

DOI 10.24412/2686-7702-2024-4-93-104

## Хакатон как полигон для создания цифровых инструментов в отечественном востоковедении

Кудакаев Родион Фидельевич<sup>1</sup>,  
Мокрецкий Александр Чеславович<sup>1</sup>  
Костыркин Александр Вячеславович<sup>2</sup>

<sup>1</sup> Институт Китая и современной Азии РАН

<sup>2</sup> Институт востоковедения РАН

**Аннотация.** При формировании нового технологического уклада и мирового соперничества за лидерство в цифровом пространстве внимание специалистов смещается в сторону роста политического, экономического и научно-технического влияния стран Востока, что предъявляет новые требования к исследовательским методикам и инструментам. Статья резюмирует опыт расширения академического научно-поискового процесса за счёт конкурсной формы подключения молодых IT-специалистов к востоковедческой проблематике. В рамках двухнедельного конкурса «Хакатон», организованного Yandex Cloud, Napoleon IT и AI Talent Hub с участием специалистов Института востоковедения и Института Китая и современной Азии РАН, была поставлена задача реализовать чат-бот с использованием генеративных моделей и машинного перевода для анализа новостных потоков в информационном пространстве стран Восточной Азии, прежде всего КНР и Японии. Обзор подходов и решений команд-финалистов вместе с принципиальной реализуемостью проекта показывает, что множество частных лингвистических и инженерных задач, актуальных ещё недавно, уже решены с достаточным качеством, и при планировании и разработке цифровых инструментов нового поколения необходимо переходить к функционально-смысловым блокам более высокого и приближенного к человеку уровня обобщения.

**Ключевые слова:** востоковедение, генеративные модели, научно-техническое развитие, большие языковые модели, машинный перевод.

**Авторы:**

*Кудакаев Родион Фидельевич*, младший научный сотрудник, Центр политических исследований и прогнозов, Институт Китая и современной Азии РАН (адрес: 117977, Москва, Нахимовский пр-т, 32). ORCID: 0000-0003-2054-8287. E-mail: rfkudakaev@gmail.com

*Мокрецкий Александр Чеславович*, старший научный сотрудник, Центр «Россия, Китай, мир», Институт Китая и современной Азии РАН (адрес: 117977, Москва, Нахимовский пр-т, 32). ORCID: 0000-0002-5150-7747. E-mail: 88am@mail.ru

*Костыркин Александр Вячеславович*, кандидат филологических наук, старший научный сотрудник, Лаборатория цифровых исследований современного Востока, Институт востоковедения РАН (адрес: 117977, Москва, Рождественка, 12). ORCID: 0009-0002-8167-1529. E-mail: kost@ivran.ru

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Финансирование.** Статья подготовлена в рамках государственных заданий Министерства науки и высшего образования Российской Федерации (тема № FMSF-2024-0020 «Государственное и правовое строительство в КНР»; тема № FMSF-2024-0019 «Современная внешняя политика КНР и взаимодействие России и КНР в политической, экономической и других практических областях, в том числе в многосторонних форматах (БРИКС, ШОС, РИК»); тема № FMNN-2024-0001 «Лаборатория цифровых исследований современного Востока»).

**Для цитирования:** Кудакеев Р.Ф., Мокрецкий А.Ч., Костыркин А.В. Хакатон как полигон для создания цифровых инструментов в отечественном востоковедении // Восточная Азия: факты и аналитика. 2024. № 4. С. 93–104. DOI 10.24412/2686-7702-2024-4-93-104

## Hackathon as a testing ground for creating digital tools in domestic Oriental studies

Kudakaev Rodion F.<sup>1</sup>,  
Mokretskiy Aleksandr Ch.<sup>1</sup>,  
Kostyrkin Aleksandr V.<sup>2</sup>

---

<sup>1</sup> Institute of China and Contemporary Asia of the Russian Academy of Sciences

<sup>2</sup> Institute of Oriental Studies of the Russian Academy of Sciences

**Abstract.** With the formation of a new technological paradigm and global competition for leadership in the digital space, the attention of experts is shifting towards the growth of political, economic and R&D influence of Eastern countries, which imposes new demands on research methods and tools. The article summarizes the experience of enhancing the academic research process through involving young IT professionals in the Oriental studies in competitive mode. During two weeks of the “Hackathon” contest organized by Yandex Cloud, Napoleon IT and AI Talent Hub in collaboration with the experts from the Institute on Oriental Studies and the Institute of China and Contemporary Asia of the Russian Academy of Sciences, contestants were challenged to develop a chat bot employing generative models and machine translation to analyze news flows of East Asian countries, primarily China and Japan. A review of the winners’ approaches and solutions proves overall feasibility of the idea and shows that many specific linguistic and engineering tasks that were relevant only recently have already been successfully solved. Therefore, when planning and developing next-generation digital tools, it is necessary to operate at functionally and semantically higher levels of generalization closer to human reasoning.

