

БЕЛОРУССКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет информационных технологий и робототехники
Кафедра «Программное обеспечение информационных систем и технологий»

ДОПУЩЕН К ЗАЩИТЕ

Старший преподаватель

Н.В. Воюш (подпись) Н.В. Воюш (инициалы и фамилия)

« 11 » 06 2026 г.

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
ДИПЛОМНОГО ПРОЕКТА

«WEB-сервис подключения больших языковых
модулей в корпоративный ИИ-чат»

Специальность 1-40 05 01 «Информационные системы и технологии (по направлениям)»

Направление специальности 1-40 05 01-04 «Информационные системы и технологии (в обработке и представлении информации)»

Обучающийся

группы 10702222
(номер)

11.02.06.26
(подпись, дата)

В.Ю. Лепесевич

Руководитель

04.06.26
(подпись, дата)

В.А. Полубок

Консультанты:

по разделу «Компьютерное
проектирование»

04.06.26
(подпись, дата)

В.А. Полубок

по разделу «Охрана труда»

04.06.26
(подпись, дата)

М.Л. Калиниченко

по разделу «Экономика»

04.06.26
(подпись, дата)

Н.В. Комина

Ответственный за нормоконтроль

08.06.26
(подпись, дата)

Н.С. Домаренко

Объем проекта:

расчетно-пояснительная записка — 97 страниц;

графическая часть — 10 листов;

магнитные (цифровые) носители — 1 единиц.

Минск 2026

РЕФЕРАТ

КОРПОРАТИВНЫЙ AI-ЧАТ, БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ, SHADOW AI, REACT, OPENROUTER, ПОТОКОВАЯ ПЕРЕДАЧА ДАННЫХ.

Объектом исследования является процесс интеграции больших языковых моделей в корпоративную информационную инфраструктуру.

Цель проекта – проектирование и разработка веб-сервиса подключения больших языковых модулей в корпоративный ИИ-чат, обеспечивающего унифицированный интерфейс взаимодействия с несколькими LLM-провайдерами через агрегирующий API.

В процессе проектирования выполнены следующие исследования: проанализирована проблема «теневого ИИ» (Shadow AI); исследованы доступные API-решения и агрегаторы языковых моделей; обоснован стек технологий (React, Vite, Tailwind CSS); спроектирована архитектура системы; разработан и протестирован прототип веб-сервиса с поддержкой потоковой передачи ответов и переключения между моделями.

Элементами практической значимости полученных результатов являются: разработанное масштабируемое веб-приложение корпоративного AI-чата на базе агрегирующего шлюза OpenRouter, обеспечивающее безопасное взаимодействие с языковыми моделями, потоковую передачу ответов, переключение между моделями и ролями ассистента, а также учёт стоимости запросов.

Областью возможного практического применения являются коммерческие предприятия и государственные организации Республики Беларусь, нуждающиеся во внедрении безопасных инструментов генеративного ИИ с соблюдением требований законодательства о защите персональных данных. Прототип создаёт контролируемую точку доступа к языковым моделям и снижает риски утечки конфиденциальной информации.

Технико-экономическое обоснование разработки показало целесообразность внедрения продукта за счёт снижения финансовых рисков от утечек данных и экономии рабочего времени сотрудников.

Студент-дипломник подтверждает, что приведённый в дипломном проекте расчётно-аналитический материал объективно отражает состояние разрабатываемого объекта, все заимствованные из литературных и других источников теоретические и методологические положения и концепции сопровождаются ссылками на их авторов.

