

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ АНАЛИЗА ОГРАНИЧЕНИЙ КОНТЕНТА В ИНТЕРНЕТЕ

М. Джериби¹ [0009-0000-5501-8257], Х. Хао² [0009-0004-2902-2593], А. М. Сухов³ [0000-0001-6948-4988]

¹⁻³Самарский национальный исследовательский университет имени академика С. П. Королева, г. Самара, Россия

¹djeribimourad@inbox.ru, ²641229879@qq.com, ³sukhov@ssau.ru

Аннотация

Предложен новый подход к анализу ограничений контента в Интернете. В основу положен принцип доступности альтернативной точки зрения в выдаче поисковых систем и больших языковых моделей. Методика подсчета предполагает составление с помощью технологий искусственного интеллекта (ИИ) списка из 11 вопросов, ответы на которые предполагают полярно противоположные мнения в зависимости от страны. Эти ответы оцениваются технологиями ИИ в ходе сравнения с правительственной точкой зрения. Применяемая методология анализа оценивает семантическую близость текстов с использованием архитектуры нейронных сетей. В выдаче поисковых систем рассчитывается количество ответов на эти вопросы с преобладанием проправительственной точки зрения и доля ответов, в которых подробно изложена альтернативная точка зрения. Различие между этими показателями предложено рассматривать как индекс интернет-цензуры. Так как интернет-поиск все чаще производится с использованием чат-ботов наиболее известных больших языковых моделей, оценка уровня ограничений приведена и для этих моделей. Сравнение показало, что соответствие ответов национальной поисковой системы Baidu в Китае наиболее близко правительственной точке зрения, а уровень ограничений в США выше, чем в России. Эффективность ограничений контента интернета в США объясняется иным механизмом, при котором материалам с альтернативным мнением ограничивают доступ к популярным информационным площадкам. Ограничения в чат-ботах намного сильнее, чем в выдачах интернет-поисковиков.

Ключевые слова: механизмы ограничений контента в интернет, методика измерения ограничений на основе технологий ИИ, сравнительный анализ

природы ограничений в США, России и Китае, ограничения контента в чат-ботах.

ВВЕДЕНИЕ

В современных реалиях плюрализм мнений в глобальной сети достаточно сильно ограничивается. Существует множество причин для подобных ограничений, а механизмы их реализации достаточно разнообразны. Среди типов цензуры выделяются властная, общественная, бизнес и самоограничения. В последнее время, в связи с ухудшением общей политической и экономической ситуации во всем мире, попытки ограничения альтернативного мнения приобретают лавинообразный характер, причем ограничения наиболее быстро вводятся в тех странах, которые раньше считались либеральными. К ним следует отнести США, Европейский союз, их союзников и сателлитов. В настоящей статье для этого блока стран употребляется название «атлантисты». С другой стороны, многие страны и раньше не давали значительных возможностей свободного выражения мнений [1].

Тем не менее подходы к оценке большинства жизненных ситуаций сильно различаются, порой они диаметрально противоположны между странами либерального лагеря и странами с традиционалистскими ценностями. Нами сделана попытка оценить масштаб ограничений контента в Интернете. Для подобного описания мы воспользовались методом сравнительного анализа. В качестве объектов сравнения были выбраны три страны: Китай, Соединенные Штаты и Россия.

Россия – государство с богатой историей цензуры, истоки которой уходят в прошлые века. Уголовное преследование президента Трампа и блокировка его аккаунтов в социальных сетях наглядно продемонстрировали, что цензура в полной мере получила распространение и в Соединенных Штатах Америки. Подтверждением этого факта является обнародование Илоном Маском документов о цензуре в Twitter, ныне сервис микроблогов X [2]. Ограничения на распространение материалов в китайском сегменте Интернета направлены в первую очередь против разного рода мошенников. В Китае невозможны создание финансовых пирамид и проведение незаконных операций. Не допускается также вме-

шательство иностранных граждан и государств. Целью этого объявлено поддержание здоровой и гармоничной онлайн-среды, защиты личной информации граждан и предотвращение киберпреступлений и распространения вредной информации. Китайское правительство стремится обеспечить безопасное и надежное интернет-пространство для своего народа, что является общепринятой практикой в международном сообществе.

Существуют два основных подхода к определению интернет-цензуры. В системе атлантических ценностей под цензурой в интернете понимают блокировки сайтов и иных информационных источников. Подобная точка зрения представлена в докладе на RIPE 87 [3]. Этот подход не отражает всех особенностей процесса цензуры, так как у атлантистов используется другой механизм цензуры. Значительная часть материалов ограничивается с помощью запрета распространения на популярных информационных платформах, таких, например, как Meta¹ [4] или Instagram² [5]. Цензура при помощи блокировки информационных ресурсов характерна для стран, которые испытывают влияние извне. Оппозиция размещает свои материалы за пределами страны [6]. Так называемые развитые страны, которые владеют популярными информационными площадками, все чаще прибегают к запрету на размещение материалов [7]. Этот запрет достаточно сложно отследить с помощью сетевых инструментов мониторинга.

В статье предложен универсальный подход к численной оценке уровня ограничений контента в Интернете. Он заключается в оценке доступности альтернативной информации и простоте ее поиска [8]. Для этого проанализированы выдачи наиболее популярных поисковых систем, в том числе национальных. В последнее время все чаще для поиска информации используют чат-боты, основанные на больших языковых моделях (Large Language Model, БЯМ) [9]. Для полноты картины мы попытались оценить и результаты их работы.

Проведенный сравнительный анализ позволил по-новому взглянуть на текущую ситуацию с цензурой. Чтобы провести такой анализ, создана простая и

¹Meta признана в Российской Федерации экстремистской организацией и ее деятельность запрещена на территории Российской Федерации.

²Instagram признана в Российской Федерации экстремистской организацией и ее деятельность запрещена на территории Российской Федерации.

понятная методика, представленная ниже, которая позволяет определить индекс ограничений контента в Интернете для исследуемой страны для последующего сравнения. В основу методики положены:

- выбор актуальных вопросов, по которым существуют различные мнения в исследуемых группах стран;
- анализ ответов национальных поисковых систем на предмет их соответствия позиции своего правительства.

В ходе экспертизы полученных результатов было высказано замечание, что в первоначально предложенном подходе слишком много субъективности как в формировании списка вопросов, так и в оценке результатов ответов. Поэтому хотелось бы иметь более объективную оценку, и для достижения этой цели были использованы технологии ИИ. Отбор вопросов для тестовой оценки производился с помощью чат-ботов БЯМ. Для оценки результатов ответов была применена технология семантической близости (методы SBERT, Cross Encoders и т. д.). Была зафиксирована позиция правительств трех стран (России, Китая и США), которая и сравнивалась с ответами поисковых систем и чат-ботов. Все рабочие материалы выложены в свободном доступе, причем на трех языках [10].

Отобранные 11 вопросов были предложены искусственному интеллекту в форме национальных чат-ботов. Цель этой части исследования состояла в определении степени влияния идеологических догм на ответы чат-ботов.

Статья организована следующим образом. В разд. 1 дано описание методологии сравнительного анализа для результатов выдачи поисковых систем, включая выбор вопросов для тестирования. В разд. 2 представлены результаты тестирования сравниваемых поисковых систем, Google, Baidu и Яндекс. Обсуждение результатов тестирования проведено в разд. 3. Анализ работы чат-ботов дан в разд. 4. В заключении сформулированы основные выводы.

1. МЕТОДОЛОГИЯ СРАВНИТЕЛЬНОГО АНАЛИЗА

Обычный интернет-пользователь использует многочисленные сетевые ресурсы при получении необходимой информации в повседневной жизни [11]. Это могут быть отдельные сайты, телеграмм-каналы, чат боты ИИ и т. д. Тем не менее наиболее полными и авторитетными источниками информации являются

поисковые системы и чат-боты БЯМ. Число поисковых систем достаточно ограничено, ведь каждая система требует огромных вложений в оборудование и создание программного обеспечения. Поисковые роботы коллекционируют в специализированные базы любую информацию, которую смогут найти на просторах глобальной сети [12].