**Keywords:** Oriental studies, generative models, research & development, large language models, machine translation.

**Authors:**

*Kudakaev Rodion F.*, Junior Researcher, Center for Political Studies and Forecasting, Institute of China and Contemporary Asia of the Russian Academy of Sciences (address: 32, Nakhimovskiy Av., Moscow, 117997, Russian Federation). ORCID: 0000-0003-2054-8287. E-mail: rfkudakaev@gmail.com

*Mokretskiy Aleksandr Ch.*, Senior Researcher, Center for “Russia, China, World” Studies, Institute of China and Contemporary Asia of the Russian Academy of Sciences (address: 32, Nakhimovskiy Av., Moscow, 117997, Russian Federation). ORCID: 0000-0002-5150-7747. E-mail: 88am@mail.ru

*Kostyrkin Aleksandr V.*, Ph.D. (Philology), Senior Researcher, Laboratory of Digital Studies of the Modern East, Institute of Oriental Studies of the Russian Academy of Sciences (address: 12, Rozhdestvenka, Moscow, 117977, Russian Federation). ORCID: 0009-0002-8167-1529. E-mail: kost@ivran.ru

**Conflict of interests.** The authors declare the absence of the conflict of interests.

**Support.** The article was carried out under the state assignment of the Ministry of Science and Higher Education of the Russian Federation (project No. FMSF-2024-0020 "State and legal construction in the PRC"; project No. FMSF-2024-0019 "The modern foreign policy of the People's Republic of China and cooperation between Russia and the People's Republic of China in political, economic and other practical fields, including in multilateral formats (BRICS, SCO, RIC)"; project No. FMNN-2024-0001 "Digital Research Laboratory of Contemporary East IOS RAS").

**For citation:** Kudakaev R.F., Mokretskiy A.Ch., Kostyrkin A.V. Khakaton kak poligon dlya sozdaniya tsifrovyykh instrumentov v otechestvennom vostokovedenii [Hackathon as a testing ground for creating digital tools in domestic Oriental studies]. *Vostochnaya Aziya: fakty i analitika* [East Asia: Facts and Analytics], 4: 93–104. (In Russian). DOI 10.24412/2686-7702-2024-4-93-104

Достижение суверенитета в научно-технической сфере является первоочередной задачей ведущих держав мира. Процесс цифровизации затрагивает все направления развития современного общества, включая обороноспособность и экономику, что находит своё отражение в доктринальных документах государств. В мае 2023 г. правительством России была утверждена Концепция технологического развития до 2030 г., согласно которой «к концу третьего десятилетия XXI в. Россия должна обладать собственной научной, кадровой и технологической базой критических и сквозных технологий»<sup>1</sup>. Концепция предусматривает вовлечение молодёжи в технологические проекты путём организации акселераторов, стартап-студий и кейсчемпионатов по решению практических научных и технологических задач<sup>2</sup>.

Китай, как и Россия, придаёт большое значение научно-техническому развитию: например, в материалах 3-го пленума ЦК КПК 20-го созыва – ключевого события внутривнутриполитической жизни КНР в 2024 г. – оно названо «стратегической опорой в осуществлении китайской модернизации» [Виноградов 2024: 13]. Если с начала политики «реформ и открытости» Китай развивался в благоприятных условиях, характеризовавшихся позитивной внутренней и внешней конъюнктурой, то начиная с «новой эпохи» (термин, закреплённый в документах XIX съезда КПК в 2017 г.) КНР последовательно сталкивается с новыми вызовами: активным этапом «торговой войны» с США, последствиями COVID-19 и т.д. В этих условиях руководство страны решилось на смену экономической модели, основой которой выступает применение инновационных технологий и подготовка нового поколения специалистов.

В это же время ключевой целью научно-технической политики США является сохранение лидирующих позиций, учитывая возрастающую конкуренцию со стороны КНР и РФ<sup>3</sup>. Соединённые Штаты первыми вступили на путь становления технологической державы: ещё при президенте Ф.Д. Рузвельте начали создаваться «уникальные аналоговые компьютеры и новые бизнес-модели» [Юдина, Линдун 2024: 118].

Несмотря на возрастающую роль негосударственных акторов – крупных технологических компаний, транснациональных объединений и т.д. – «только государство способно обеспечить концентрацию инвестиций и ресурсов на приоритетных направлениях научно-технического развития» [Виноградов 2024: 13].

