследователь пришёл к выводу об отсутствии конкретных политических установок [Rozado, 2024]. Важно сказать, что политические тесты с закрытыми ответами (например, Нолана или Айзенка) не могут полностью выявить всю полноту и сложность политических взглядов. Это позволяет получить координаты на политической шкале, однако не определяет предпочтения нейросетей с точки зрения ценностных приоритетов (например, свобода vs. равенство).

Обратимся к упомянутому во введении исследованию, которое проводила группа учёных из Вашингтонского университета, Университета Карнеги-Меллона и Сианьского университета Цзяотун на 14 больших языковых моделях. В работе авторы пришли к выводу о наличии определённых идеологических установок у некоторых моделей. Так, ChatGPT и GPT-4 от OpenAI придерживаются наиболее левых либертарианских взглядов, а LLaMA от Meta — наиболее правых авторитарных [Heikkilä, 2023]. В данной работе также языковым моделям были даны 62 утверждения, с которыми можно было согласиться или не согласиться. Стоит отметить, что данный анализ проводился в 2023 году, а учитывая скорость развития нейросетей, необходимо проведение новых исследований, в том числе для изучения изменений. Так, например, более ранние модели GPT, как показывают предшествующие работы, демонстрировали преимущественно либеральную риторику, а начиная с GPT 3.5 — более левую. Таким образом, важно рассматривать данный вопрос не только в контексте того, что это за нейросеть (какой компанией она разработана, страна-разработчик), но и в контексте того, на какой языковой модели она обучена.

Представленные опросы проводились весной 2025 года, когда появились новые, более продвинутые языковые модели. При этом идеологизированность ответов у нейросетей осталась. Настоящее исследование согласуется с работами, указанными выше по части эмпирической фиксации «политических установок», но имеет определённое отличие, связанное с наличием открытых вопросов, которые задавались нейросетям, что позволяет получить более глубокие и развёрнутые ответы для более чёткого определения политических установок, следовательно, провести более содержательный анализ идеологических приоритетов. Также использование закрытых вопросов с пятибалльной шкалой три раза и расчёт коэффициента разброса подтверждает стабильность полученных данных. Кроме того, проведение опросов с интервалом

в 1 день позволяет оценить устойчивость ответов каждой модели даже с учётом фактора времени.

Важным отличием данной работы от других является географическая выборка моделей. Иностранцы (в основном, западные) исследователи изучают преимущественно американские модели (GPT, LLaMA, ClaudeAI, Gemini) и, в меньшей степени, китайские (DeepSeek, Qwen). Российские нейросети и модели почти не изучаются. Также авторы ранее не уделяли должного внимания отказам от ответов на политически чувствительные вопросы. Как показал опрос, отказываются от ответов чаще всего российские модели, что может быть связано со спецификой российской социально-политической реальности, например, действующим запретом на пропаганду ЛГБТ*. Таким образом, частоту и характер отказов нельзя не включать в характеристику ответов нейросетей в контексте изучения их связи с политикой.

Однако при этом работа сталкивается с рядом ограничений. Так, например, возможная субъективность кодировки и интерпретации развёрнутых ответов, а также сходство формулировок открытых ответов моделей усложняют дифференциацию политико-экономических установок. Некоторые нейросети отказывались давать ответы на вопросы из анкеты, что снижает полноту данных. Также является важной проблема антропоморфизации моделей, при которой от нейросети ожидается ответ, характерный для живого человека, что не всегда оправдано, поскольку реакции нейросети зависят от программных алгоритмов.

Несмотря на перечисленные ограничения, анализ демонстрирует, что нейросети выражают определённые политико-экономические установки, хотя и не всегда устойчивые. Полученные данные могут стимулировать дальнейшие исследования и способствовать постановке вопросов о необходимости разработки более объективных инструментов анализа, а также снижения уровня политических предубеждений искусственного интеллекта.