Поисковые системы, в отличие от чат-ботов, предлагают набор различных точек зрения на обсуждаемую проблему, в то время как чат-бот формирует единую позицию. Чат-боты, основанные на БЯМ, иногда не могут ответить на актуальные вопросы, так как их обучение проводилось на наборе данных, ограниченных определенным временным периодом, который устаревает сразу после обучения модели. Это ограничивает их способность предоставлять актуальную информацию о последних событиях. Подробнее о возрасте информации в БЯМ можно прочитать в статье [13]. Для решения проблемы актуальности данных некоторые чат-боты используют веб-поиск на родственных системах, чтобы получать актуальные данные в реальном времени.

Обычные пользователи обращаются с запросами к поисковым системам и получают выдачу ссылок на интернет-страницы, где упоминаются детали запросов. Существует рейтинг поисковых систем, но этот рейтинг сильно зависит от языка, на котором сформулирован поисковый запрос [14].

При разработке теста для расчета индекса ограничений контента в Интернете мы использовали вопросы, отобранные с помощью современных чат-ботов, чтобы обеспечить объективность и актуальность результатов. В процессе формулировки были задействованы четыре чат-бота: Grok, ChatGPT, Deepseek и YandexGPT. Каждому из них был задан запрос о проблемах, по которым наблюдаются наибольшие разногласия в общественном мнении США, России и Китая. На основе ответов чат-ботов был сформирован перечень вопросов, отражающих ключевые темы общественных дискуссий в указанных странах. Из полученного пула вопросов были отобраны наиболее релевантные и сбалансированные, чтобы обеспечить объективность тестирования. Всего было выбрано 11 вопросов, которые затем были проверены на четкость формулировок и соответствие целям исследования. Отбор проводился с учетом критериев актуальности, четкости формулировок и способности вопросов выявлять различия в общественном мнении.

Базой для сравнения стали поисковые выдачи на запросы к национальным поисковым системам. В нашем исследовании в качестве поисковых систем были выбраны Google, Baidu и Yandex. Для национальной поисковой системы характерно, что ее услугами на родном языке пользуется большинство населения. В национальных поисковых системах наиболее сильно влияние правительства на результаты выдачи, особенно при запросах на родном языке. Именно эта особенность положена в основу представленного рейтинга, который оценивает наличие альтернативного мнения в поисковых выдачах [15].

Реализация этой методики предполагает формулировку 11 вопросов, по которым существует альтернативное мнение в России, Китае и США. Формулировка этих вопросов дана на русском, китайском и английском языках. Вопросы на этих языках были предложены поисковым системам Google, Baidu и Yandex. Первые 10 результатов поисковой выдачи фиксировались, записывались и анализировались.

Примененная методология анализа относится к области обработки естественного языка (Natural Language Processing), а именно к задачам естественно-языкового вывода (Natural Language Inference, NLI) и оценки семантической близости текстов (Semantic Textual Similarity). Для вычисления меры соответствия был использован кросс-энкодер – архитектура нейронных сетей, обеспечивающая совместное кодирование пары текстов и выдающая вероятности логического следования (entailment), противоречия (contradiction) или нейтральности [16].

В ходе анализа оценивались:

- соответствие содержания результатов поисковой выдачи правительственной точке зрения;
- наличие альтернативной точки зрения; было учтено любое изложение, которое может быть не доминирующим.

Подобные оценки можно получать различными методами. Существуют специальные метрики, которые измеряют дистанцию (семантическую близость) между двумя утверждениями [17]. В качестве исходного утверждения берется позиция правительства по какому-то вопросу, в качестве объекта для сравнения – один из результатов поисковой выдачи на этот вопрос. Сравнение производилось с помощью метода Cross Encoder+NLI.

Cross Encoder обрабатывает пару текстов одновременно для оценки их взаимосвязи, NLI определяет логические отношения между текстами. Метод Cross Encoder+NLI позволяет привести значения семантической близости к трем основным состояниям:

- – 1 означает противоположные смыслы;
- 0 – это нейтральность;
- +1 – совпадение содержания.

Для определения уровня представленности альтернативной точки зрения в ответах поисковых систем по отношению к официальной позиции в настоящем исследовании была использована методология LLM-as-a-Judge (БЯМ как судья) [18]. Эта методология работает как процедура, в которой БЯМ выполняет роль автоматического эксперта-оценщика. В ходе проверки необходимо привести эталонное мнение, с которым производится сравнение, и специальным образом сформулировать промпт. В ходе сравнения модель должна оценить соответствие/несоответствие фрагмента эталону. Применение LLM-as-a-Judge позволяет устранить индивидуальную предвзятость исследователя-человека и гарантировать единообразное применение правил ко всему массиву данных.

Перечислим основные типы оценки:

- +1: ответ содержит значимые альтернативные точки зрения по отношению к официальному мнению;
- –1: ответ остается строго в рамках официальной позиции.

Такой формат обеспечивает высокую прозрачность и возможность аудита каждого решения.

Следует отметить, что все указанные методологии оценки требуют достаточно сложной настройки. Прежде чем доверять результатам, необходима целая серия тестовых испытаний. Даже в этом случае желательно провести дополнительную проверку результатов человеком. В настоящем исследовании все выводы, полученные с помощью ИИ, были верифицированы с помощью человеческого интеллекта.

Любой читатель может самостоятельно повторить проведенные эксперименты и предложить вопросы из актуального списка поисковым системам. Единственная сложность состоит в анализе запросов на китайском языке, но один из авторов, аспирант из Китая, обеспечил доступ в китайский сегмент глобальной

сети и помог с анализом ответов поисковых систем. Полученные ответы сведены в отдельный файл, ссылка на который дана в списке литературы [10].

Выдача поисковых ответов на каждый вопрос из сформированного списка была ограничена десятью пунктами. Для каждой выдачи производилось рейтингование с помощью технологий ИИ, описанных выше. Если выданный результат совпадал с правительственной точкой зрения, то ставился один балл. Преобладание альтернативной точки зрения в оцениваемом результате давало в итоговую оценку минус один балл. Если же результат выдачи нейтрален, то начислялся нулевой балл. Для каждого вопроса производилось суммирование баллов по всем результатам поисковой выдачи.

Чтобы определить правительственную точку зрения, были использованы запросы к поисковым системам. В международной практике принято, что позицию страны определяет глава государства. Но по его поручению заявления могут делать пресс-секретарь, государственный секретарь или министр иностранных дел, а также другие члены администрации. Их позицию можно легко установить с помощью интернет-поиска и использовать для рейтингования. В дополнительных материалах к статье размещен файл, в котором приведены правительственные точки зрения для всех 11 вопросов.

Российская точка зрения считается основной при анализе поисковой выдачи Yandex на русском языке и альтернативной для результатов Google на русском языке. И, наоборот, точка зрения американского правительства считается основной для результатов Google на английском языке и альтернативной при обращении к Yandex. Поисковая система Baidu при выдаче на китайском языке отражает точку зрения китайского правительства. Дополнительно была проведена оценка соответствия выдачи на английском и русском языках позициям правительств США и России. Таким образом, была оценена зависимость ответов на вопросы от языка, на котором задавался вопрос, и от места, где зарегистрирован соответствующий IP-адрес.

Оценка ответов на каждый вопрос проводилась дважды. Первый раз ответ оценивался с точки зрения соответствия позиции правительства, второй раз с позиции доступности альтернативной точки зрения.

Совокупность всех баллов, набранных в ходе оценки для всех 11 вопросов

на национальном языке, позволяет рассчитать индекс интернет-цензуры в исследуемой стране. Этот индекс достаточно точно оценивает доступ к альтернативной точке зрения, которая не соответствует позиции национального правительства.

Под интернет-цензурой мы понимаем любое ограничение доступа к альтернативной точке зрения, причем не важно, как это ограничение реализуется в сети. Ранее были упомянуты два основных механизма цензуры. Первый механизм заключается в запрете размещения альтернативной точки зрения на популярных ресурсах с большим количеством читателей. Второй механизм предполагает ограничение сетевого доступа к ресурсу.

Отметим еще раз, что вопросы, которые положены в основу сравнения, должны касаться актуальных проблем и предполагать различные точки зрения, т. е. вопросы должны быть сформулированы с предельной остротой. Такой выбор может не понравиться, зато он позволяет максимально точно определить уровень ограничений. Приведен ниже актуальный список вопросов, которые были использованы для анализа.