---

<sup>1</sup> Правительство утвердило Концепцию технологического развития до 2030 года. *Правительство России*, 25.05.2024. URL: <http://government.ru/news/48570/> (дата обращения: 30.08.2024).

<sup>2</sup> Там же.

<sup>3</sup> Толстухина А. Технологическая политика США в условиях соперничества с Китаем. *РСМД*, 30.11.2023. URL: <https://russiancouncil.ru/analytics-and-comments/analytics/tekhnologicheskaya-politika-ssha-v-usloviyakh-sopernichestva-s-kitaem/> (дата обращения: 30.09.2024).

### Новый технологический уклад

Человечество вступает в новый этап развития промышленности, характеризующийся слиянием технологий и размытием границ между физической, цифровой и биологической сферами. Одним из ключевых направлений НТР является искусственный интеллект (ИИ), что в полной мере подтверждают последние достижения в этой области, а именно генеративный ИИ – передовая нейросетевая технология на основе больших языковых моделей, прошедших глубокое машинное обучение и способных проводить семантический анализ имеющихся данных и создавать новые данные (текст, изображения, видео и т.д.)<sup>4</sup> [Кудакаев 2024: 184]. Основным конкурентным преимуществом становится создание и использование интеллектуального капитала – знаний, навыков и компетенций людей, а также нематериальных активов (патентов, авторских прав, баз данных) – маркера постиндустриальной эпохи.

Учитывая возросшие объёмы информации и появление её новых видов, в научной сфере выдвигаются особые требования к инструментам сбора, хранения, обработки и поиска данных. Основой инструментария современного исследователя являются сайты-агрегаторы новостей (Информационный ресурс по Китаю: <https://1hut.ru>)<sup>5</sup> [Кудакаев 2024: 190], электронные библиотеки (РГБ, КиберЛенинка), а также электронные библиографические и реферативные базы научных публикаций (Scopus, eLibrary, CNKI) [Цифровая трансформация: ожидания и реальность 2022: 144]. Их дополняют аналитические наукометрические онлайн-сервисы, например InCites и SciVal, позволяющие визуализировать, анализировать и экспортировать запрашиваемые данные<sup>6</sup>. Среди иных исследовательских инструментов можно выделить цифровые репозитории (Social Science Research Network), патентные базы (China Patents), облачные платформы (Amazon Web Services, Google Cloud Platform) и т.д. Однако большинство подобных ресурсов реализуется по подписочной модели, что требует от пользователей как больших финансовых затрат, так и прохождения обучения для использования платформ, включая подготовку команды технических специалистов. Очевидно, что для многих исследователей, особенно молодых учёных, работающих вне сложившихся исследовательских коллективов и пока не участвующих в грантовой деятельности, применение некоторых из перечисленных инструментов на постоянной основе не представляется возможным.

Другой преградой на пути исследователя может стать языковой барьер. В мировой науке растёт влияние западных стран, и чтобы быть в курсе мировых научно-технических достижений, сегодня необходимо работать с источниками на восточных языках. По данным японского Центра стратегии исследований и разработок (研究開発戦略センター), по расходам на НИОКР в числе мировых лидеров из восточных стран (как по абсолютным числам, так и по доле от ВВП) находятся КНР, Япония и Корея, причём названные показатели стабильно растут на протяжении последних 20 лет, а доля

---

<sup>4</sup> Тема подробно освещена в докладе директора проектов Департамента стратегии и развития ПАО «Сбербанк» В.А. Лемутова «Генеративный искусственный интеллект в КНР: текущее состояние и перспективы развития».

<sup>5</sup> Описание ресурса представлено в докладе ведущего специалиста ЦИСК ИВ РАН Е.А. Спиридонова.

<sup>6</sup> Савина Т.Ф., Стерлигов И.А. InCites vs SciVal: сравниваем продвинутые наукометрические пакеты из подписки НИУ ВШЭ. *Окна роста*, 29.09.2016. URL: <https://okna.hse.ru/news/191799062.html> (дата обращения: 12.07.2024).

исследователей от общего населения в РК и КНР за это время показала кратный рост<sup>7</sup>. В частности, в 2023 г. КНР увеличила финансирование НИОКР до 2,65 % ВВП, в том же году в стране насчитывалось 37,8 млн студентов, 3,9 млн аспирантов [Островский 2024: 78–79]. Усилия, предпринятые руководством КНР, выражаются в завоевании лидирующих позиций по таким показателям, как количество научных публикаций в области естественных наук<sup>8</sup>, количество заявок на патенты (по данным на 2024 г. оно составило 74 тыс.) [Островский 2024: 78–79], количество произведённых высокотехнологичных продуктов (например, 351,4 млрд интегральных схем)<sup>9</sup>.