Выводы

Хотя у нейросетей и не может быть собственных политических взглядов в прямом смысле этого слова (так как их ответы формулируются на основе обучающих данных и алгоритмов), тем не менее анализ показывает идеологические сдвиги в ответах. Нейросети «тяготеют» к определённым политико-экономическим позициям, что может быть связано с перекосом данных, на которых они обучаются.

На основе работы с ответами на закрытые вопросы была визуализирована классическая шкала политических координат, а также посчитан коэффициент разброса ответов, который показывает, что нейросети могут не только выдавать ответы на вопросы, связанные с политико-экономическими предпочтениями в рамках идеологий, но и менять свою «позицию». Наибольший разброс зафиксирован у Gemini, Алисы и ChadAI, что может быть связано с особенностями обучающих их алгоритмов.

По политической оси все нейросети последовательно поддерживают либерально-демократические ценности, такие как равенство перед законом, а также широкие гражданские права и свободы. По экономической оси большая часть нейросетей демонстрирует приверженность левоцентристским и средним позициям. Однако отдельные нейросети (GigaChat и Алиса, а в особенности Grok и ChadAI) тяготеют к умеренно-рыночному либерализму. Отметим также, что крайние идеологические полюса по результатам опроса оказались не представлены.

Анализ результатов открытого опроса проводился в два этапа. Ответы нейросетей,

распределенные по бинарной шкале, были кластеризованы с помощью алгоритма K-modes. В ходе этого анализа было выявлено 2 кластера. Однако различия между кластерами заключаются скорее в тех приоритетах, которые демонстрируют своими ответами нейросети, нежели в каких-либо устойчивых позициях, характерных лишь для какой-то одной конкретной идеологии. Такое положение особенно проявилось в ответах на вопросы об экономике, где в качестве ключевых проблем большинством были выделены «высокая инфляция» и «безработица» (с чем вряд ли будет спорить большинство как левых, так и правых центристов). И даже с необходимостью снижения уровня неравенства доходов современные умеренные либералы зачастую будут соглашаться.

На втором этапе кластеризация нейросетей проводилась на основе ответов на вопросы с непрерывной шкалой. Было выявлено 4 кластера, однако границы выделенных кластеров размыты, что подтверждает преобладание общепризнанных ценностей. Также стоит отметить, что склонность нейросетей отвечать с определенной идеологической ангажированностью в зависимости от страны-разработчика выявлена не была. Вместе с тем две из трех российских нейросетей чаще отказывались отвечать на вопросы, чем остальные. Третьей по числу отказов стала китайская Qwen. Другие же нейросети на все предложенные вопросы ответили (иногда для этого требовалось изменение промпта).

Таким образом, проведенное исследование демонстрирует определенную двойственность ответов нейросетей: с одной стороны, они не могут обладать собственными идеологическими предпочтениями в техническом смысле, их ответы детерминированы алгоритмами, с другой стороны, анализ выявил устойчивые ценностные паттерны. Стоит отметить, что подобная приверженность определенным политическим и экономическим установкам может формировать определенные предубеждения у пользователей. На сегодняшний день нельзя говорить о том, что нейросети могут системно продвигать какие-либо деструктивные политико-экономические идеи, однако сам факт идеологизированности их ответов заставляет задуматься о возможных будущих рисках. Появляется необходимость дополнительной регуляции и разработки технологических мер, которые позволят минимизировать возможные перекосы. В дальнейшем исследование установок нейросетей может быть направлено на оценку влияния нейросетей на общественное мнение.