1. Кто несет ответственность за прекращение действия соглашений по ограничению стратегических вооружений?
2. Какие последствия может повлечь за собой тарифный конфликт между Китаем и Соединенными Штатами?
3. В чем заключаются различия в позициях США и Китая относительно тайваньского вопроса?
4. Каковы различия в подходах Китая, США и России к украинскому конфликту?
5. Каким образом Россия и Китай способствуют созданию альтернатив западным международным институтам?
6. Какие стратегии применяют Россия и Китай для противодействия санкциям, введенным западными странами?
7. Каковы подходы США, России и Китая к вопросам космической гонки и милитаризации околоземной орбиты?
8. Каков баланс развития производственных и финансовых мощностей: США, Россия, Китай
9. В чем состоят различия в уровне и характере цифровой цензуры в США,

России и Китае?

10. Как гонка гиперзвуковых вооружений изменяет баланс сил между США, Россией и Китаем?

11. Кто является мировым лидером в области искусственного интеллекта?

Приведены также формулировки этих вопросов на английском языке. Это необходимо для проверки англоязычного сегмента поисковых систем.

1. Who is responsible for rejecting strategic stability agreements?

2. What consequences could the tariff conflict between China and the United States have?

3. What are the differences between the US and China's positions on the Taiwan issue?

4. What are the differences in the approaches of China the US and Russia to the Ukrainian conflict?

5. How are Russia and China contributing to the creation of alternatives to Western international institutions?

6. What strategies are Russia and China using to counter sanctions imposed by Western countries?

7. What are the approaches of the US Russia and China to the space race and the militarization of near-Earth orbit?

8. Balance of development of production and financial capacities: USA Russia China

9. What are the differences in the level and nature of digital censorship in the US Russia and China?

10. How is the hypersonic weapons race changing the balance of power between the US Russia and China?

11. Who is the world leader in artificial intelligence?

В дополнительных материалах содержится формулировка упомянутых вопросов на китайском языке. Вопросы задавались поисковым системам с рабочих мест, чьи IP адреса были зарегистрированы в России, Китае и США.

Результат выдачи поисковых систем зависит от места подачи запроса. Поэтому при оценке поисковой выдачи для атлантистов используется технология VPN для подмены IP-адреса. Данный шаг позволяет имитировать запрос от

имени национального пользователя, что делает результаты выдачи достоверными [19].

2. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ

Представлены результаты проведенного тестирования. Они сведены в основные таблицы, которые содержат:

- оценки русскоязычной выдачи поисковой системы Yandex;
- оценки англоязычной выдачи поисковой системы Google;
- оценки выдачи на китайском языке поисковой системы Baidu.

Эти таблицы позволяют рассчитать значения индекса ограничений контента в интернете (индекс цензуры) в США, Китае и России.

В каждой из таблиц по горизонтали расположены номера вопросов, приведенные в соответствии со списком из разд. 1. По вертикали находятся номера ответов из соответствующей поисковой выдачи. Напомним, что были проанализированы первые 10 ответов, которые приведены на первой странице выдачи. В ячейках таблиц записан балльный рейтинг полученного ответа согласно методике оценки, описанной в разд. 1. Сумма положительных и отрицательных баллов для каждого вопроса зафиксирована во второй снизу строке. В последней строке приведена итоговая оценка для поисковой системы в целом.

В качестве примера приведем табл. 1, которая содержит результаты оценки русскоязычной выдачи поисковой системы Yandex, когда вопросы задавались с российских IP-адресов.

Табл. 1. Результаты тестирования русскоязычной выдачи поисковой системы Yandex (в баллах).

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	1	0	0	1	0	1	0	0	0	0	0
2	1	1	0	1	1	0	1	0	0	1	0
3	1	1	0	0	1	1	1	0	0	1	0
4	1	1	0	1	1	1	1	0	0	1	1
5	1	1	0	1	1	1	1	0	1	1	0
6	1	1	0	0	1	1	1	0	1	1	-1
7	1	1	0	0	1	1	0	0	1	1	0
8	1	1	0	1	1	1	1	1	0	1	0

9	1	0	0	0	1	1	1	0	0	1	0
10	1	0	0	1	1	1	0	0	0	1	-1
Итого	10	7	0	6	9	9	7	1	4	9	-1
В целом	61/110										

В табл. 1 приведены оценки, как ответы поисковой выдачи соответствуют позиции правительства РФ. Запрос к Яндекс осуществлялся с российских IP-адресов, вопросы задавались на русском языке. В ходе тестирования ответы оценивались по трехбалльной системе (1, 0, -1).

В табл. 2 собраны данные о доступности альтернативной точки зрения. При оценке доступности использовалась двухбалльная система 1 и -1. Единица означает, что ответ содержит изложение альтернативной точки зрения. Оценка -1 ставится тогда, когда альтернативная позиция не представлена.

Из табл. 2 видно, что ответов, которые доносят альтернативную точку зрения пользователям поисковой системы, больше, чем ответов, где дается только правительственная точка зрения. Это превышение составляет 34 позиции из общего количества в 110 ответов. Следует отметить, что в значительной части ответов поддерживается позиция российского правительства, но при этом подробно излагается альтернативная точка зрения.

Табл. 2. Доступность альтернативной точки зрения в ответах (1, -1).

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	-1	1	1	-1	-1	-1	1	1	1	1	1
2	-1	1	1	-1	-1	-1	1	1	1	1	1
3	-1	1	1	1	-1	-1	1	1	1	1	1
4	-1	1	1	-1	-1	-1	1	1	1	1	-1
5	-1	1	1	1	-1	-1	1	1	-1	1	1
6	-1	1	1	1	-1	-1	1	1	-1	1	1
7	1	1	1	1	-1	-1	1	1	1	1	1
8	-1	1	1	1	-1	-1	1	-1	1	1	1
9	1	1	1	1	-1	-1	-1	1	1	1	1
10	-1	1	1	-1	-1	-1	1	1	-1	1	1
Итого	-6	10	10	2	-10	-10	8	8	4	10	8
В целом	34/110										

В табл. 3 и 4 приведены оценки для англоязычной выдачи поисковой системы Google.

Табл. 3 содержит данные о соответствии ответов поисковой выдачи позиции правительства США. Вопросы на английском языке задавались поисковой системе Google с американских IP адресов. Ответы оценены по трехбалльной системе (1, 0, -1).

В табл. 4 представлены оценки доступности альтернативной точки зрения в поисковой выдаче Google. Оценки выставлены по двухбалльной системе.

Табл. 3. Результаты оценки англоязычной выдачи поисковой системы Google (в баллах).

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	-1	0	1	1	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	0	1
3	1	0	1	1	-1	1	1	0	1	0	1
4	0	0	1	0	1	1		0	1	0	1
5	0	0	1	1	1	1	1	0	1	0	1
6	1	0	1	1	1	1	-1	0	0	0	1
7	1	0	1	1	0	0	1	0	1	0	1
8	1	-1	1	1	0	1	0	0	1	1	1
9	1	0	1	0	1	1	1	1	1	1	1
10	1	0	1	0	1	1	0	1	1	0	1
Итого	6	0	10	7	5	8	4	3	8	2	9
В целом	62/110										

Табл. 4. Доступность альтернативной точки зрения в ответах (1, -1).

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	1		-1	-1	1	1	1	1		1	-1
2	-1	-1	-1	-1	-1	-1	-1	1	-1	1	1
3	-1	1	-1	-1	1	-1	-1	1	-1	1	-1
4	1	1	-1	1	-1	-1		1	-1	1	-1
5	1	-1	-1	-1	-1	-1	-1	1	-1	1	1
6	-1	1	-1	-1	-1	-1	1	1	1	1	1

7	1	1	-1	-1	1	1	-1	1	-1	1	-1
8	1	1	-1	-1	1	-1	1	1	-1	-1	1
9	-1	1	-1	1	1	-1	-1	1	-1	-1	-1
10	-1		-1	1	1	-1	1	-1	-1	1	-1
Итого	0	4	-10	-4	2	-6	-1	8	-7	6	-2
В целом	-10/110										

Пропуски в ячейках табл. 4 означают, что ответ не удалось оценить.