Освоение иероглифических языков до уровня чтения общественно-политических и академических текстов требует продолжительного обучения. Однако развитие технологий позволяет сократить путь по крайней мере на уровне поиска и первичного анализа информации. Интересный опыт был получен в этом направлении на стыке востоковедческой проблематики и IT-технологий.

#### 4-й Хакатон (AI Product Hackathon)

С 31 мая по 16 июня 2024 г. прошёл 4-й Хакатон (AI Product Hackathon). Это конкурс для разработчиков программного обеспечения, программистов, дизайнеров, менеджеров и других IT-специалистов, во время которого команды решают представленные «кейсодателями» бизнес-задачи с использованием AI-технологий. Организаторами выступили Yandex Cloud, Napoleon IT и AI Talent Hub.

Каждый из четырёх партнёров Хакатона – Лемана ПРО (Леруа Мерлен), Ситилинк, Институт востоковедения РАН, Группа НЛМК – предоставили участникам темы своих проектов, которые предлагалось воплотить в жизнь с помощью генеративной модели YandexGPT или других аналогичных решений машинного обучения (Machine Learning). В течение двух недель 65 команд из 20 регионов России и стран СНГ соревновались в четырёх выбранных направлениях. В итоге 28 команд смогли представить свои решения экспертам, пройдя 26 часов консультаций, QA-сессии (период «вопросов и ответов») и промежуточные этапы проверки работы.

ИБ РАН поставил задачу создать русскоговорящий чат-бот для анализа СМИ на китайском и японском языках – систему с использованием генеративных моделей, в том числе машинного перевода, для анализа новостных сообщений в информационном пространстве стран Восточной Азии, прежде всего КНР и Японии. При этом были оговорены основные условия: запросы принимаются на русском языке, ответы или цепочки логических выводов также выдаются на русском языке; каждый факт или утверждение, предоставляемые в ответе, должны сопровождаться ссылками на источники; модель должна дообучаться, учитывая ежедневно динамически поступающие сообщения, и отвечать с учётом самых последних новостей. Для эксперимента участникам были предоставлены японский и китайский новостные датасеты объёмом 1,7 млн сообщений каждый.

---

<sup>7</sup> 研究開発の俯瞰報告書 論文・特許データから見る研究開発動向 (2024 年) [Обзорный отчёт о развитии исследований за 2024 г. Исследовательские бюджеты и число исследователей]. *Center for Research and Development Strategy*, 21.06.2024. URL: [https://www.jst.go.jp/crds/pdf/2024/FR/CRDS-FY2024-FR-01/CRDS-FY2024-FR-01\\_10000.pdf](https://www.jst.go.jp/crds/pdf/2024/FR/CRDS-FY2024-FR-01/CRDS-FY2024-FR-01_10000.pdf) (accessed: 14.07.2024). (На яп.).

<sup>8</sup> Китай впервые опередил США по количеству научных публикаций. *TASS*, 20.09.2023. URL: <https://nauka.tass.ru/nauka/18797069> (дата обращения: 14.07.2024).

<sup>9</sup> Там же.

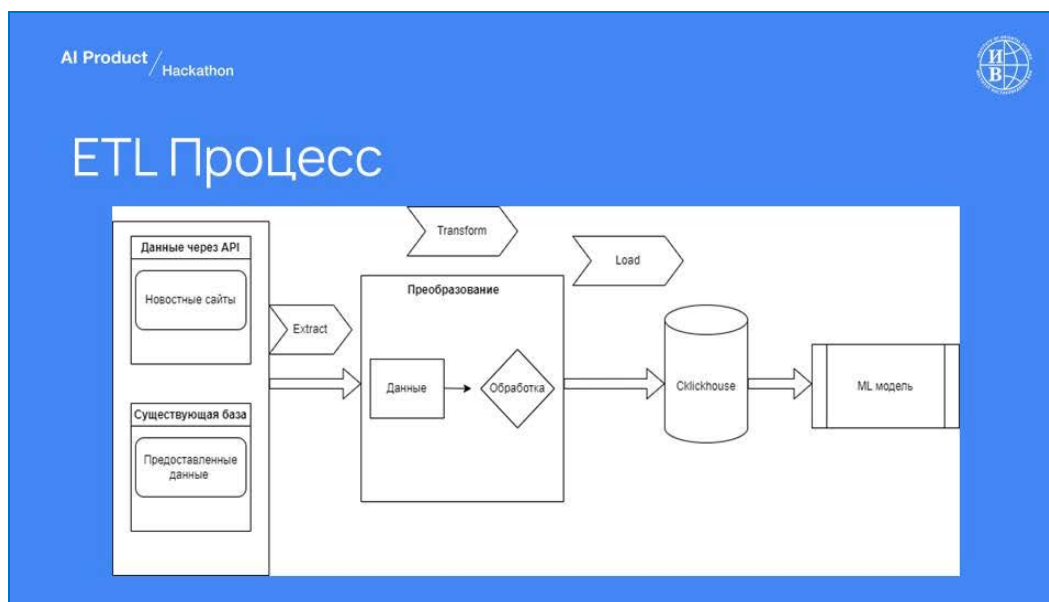
В состав экспертов, помимо технических специалистов от организаторов Хакатона, вошли авторы данной статьи.

