

## РЕЗЮМЕ

август 2024

### О компании

- Зарегистрирована в 2023г.
- Офисы в Москве, Санкт-Петербурге и Кирове
- Резидент i.moscow
- В реестре ИТ-компаний за № 65089

### Специализация

- Разработка интеллектуальных систем анализа и генерации текстов.

### Правообладание – интеллектуальная поисковая система «Минч»

- Свидетельство на ПО №2023660092 от 17.05.2023
- Запись № 19231 о внесении в реестр российского ПО от 23.09.2023

### Команда

Ядро «Минч ИИ» состоит из команды ученых Лаборатории интеллектуальных систем ВятГУ и Европейского университета

12 ML-специалистов и разработчиков, в т.ч. 1 д.т.н., 3 к.т.н., 4 аспиранта:

- Научный руководитель, Team-lead (д.т.н.);
- NLP-engineer (к.т.н.);
- NLP-developer (аспирант);
- Backend-разработчик (к.т.н.);
- Backend-разработчик (аспирант);
- frontend-разработчик;
- Data Scientist (к.т.н.);
- Data Scientist (аспирант);
- MLOps/DevOps;
- BD-разработчик (аспирант);
- системный аналитик;
- бизнес-аналитик

### Научный руководитель

[д.т.н., профессор Евгений Котельников](#)

- в 2019 год вошел в Топ-6 российских исследователей в области

- компьютерной лингвистики по версии альманаха «Искусственный интеллект»;
- руководит магистратурой «Машинное обучение и анализ данных»;
  - соорганизатор международных и всероссийских открытых соревнований по компьютерной лингвистике SentiRuEval-2015, SemEval-2016, RuArg-2022;
  - постоянный рецензент ведущих конференций по анализу текстов и социальных медиа:
    - Международная конференция по компьютерной лингвистике и интеллектуальным технологиям «Диалог»;
    - International Conference on Text, Speech and Dialogue (TSD);
    - International Conference on Artificial Intelligence and Natural Language (AINL);
    - International Conference «Analysis of Images, Social networks and Texts» (AIST);
    - International Conference «Digital Transformations & Global Society» (DTGS);
    - International Conference «Electronic Governance and Open Society: Challenges in Eurasia» (EGOSE).
  - стажировался в Потсдамском университете (2018) и Дармштадтском техническом университете (2021).

### **Области исследований**

- обработка естественных языков (NLP): автоматическая обработка и генерация текстов на русском, английском и других языках;
- исследование и применение больших языковых моделей;
- машинное обучение и анализ данных.

Более 100 научных работ, в том числе более 50 из них в изданиях, индексируемых Web of Science и Scopus по направлениям:

### **Сотрудничество с командами**

МГУ, ВШЭ, ИСП РАН, ИТМО, НГТУ

### **Некоторые гражданские проекты за последние 5 лет**

- Разработка автоматизированной системы интеллектуального поиска с генерацией ответов на основе дополненной выборки (дочерняя компания Евраз, 2024).
- Разработка интеллектуального поиска по документации (подведомственное учреждение Ростехнадзора).
- Развертывание и эксплуатация Базы знаний с интеллектуальным поиском по Базе корпоративной документации с интеграцией в базу знаний InKnowledge (дочерняя компания РЖД, завод, совместный проект с вендором базы знаний InKnowledge).
- Разработка системы семантического поиска для платформы ELMA365 (для

компании Элма).

- Разработка модуля проверки соответствия заполнения документов (антифрод) для платформы ELMA365 (для международного ритейлера строительных материалов).
- Разработка системы информационной поддержки с интеллектуальным поиском и чат-ботом для клиентского отдела банка (для регионального банка).
- Разработка вопросно-ответной системы по базе нормативных документов с интеграцией в ИС заказчика с интеграцией в 1С:Документооборот (фармацевтическая компания).
- Разработка и исследование моделей классификации товаров на основе текстового описания.
- Автоматический анализ выпускных квалификационных работ, включая анализ заимствований, оценку новизны и актуальности, анализ цитирований, тематической связности, качества терминологии, фильтрацию искусственных текстов (госзадание Минобрнауки).

#### **Некоторые гранты за последние 5 лет**

- Разработка и исследование метода управляемой генерации текстов по сюжетной линии (грант РФФ, 2023–2024).
- Разработка учебного курса «Машинное обучение на основе больших данных» (грант фонда Потанина, 2023–2024).
- Разработка и исследование методов генерации текстов с аргументацией (грант РФФ, 2022–2023).
- Разработка учебного курса «Современные нейросетевые языковые модели» (грант фонда Потанина, 2022–2023).
- Generating creative language (грант Минобрнауки в рамках международного научно-образовательного сотрудничества по программе «Михаил Ломоносов», 2021).
- Обзор современных словарей оценочной лексики для анализа мнений на русском и английском языках (грант РФФИ, 2020).
- Разработка и исследование аннотированного русскоязычного текстового корпуса для анализа аргументации (грант Минобрнауки в рамках международного научно-образовательного сотрудничества по программе «Михаил Ломоносов», 2018).
- Разработка и исследование словарей оценочной лексики для анализа тональности текстов (грант Минобрнауки, 2017–2019).

#### **Стек**

- LLM (TOP-5 opensource model Leaderboard Chatbot Arena),
- embedding-model (TOP-5 opensource model MTEB-Leaderboard) + finetune, LoRA, p-tuning;
- RAG (Retrieval Augmented Generation)

- Python and NLP library (spaCy, PyTorch, Transformers, vLLM, Flask)
- Gunicorn/uvicorn
- Docling
- REST API, FastAPI, Swagger
- JavaScript/TypeScript (Node.js)
- Java/Kotlin
- PostgreSQL, Opensearch, Milvus, MongoDB
- CI/CD
- Docker, K8s, Kafka
- Gitlab