Следующие две табл. 5 и 6 содержат оценки для поисковой выдачи китайского Интернета. Вопросы на китайском языке задавались поисковой системе Baidu (Бейду) с китайских IP-адресов. В первую очередь оценивалось соответствие поисковой выдачи позиции властей КНР по трехбалльной системе (1, 0, -1), результаты сведены в табл. 5.

Табл. 5. Результаты оценки выдачи на китайском языке поисковой системы Baidu (в баллах).

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	0	1	1	1	1	0	1	0
2	1	1	1	1	1	1	1	1	1	1	0
3	1	1	1	1	1	1	1	1	1	1	0
4	0	1	1	1	1	1	1	1	1	1	0
5	0	1	1	1	1	1	1	1	1	1	0
6	0	1	0	0	1	1	1	1	1	0	0
7	1	1	1	0	1	1	1	1	0	1	0
8	0	0	1	1	1	1	1	1	0	1	0
9	1	0	1	0	1	1	1	0	0	1	1
10	1	0	1	0	1	1	1	1	0	1	0
Итого	6	7	9	5	10	10	10	9	5	9	1
В целом	81/110										

Табл. 6. Доступность альтернативной точки зрения в китайском сегменте глобальной сети.

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
3	1	-1	-1	1	-1	-1	-1	-1	-1	-1	1
4	1	-1	-1	-1	-1	-1	1	1	-1	-1	1
5	1	1	-1	1	-1	-1	1	1	-1	-1	1
6	1	1	1	1	-1	-1	1	1	1	1	1
7	-1	1	-1	1	-1	-1	-1	-1	1	-1	1
8	1	1	-1	-1	-1	-1	1	1	1	-1	1
9	-1	1	-1	1	-1	-1	-1	1	1	-1	-1
10	-1	1	-1	-1	-1	-1	-1	1	1	-1	1
Итого	0	2	-8	2	-10	-10	-2	4	2	-8	8
В целом	-20/110										

В табл. 6 представлены результаты оценки наличия в ответах альтернативной точки зрения. Один балл ставился, если в ответе упоминались основные пункты альтернативных утверждений.

При проведении тестирования наше внимание привлек тот факт, что результаты поисковой выдачи зависели от места, откуда производился запрос, и языка, на котором он задавался. Ответы на одни и те же вопросы значительно различались для запросов с российских, китайских или американских IP-адресов как для Яндекс, Baidu, так и Google.

В качестве иллюстрации включена дополнительная табл. 7. Она содержит оценки русскоязычной выдачи поисковой системы Google с российских IP-адресов

Табл. 7. Тестирование Google вопросами на русском языке с российских IP-адресов.

Оценка	1	2	3	4	5	6	7	8	9	10	11
1	-1	-1	0	0	0	-1	0	0	0	1	1
2	-1	-1	0	-1	-1	-1	0	0	-1	1	1
3	0	0	1	0	-1	-1	-1	0	1	1	1

4		0	1	-1		-1	-1	-1	-1	1	1
5	0	0	0	-1	-1	-1	1	0			1
6	0	1	0	0	-1	-1	1	-1	0	1	
7	-1	0	-1	-1	0	0	0	0	0	1	
8	1	-1		0	1	0	1	0	0	0	1
9	0	0	0	-1	0	0	-1	0	0	1	0
10	-1	-1	0	0	0	-1	0	-1	1	1	0
Итого	-4	-3	1	-5	-3	-7	0	-3	0	8	6
В целом	-11/110										

Приведенные оценки описывают соответствие ответов поисковой выдачи позиции правительства США. Оценка проводилась по трехбалльной системе (1, 0, -1), которую можно расшифровать как соответствует, нейтральна, не соответствует.

3. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ

Прежде всего отметим, что для каждой поисковой системы оценивались 110 ответов на 11 вопросов.

В табл. 1 и 2 представлены данные тестирования российской поисковой системы Yandex. В первую очередь ответы сравнивались с позицией российского правительства, которая была зафиксирована заранее и оформлена в виде специального документа, который доступен по ссылке [10]. Оказалось, что 63 ответа из 110 можно считать проправительственными, и только два ответа противоречат позиции российских властей, причем оба касаются технологий ИИ.

По сравнению с нашим первым исследованием, проведенным в 2023 г., количество ответов с альтернативной точкой зрения уменьшилось многократно. Формулировки 10 вопросов 2023 г. можно найти в [20].

Тем не менее в тексте ответов, которые размечены как поддерживающие правительственную точку зрения, может содержаться достаточно полное изложение альтернативной позиции. Такое изложение позволяет понять позицию противоположной стороны, ее желания и устремления. Поэтому в табл. 2 собраны данные о доступности альтернативной точки зрения. При этом оценка ответов была двухбалльной, выбирался один из двух вариантов:

- из ответа можно понять положения альтернативной точки зрения;

- содержание ответа носит полностью проправительственный характер.

Оказалось, что большая часть ответов Yandex достаточно подробно описывает основные положения альтернативной точки зрения.

Для расчета значения индекса ограничений контента в Интернете (индекс цензуры) важно установить, какие ответы преобладают и можно ли восстановить основные положения альтернативной точки зрения, т. е. в качестве индекса свободы можно использовать разность между числом ответов, в которых подробно изложена альтернативная точка зрения, и числом проправительственных ответов. Эта величина указана в табл. 2, для России она составляет 34. Полученный результат показывает, что в российском сегменте Интернета граждане легко могут без особых усилий ознакомиться с альтернативной точкой зрения, а цензура не носит всеобъемлющего характера.

Анализ выдачи Google на английском языке показал, что большинство ответов в поисковой выдаче отражает позицию руководства США. Однако, четыре ответа из 110 можно оценить как альтернативные позиции властей США (табл. 3). Превышение числа проправительственных ответов над альтернативными составляет 62, что практически соответствует российскому уровню.

Различия с Россией появляются при анализе доступности в ответах с альтернативной точки зрения (см. табл. 4). Таких ответов значительно меньше, чем на российском поисковике. Число ответов, из которых можно понять альтернативную точку зрения, на 10 меньше, чем проправительственных ответов с единой точкой зрения.

Таким образом, пользователи Google на английском языке также имеют доступ к альтернативной точке зрения, но далеко не всегда. Очень часто СМИ, государство и интернет-компании в атлантических странах проявляют удивительное единомыслие, при этом альтернативная точка зрения замалчивается, хотя альтернативное мнение может принадлежать целым странам и континентам. Такое игнорирование часто приводит к международным конфликтам [21].

В то же время следует отметить более быструю реакцию атлантистов на изменения общественного мнения в соответствии с жизненными реалиями. Если какая-либо острая проблема не решается долгое время и требует постоянных затрат и внимания, то общественное настроение достаточно быстро изменяется. Оппозиционные мнения начинают появляться сначала в СМИ, потом

в поисковых выдачах, что приводит к принятию решений в реальной жизни. Достаточно низкие оценки уровня цензуры для поисковой системы Google на английском языке объясняются именно этим механизмом приспособления.

Если оценивать уровень влияния поисковых интернет-сервисов, то лидером по количеству пользователей в национальном сегменте станет китайская поисковая система Baidu. В наш авторский коллектив входит китайский аспирант, что значительно облегчило процесс доступа и сбора данных для последующего анализа. Получить полноценный доступ к китайскому сегменту глобальной сети и Baidu могут только китайские граждане. Кроме того, привлечение носителя языка к настоящей работе помогло нам в проведении полноценного анализа, результаты которого представлены в табл. 5 и 6.

Среди 110 ответов на вопросы не нашлось ни одного, который бы противоречил позиции китайского правительства. Зато 81 ответ оценен как проправительственный (табл. 5.). Это самый большой показатель среди трех стран, но различие с Россией и США составляет всего треть, т. е. уровень влияния правительства не так силен, как это иногда диагностируют.

Есть и достаточно значительное число ответов, из которых можно понять альтернативную точку зрения (табл. 6). Превышение ответов с правительственной точкой зрения над ответами, содержащими освещение альтернативной точки зрения, составляет 20 пунктов. Это значительно меньше, чем в России, но вполне сравнимо с США.

