



Murat APISHEV

Moscow / Russia | +7-916-769-82-75

mel-lain@yandex.ru

[github](#) | [My page \(in russian\)](#)

SUMMARY

I am an expert in data science, machine learning, natural language processing and software development. My skills include data analysis, training ML and DL models, designing and writing efficient code, team leading. Also I teach a lot and like to be a public speaker.

WORK EXPERIENCE AND PROJECTS

SBERDEVICES

Since May 2023

NLP-ENGINEER (SENIOR)

- Research and experiments with LLM training, fine-tuning and measuring in GigaChat project

JUST AI

October 2021 - April 2023

NLP R&D TEAM LEADER

- NLP core of cloud text processing engine (*Team leading / Project management / NLP / Python*)
- Project on training and distilling models for text vectorization and classification (*NLP / Python*)
- Project on training models for controllable text paraphrasing (*NLP / Research / Python*)
- Extensive research on LLM development with the publication of a review on habr.com (*NLP / Research*)

TALKMART42

April 2020 - October 2022

DATA SCIENTIST (CO-FOUNDER)

- Prototype of voice control system for self-service checkout for retail company (*Python / NLP*)
- English neural QA-system for specific domain for advertising agency (*Python / NLP*)
- Advertisements classification system for advertising and media research company (*Python / NLP*)
- Voice games for Yandex Alice and Sber Salut (*Python*)
- Voice and visual business apps in Sber Salut (*TypeScript / Python*)

ALFA BANK

August 2021 - October 2021

DATA SCIENTIST (SENIOR)

- ML and DL models for chat-bot intent detection (*Python / NLP*)
- Project management of several NLP tasks with reporting to senior officials (*Team leading / Project management*)

DIGITAL DECISIONS (AITHEA)

April 2019 - April 2020

NLP TEAM LEADER / DATA SCIENTIST (SENIOR)

- Spell checker for government organization (*C++ / Python / NLP*)
- English NER in application onboarding for B2C startup (*Project management*)
- Search and recommender system for English scientific articles (*Project management / Team leading*)
- Trend detection in Twitter and YouTube data for Instinct BBDO (*Python / NLP / Team leading*)
- Collecting NLP-data for Huawei voice assistant (*Python / NLP / Team leading*)
- Tickets ordering chatbot for B2C company (*NLP / Team leading / Project management*)

YANDEX ZEN

September 2017 - April 2019

ML-ENGINEER (MIDDLE)

- System for evergreen content detection within Zen recsys (*Java / Python / SQL / ML*)
- System for one-day/yellow/black content classification within Zen recsys (*Java / Python / NLP*)
- Zen recsys channel catalog ranking system (*Java / SQL*)
- Zen recsys fresh content usage strategies algorithms (*Java / SQL*)

YANDEX SEARCH

April 2016 - September 2017

SOFTWARE ENGINEER (JUNIOR)

- Yandex Search antirobot development (*C++*)

COMPUTING CENTRE OF RUSSIAN ACADEMY OF SCIENCES

July 2014 - April 2016

SOFTWARE ENGINEER

- BigARTM - open-source library for topic modeling (*C++ / Python / NLP*)
- Ethnic content exploration in social networks data (*Python / NLP*)

FREELANCE

DATA SCIENTIST

- Clustering and sentiment analysis of bank employees reviews (*Python / NLP*)
- Retail data analysis and parsing for scientific nutritional project (*Python / NLP*)
- Development of recommender system for online food store (*Python*)

EDUCATION

MOSCOW STATE UNIVERSITY (CMC) / FRC IM OF THE RUSSIAN ACADEMY OF SCIENCES 2017-2020

PH.D IN COMPUTER SCIENCE

- Effective Implementation of Topic Modeling Algorithms with Additive Regularization

MOSCOW STATE UNIVERSITY (CMC) 2015-2017

MASTER OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

MOSCOW STATE UNIVERSITY (CMC) 2011-2015

BACHELOR OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

MAIN TEACHING EXPERIENCE

ONLINE EDUCATION COMPANY NETOLOGY 2022

LECTURER OF NLP / DL COURSES

MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY 2019-2023

LECTURER OF PYTHON COURSE

HIGHER SCHOOL OF ECONOMICS 2020

LECTURER OF NLP ONLINE COURSE

MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY / MOSCOW STATE UNIVERSITY 2017-2023