Победителем стала команда `backpropagation.ai`, использовавшая в своём решении схему Retrieval Augmented Generation, которая позволяет по запросу пользователя найти необходимые документы по предварительно сформированной из новостных сообщений векторной базе данных, после чего передаёт исходный пользовательский запрос в сочетании с найденными сообщениями в виде специального «промпта» на вход генеративной модели для составления резюмированного ответа. Эти и идейно близкие решения других финалистов могут быть масштабированы для включения других восточных языков и расширения экспериментального датасета потоковыми онлайн-данными.

Ниже подробнее представим проекты пяти финалистов в кейсе № 2 (ИБ РАН) по созданию чат-бота для анализа японских и/или китайских СМИ.

### **backpropagation.ai**

Правильно поняв поставленную проблему и текущие задачи, команда предложила четырёхэтапный процесс разработки проекта: загрузку и очистку данных, фрагментарное дообучение на материале новостных текстов и запросов большой языковой (LLM) модели (fine-tuning), разработку интерфейса чат-бота в социальной сети Telegram и динамическое обновление данных и модели (рис. 1). Для парсинга<sup>10</sup> и перевода поступающих данных были использованы следующие технологии: библиотека BeautifulSoup для парсинга исходных HTML текстов и извлечения из них содержимого новостных сообщений, Clickhouse Connect для записи извлечённых сообщений в базу данных и Yandex Translate API для их перевода на русский.



**Рис. 1.** Процесс разработки.

**Fig. 1.** Development process.

*Источник:* материалы 4-го Хакатона (AI Product Hackathon)<sup>11</sup>.

<sup>10</sup> Автоматизированное получение информации с веб-сайтов.

<sup>11</sup> Здесь и далее представлены фрагменты презентации команд, принимавших участие в Хакатоне. Ранее данные материалы в открытый доступ не выкладывались.

Основной платформой для создания и функционирования чат-бота была выбрана популярная социальная сеть Telegram: дата во время поиска статей определялась при помощи `telebot_calendar`.

На начальном этапе Хакатона команда `backpropagation.ai` пыталась переобучить модель на собственных данных, используя модели BERT для обработки немаркированного датасета и применяя библиотеки `Trasformers` и `PyTorch` для переобучения без учителя. Тем не менее, возникли проблемы с качеством модели, в частности, полученные результаты были сравнимы с самодельной моделью. В процессе настройки температуры модели было обнаружено, что получаемые при минимальной температуре результаты бессмысленны. В ходе консультации с организаторами и экспертами было решено перейти к `fine-tuning` и создать RAG-архитектуру, повысить температуру настройки для получения «читабельных» данных. Более того, конечной рабочей моделью стала `YandexGPT Pro`, библиотеками – `Langchain`, `Yandex-chain`, `FAISS` (для сравнения векторов) и др., базами данных – `Polars` и `Clickhouse_connect`; процесс векторизации проходил на `Colab`.

Во время защиты проекта команда `backpropagation.ai` продемонстрировала рабочий продукт в виде телеграм-бота с прилагаемым руководством к использованию, а также некоторые расчётные данные: план развития проекта (масштабирование) и необходимый бюджет.

### CatBot (CAIts)

Команда CAIts пошла иным путём: она решила работать с датасетами на японском языке, исключив китайский. Дообучив модель `YandexGPT` и протестировав базы данных `Elasticsearch`, `Clickhouse`, был проанализирован метод RAG на базе сервисов `YandexCloud` с использованием `BertBaseJapanese` для векторизации и создан телеграм-бот `CatBot` (рис. 2).

В отличие от предоставленного командой `backpropagation.ai` предварительного бюджета проекта CAIts провела соцопрос среди респондентов, интересующихся восточной тематикой, и смогла сделать предпродажу, попутно рассмотрев введение корпоративной подписки.

