

# ANDREI BELIAEV

[✉ andrewbelyaev2164@gmail.com](mailto:andrewbelyaev2164@gmail.com) [📞 +1 \(217\) 419-9945](tel:+1(217)419-9945) [🌐 andreibeliaev.github.io](http://andreibeliaev.github.io) [linkedin andreibeliaev](https://www.linkedin.com/in/andreibeliaev) [siiyayu](https://www.siiyayu.com)

## EDUCATION

### University of Illinois Urbana-Champaign

Master's in Computer Science | GPA 3.8/4

Champaign, IL

Aug 2024 – May 2026

Moscow, Russia

### Bauman Moscow State Technical University

Specialist in Autonomous Informational and Control Systems | GPA 4.7/5

Sep 2016 – Jun 2022

## EXPERIENCE

### AI/ML Software Engineer

*Epivara, Inc., BioTech*

Champaign, IL

Oct 2024 – Present

- (In progress) Adapting foundational CV models to support analysis & interactive annotation of reproductive tissue samples, training on NCSA Delta AI cluster, building a web platform to work with whole slide images.
- Built an AI powered full stack conversational data platform that connects to the research database, allowing researchers to generate analytical reports, create visualizations, and run statistical analysis using natural language.
- Integrated an AI assisted data ingestion workflow that parses, validates, and prepares uploaded data for approved insertion into the research database, bringing ingestion time down from hours to minutes.
- Conducted statistical study design ensuring the most powerful and appropriate methods are used to reduce costs, sample size, and generate reliable statistics.

### Data Analyst (Compliance)

*Tinkoff, FinTech*

Moscow, Russia

Feb 2022 - Oct 2022

- Contributed to feature development for the new Anti-Money Laundering model, which increased the share of fully automated task resolutions from 45% to 60% and reduced annual operational and other costs by \$1.7 million.
- Delivered insights and identified issues by creating dashboards in Tableau and aggregating data using SQL, Python, and Spark. Enhanced the analysis of financial fraud and compliance with Federal Law, saving 1000 hours annually.

### Data Analyst (Insurance)

*Tinkoff, FinTech*

Moscow, Russia

Mar 2021 - Feb 2022

- Created a system of 25+ metrics to measure UX, evaluate system robustness, and better assess customer service by leveraging SQL (Postgres), Python, and Spark to aggregate front- and back-end logs alongside table data.
- Collaborated with business and software teams to address product issues found. Established an iterative analysis and quality improvement process, increasing the product quality metric by 18% and saving \$60k monthly.
- Prepared 14 datasets with UX metrics for the ML team to enable churn prediction and proactive user engagement.
- Designed and conducted an A/B test to evaluate the impact of a new web landing page on conversion rates.

## PROJECTS

### Entropy-aware sampling in vLLM [\[Link\]](#)

- Built adaptive token selection mechanism for vLLM using GPU-batched lookahead to control generation diversity by dynamically penalizing tokens with high entropy reduction.
- Tracked across 30+ metrics (self-BLEU, perplexity, entropy dynamics) via Weights & Biases.

### Energy-Based Transformers [\[Link\]](#)

- Performed ablation study of MCMC sampling strategies to improve performance and scaling.

### Protecting Images from Generative AI Editing [\[Link\]](#)

- Implemented a semantic attack using the cross-attention layers of the U-Net to make images resistant to editing at the James M. Rehg lab, UIUC.

### Real-Time Trade Mirroring System

- Developed a real-time futures trade mirroring service using WebSockets and async Python, replicating orders and state updates between accounts with robust synchronization and failure handling achieving 56 ms latency.

## SKILLS

**Competencies:** ML, Deep Learning, LLMs, Computer Vision, Statistical Modeling, A/B Testing, Full Stack Development, HPC

**Programming Languages:** Python, R, C++, Java, TypeScript, SQL, MATLAB

**Tools & Frameworks:** PyTorch, LangGraph, vLLM, AWS, Docker, React, Zustand, Tableau, Grafana, PySpark, Neo4j, Wandb.ai