



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.

## EDUCATION

<b>MOSCOW STATE UNIVERSITY (CMC) / FRC IM OF THE RUSSIAN ACADEMY OF SCIENCES</b>	2017-2020
<u>PH.D IN COMPUTER SCIENCE</u>	
•Effective Implementation of Topic Modeling Algorithms with Additive Regularization	
<b>MOSCOW STATE UNIVERSITY (CMC)</b>	2015-2017
<u>MASTER OF APPLIED MATHEMATICS AND COMPUTER SCIENCE</u>	
<b>MOSCOW STATE UNIVERSITY (CMC)</b>	2011-2015
<u>BACHELOR OF APPLIED MATHEMATICS AND COMPUTER SCIENCE</u>	

## WORK EXPERIENCE AND PROJECTS

<b>TALKMART42</b>	April 2020 - Now
<u>DATA SCIENTIST (CO-FOUNDER)</u>	
•Prototype of voice control system for self-service checkout for retail company ( <i>Python / NLP</i> )	
•English neural QA-system for specific domain for advertising agency ( <i>Python / NLP</i> )	
•Advertisements classification system for advertising and media research company ( <i>Python / NLP</i> )	
•Voice games for Yandex Alice and Sber Salut ( <i>Python</i> )	
•Voice and visual business apps in Sber Salut ( <i>Typescript (React) / Python</i> )	
<b>JUST AI</b>	October 2021 - Now
<u>NLP R&amp;D TEAM LEADER</u>	
•NLP core of cloud text processing engine for internal usage and external marketplace ( <i>Team leading / Project management / NLP / Python</i> )	
<b>ALFA-BANK</b>	August 2021 - October 2021
<u>DATA SCIENTIST (LEAD)</u>	
•ML and DL models for chat-bot intent detection ( <i>Python / NLP</i> )	
•Project management of several NLP tasks with reporting to senior officials ( <i>Team leading / Project management</i> )	
<b>DIGITAL DECISIONS (AITHEA)</b>	April 2019 - April 2020
<u>NLP TEAM LEADER / DATA SCIENTIST (SENIOR)</u>	
•Spell checker for government organization ( <i>C++ / Python / NLP</i> )	
•English NER in application onboarding for B2C startup ( <i>Project management</i> )	
•Search and recommender system for English scientific articles ( <i>Project management / Team leading</i> )	
•Trend detection in Twitter and YouTube data for Instinct BBDO ( <i>Python / NLP / Team leading</i> )	
•Collecting NLP-data for Huawei voice assistant ( <i>Python / NLP / Team leading</i> )	
•Tickets ordering chatbot for B2C company ( <i>NLP / Team leading / Project management</i> )	
<b>YANDEX ZEN</b>	September 2017 - April 2019
<u>ML-ENGINEER (MIDDLE)</u>	
•System for evergreen content detection within Zen recsys ( <i>Java / Python / SQL / ML</i> )	
•System for one-day/yellow/black content classification within Zen recsys ( <i>Java / Python / NLP</i> )	
•Zen recsys channel catalog ranking system ( <i>Java / SQL</i> )	
•Zen recsys fresh content usage strategies algorithms ( <i>Java / SQL</i> )	
<b>YANDEX SEARCH</b>	April 2016 - September 2017
<u>SOFTWARE ENGINEER (JUNIOR)</u>	
•Yandex Search antirobot development ( <i>C++</i> )	

**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)

**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)
- Clustering and sentiment analysis of bank employees reviews (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)

**MAIN TEACHING EXPERIENCE**

**MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY**

2019-2021

LECTURER OF PYTHON COURSE

**HIGHER SCHOOL OF ECONOMICS**

2020

LECTURER OF NLP ONLINE COURSE

**MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY / MOSCOW STATE UNIVERSITY**

2017-2020

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**

<b>PROGRAMMING LANGUAGES</b>	<b>Experienced:</b> Python 2/3   C++11 <b>Familiar:</b> Java   JavaScript   SQL
<b>SOFTWARE DEVELOPMENT</b>	Algs and Data Structures   OOP Patterns   Multi-Threaded Programming   MapReduce
<b>GENERAL ML AND DS</b>	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
<b>NON-DEEP NLP</b>	Text Preprocessing   Regexes   TF-IDF   Word2vec   GloVe   FastText   LDA   ARTM
<b>DEEP NLP</b>	FFNN   CNN   GAN   RNN (BiLSTM)   Seq-to-seq (Attention)   Transformer   BERT   RoBERTa   ALBERT   ERNIE   DistilBERT   XLM   GPT-2   XLNet   T5   BART
<b>TECHNICAL SKILLS</b>	Linux Terminal (Ubuntu)   VC (Git, SVN)   Unit Testing   CI   LaTeX
<b>SOFT/MANAGEMENT SKILLS</b>	Teaching   Agile (Scrum, Kanban)   Waterfall   Fluent Speech (Russian)
<b>LANGUAGES</b>	Russian (Native)   English (Intermediate)

**TOOLS, LIBRARIES, FRAMEWORKS**

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

## MAIN PUBLIC ACTIVITIES

---

- 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

---

- 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.**, 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