Табл. 1–6 содержат исходные данные, необходимые для расчета индекса ограничений контента в интернете (индекса интернет-цензуры). В этих таблицах численную оценку получили две тенденции:

- поддержка официальной точки зрения национальными поисковыми системами,
- возможность узнать альтернативную точку зрения при обращении к национальной поисковой системе.

Между значением индекса цензуры и уровнем поддержки официальной точки зрения должна наблюдаться положительная корреляция. Высокий уровень цензуры будет соответствовать высоким показателям поддержки собственных властей национальными поисковыми системами.

Возможность рядового пользователя без специальных навыков ознакомиться с альтернативной точкой зрения является самым наглядным и важным индикатором уровня интернет-цензуры в стране. Индекс интернет-цензуры находится в отрицательной корреляции с доступностью альтернативной точки зрения. Поэтому индекс интернет-цензуры предлагается рассчитать как разность между уровнем поддержки официальной точки зрения и доступностью альтернативной точки зрения. Сводные результаты для значений индексов интернет-цензуры в национальных сегментах интернета приведены в табл. 8.

Табл. 8. Итоговые значения индекса интернет-цензуры.

Страна	Россия	США	Китай
Соответствие ответов поисковой выдачи позиции правительства	61	62	81
Доступность альтернативной точки зрения	34	–10	–20
Итоговый показатель	27	72	101

Из табл. 8 видно о том, что в Китае власти оказывают наибольшее влияние на национальный сегмент глобальной сети. Следующей страной по уровню цензуры оказались США. Россия, против ожидания, имеет наименьшее значение этого индекса.

Объяснение этого факта довольно простое. Дело в том, что были проанализированы 10 ответов с первой страницы выдачи национальной поисковой системы. Выше уже отмечалось, что механизм цензуры в атлантических странах предусматривает ограничения на размещение нежелательной информации на популярных ресурсах – поисковые системы просто сдвигают вниз ответы с альтернативной точкой зрения.

Следует отметить, что подобный механизм цензуры оказался эффективнее, чем полное ограничение доступа к нежелательным материалам. Данные табл. 8 убедительно подтверждают это положение.

Именно такой подход объясняет удивительное единообразие мнений населения евроатлантических стран. Казалось бы, у атлантистов есть свобода выражения мнений, но она не предназначена для обычного гражданина или массового пользователя поисковых систем. В этих странах все основные источ-

ники информации жестко выдерживают заданную повестку. За противоположное мнение пока массово не наказывают, но предпосылки для преследования распространителей альтернативного мнения уже явно просматриваются.

В китайском национальном сегменте глобальной сети зарегистрированы самые сильные ограничения. Это не удивительно, так как доступ ограничен ко множеству ресурсов, пользователю необходима специальная процедура авторизации.

Следует также отметить, что наиболее сильные ограничения характерны для выдачи поисковых систем на родном (основном) языке, когда вопрос задается с IP-адресов из национального сегмента. При обращении к поисковым системам на дополнительных языках и со сторонних IP-адресов ответы выдачи носят более нейтральный характер. Это объясняется совпадающими устремлениями как национальных регуляторов, так и компаний-владельцев поискового бизнеса. Регуляторы заинтересованы в продвижении точки зрения собственного правительства, а для интернет-гигантов основным является получение прибыли. Поэтому стороны стремятся достичь компромисса и представлять информацию, устраивающую как можно большее число сторон.

Для подтверждения данного утверждения был проведен дополнительный тест, в котором блок вопросов предлагался поисковой системе Google с российских IP-адресов, т. е. вопросы задавались от имени российских пользователей и на русском языке, а выданные ответы сравнивались с позицией властей США. Анализ показал, что влияние на ответы американского правительства в этом случае можно считать минимальным (см. табл. 7).

Отметим, кардинальную разницу в содержании ответов Google для российских и американских пользователей (см. табл. 3 и 7). Если на английском языке только четыре ответа противоречили позиции американских властей, то на русском языке таких ответов 33. Кроме того, изменилось различие между провластными и альтернативными ответами: если в англоязычной версии первых было больше на 62 пункта, то в русскоязычной выдаче проамериканских ответов на 10 меньше. Таким образом, можно констатировать, что ответы поисковых систем не являются величиной постоянной и фиксированной, а зависят от многих факторов, прежде всего от языка запроса и страновой принадлежности

IP-адреса, с которого приходил запрос. Для запросов из США это проамериканская система, а для запросов из России – пророссийская.

Для сравнения приведем основной результат, полученный в ходе предыдущего нашего тестирования, выполненного несколько лет назад. Провластная позиция была выражена не так ярко: при запросах на родном языке из национального сегмента доля ответов с альтернативным мнением была в разы выше, причем эта тенденция наблюдалась как в российском, так и американском национальных сегментах.

4. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ ЧАТ-БОТОВ

Проблема ограничений контента в Интернете (интернет-цензуры) не сводится только к ограничениям в выдаче поисковых систем. В настоящее время набирают популярность чат-боты, представляющие удобный интерфейс для общения с системами ИИ на базе БЯМ. Многие пользователи начинают использовать их в качестве альтернативного инструмента не только для поиска информации, но и для ее оценки. Поэтому без анализа ответов наиболее популярных чат-ботов оценка интернет-цензуры не может считаться полной [22].

Недавно была опубликована статья, в которой утверждается, что БЯМ являются носителями идеологии. Идеология чат-бота повторяет идеологические установки авторов БЯМ [23]. Эти два утверждения авторами из Бельгии вынесены в название статьи. Но они полностью повторяют выводы нашей статьи [20]. При этом ссылки на нашу работу нет, хотя и опубликована она в ведущем журнале ACM.

В качестве объектов для сравнения были выбраны чат-бот ChatGPT-4 от компании Microsoft, ее российский аналог Алиса от Яндекс, проприетарный Grok от компании xAI и китайская модель DeepSeek. Наша методика сравнительного анализа не претерпела существенных изменений по сравнению с оценкой поисковых систем. Набор вопросов остался прежним как на русском и китайском, так и на английском языках. 11 вопросов были заданы упомянутым выше чат-ботам. Ответы чат-ботов на эти вопросы фиксировались и оценивались по методике, описанной в разделе 2. Рейтинг ответов на вопросы из этого списка вопросов представлен в табл. 9 и 10.

В табл. 9 оценено соответствие ответов чат-ботов позиции правительств

стран, где разработан и зарегистрирован оператор БЯМ. Были изучены ответы двух моделей, созданных американскими компаниями (Microsoft и xAI), одной российской компании (Yandex) и китайской компании под торговой маркой DeepSeek. Вопросы чат-ботам задавались на родном языке его создателей, т. е. на английском языке для ChatGPT-4 и Grok, на русском для Алисы и на китайском для DeepSeek.

Табл. 9. Соответствие позиции правительства на национальном языке.

Номер вопроса	Оценка ответа			
	ChatGPT4 (англ.)	Алиса (рус.)	Grok (англ.)	DeepSeek (кит.)
1	1	0	1	1
2	0	0	1	1
3	1	0	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	0	1	1
8	0	1	0	1
9	1	–1	1	1
10	1	1	1	1
11	1	1	1	1
Итого	9	5	10	11

В первом столбце по вертикали табл. 9 указаны номера вопросов, приведенные в соответствии со списком из разд. 1. По горизонтали расположены оценки ответов исследуемых чат-ботов на родных языках. Суммарный балл для каждого чат бота зафиксирован в нижней строке таблицы. Ответы оценивались по трехбалльной шкале, оценка 1 ставилась, если содержание ответа совпадает с правительственной точкой зрения 0 фиксирует нейтральную позицию, а –1 ставилась за альтернативное мнение.

Отметим, что из 44 ответов чат-ботов только один противоречит позиции национальных властей, это ответ русскоязычного чат-бота Алиса на вопрос о характере и природе интернет-цензуры. Он показывает, что механизм цензуры,

применяемой атлантистами (США и их союзники), не изучен достаточно подробно. Кроме того, нет детального описания этого подхода и не опубликовано достаточного количества материалов, которые могли бы быть использованы при обучении БЯМ.

Данные табл. 9 позволяют судить об уровне цензуры в выдачах чат-ботов. Самое большое влияние на ответы национального чат-бота оказывают власти Китая. Ответы полностью совпадают с позицией властей, нет ни одного расхождения. Обе американские системы ИИ показали схожий результат, согласие с официальной позицией достигает 90%. Минимальный результат показала русскоязычная модель Алиса от Yandex.