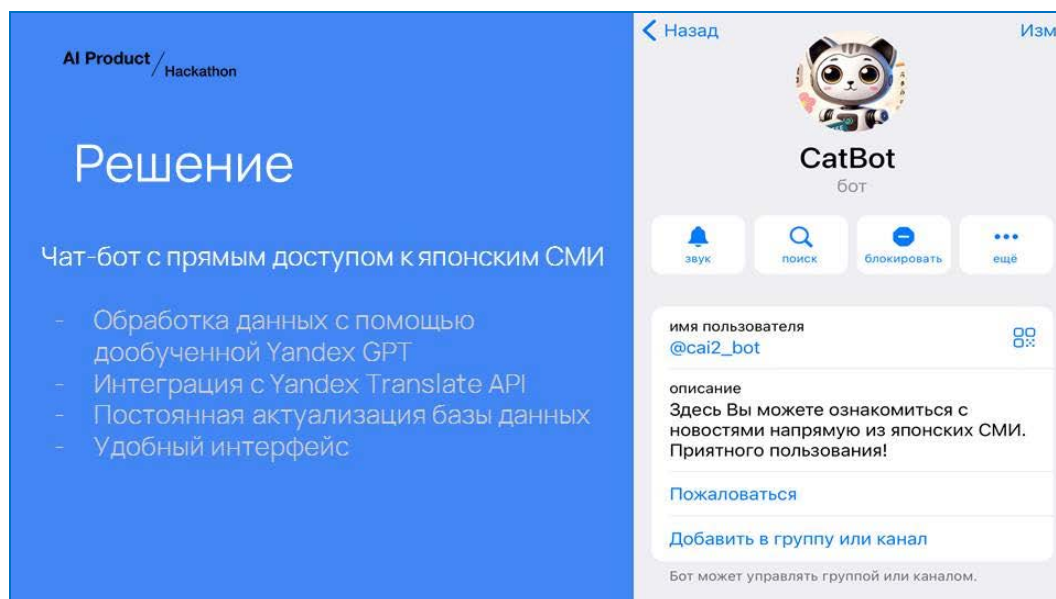


Рис. 2. Телеграм-бот, созданный командой CatBot.

Fig. 2. Telegram-bot created by the CatBot team.

News.ru

Международная команда News.ru, объединившая бакалавров и магистров, проходящих обучение в России, включая носителя китайского языка, представила незаконченный прототип модели чат-бота (рис. 3).

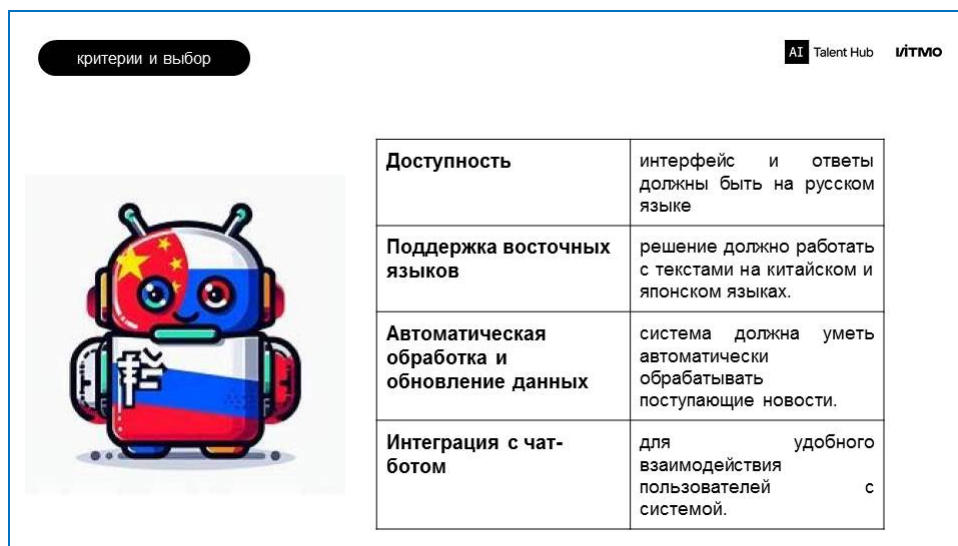


Рис. 3. Чат-бот, созданный командой News.ru.  
 Fig. 3. Chatbot created by the News.ru team.

«Ошибки распределения»

Команда «Ошибки распределения» (рис. 4) творчески подошла к решению задачи, сконцентрировавшись на четырёх основных компонентах разработки: масштабируемость проекта (обработка больших объёмов данных), гибкость (адаптация продукта под разные сценарии пользования), точность (высокая точность анализа и предсказаний) и простота (лёгкость интеграции с существующими системами).

