

# Anton BELY

anton.belyy@gmail.com · avbelyy.github.io

LinkedIn · GitHub · Scholar

## EDUCATION

AUG '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) Research projects: knowledge graph completion, controlled text generation, schema induction Research internship on <b>Conversational AI</b> at <b>Microsoft Research</b> in Summer '21: <a href="#">paper</a>
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) Research projects: plagiarism detection, hierarchical topic modeling 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 '24 — PRESENT	<b>Senior Machine Learning Engineer in Credit Risk Detection</b> , Stripe, USA <ul style="list-style-type: none"><li>• <b>Default prediction</b>: I lead project Voyager, which scores millions of Stripe merchants to predict whether they might go default in the near future and lead to Stripe-liable losses. Voyager is a PyTorch DNN-based model, which saves tens of millions in credit losses yearly.</li><li>• <b>Model explainability</b>: I implemented and deployed scalable post-hoc explanation pipelines for several ML models at Stripe, leading to higher model trust during review-based actioning.</li></ul>
JUN '22 — MAY '24	<b>Machine Learning Engineer in Knowledge Graphs and LLM Datasets</b> , Apple, USA <ul style="list-style-type: none"><li>• <b>KG</b>: I am a founding engineer of the ODKE project, which transforms unstructured Web data into millions of facts every week to continuously improve coverage and freshness of Apple KG. I implemented scalable fact extractors using in-house and OpenAI LLMs, built fact verification model using Spark MLlib, built CI/CD pipelines and oversaw integrating ODKE with Apple KG.</li><li>• <b>LLM Datasets</b>: I used MapReduce, Spark and ML to turn petabytes of multimodal Web data into billion-scale, high-quality datasets used to train LLMs for Siri and Apple Intelligence.</li></ul>
SEP '17 — AUG '19	<b>Machine Learning Engineer in Risks and ML Automation</b> , Tochka Bank, Russia <ul style="list-style-type: none"><li>• <b>Risk scoring</b>: using Hadoop and Spark, enriched internal data for 200K+ bank clients and 50M+ transactions with Web-scraped public data to improve risk scoring accuracy by 10%.</li><li>• <b>ML automation</b>: built ML models to detect intent for the top-90% client questions in chat. Built systems to automate scheduling call center shifts while improving accuracy by 10%.</li></ul>
OCT '15 — SEP '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 (advanced); Golang, Scala, JavaScript, SQL, C++, bash (proficient)
TECHNOLOGIES	Airflow, Spark, Iceberg (advanced); XGBoost, PyTorch, AWS, Docker, K8s (proficient)