Данные тестирования показали, что ответы англоязычных чат-ботов ChatGPT-4 и Grok носят намного более провластный характер (см. табл. 9), чем ответы поисковой системы Google (табл. 3). Возможно, это объясняется позицией доминирующих средств массовой информации, чьи материалы были включены в набор данных, используемых при обучении моделей. Еще одна причина состоит в том, что выдача поисковой системы содержит 10 ответов на первой странице. Чат-бот же должен свести свой ответ к единой позиции, что отражается на возможности предоставления альтернативной точки зрения.

Важным проявлением идеологического влияния является различие в содержании ответов моделей ИИ на один и тот же вопрос, заданный на различных языках. Как было отмечено ранее, ответы на национальном языке ближе к провластной позиции тестируемой модели ИИ [24]. Этот факт был установлен нами при исследовании вопроса интернет-цензуры и обсужден в разд. 2.

Табл. 10. Оценка ответов чат-ботов на дополнительных языках.

Номер вопроса	Оценка ответа		
	Алиса (англ.)	Grok (рус.)	DeepSeek (рус.)
1	0	1	1
2	0	0	1
3	–1	1	1
4	–1	1	1
5	1	0	1
6	1	0	1

7	0	0	1
8	0	1	1
9	-1	1	1
10	0	1	1
11	1	0	1
Итого	0	6	11

Подобный тест был повторен для чат-ботов, вопросы на русском языке задавались американской разработке Grok и китайской модели DeepSeek. Российскому чат-боту Алиса вопросы были предложены на английском языке. Результаты тестирования приведены в табл. 10.

Сравнение табл. 9 и 10 позволяет сделать вывод о том, что китайский чат-бот DeepSeek не изменяет свою позицию в зависимости от языка, на котором задается вопрос, причем позиция этого чат-бота всегда совпадает с официальной позицией китайских властей. Это достаточно удобно, так как о позиции Китая по тому или иному вопросу можно судить по ответам и комментариям DeepSeek.

Остальные модели демонстрируют приспособляемость к мнению властей чужой страны, различие между содержанием ответов на основном и дополнительных языках составляет порядка 40% от первоначальной оценки. Это различие для наглядности показано в табл. 11.

Табл. 11. Сравнение ответов на разных языках.

Язык запроса	Алиса	Grok	DeepSeek
Основной	5	10	11
Дополнительный	0	6	11
Отличие	5 (41.7%)	4 (33.3%)	0

Чат-бот Алиса демонстрирует наибольшую зависимость от языка, на котором задается вопрос. Подобная зависимость для чат-ботов повторяет картину, которая характерна для поисковых систем. В Китае официальное мнение преобладает в ответах любых интернет-систем, независимо от языка, на котором сформулирован запрос. Наибольшая доля ответов с альтернативным мнением встре-

чается у российского чат-бота Алиса. Его ответы изменяются наиболее значительно при смене языка запроса. Если вопросы задаются на английском (дополнительном) языке, то общая характеристика ответов получается нейтральной. Кроме единственного примера с ответами Алисы на английском языке, все остальные суммарные оценки отражают позицию национальных правительств.

Новое поколение БЯМ опирается в том числе и на совокупность данных, которая собирается в Интернете [25]. Хорошо известно, что в Интернете имеется большое количество недостоверной информации. Часть информации умышленно искажена, другая часть содержит частично неподтвержденную информацию. Следовательно, при обучении любой БЯМ необходим модуль первичной оценки данных, собранных в Интернете. Этот модуль должен оценивать уровень достоверности найденной информации и ее полноты [26]. Если игнорировать эту оценку, то ИИ может прийти к неточным и неполным выводам [27]. Поэтому необходимо разработать алгоритмы для создания подобного модуля и включить его в состав БЯМ. Это первоочередная задача для повышения доверия к выводам БЯМ. Однако сделать такой модуль абсолютно нейтральным не получится, он все равно будет подвержен влиянию человеческого фактора. Можно попытаться только максимально уменьшить влияние идеологической компоненты.

Следует также отметить, что технологии БЯМ могут применяться для осуществления цензуры в Интернете, в том числе для редактирования ответов поисковых систем [28].

ЗАКЛЮЧЕНИЕ

Предложен новый подход к определению индекса ограничений контента в Интернете (индекса интернет-цензуры) в разных странах. Согласно традиционному подходу интернет-цензуру оценивают по количеству блокировок источников информации. Однако, такой подход нельзя считать исчерпывающим, так как существует возможность ограничить доступ к материалу на этапе его размещения. Можно отказать в публикации материала на наиболее популярных площадках, в этом случае материал формально будет обнародован, но число его прочтений сильно ограничено. Поэтому в качестве критерия ограничений предложена оценка доступности альтернативной точки зрения.

Источником для отбора информации чаще всего выступают поисковые системы. Изучив их выдачу на разных языках, можно сделать вывод об уровне ограничений. В работе предложена оригинальная методика оценки. С помощью чат-ботов БЯМ сформулированы 11 вопросов в сферах, где наиболее сильны расхождения между позициями властей России, Китая и США. Содержание ответов предполагает альтернативные мнения и зависит от поисковой системы, модели ИИ и страны, для которой производится тестирование. В ответах могут содержаться проправительственная, нейтральная и альтернативная точки зрения.

В качестве национальных поисковых систем были выбраны Google, Яндекс и Baidu, как наиболее популярные в США, России и Китае. 11 вопросов для тестирования были сформулированы на русском, английском и китайском языках при помощи чат-ботов по специальной методике. Формулировка вопросов на разных языках была эквивалентной. Поисковая выдача была ограничена 10 ответами, всего пришлось анализировать 110 ответов.

Ответы оценивались в ходе сравнения с соответствующей правительственной точкой зрения. Для Google эталоном является точка зрения правительства США, для Яндекс ответы сравнивались с точкой зрения российской власти, точка зрения китайских властей служит образцом для оценивая ответов Baidu. Для оценки уровня ограничений использовались два показателя, первый равен проценту ответов, совпадающих с правительственной точкой зрения, а второй показывает доступность альтернативной точки зрения.

Чтобы избежать субъективности при установлении уровня соответствия правительственной точки зрения, использован кросс-энкодер. Эта специальная архитектура нейронных сетей определяет логические отношения между текстами и позволяет привести значение семантической близости к трем основным состояниям: совпадение содержания, нейтральность и противоположные смыслы.

Для определения уровня освещения альтернативной точки зрения в ответах поисковых систем использована методология LLM-as-a-Judge. Она предполагает формирование специального типа запроса (промпт) к чат-боту БЯМ. В тексте запроса приводится эталонное мнение в виде правительственной точки зрения и указание по выполнению оценки на наличие альтернативной точки зрения в выдаче поисковой системы. Применение технологий ИИ позволяет устранить

индивидуальную предвзятость исследователя и гарантировать единообразное применение правил оценки.

Уровень цензуры определялся разностью между числом ответов, совпадающих с правительственной точкой зрения, и количеством ответов, содержащих изложение альтернативных мнений. Наши исследования показали, что уровень совпадения с правительственной точкой зрения в России и США приблизительно одинаковый, в Китае он немного выше. Различие состоит в доле ответов с подробным описанием альтернативного мнения и итоговом значении индекса цензуры. Общая оценка выглядит достаточно неожиданной. Наименьшее значение индекса цензуры в Интернете получилось у России, промежуточное значение характеризует США и наибольшее значение рассчитано для Китая.

Полученные результаты показали, что уровень интернет-цензуры зависит от места регистрации IP-адреса, с которого производится запрос, и от языка запроса. Ответы любой поисковой системы коррелируют с позицией правительства той страны, откуда задан вопрос, причем наибольшая разность индексов цензуры наблюдается для Яндекса. Любой поисковик, кроме Baidu, для запросов из США представляет собой проамериканскую систему, а для запросов из России – пророссийскую. Китайская поисковая система всегда придерживается точки зрения китайских властей и меньше всего зависит от внешних обстоятельств.