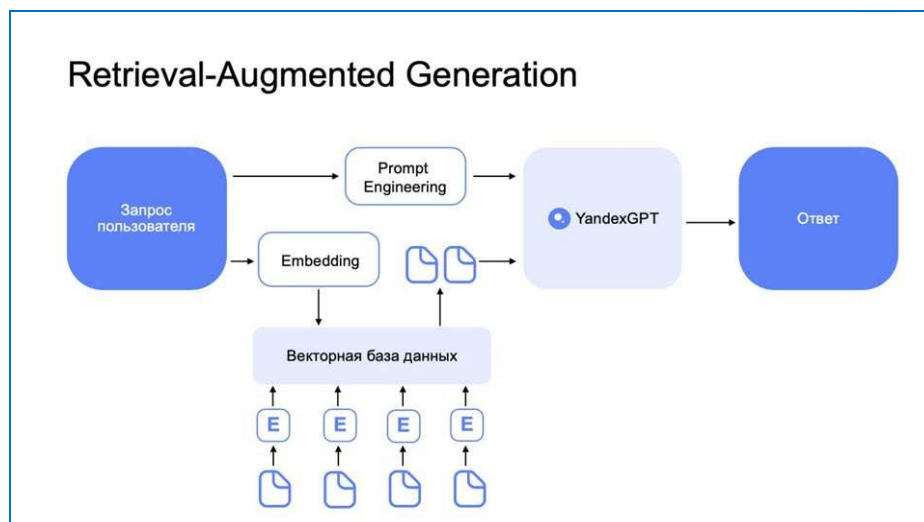


Рис. 4. Рабочая модель чат-бота команды «Ошибки распределения».  
 Fig. 4. Working model of the chatbot of the “Oshibki raspredeleniya” [Distribution Errors] team.



## «Серсо»

Команда «Серсо» была единственным финалистом, который представил свой подход решения проблемы не через чат-бот в мессенджере Telegram, а через сайт-агрегатор (рис. 5).



Рис. 5. Требования и решение, представленные командой «Серсо».

Fig. 5. Requirements and solution presented by the “Serso” team.

Довольно простая и удобная модель предлагала пользователю выбор года новости с 2016 до 2023 г. и поиск доступных данных (рис. 6).

В ходе работы над проектом была проведена очистка данных. Во время первичной очистки были произведены следующие операции: обработка тегов (начало и конец контента), определение кодировки (библиотека `chardet`), использование регулярных выражений. Закончив выше названные операции, команда столкнулась с необходимостью проведения вторичной очистки: удаление строк короче 30/20 символов, фильтрация на основе количественного соотношения китайской или японской иероглифики к латинице и остальным символам, удаление текстов длиннее 10 тыс. символов.

### Чатбот-консультант по восточным СМИ (команда 'Серсо')

Выберите год от 2016 до 2023. База будет проанализирована по этот год включительно:

Введите ваш вопрос:

Дай определение инициативе «Один пояс, один путь».

**Отправить запрос**

«Один пояс, один путь» — это международная инициатива Китая, направленная на развитие торговых и инфраструктурных проектов в Евразии. Она предполагает создание экономических коридоров между Китаем и странами Европы через Центральную Азию, а также укрепление экономических связей и сотрудничества в регионе.

Инициатива включает проекты по строительству дорог, портов, железнодорожных путей, трубопроводов и других объектов инфраструктуры. Это призвано улучшить логистику, снизить транспортные издержки и расширить торговлю между странами, участвующими в инициативе.

«Один пояс, один путь» также нацелен на углубление культурного обмена и сотрудничество в сфере образования, науки и техники, чтобы способствовать формированию общего экономического пространства и процветанию региона.

Вас могут заинтересовать следующие источники:

1. <https://digital.hakoshin.jp/special/index/16196>
2. <https://news.mynavi.jp/article/teiji-147/>
3. <https://www.chinatimes.com/realtimenews/20220611003700-260405>
4. <https://www.chinatimes.com/opinion/20220601001097-262103>

**Рис. 6.** Пример работы чат-бота-консультанта команды «Серсо».  
**Fig. 6.** An example of the “Serso” team’s chatbot-consultant operation.

\* \* \*

В ходе двухнедельного Хакатона были выявлены достоинства и недостатки представленного для работы проекта. Предполагалось, что незнание восточного языка (китайского и японского) заметно снизит эффективность решения проблемы, и это частично подтвердилось на практике: возникали проблемы с вводом имён собственных и дальнейшим переводом; требовалась дополнительная проверка файлов компетентным специалистом (экспертом-востоковедом), учитывая особенности модели «забывать» условия и др. Существенным барьером в разработке продукта стали большой объём «сырых» данных, а также трудоёмкость процесса за ограниченный промежуток времени при необходимости постоянно дообучать модель. Тем не менее, команды смогли найти приемлемые ответы на

возникшие трудности, помня изначальные установки в расчёте на целевую аудиторию продукта: не только профессиональные востоковеды, эксперты-международники, но и широкий круг заинтересованных лиц.

