# Anton Karazeev anton.karazeev@gmail.com



### **EDUCATION**

# Master of Science Moscow Institute of Physics and September 2019 — July 2021 <u>Technology</u> (expected)

• M.Sc. in Computer Science and Physics, <u>Department of Innovation and High Technologies</u>

Applied Mathematics and Physics

### <u>Bachelor of Science</u> <u>Moscow Institute of Physics and</u> September 2014 — July 2019 Technology

- B.Sc. in Computer Science and Physics, Department of Innovation and High Technologies
- Coursework for the state qualification exam in Physics at MIPT: "Molecular Dynamics" [Code]
- Intermediate Coursework: "Advanced Parser for Biomedical Texts" [Poster at MCCMB'17]
- Undergraduate Coursework: "Development of a mechanism for anomaly detection" [Presentation] [Code]

### **EXPERIENCE**

# R&D Data Scientist <u>Information Systems</u> July 2019 — Present Development Center

- Laboratory of Artificial Intelligence and Neural Networks.
- EDA and implementation of ML subsystems in different fields of interest of <u>Russian Railways</u>. Employee Turnover Prediction.
- Responsible for Natural Language Processing, processing of legal documents.

## Research Fellow <u>Laboratory for Digital Business</u>

March 2019 — Present

- Responsible for research on Anomalies and Outliers Detection.
- Found and fixed a bug concerning model based on Generative Adversarial Active Learning (GAAL) in PyOD toolkit for outlier detection.
- Developed a system for anomaly detection. Used Flask and SQLAlchemy frameworks.

### R&D Data Scientist ChatFirst September 2018 — April 2019

- Implementing different deep learning models to improve performance of chatbots, reading papers on related topics.
- Responsible for Natural Language Processing.
- Used BERT model to improve performance of production system in multiple aspects. Fine-tuned the model for downstream tasks.

### Research Fellow

### **Russian Quantum Center**

January 2018 — March 2019

- Responsible for Quantum Computing, Machine Learning.
- Developed <u>a series of seminars</u> concerning Quantum Computing. They include jupyter notebooks with basics of linear algebra, quantum mechanics and also work with QISKit (IBM), pyQuil (Rigetti) and Q# (Microsoft) was demonstrated.
- Supervised research student's bachelor degree diploma "Precision-Guaranteed Single-Qubit Process Tomography".
- Became a member of IBM Qiskit on GitHub.
- "Revealing quantum chaos with machine learning" arXiv preprint.

### **Teaching Assistant**

<u>Laboratory of Neural Networks</u> and Deep Learning

March — December 2017

• Responsible for preparing practical and theoretical assignments for the course of <u>Reinforcement Learning</u> and theoretical assignments for the course of <u>Natural Language Processing</u> with the number of 100+ enrolled students each.

### **Research Assistant**

### <u>Laboratory of Functional</u> <u>analysis of the Genome</u>

June 2016 — December 2017

- Research on protein function analysis.
- Text mining, Natural language processing, Keyword extraction, Machine learning algorithms. As an intermediate result the new method of keywords extraction using Information Theory proposed (ResearchGate).
- Participated in development of NLP package <u>SciLK</u> which was designed specifically for text-mining in natural sciences like biology and chemistry.

### **Data Scientist**

### Sberbank-Technology

August — October 2017

- Responsible for Natural Language Processing projects.
- Participated in preparing the datasets and building baselines for competition <u>Sberbank Data Science</u> Journey which is based on SQuAD.
- Developed an analogue of Amazon Mechanical Turk to improve experience of colleagues who evaluated the quality of collected datasets (Python, Flask).

R&D Data Scientist HiQE Group March — June 2017

- Negotiated with IBM engineers and applied some of the IBM Watson's services in tasks of signal processing.
- The system of baby cry recognition was built.