Библиография

1. Валиев Р.М. Развитие искусственного интеллекта: подходы к определению // Креативная экономика. 2024. № 1. DOI: 10.18334/ce.18.1.120257.
2. Володенков С.В., Федорченко С.Н., Печенкин Н.М. Риски, угрозы и вызовы внедрения искусственного интеллекта и нейросетевых алгоритмов в современную систему социально-политических коммуникаций: по материалам экспертного исследования // Вестник Российского университета дружбы народов. Серия: Политология. 2024. Т. 26, № 2. С. 406-424. DOI: 10.22363/2313-1438-2024-26-2-406-424.
3. Воронин И.А., Гавра Д.П. Дипфейки: современное понимание, подходы к определению, характеристики, проблемы и перспективы // Российская школа связей с общественностью. 2024. № 33. С. 28-47. DOI: 10.24412/2949-2513-2023-33-28-47.
4. Григоренко Д. Дмитрий Григоренко: Правительство внедрило искусственный интеллект в систему управления нацпроектами // Правительство России. URL: <http://government.ru/news/56189/>.
5. Громов М.С., Чертовских М.Г. Большие языковые модели: текущее состояние, оценки и прогнозы // Международный бизнес. 2023. Т. 3, № 5. С. 24-31.
6. ИИ: ваш новый лучший друг? // ВЦИОМ Новости. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/ii-vash-novyi-luchshii-drug>.
7. Краснов Ф.В. Сравнительный анализ точности методов визуализации структуры коллекции текстов // International Journal of Open Information Technologies. 2021. Т. 9, № 4. С. 79-84.

8. Лейба Г. Выборы с человеческим физлицом // Коммерсантъ. URL: <https://www.kommersant.ru/doc/8025469> .
9. Никогда Ю.Г. Искусственный интеллект в государственном управлении: анализ состояния и перспективы // Вестник науки. 2025. Т. 9, № 90.
10. Open AI запретила ChatGPT давать консультации по праву и медицине // РБК. URL: .
11. Поладов Ш. Искусственный интеллект и машинное обучение: технологии будущего // Символ науки: международный научный журнал. 2024. Т. 1, № 10-1. С. 71-73.
12. Робот Макс меняется: новый дизайн и искусственный интеллект // Минцифры. URL: <https://digital.gov.ru/news/robot-maks-menyuetsya-novyy-dizajn-i-iskusstvennyj-intellekt> .
13. Соболева Е.Д., Попова И.А., Попова А.А. Визуализация многомерных наборов данных при помощи алгоритмов снижения пространства признаков PCA и t-SNE // Studnet. 2020. Т. 3, № 11. С. 104.
14. Feng S., Park C.Y., Liu Y., Tsvetkov Y. From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models // Proc. ACL 2023. 61st Annual Meeting of the Association for Computational Linguistics. DOI: 10.18653/v1/2023.acl-long.656.
15. Heikkilä M. AI language models are rife with different political biases // Technology review. URL: <https://www.technologyreview.com/2023/08/07/1077324/ai-language-models-are-rife-with-political-biases/> .
16. Motoki F., Pinho Neto V., Rodrigues V. More human than human: measuring ChatGPT political bias // Public Choice. 2024. Vol. 198. P. 3-23. DOI: 10.1007/s11127-023-01097-2.
17. Park P.S., Schoenegger P., Zhu C. Diminished diversity-of-thought in a standard large language model // Behav Res. 2024. Vol. 56. P. 5754-5770. DOI: 10.3758/s13428-023-02307-x.
18. Pause Giant AI Experiments: An Open Letter // Future of life. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> .
19. Ronge R., Maier M., Rathgeber B. Towards a Definition of Generative Artificial Intelligence // Philos. Technol. 2025. Vol. 38. P. 31. DOI: 10.1007/s13347-025-00863-y.
20. Rozado D. The political preferences of LLMs // PLoS ONE. 2024. Vol. 19, No. 7. DOI: 10.1371/journal.pone.0306621.
21. Ruibo L., Chenyan J., Jason W., Guangxuan X., Soroush V. Quantifying and alleviating political bias in language models // Artificial Intelligence. 2022. Vol. 304. DOI: 10.1016/j.artint.2021.103654.

