

# Anton BELY

[anton.belyy@gmail.com](mailto:anton.belyy@gmail.com) · [avbelyy.github.io](https://avbelyy.github.io)

[LinkedIn](#) · [GitHub](#) · [Scholar](#)

## EDUCATION

SEP '19 — MAY '22	<b>M.Sc.Eng. in Computer Science (NLP)</b> , Johns Hopkins University, USA GPA: 4.0. Advisors: Benjamin Van Durme (Primary), Vladimir Braverman (Secondary)
SEP '13 — MAY '18	<b>B.Sc. in Informatics and Applied Mathematics</b> , ITMO University, Russia GPA: 3.6. Advisors: Andrey Filchenkov (Primary), Konstantin Vorontsov (Secondary) Thesis: <a href="#">Construction and Quality Evaluation of Heterogeneous Hierarchical Topic Models</a> <ul style="list-style-type: none"><li>• <b>Distinguished thesis award</b> (given to 2 out of approx. 50 undergraduates)</li></ul>

## EMPLOYMENT

JUN '22 — PRESENT	<b>Machine Learning Engineer in Knowledge Platform</b> , Apple, USA <ul style="list-style-type: none"><li>• Building ML systems for large-scale knowledge extraction from unstructured data feeds</li></ul>
SEP '19 — MAY '22	<b>Graduate Research Assistant</b> , Johns Hopkins University, USA <ul style="list-style-type: none"><li>• <b>Semantic data mining</b>: proposed novel ARM-based algorithm for script induction <a href="#">[paper]</a>, built <a href="#">SchemaBlocks</a>, Scratch-like annotation interface for complex event scenarios <a href="#">[preprint]</a></li><li>• <b>Knowledge graph completion</b>: building a human-in-the-loop KG completion system using entity linking, rule learning and data mining over million-scale knowledge graphs</li><li>• <b>Text generation</b>: built <a href="#">demo</a> for InFillmore, our FrameNet frame-guided NLG model <a href="#">[paper]</a></li></ul>
JUN '21 — AUG '21	<b>Research Intern in Semantic Machines</b> , Microsoft Research, USA <ul style="list-style-type: none"><li>• Built guided annotation interface to help label semantic parsing data 35% faster <a href="#">[preprint]</a></li></ul>
SEP '17 — AUG '19	<b>Senior Data Scientist in Compliance Risks and AI lab</b> , Tochka Bank, Russia <ul style="list-style-type: none"><li>• <b>Risk scoring</b>: vectorized new data sources for 200K+ bank clients and 50M+ transactions, generated temporal/spatial features using Hadoop/Spark to improve scoring accuracy by 10%</li><li>• <b>Communication analysis</b>: built intent recognition models to classify 90% customer inquiries</li><li>• <b>Call center planning</b>: using <a href="#">OR-Tools</a>, automated CC planning and improved accuracy by 10%</li><li>• <b>ML culture</b>: interviewed and mentored 3 junior ML engineers, designed internal ML guidelines</li></ul>
MAR '17 — AUG '17	<b>Machine Learning Engineer</b> , Antirutina, Russia <ul style="list-style-type: none"><li>• <b>Tender anomaly detection</b>: developed clustering algorithms to identify bidding anomalies, allowing to discover bid-rigging behavior on auctions with contract amount exceeding \$3B</li><li>• <b>Precise IE</b>: designed information extraction pipelines for precise identification of vendor codes, volumes and quantities of goods from unstructured and diverse vendors' price lists</li></ul>
OCT '15 — OCT '16	<b>Software Engineer in Ads</b> , VK.com, Russia <ul style="list-style-type: none"><li>• <b>URL fraud</b>: built service to periodically detect malicious URL redirect changes in VK ads</li><li>• <b>Click fraud</b>: built ML models to detect users that generate fraudulent clicks in VK ad network. Model was deployed semi-automatically and helped recover up to 3% monthly ad revenue</li><li>• <b>Ads search</b>: launched moderator search interface (incl. full-text search) over 30M+ VK ads</li><li>• <b>Ads scoring</b>: implemented advertiser ranking for faster moderation of top-10% clients</li></ul>

## TEACHING EXPERIENCE

JAN '21 — MAY '21	<b>Introduction to Algorithms 601.433/633 (Head TA)</b> , JHU (100+ students) <ul style="list-style-type: none"><li>• Managed 9 CAs and 1 TA, created homework and exam problems, held weekly office hours</li></ul>
SEP '17 — AUG '19	<b>Natural Language Processing (TA)</b> , Coursera (40,000+ students by Sep '19) <ul style="list-style-type: none"><li>• Answered 200+ students' questions, helped create homework and project assignments</li></ul>

## LANGUAGES AND TECHNOLOGIES

LANGUAGES	Python (proficient); JavaScript, bash (intermediate); C#, C++, Haskell, x86 assembly (coursework)
TECHNOLOGIES	pandas, sklearn, XGBoost, pytorch, faiss, nmslib; Docker, *SQL, MongoDB, Lucene, Hadoop/Spark