### **TECHNICAL EXPERIENCE**

### **Projects**

- **Service for Reading** (2019). Service has a web interface and an application for Android. It helps to read texts in foreign languages and easily add unknown words to the wordlist to further studying.
- Quantum Computing Bot (2018). Monitoring the load of IBM Q processors from IBM Quantum
   Experience. Bot is already available inside QISKit workspace in Slack (soon it will be uploaded here:
   <a href="https://github.com/Qiskit/qiskit-bot">https://github.com/Qiskit/qiskit-bot</a>). And it's available on Telegram: <a href="https://t.me/QuantumComputingBot">https://t.me/QuantumComputingBot</a>.
- **Quantum Keypad** (2018). This keypad allows to easily compose quantum circuits of different kinds. Besides keypad itself, Quantum Keypad consists of a power bank and Raspberry Pi Zero W. As a simulator I used QISKit package for Python. Inspired by Model Q.
- Reverse Engineering in Dispersion Engineering (2018). With my friend from EPFL we have a project on Dispersion Engineering. Our model predicts parameters of resonator system's simulation.
- <u>Frontopolar</u> (2017). Applied Reinforcement Learning for Stock Trading. State of the art results were reached. Different approaches were tested including Q-learning and Recurrent Reinforcement Learning.

### **Contributed to Open source**

- **PyOD** PR #108
- **QISKit** PR #366
- **pyQuil** PR #371
- SimulaQron PR#90
- Gensim fixed issue #671
- yandexdataschool/Practical RL PR #12
- My projects on GitHub

### **SKILLS**

- Russian: native, English: fluent, German: basics (A2)
- Programming languages: Python, C/C++, bash, R, SQL; experienced with JavaScript, HTML and CSS
- Python libraries: numpy, scikit-learn, pandas; for NLP: NLTK, Gensim; for Deep Learning: PyTorch, TensorFlow; for Quantum Computing: QISKit, pyQuil, Q#; for Web: Flask; for databases: peewee, SQLAlchemy
- DevOps: containers (Docker), cloud computing (AWS, GCP), code testing, source control (git)
- Experimented with RaspberryPi and Arduino. Projects
- Founded "MIPT SciTech Club" to discuss and share ideas on deep learning topics. Led a few seminars on topics such as "Introduction to bayesian methods"
- Experienced with **3D modeling** (FreeCAD, Blender) and **3D printing** (Ultimaker Cura, Ender 3)

### **TEACHING**

# Programming Existing Quantum Computers Cryptography course at Yandex School of Data Analysis Deep Reinforcement Learning course at MIPT, based on rll.berkeley.edu/deeprlcourse/ Practical assignments Deep Learning in Natural Language Processing course at MIPT, based on cs224n.stanford.edu Practical assignments PUBLICATIONS May 8, 2018 May 8, 2018 October — December 2017 March — December 2017

PUBLICATIONS	
"Revealing Quantum Chaos with Machine Learning"  arXiv preprint	February 2019
"Neural Network Quantum State Tomography"  Superconducting Quantum Technologies (SQT)  Poster (based on <a href="https://github.com/RQC-QApp/NNQST">https://github.com/RQC-QApp/NNQST</a> )	July 30 — August 3, 2018
"Precision-guaranteed quantum process tomography: Application to IBM Quantum Experience" Central European Workshop on Quantum Optics (CEWQO) Poster	May 21 — 25, 2018
"Generative Adversarial Networks (GANs): Engine and Applications"  Medium Story	August 2017
"Advanced Parser for Biomedical Texts"  Moscow Conference on Computational Molecular Biology (MCCMB)	July 27 — 30, 2017

### ADDITIONAL EDUCATION

Poster, Thesis

"Quantum Computing" course	Quantum Computing	February 1 — March 16, 2018
at Skoltech		

• Final Project - Quantum walks and Variational algorithm for 3- and 4-level systems.

"Summer school on Bayesian <u>DeepBayes Summer School</u> August 26 — 30, 2017 Methods in Deep Learning"

"Big Data in Bioinformatics"

**Bioinformatics Summer School** 

July 31 — August 5, 2017

• Participated in a hackathon during the school. Project.

"Natural Language Processing" course (based on cs224d.stanford.edu)

DeepHack Lab

September — December 2016

• Accepted a proposal to become a Teaching Assistant after the end of the course.

"Supercomputer technologies for atomistic modelling" course

Igor Morozov (<u>IHED RAS</u>)

September — December 2015

• Final Project - <u>Molecular Dynamics</u> is a program written in C using OpenMP framework for parallel computing. Used VMD for visualisation.