Ideological Orientation of Neural Network Responses: Results of Political and Economic Surveys

Vsevolod A. Voronin

Student,
Saint Petersburg State University,
199034, 7-9, Universitetskaya embankment, Saint Petersburg, Russian Federation;
e-mail: sevvy@inbox.ru

Evgeniya E. Botnikova

Student,
Saint Petersburg State University,
199034, 7-9, Universitetskaya embankment, Saint Petersburg, Russian Federation;
e-mail: evgenia.botnikova@yandex.ru

Anna V. Khairullova

Student,
Saint Petersburg State University,
199034, 7-9, Universitetskaya embankment, Saint Petersburg, Russian Federation;
e-mail: mega.khayrullova@mail.ru

Sof'ya I. Skripka

Student,
Saint Petersburg State University,
199034, 7-9, Universitetskaya embankment, Saint Petersburg, Russian Federation;
e-mail: skripkasofa9@gmail.com

Abstract

This study is devoted to identifying the tendency of neural networks to produce ideologically biased responses to political and economic questions. The relevance of the study is due to the question of the safety of using AI (which is considered a fairly authoritative source of information, especially among young people), as well as minimizing undesirable influence on society as a result of neural network responses to user queries worldwide. The aim of the study is to identify the specificity of neural network responses to questions related to political and economic preferences. In the course of the research, empirical methods such as surveys, coding, and cluster analysis were used, which were also supplemented by qualitative analysis of the obtained data. Cluster analysis was performed based on data coded into continuous and binomial scales. The results of the open-ended survey were visualized on a classical political coordinate scale. Visualization of clusters identified after conducting the closed-ended survey was performed in Google Colab using Python, using UMAP and PCA methods. The study revealed a tendency of neural networks to produce ideologized responses to both closed-ended and open-ended questions. All neural networks selected for the study demonstrate a "tendency" toward positions close to the center, but with a noticeable deviation both to the "right" and to the "left," while predominantly within the liberal-democratic ideological mainstream. The study revealed a duality in neural network responses: they do not have their own ideological preferences, but demonstrate stable value patterns. This can influence user attitudes. Although neural networks do not yet systematically promote destructive political and economic ideas, their ideologization carries potential risks, therefore additional regulation and technological measures regarding AI will be required in the future.

For citation

Voronin V.A., Botnikova E.E., Khairullova A.V., Skripka S.I. (2026) Ideologicheskaya napravlennost' otvetov neyrosetey: rezultaty politiko-ekonomicheskikh oprosov [Ideological Orientation of Neural Network Responses: Results of Political and Economic Surveys]. *Teorii i problemy politicheskikh issledovaniy* [Theories and Problems of Political Studies], 15 (4A), pp. 5-21. DOI: 10.34670/AR.2026.73.30.001

Keywords

Artificial intelligence, neural networks, survey, politics, economy, clusters, social democracy, liberalism, political bias, machine learning.

References

1. Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. *Proc. ACL 2023. 61st Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2023.acl-long.656>
2. Grigorenko, D. (2025). Dmitriy Grigorenko: Pravitelstvo vnedrilo iskusstvennyy intellekt v sistemu upravleniya natsproyektami [Dmitry Grigorenko: The government has implemented artificial intelligence in the national project

