

## GENERATIVE AI / ML SOFTWARE ENGINEER

*Multidisciplinary self-starter and team-player, transforming ideas into products.*

### Senior Data Scientist

QuantumStreet AI

July 2024 –

- Exploring better ways to capture unstructured and structured data.
- Driving research and development of data processing methodologies.
- Driving R&D of ML model to improve predictive outcomes.

### SKILLS

#### Programming

### Generative AI Engineer, Volunteer

National Science Foundation

April 2024

- Collaborated on building Prototype Open Knowledge Network (Proto-OKN.net).
- Constructed Knowledge Graph from OpenAlex API with **generic code**.
- Developed **RAG chat-bot** summarizing KG Cypher results for user questions.

Python  
bash  
Java  
PHP  
Multi-Agent-Systems

### Data Scientist, Affiliate Faculty

George Mason University

2022 – 2024

- Built a platform extensible to new sources with **ad hoc faceted searches**.
- Analyzed **citation patterns** between U.S. Congress and National Academies.
- Utilized **data mining** to identify trends in congressional documents.
- Gave a gentle **introduction to ML and NLP** to two university students.

#### DB / Storage

MySQL  
Postgres  
MongoDB  
Elastic  
SOLR  
Faiss  
Redis  
Neo4j

### ML Engineer, Contractor

Inverudio

2005 – 2024

#### *Natural Language Processing + GenAI*

- Generative AI **discrepancy detection** in insurance invoices / employee data.
- Build video transcripts **summarization**, acquired by *Pictory.ai*. \* *indicates team work*
- Mapped **recurring ideas** from 500k multi-lingual books for **IR / LLM / RAG**.
- Net Neutrality **opinion mining** - distilled 300k comments to 300 arguments.
- Applied prompt engineering for **RAG** based resume builder for *CVJury.com*. \*
- Earnings calls **sentiment analysis**: Watson, Spacy, Vader & BERT **ensemble**.
- Legal inquiries **clustering** to streamline daily data pipeline workflow.

#### Web Dev.

LAMP  
DevOps  
Docker  
Swarm  
Flask  
D3.js  
API  
SaaS  
AWS  
Micro-services

#### *Structuring Unstructured Data*

- Developed a locale **language agnostic Wiktionary parser**.
- Cleaned data and **standardized drug names** from free texts for *Inspire*.
- **NER algorithm for ingredient parsing** with very high precision and recall.

#### *Recommender Systems*

- Built an advanced **recommendation system** for *Macrovo.com*. \*Recipe suggestions based on impact on user **health conditions**.
- Developed SEO insight based algorithm to identify **diverse perspectives**.

#### *Business Intelligence*

- **Forecasting model** for tourism industry based on weather, local events...
- **Automated labeling** of open-ended user responses for *EyeSee-Research*.

- Data mined Amazon reviews for **missing product features**.
- Mined unstructured data to identify **customer segmentation topics**.

*Medical*

- Used **medical ontologies** for **data augmentation** and retrieval.
- Linked pharmacological **drugs and medical data** sources for *PharmaDir* \*
- Built a **knowledge graph** of health effects of food ingredients.

*University clients (Harvard Business School, University of Chicago, U. of Michigan...)*

- **Multi-modal transportation** route optimization.
- **Fuzzy matching** for product and company mention disambiguation.
- **Record linkage** of land record database names to external databases. \*
- **Location prediction** from unstructured non-lethal violence reports. \*
- LSTM based network **intrusion detection system (IDS)**. \*
- GUI based **multi-agent infectious disease simulation**, i.e. COVID-19.

*Search Engine Optimization*

- **Increased profit 10x in 2 months with A/B testing**, a top 3 site in its niche.
- Applied on-page SEO to **double the traffic** in two weeks for *Links.com*.
- **Semi-automated** the SEO audit process, drastically increasing productivity.
- **ML powered SEO tools** for keyword suggestion and competition analysis.
- Advised on application of ML for SEO agency **workflow improvements**.
- Brought several sites from 0 to 100k visits per month with **on-page SEO**.

*Search Engine Development*

- Architected a **selective indexing** search engine inspired by Blekko.
- Introduced Semantic flow measure for **rapid niche resource identification**.
- Co-developed a ML-driven **topical crawling SaaS** for *SemanticJuice.com*. \*
- Implemented **content farms detection** with graph based measures.
- Developed a search engine with efficient **subset search** with replacements.

*Algorithm Design*

- **CPU based 1ms sentence alignment**, 30x faster than alternatives on GPU.
- Invented **data structures** for multidimensional maxima estimation.
- Developed **linguistic similarity** algorithms for *EZGlot.com*. \*
- **Benchmarked** & optimized system performance, including ports to Go, C++.

*Miscellaneous*

- Applied ML to **industrial process automation** in industrial PID controller.
- Docker Swarm based **REST API NLP micro-services** behind HAProxy.
- **Pair programming** with a professor during SaaS core development. \*
- Designed many DB schemas and wrote **optimized SQL & NOSQL queries**.
- **Faceted dictionary** with epoch, subject, and author gender stats.

Publications: [Google Scholar list](#).

**Research Associate**                      **Chicago Center for Creative Dev.**                      **2003 – 2006**

- **Interdisciplinary researcher**, programmer, coauthor of peer review papers.
- GUI desktop applications for **recurrence plots & time series analysis (TSA)**.
- Collaborated on RRI based **mental health classification**.

Publications: [ResearchGate list](#).

**Master of Science**                      **Univ. of Belgrade Faculty of Electrical Eng.**                      **1997 – 2002**

ML / NLP

NumPy  
Pandas  
sklearn  
Matplotlib  
Jupyter  
LightGBM  
RF  
Kmeans  
SVM  
PCA  
LDA  
ANN  
PyTorch  
SpaCy  
AllenNLP  
Gensim  
SBERT  
NLTK  
Splink  
WordNet  
HuggingFace  
OpenAI  
GPT  
LLM  
Langchain  
Regression  
TSF

IR

Scraping  
ETL  
NLP  
NEL  
NER  
ML  
AI  
DL  
LTR  
KG  
RAG  
LSH  
SQL