LECTURER OF NLP COURSE (EARLIER - ASSISTANT)

YANDEX SCHOOL OF DATA ANALYSIS 2016-2019

ASSISTANT IN ML COURSE

HIGHER SCHOOL OF ECONOMICS 2017-2018

ASSISTANT IN NLP COURSE

SKILLS, MODELS, APPROACHES

PROGRAMMING LANGUAGES	Experienced: Python 2/3 C++11 Familiar: TypeScript Java SQL
SOFTWARE DEVELOPMENT	Algs and Data Structures OOP Patterns Multi-Threaded Programming MapReduce
GENERAL ML AND DS	Log/Linear Regression kNN RF GB PCA t-SNE K-means GMM DBSCAN SMOTE Time-Series Exp. Smoothing ARIMA A/B Tests Stat. Hypothesis Testing
NON-DEEP NLP	Text Preprocessing Regexes TF-IDF Word2vec GloVe FastText LDA ARTM
DEEP NLP	FFNN CNN GAN RNN Seq-to-seq (Attention) Transformer BERT-like models GPT-like models T5-like models
TECHNICAL SKILLS	Linux Terminal (Ubuntu) VC (Git, SVN) Unit Testing CI LaTeX
SOFT/MANAGEMENT SKILLS	Teaching Agile (Scrum, Kanban) Waterfall Fluent Speech (Russian)
LANGUAGES	Russian (Native) English (Intermediate)

TOOLS, LIBRARIES, FRAMEWORKS

SOFTWARE DEVELOPMENT	Experienced: STL Pyenv & Venv Pipenv Docker Redis Flask Google Protobuf Familiar: Hadoop PySpark MPI CUDA Cython Numba Android SDK (Java) QT React GDB CMake MongoDB S3 Travis CI AWS
GENERAL ML	NumPy Pandas Sklearn Matplotlib SciPy XGBoost CatBoost Vowpal Wabbit Statsmodels Faiss ScaNN Jupyter Notebook Colab Yandex Toloka Doccano
NON-DEEP NLP	Re Pymorphy2/Pymystem3 NLTK SpaCy FastText Gensim BigARTM UDPipe
DEEP NLP	PyTorch Huggingface ONNX Sentence Transformers Haystack DrQA EasyNMT
VOICE ASSISTANTS, ASR/TTS	Yandex SpeechKit Vosk (Kaldi) Yandex Dialogs Google Actions RHVoice
MANAGEMENT/OTHERS	Yandex Tracker Emacs MS Office Google Docs Overleaf Beamer

MAIN PUBLIC ACTIVITIES

- OpenTalks.AI Conference, 2023 (single speech)
- Conversetions, 2022 (single speech)
- Alice Urban Skills Hackathon, 2021 (with team, 3rd place)
- VirusHack Hackathon, 2020 (with team, 1st place)
- DataStart Conference, 2020 (two paired speeches, Instinct BBDO and Huawei)
- OpenTalks.AI Conference, 2020 (paired speech, Huawei)

MAIN SCIENTIFIC PAPERS

- Evgeny O., **Apishev M.** Paraphrasers and Classifiers: Controllable Text Generation for Text Style Transfer. AIST, 2023
- Artemova E., **Apishev M.**, Sarkisyan V., Aksenov S., Kirjanov D., Serikov O. Teaching a Massive Open Online Course on Natural Language Processing. NAACL Workshop, 2021
- **Apishev M.**, Vorontsov K. Learning Topic Models With Arbitrary Loss. FRUCT, 2020
- **Apishev M.**, Koltcov S., Koltsova O., Nikolenko S., Vorontsov K. Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts. MICAI, 2016
- Frei O., **Apishev M.** Parallel Non-blocking Deterministic Algorithm for Online Topic Modeling. AIST, 2016
- Vorontsov K., Frei O., **Apishev M.**, Romov P., Suvorova M., Yanina A. Non-Bayesian Additive Regularization for Multimodal Topic Modeling of Large Collections. CIKM Workshop, 2015