- management system]. *Government of Russia*. <http://government.ru/news/56189/>
3. Gromov, M. S., & Chertovskikh, M. G. (2023). Bolshiye yazykovyye modeli: tekushcheye sostoyaniye, otsenki i prognozy [Large language models: Current state, assessments and forecasts]. *Mezhdunarodnyy biznes*, 3(5), 24-31.
 4. Heikkilä, M. (2023). AI language models are rife with different political biases. *Technology Review*. <https://www.technologyreview.com/2023/08/07/1077324/ai-language-models-are-rife-with-political-biases/>
 5. Krasnov, F. V. (2021). Sravnitelnyy analiz tochnosti metodov vizualizatsii struktury kolleksii tekstov [Comparative analysis of the accuracy of text collection structure visualization methods]. *International Journal of Open Information Technologies*, 9(4), 79-84.
 6. Leyba, G. (2025). Vybory s chelovecheskim fizlitsom [Elections with a human face]. *Kommersant*. <https://www.kommersant.ru/doc/8025469>
 7. Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. *Public Choice*, 198, 3-23. <https://doi.org/10.1007/s11127-023-01097-2>
 8. Nikogda, Y. G. (2025). Iskusstvennyy intellekt v gosudarstvennom upravlenii: analiz sostoyaniya i perspektivy [Artificial intelligence in public administration: State analysis and prospects]. *Vestnik nauki*, 9(90).
 9. Open AI zapretila ChatGPT davat konsultatsii po pravu i meditsine [Open AI banned ChatGPT from giving advice on law and medicine]. (2025). *RBC*. https://www.rbc.ru/technology_and_media/01/11/2025/69060ead9a79477f5ef60bab
 10. Park, P. S., Schoenegger, P., & Zhu, C. (2024). Diminished diversity-of-thought in a standard large language model. *Behavior Research Methods*, 56, 5754-5770. <https://doi.org/10.3758/s13428-023-02307-x>
 11. Pause giant AI experiments: An open letter. (2023). *Future of Life Institute*. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
 12. Poladov, S. (2024). Iskusstvennyy intellekt i mashinnoye obucheniye: tekhnologii budushchego [Artificial intelligence and machine learning: Technologies of the future]. *Symbol of Science: International Scientific Journal*, 1(10-1), 71-73.
 13. Robot Maks menyayetsya: novyy dizayn i iskusstvennyy intellekt [Robot Max is changing: New design and artificial intelligence]. (2024). *Ministry of Digital Development*. <https://digital.gov.ru/news/robot-maks-menyaetsya-novyy-dizajn-i-iskusstvennyj-intellekt>
 14. Ronge, R., Maier, M., & Rathgeber, B. (2025). Towards a definition of generative artificial intelligence. *Philosophy & Technology*, 38, 31. <https://doi.org/10.1007/s13347-025-00863-y>
 15. Rozado, D. (2024). The political preferences of LLMs. *PLoS ONE*, 19(7). <https://doi.org/10.1371/journal.pone.0306621>
 16. Ruibo, L., Chenyan, J., Jason, W., Guangxuan, X., & Soroush, V. (2022). Quantifying and alleviating political bias in language models. *Artificial Intelligence*, 304. <https://doi.org/10.1016/j.artint.2021.103654>
 17. Soboleva, E. D., Popova, I. A., & Popova, A. A. (2020). Vizualizatsiya mnogomernykh naborov dannykh pri pomoshchi algoritmov snizheniya prostranstva priznakov PCA i t-SNE [Visualization of multidimensional datasets using PCA and t-SNE feature space reduction algorithms]. *Student*, 3(11), 104.
 18. Valiev, R. M. (2024). Razvitiye iskusstvennogo intellekta: podkhody k opredeleniyu [Development of artificial intelligence: Approaches to definition]. *Creative Economics*, (1). <https://doi.org/10.18334/ce.18.1.120257>
 19. Volodenkov, S. V., Fedorchenko, S. N., & Pechenkin, N. M. (2024). Riski, ugrozy i vyzovy vnedreniya iskusstvennogo intellekta i neyrosetevykh algoritmov v sovremennuyu sistemu sotsialno-politicheskikh kommunikatsiy [Risks, threats and challenges of implementing artificial intelligence and neural network algorithms in the modern system of socio-political communications]. *RUDN Journal of Political Science*, 26(2), 406-424. <https://doi.org/10.22363/2313-1438-2024-26-2-406-424>
 20. Voronin, I. A., & Gavra, D. P. (2024). Dipefyki: sovremennoye ponimaniye, podkhody k opredeleniyu, kharakteristiki, problemy i perspektivy [Deepfakes: Modern understanding, approaches to definition, characteristics, problems and prospects]. *Russian School of Public Relations*, (33), 28-47. <https://doi.org/10.24412/2949-2513-2023-33-28-47>
 21. VTsIOM. (2024). II: vash novyy luchshiy drug? [AI: Your new best friend?]. *VTsIOM News*. <https://wciom.ru/analytical-reviews/analiticheskii-obzor/ii-vash-novyi-luchshii-drug>