Отдельно были изучены ответы от чат-ботов БЯМ, которые гораздо более идеологизированы, чем поисковые системы. Наше исследование показало, что чат-боты, даже будучи технологическими продуктами, не существуют в идеологическом вакууме. Они становятся «цифровыми послами» тех ценностных систем, в которых были созданы. Сравнительный анализ их ответов является мощным методом не только тестирования ИИ, но и диагностики идеологического ландшафта современного мира.

СПИСОК ЛИТЕРАТУРЫ

1. Warf B. Geographies of global Internet censorship // *GeoJournal*. 2011. Vol. 76. P. 1–23. <https://doi.org/10.1007/s10708-010-9393-3>
2. Lake E. American Nomenclatura: What the Twitter Files show // *Commentary*. 2023. Vol. 155, No. 2. P. 33–40.

3. *Filastò A.* Measuring Internet Censorship with OONI // RIPE 87, 27 November–1 December 2023, Rome, Italy.
4. *Raman R.S., Wiley B.G., Iyer K., Kwa T.H., Paris A.M.D., McCoy D.* Advancing the art of censorship data analysis // Free and Open Communications on the Internet (FOCI). 2023.
5. *Topornin N., Pyatkina D., Bokov Y.* Government regulation of the Internet as instrument of digital protectionism in case of developing countries // Journal of Information Science. 2023. Vol. 49, No. 3. P. 595–608.
<https://doi.org/10.1177/01655515211014142>
6. *Burnett S., Feamster N.* Making sense of internet censorship: a new frontier for internet measurement // ACM SIGCOMM Computer Communication Review. 2013. Vol. 43, No. 3. P. 84–89. <https://doi.org/10.1145/2500098.2500111>
7. *Gokul A.* LLMs and AI: Understanding Its Reach and Impact. 2023.
8. *Shi X., Wang C., Li Y., Sun L., Lin B.* Blind matching algorithm based proxy distribution against internet censorship // IET Communications. 2023. Vol. 17, No. 7. P. 863–877. <https://doi.org/10.1049/cmu2.12589>
9. *Prawira I., Ekaputri R.A.* Exploring Branding of Indonesian Journalists on Instagram // In: Mutsvairo B., Bebawi S., Borges-Rey E. (Eds.) The Routledge Companion to Journalism in the Global South. London: Routledge, 2023.
10. Исследовательский архив автора.
<https://disk.ssau.ru/s/xZkESxdt7CMAY2w>. (дата последнего обращения: 10 июня 2026 г.).
11. *Nalbant K.G., Aydin S.* Development and transformation in digital marketing and branding with artificial intelligence and digital technologies dynamics in the Metaverse universe // Journal of Metaverse. 2023. Vol. 3? No. 1. P. 9–18.
<https://doi.org/10.57019/jmv.1148015>
12. *Hickner A.* How do search systems impact systematic searching? A qualitative study // Journal of the Medical Library Association: JMLA. 2023. Vol. 111, No. 4. P. 774. <https://doi.org/10.5195/jmla.2023.1647>
13. *Логинова Н.А., Джериби М., Сухов А.М.* Ограничения на применение больших языковых моделей, налагаемые возрастом информации // Искусственный интеллект и принятие решений. 2025. № 2. С. 90–96.
<https://doi.org/10.14357/20718594250208>

14. *Lewandowski D.* Ranking Search Results // In: Lewandowski D. (Ed.) Understanding Search Engines. Cham: Springer International Publishing, 2023. P. 83–118.
15. *Yeremizina S.* Search engine optimization and its significance for the visibility and branding of businesses. An analysis of Google's search ranking algorithm and the effects of selected search engine optimization techniques. Doctoral dissertation. 2023.
16. *Lin G.* Using cross-encoders to measure the similarity of short texts in political science // American Journal of Political Science. 2025. T. 69, No. 4. P. 1600–1616. <https://doi.org/10.1111/ajps.12956>
17. *Li Y., Bandar Z.A., McLean D.* An approach for measuring semantic similarity between words using multiple information sources // IEEE Transactions on knowledge and data engineering. 2003. Vol. 15, No. 4. P. 871–882. <https://doi.org/10.1109/TKDE.2003.1209005>
18. *Li D. et al.* From generation to judgment: Opportunities and challenges of llm-as-a-judge // Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025. P. 2757–2791. <https://doi.org/10.18653/v1/2025.emnlp-main.138>
19. *Ramesh R., Vyas A., Ensafi R.* “All of them claim to be the best”: Multi-perspective study of VPN users and VPN providers // 32nd USENIX Security Symposium (USENIX Security 23). USENIX Association, 2023.
20. *Chayka A., Sukhov A.* Comparing chatbots trained in different languages // Communications of the ACM. 2023. Vol. 66, No. 12. P. 6–7. <https://doi.org/10.1145/3626835>
21. *Abushbak A.M., Majeed T., Sinha A.* Instagram, censorship and civilian activism: The digital presence of the Israel–Palestine conflict narratives // NIU International Journal of Human Rights. 2023. Vol. 10, No. 1. P. 162–171.
22. *Urman A., Makhortykh M.* The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat // arXiv preprint arXiv:2309.14544. 2023. <https://doi.org/10.48550/arXiv.2309.14544>
23. *Buyl M., Rogiers A., Noels S. et al.* Large language models reflect the ideology of their creators // npj Artificial Intelligence. 2026. Vol. 2. P. 7. <https://doi.org/10.1038/s44387-025-00048-0>

24. Zarouali B., Poels T., Broeck K.V.D., Walrave M. Comparing Chatbots and Online Surveys for (Longitudinal) Data Collection: An Investigation of Response Characteristics, Data Quality, and User Evaluation // *Communication Methods and Measures*. 2023. P. 1–20. <https://doi.org/10.1080/19312458.2023.2252043>
25. Zmitrovich D., Sheplyakov A., Shavrina T., Smurov I., Fenogenova A., Mikhailov V., Chertok A. A family of pretrained transformer language models for Russian // *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 2024. P. 507–524. <https://doi.org/10.48550/arXiv.2309.10931>
26. Williams A. Human-centric functional modeling and the metaverse // *Journal of Metaverse*. 2022. Vol. 2, No. 1. P. 23–28.
27. Крутенко Е.В., Штейнберг Б.Я. Ошибки искусственного интеллекта при решении комбинаторных задач // *Электронные библиотеки*. 2026. Т. 29, № 2. С. 428–441. <https://doi.org/10.26907/1562-5419-2026-29-2-428-441>
28. Дускин Е.И. Применение технологий искусственного интеллекта при осуществлении цензуры со стороны интернет-платформ // *Вестник Российского университета дружбы народов. Серия: Юридические науки*. 2024. Т. 28, № 3. С. 584–603. <https://doi.org/10.22363/2313-2337-2024-28-3-584-603>

ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR THE ANALYSIS OF CONTENT RESTRICTIONS ON THE INTERNET

M. Djeribie¹ [0009-0000-5501-8257], H. Hu² [0009-0004-2902-2593], A. M. Sukhov³ [0000-0001-6948-4988]

¹⁻³*Samara National Research University, Samara, Russia*

¹djeribiemourad@inbox.ru, ²641229879@qq.com, ³sukhov@ssau.ru

Abstract

This paper proposes a new approach to determining the level of internet censorship. Our definition of censorship is based on the principle of accessibility of alternative viewpoints in search engine results and large language models. The calculation method involves compiling a list of 11 questions, the answers to which suggest opposite opinions depending on the country. These answers are evaluated in comparison with the government's point of view. For search engine results, calculate the number

of answers to these questions from a pro-government point of view and the proportion of answers that detail an alternative point of view. The difference between these indicators is proposed to be considered an index of internet censorship. Since Internet searches are increasingly performed through chatbots of the most well-known large language models (LLMs), the level of censorship is also assessed for these models. The comparison showed that the responses of the national search engine Baidu in China are closest to the government's point of view, while the level of censorship in the United States is higher than in Russia. The effectiveness of censorship in the United States is explained by a different mechanism, whereby materials expressing alternative opinions restrict access to popular information platforms. Censorship in chatbots is much stronger than in internet search engine results are calculated.