Разработка цифрового инструментария востоковедов должна происходить с учётом следующих новых обстоятельств. Большие вычислительные мощности, удешевление хранения данных и облачные сервисы сделали большие данные доступными для рядовых исследователей. Агрегаторы веб-данных, такие как Common Crawl, избавляют учёных от необходимости создавать собственные системы веб-скрейпинга, т.е. обхода сайтов и скачивания с них данных. Большие языковые модели, в том числе размещённые в открытом доступе, в сочетании с соответствующими программными библиотеками позволяют оперировать с текстами на разных языках, осуществляя перевод, аннотирование, семантический поиск. Доступность разнообразных программных библиотек для обработки текстов на разных языках избавляет от необходимости самостоятельно реализовывать базовые функции, такие как сегментация и лемматизация текста. Всё это смещает фокус с частных технических или лингвистических задач в область поиска необходимых конфигураций из готовых компонентов, и это открывает новые возможности для конструирования специализированных инструментов под конкретные исследовательские задачи.

4-й Хакатон (AI Product Hackathon) в полной мере соответствует духу Концепции технологического развития РФ до 2030 г. Результаты, полученные во время проведения мероприятия, заложили основу для дальнейших исследований не только в области китаистики и японоведения, но и востоковедения в целом. Синтез гуманитарных и технических наук видится ключом к достижению Россией суверенитета в научно-технической сфере.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Виноградов А.В.* 3-й пленум ЦК КПК 20-го созыва: факторы нестабильности и перспективы развития // Проблемы Дальнего Востока. 2024. № 4. С. 7–19. DOI 10.31857/S0131281224040014
- Кудакаев Р.Ф.* Ежегодная всероссийская научная конференция «Современное китайское государство» // Проблемы Дальнего Востока. 2024. № 2. С. 182–192. DOI 10.31857/S0131281224020144
- Островский А.В.* Экономика Китая: новые горизонты развития до 2025 года // Проблемы Дальнего Востока. 2024. № 2. С. 73–85. DOI 10.31857/S0131281224020062
- Цифровая трансформация: ожидания и реальность: докл. к XXIII Ясинской (Апрельской) междунар. науч. конф. по проблемам развития экономики и общества, Москва, 2022 г. / Рук. авт. кол. П.Б. Рудник. — М.: Изд. дом Высшей школы экономики, 2022. — 221 с. ISBN 978-5-7598- 2658-3
- Юдина Т.Н., Яо Линдун.* Становление Китая как кибернетической сверхдержавы // Проблемы Дальнего Востока. 2024. № 2. С. 117–132. DOI 10.31857/S0131281224020094

## REFERENCES

- Kudakaev R.F. (2024). Ezhegodnaya vserossiyskaya nauchnaya konferentsiya «Sovremennoye kitayskoye gosudarstvo» [Annual All-Russian Conference “Modern Chinese State”]. *Problemy Dal'negu Vostoka* [Far Eastern Affairs], 2: 182–192. (In Russian). DOI 10.31857/S0131281224020144
- Ostrovskiy A.V. (2024). Ekonomika Kitaya: novyye gorizonty razvitiya do 2025 goda [The PRC Economy: New Horizons of its Development till 2025]. *Problemy Dal'negu Vostoka* [Far Eastern Affairs], 2: 73–85. (In Russian). DOI 10.31857/S0131281224020062

- Tsifrovaya transformatsiya: ozhidaniya i real'nost' [Digital transformation: expectations and reality]. Dokl. k XXIII Yasinskoy (Aprel'skoy) mezhdunar. nauch. konf. po problemam razvitiya ekonomiki i obshchestva [Report to the XXIII Yasinskaya (April) International Scientific Conference on problems of economic and social development]. Head of the author's team P.B. Rudnik. HSE University Publishing House. Moscow, 2022. 221 p.
- Vinogradov A.V. (2024). 3-y plenum TsK KPK 20-go sozyva: faktory nestabil'nosti i perspektivy razvitiya [The Third Plenum of the CCP Central Committee: Factors of Instability and Development Prospects]. *Problemy Dal'nego Vostoka* [Far Eastern Affairs], 4: 7–19. (In Russian). DOI 10.31857/S0131281224040014
- Yudina T.N., Yao Lingdong (2024). Stanovleniye Kitaya kak kiberneticheskoy sverkhderzhavy [The Emergence of China as a Cybernetic Superpower]. *Problemy Dal'nego Vostoka* [Far Eastern Affairs], 2: 117–132. (In Russian). DOI 10.31857/S0131281224020094

---

Поступила в редакцию: 08.10.2024

Принята к публикации: 10.10.2024

Received: Oct 8, 2024

Accepted: Oct 10, 2024