

Bekbol Shiribaiev

bekbolnyc@gmail.com linkedin.com/in/shiribaiev github.com/bshiribaiev bektbolsh.com

EDUCATION

The City College of New York - CUNY

Bachelor of Science in Computer Science & Economics

New York, NY

Expected May 2027

- **Coursework:** Data Structures, Algorithms, Software Engineering, Computer Organization, Operating Systems
- **Hackathon wins:** Stony Brook University Hacks 2025, Wellesley Hacks 2025, CUNY Tech Hacks 2025
- **Activities:** MIT - CCNY AI & Entrepreneurship Fellow, Colin Powell Fellow in Leadership and Public Service

EXPERIENCE

Software Engineer Intern

Stella Rising

Aug 2025 – Present

New York, NY

- Automated earnings-call transcript search by building a Selenium scraper with Gemini API for analysis, deployed on Google Cloud Run with AWS SES for weekly emails - saving 10 hours of sales research time per week
- Developed a QA system to validate Salesforce data with Pandas, AWS S3, improving data accuracy by 30%
- Engineered SEO reporting pipeline to fetch data from Google Search Console, Semrush, and Looker Studio APIs, generates PowerPoint slides with python-pptx, and generates analysis with Gemini, saving \$80,000 per year

Software Engineer Intern

Evergreen Investments

Feb 2025 – May 2025

New York, NY

- Built an agent-messaging backend with Express.js/PostgreSQL, that increased lead conversion by **30%**
- Deployed a Java Spring Boot microservice for agent messaging, horizontally scalable behind a load balancer, reducing API latency from 300ms to 100ms and improving reliability under growing traffic
- Implemented a responsive property-search UI in React/TypeScript with optimized rendering and search filters

PROJECTS

AI Video Summarizer | *Python, FastAPI, FFmpeg, Whisper, Gemini*

[Github](#) | cbmeetings.vercel.app

- Deployed an AI-powered app that turns 2-hour NYC community board meeting videos into 5-minute briefs
- Built a FastAPI, FFmpeg pipeline with Whisper transcription, SQLite storage, and Gemini for analysis, reducing processing time by 2 hours through asynchronous processing
- Integrated Bright Data proxy routing, enabling 99% reliability to fetch YouTube videos in cloud environment

Prompt Engineering Tool | *Java Spring Boot, Tauri, Rust*

[Github](#) | promptly-eight-zeta.vercel.app

- Engineered a voice-enabled system using Java Spring Boot to transform vague ideas into precise AI prompts
- Developed a Next.js UI and Tauri desktop app enabling users to upload PDF/DOCX files and generate prompts
- Implemented a reactive WebFlux API with WebSocket streaming for real-time speech-to-text conversion

Multi-Model LLM Router | *Python, LangGraph, Tauri*

[Github](#)

- Designed a desktop overlay using Tauri and LangGraph for intelligent AI model routing across 5 models
- Implemented task classification system that analyzes user queries and automatically chooses optimal models
- Achieved 30% faster responses and 25% cost reduction with LangGraph decision graphs for LLM selection

SKILLS

Languages: Python, TypeScript, C++, Java, SQL, HTML, CSS

Frameworks & Libraries: Spring Boot, Node.js, Express.js, Next.js, React Native

AI/ML: LangGraph, PyTorch, scikit-learn, TensorFlow, NumPy, Pandas

DevOps & Infra: Git, AWS, Docker, Google Cloud, PostgreSQL