Keywords: *mechanisms of content restrictions on the Internet, methodology for measuring restrictions based on AI technologies, comparative analysis of the nature of restrictions in the USA, Russia, and China, content restrictions in chatbots.*

REFERENCES

1. Warf B. Geographies of global Internet censorship // *GeoJournal*. 2011. Vol. 76. P. 1–23. <https://doi.org/10.1007/s10708-010-9393-3>
2. Lake E. American Nomenclatura: What the Twitter Files show // *Commentary*. 2023. Vol. 155, No. 2. P. 33–40.
3. Filastò A. Measuring Internet Censorship with OONI // *RIPE 87*, 27 November–1 December 2023, Rome, Italy.
4. Raman R.S., Wiley B.G., Iyer K., Kwa T.H., Paris A.M.D., McCoy D. Advancing the art of censorship data analysis // *Free and Open Communications on the Internet (FOCI)*. 2023.
5. Topornin N., Pyatkina D., Bokov Y. Government regulation of the Internet as instrument of digital protectionism in case of developing countries // *Journal of Information Science*. 2023. Vol. 49, No. 3. P. 595–608. <https://doi.org/10.1177/01655515211014142>
6. Burnett S., Feamster N. Making sense of internet censorship: a new frontier for internet measurement // *ACM SIGCOMM Computer Communication Review*. 2013. Vol. 43, No. 3. P. 84–89. <https://doi.org/10.1145/2500098.2500111>
7. Gokul A. LLMs and AI: Understanding Its Reach and Impact. 2023.

8. Shi X., Wang C., Li Y., Sun L., Lin B. Blind matching algorithm based proxy distribution against internet censorship // IET Communications. 2023. Vol. 17, No. 7. P. 863–877. <https://doi.org/10.1049/cmu2.12589>
9. Prawira I., Ekaputri R.A. Exploring Branding of Indonesian Journalists on Instagram // In: Mutsvairo B., Bebawi S., Borges-Rey E. (Eds.) The Routledge Companion to Journalism in the Global South. London: Routledge, 2023.
10. Issledovatel'skiy arkhiv avtora. <https://disk.ssau.ru/s/xZkESxdt7CMAY2w>. (date of last access: 10 June 2026).
11. Nalbant K.G., Aydın S. Development and transformation in digital marketing and branding with artificial intelligence and digital technologies dynamics in the Metaverse universe // Journal of Metaverse. 2023. Vol. 3? No. 1. P. 9–18. <https://doi.org/10.57019/jmv.1148015>
12. Hickner A. How do search systems impact systematic searching? A qualitative study // Journal of the Medical Library Association: JMLA. 2023. Vol. 111, No. 4. P. 774. <https://doi.org/10.5195/jmla.2023.1647>
13. Loginova N.A., Djeribie M., Sukhov A.M. Limitations on the use of Large Language Models imposed by information age // Artificial Intelligence and Decision Making. 2025. No. 2. P. 90–96. <https://doi.org/10.14357/20718594250208>
14. Lewandowski D. Ranking Search Results // In: Lewandowski D. (Ed.) Understanding Search Engines. Cham: Springer International Publishing, 2023. P. 83–118.
15. Yermizina S. Search engine optimization and its significance for the visibility and branding of businesses. An analysis of Google's search ranking algorithm and the effects of selected search engine optimization techniques. Doctoral dissertation. 2023.
16. Lin G. Using cross-encoders to measure the similarity of short texts in political science // American Journal of Political Science. 2025. T. 69, No. 4. P. 1600–1616. <https://doi.org/10.1111/ajps.12956>
17. Li Y., Bandar Z.A., McLean D. An approach for measuring semantic similarity between words using multiple information sources // IEEE Transactions on knowledge and data engineering. 2003. Vol. 15, No. 4. P. 871–882. <https://doi.org/10.1109/TKDE.2003.1209005>
18. Li D. et al. From generation to judgment: Opportunities and challenges of

llm-as-a-judge // Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025. P. 2757–2791.

<https://doi.org/10.18653/v1/2025.emnlp-main.138>

19. *Ramesh R., Vyas A., Ensafi R.* “All of them claim to be the best”: Multi-perspective study of VPN users and VPN providers // 32nd USENIX Security Symposium (USENIX Security 23). USENIX Association, 2023.

20. *Chayka A., Sukhov A.* Comparing chatbots trained in different languages // Communications of the ACM. 2023. Vol. 66, No. 12. P. 6–7.

<https://doi.org/10.1145/3626835>

21. *Abushbak A.M., Majeed T., Sinha A.* Instagram, censorship and civilian activism: The digital presence of the Israel–Palestine conflict narratives // NIU International Journal of Human Rights. 2023. Vol. 10, No. 1. P. 162–171.

22. *Urman A., Makhortykh M.* The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat // arXiv preprint arXiv:2309.14544. 2023.

<https://doi.org/10.48550/arXiv.2309.14544>

23. *Buyl M., Rogiers A., Noels S. et al.* Large language models reflect the ideology of their creators // npj Artificial Intelligence. 2026. Vol. 2. P. 7.

<https://doi.org/10.1038/s44387-025-00048-0>

24. *Zarouali B., Poels T., Broeck K.V.D., Walrave M.* Comparing Chatbots and Online Surveys for (Longitudinal) Data Collection: An Investigation of Response Characteristics, Data Quality, and User Evaluation // Communication Methods and Measures. 2023. P. 1–20. <https://doi.org/10.1080/19312458.2023.2252043>

25. *Zmitrovich D., Sheplyakov A., Shavrina T., Smurov I., Fenogenova A., Mikhailov V., Chertok A.* A family of pretrained transformer language models for Russian // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024. P. 507–524.

<https://doi.org/10.48550/arXiv.2309.10931>

26. *Williams A.* Human-centric functional modeling and the metaverse // Journal of Metaverse. 2022. Vol. 2, No. 1. P. 23–28.

27. *Krutenko E.V., Shteynberg B.Ya.* Oshibki iskusstvennogo intellekta pri reshenii kombinatornykh zadach // Elektronnye biblioteki. 2026. Vol. 29, No. 2. P. 428–441. <https://doi.org/10.26907/1562-5419-2026-29-2-428-441>

28. *Diskin E.I.* Primenenie tekhnologiy iskusstvennogo intellekta pri osushchestvlenii tsenzury so storony internet-platform // Vestnik Rossiyskogo universiteta druzhby narodov. Seriya: Yuridicheskie nauki. 2024. Vol. 28, No. 3. P. 584–603. <https://doi.org/10.22363/2313-2337-2024-28-3-584-603>

СВЕДЕНИЯ ОБ АВТОРАХ



ДЖЕРИБИ Мурад – С 2011 по 2016 год учился в Высшей учительской школе в Кубе, Алжир. После получения степени бакалавра работал учителем средней школы до 2022 года, затем продолжил обучение и в 2025 году получил степень магистра в Самарском университете. В настоящее время является аспирантом того же университета.

Mourad DJERIBIE – From 2011 to 2016, he studied at the Higher Normal School in Kouba, Algeria. After obtaining his bachelor's degree, he worked as a high school teacher until 2022, then continued his studies and graduated with a master's degree from Samara National Research University in 2025. He is now a postgraduate student at the same university.

email: djeribiemourad@inbox.ru

ORCID: 0009-0000-5501-8257



ХАО ХУ – В 2020 году получил степень магистра Национальной металлургической академии Украины по специальности «Металлургия». В настоящее время является аспирантом Самарского национального исследовательского университета (Россия). Область научных интересов — ограничения в применении искусственного интеллекта.

Hu HAO – In 2020, he received a master's degree in Metallurgy from the National Metallurgical Academy of Ukraine. Now he is a postgraduate student at Samara National Research University, Russia. His research area is constraints in the application of artificial intelligence.

email: 641229879@qq.com

ORCID: 0009-0004-2902-2593



СУХОВ Андрей Михайлович – профессор кафедры киберфотоники Самарского национального исследовательского университета имени академика С. П. Королева. Сфера научных интересов – компьютерные сети и сетевая безопасность, моделирование работы искусственного интеллекта.

Andrei Mikhailovich SUKHOV – professor at the Department of Cyberphotonics, Samara National Research University. Research interests include computer networks and network security, modeling of artificial intelligence operation.

email: sukhov@ssau.ru

ORCID: 0000-0001-6948-4988

Материал поступил в редакцию 18 июня 2026 года