Дипломный проект: 97 с., 13 рис., 18 табл., 29 источник, 1 прил.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Menlo Security. 2025 Report: How AI is Shaping the Modern Workspace [Электронный ресурс]. – Режим доступа: <https://www.menlosecurity.com/press-releases/menlo-securitys-2025-report-uncovers-68-surge-in-shadow-generative-ai-usage-in-the-modern-enterprise>. – Дата доступа: 20.04.2026.
- 2 UpGuard. Shadow AI is widespread – and executives use it the most [Электронный ресурс]. – Режим доступа: <https://www.cybersecuritydive.com/news/shadow-ai-employee-trust-upguard/805280/>. – Дата доступа: 20.04.2026.
- 3 IBM. Cost of a Data Breach Report 2025 [Электронный ресурс]. – Режим доступа: <https://www.ibm.com/reports/data-breach>. – Дата доступа: 20.04.2026.
- 4 Knostic AI. Detect and Control: Shadow AI in the Enterprise [Электронный ресурс]. – Режим доступа: <https://www.knostic.ai/blog/shadow-ai>. – Дата доступа: 20.04.2026.
- 5 Gallup. State of the Global Workplace 2026 Report [Электронный ресурс]. – Режим доступа: <https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx>. – Дата доступа: 20.04.2026.
- 6 SQ Magazine. Shadow AI Usage Statistics 2026 [Электронный ресурс]. – Режим доступа: <https://sqmagazine.co.uk/shadow-ai-usage-statistics/>. – Дата доступа: 20.04.2026.
- 7 Stack Overflow. Developer Survey 2024 [Электронный ресурс]. – Режим доступа: <https://survey.stackoverflow.co/2024/>. – Дата доступа: 20.04.2026.
- 8 Закон Республики Беларусь от 7 мая 2021 г. № 99-З «О защите персональных данных» [Электронный ресурс] // Национальный правовой интернет-портал Республики Беларусь. – Режим доступа: <https://pravo.by/document/?guid=12551&p0=H12100099>. – Дата доступа: 20.04.2026.
- 9 OpenRouter Documentation [Электронный ресурс]. – Режим доступа: <https://openrouter.ai/docs>. – Дата доступа: 20.04.2026.
- 10 Anthropic. API Privacy Policy [Электронный ресурс]. – Режим доступа: <https://www.anthropic.com/privacy>. – Дата доступа: 20.04.2026.
- 11 React. React Documentation [Электронный ресурс]. – Режим доступа: <https://react.dev/>. – Дата доступа: 20.04.2026.
- 12 Vite. Vite Documentation [Электронный ресурс]. – Режим доступа: <https://vitejs.dev/>. – Дата доступа: 20.04.2026.
- 13 Tailwind CSS. Tailwind CSS Documentation [Электронный ресурс]. – Режим доступа: <https://tailwindcss.com/docs>. – Дата доступа: 20.04.2026.
- 14 OpenAI. OpenAI Platform Documentation [Электронный ресурс]. – Режим доступа: <https://platform.openai.com/docs>. – Дата доступа: 20.04.2026.
- 15 Google. Gemini API Documentation [Электронный ресурс]. – Режим доступа: <https://ai.google.dev/gemini-api/docs>. – Дата доступа: 20.04.2026.

- 16 Yandex Cloud. YandexGPT API Documentation [Электронный ресурс]. – Режим доступа: <https://yandex.cloud/ru/docs/foundation-models/>. – Дата доступа: 20.04.2026.
- 17 Mozilla Developer Network. Using server-sent events [Электронный ресурс]. – Режим доступа: https://developer.mozilla.org/en-US/docs/Web/API/Server-sent_events/Using_server-sent_events. – Дата доступа: 20.04.2026.
- 18 Mozilla Developer Network. Fetch API [Электронный ресурс]. – Режим доступа: https://developer.mozilla.org/en-US/docs/Web/API/Fetch_API. – Дата доступа: 20.04.2026.
- 19 Mozilla Developer Network. AbortController [Электронный ресурс]. – Режим доступа: <https://developer.mozilla.org/en-US/docs/Web/API/AbortController>. – Дата доступа: 20.04.2026.
- 20 Google. Web Vitals [Электронный ресурс]. – Режим доступа: <https://web.dev/vitals/>. – Дата доступа: 20.04.2026.
- 21 ГОСТ 19.102-77 Единая система программной документации. Стадии разработки. – М.: Издательство стандартов, 1980. – 5 с.
- 22 ГОСТ 34.601-90 Информационная технология. Автоматизированные системы. Стадии создания. – М.: Издательство стандартов, 1992. – 8 с.
- 23 Закон Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации» [Электронный ресурс] // Национальный правовой интернет-портал Республики Беларусь. – Режим доступа: <https://pravo.by/document/?guid=3871&p0=h10800455>. – Дата доступа: 20.04.2026.
- 24 Brown, T. B. [и др.]. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 1877–1901.
- 25 Wei, J. [и др.]. Emergent Abilities of Large Language Models // Transactions on Machine Learning Research. – 2022. – № 1. – С. 1–35.
- 26 Gartner. Hype Cycle for Artificial Intelligence 2024 [Электронный ресурс]. – Режим доступа: <https://www.gartner.com/en/articles/gartner-hype-cycle-for-artificial-intelligence>. – Дата доступа: 20.04.2026.
- 27 OWASP Foundation. OWASP Top 10 for Large Language Model Applications 2025 [Электронный ресурс]. – Режим доступа: <https://owasp.org/www-project-top-10-for-large-language-model-applications/>. – Дата доступа: 20.04.2026.
- 28 Lewis, P. [и др.]. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 9459–9474.
- 29 Microsoft. Active Directory Domain Services Overview [Электронный ресурс]. – Режим доступа: <https://learn.microsoft.com/en-us/windows-server/identity/ad-ds/>. – Дата доступа: 20.04.2026.