### **MOOCs**

- Neural Networks and Deep Learning by deeplearning.ai (2019)
- Mathematics and Python for Data Analysis by MIPT & Yandex (2017)
- Molecular Biology and Genetics by Bioinformatics Institute (2016)
- Neural Networks by Bioinformatics Institute (2016)

### **HACKATHONS**

### Kasárne/Kulturpark, Košice

### **Hack Kosice**

March 30 - 31, 2019

- Efficient and Faster Care challenge
- Implemented <u>healthcare system</u> using Zebra wristband printer and QR code scanner to identify patients easily
- Reduced the amount of time needed to register a new patient
- Presentation

### **Aalto University, Helsinki**

### **Junction**

November 23 — 25, 2018

- Applications with Bluetooth Mesh challenge
- Worked with Zephyr RTOS and reel board that has built-in Bluetooth transceiver
- Developed simple Industrial Internet of Things (IIoT) project which demonstrates the advantage of Bluetooth Mesh network

### Tochka Kipeniya, Moscow

### **Space Apps Challenge**

October 20 - 21, 2018

- Firefighter Bot for Telegram
- Implemented during <u>Space Apps 2018 Challenge</u> using data from NASA including <u>Active Fire Data</u> by NASA
- Python, python-telegram-bot
- Presentation

### Volkshotel, Amsterdam

### **Quantum Internet Hackathon**

October 13 — 14, 2018

- Worked with framework for Quantum Internet called SimulaQron
- Contributed to the SimulaQron project on GitHub
- <u>Implemented</u> quantum leader-election algorithm

### Skoltech, Moscow

### **Quantum Hackathon**

May 18, 2018

- 1st place
- There were problems on (1) quantum process tomography and (2) solving 3-SAT problem with QAOA
- Python, Quantum Information Toolkit (QIT)

### Aalto University, Helsinki

**Junction** 

November 24 - 26, 2017

- <u>LegalEngine</u> website/<u>telegram chat-bot</u>/email notification system, "qqmbr" team member, challenge by Castrén & Snellman
- Our solution makes the client-attorney interaction easier with the use of telegram chat-bot and email notifications, the attorney's work and billing more transparent to the client.
- Python, Flask library, html, css

EPFL, Lausanne <u>LauzHack</u> November 11 — 12, 2017

- 1st place in challenge by <u>SGS</u>, "NN:Nerds" team member, <u>Presentation</u>
- Solution allows quick access to the main concepts found in documents.
- Responsible for development of telegram-bot and processing documents using IBM Watson service for Natural Language Understanding. <u>Devpost</u>.
- Python, IBM Watson API, Telegram API

### Phystechpark, Moscow

### **mABBYYlity**

October 7 — 8, 2017

- 4th place, "App in the Restaurant" iOS application, <u>Demo</u>, <u>Presentation</u>
- App allows to recognise entities from restaurant menus using smartphone's camera and translates them.
   ABBYY Real-Time Recognition SDK, ABBYY Lingvo API and Spoonacular API were used.
- Python, Flask library

# **Skolkovo Moscow School of Management, Moscow**

### **Neurocampus**

September 22 - 24, 2017

- 2nd place, @SenseOfSpeech\_bot telegram-bot, Presentation
- Solution allows to extract emotions from user's recorded speech. Also it helps to train selected emotion with samples from TED talks.
- Speech Emotion Recognition (SER) module by <u>Vokaturi</u> was used as a core for telegram-bot based system to help users improve speech during performances.
- Python, Telegram API

### **MIPT, Moscow**

### **Bioinformatics Summer School**

August 3 - 4,2017

- "Prediction of Experimental Metadata from Gene Expression"
- Used Machine learning algorithms to predict phenotype by gene expression. Distinguish with high
  accuracy samples of male and female tissues of <u>Mus musculus</u> organism. Datasets from Gene Expression
  Omnibus were used. <u>Project</u>.

### ITMO, Saint Petersburg

### **BioHack**

March 3 - 5, 2017

- Text Mining, parsing the records from PubMed and UMLS.
- Analysis of research trends of chemical compounds and diseases during period of 1990-2015 using parsed information from PubMed database. <u>Project</u>.
- Python

### Wanha Satama, Helsinki

**Junction** 

November 25 - 27, 2016

• Used a python wrapper around the Twitter API and Topic Modeling of tweets (gensim).