KIRILL GOLTSMAN

software developer / tech writer

BASIC INFORMATION

Malaga, Spain
goltsmank@gmail.com
Website: golkir.github.io
LinkedIn: linkedin.com/in/kirillgoltsman

OVERVIEW

I have over 7 years of software engineering experience, of which 3 years I've been working as a data scientist and an AI/ML engineer. My background includes hands-on experience in data pipeline management, ML model training, optimisation, and deployment, MLOps, model fine-tuning (LLMs, speech recognition), data visualization, and integrating third-party AI/ML tools. Additionally, in my previous roles as a software developer I've developed solid knowledge of cloud platforms (AWS, Azure), containerisation technologies, microservices architecture, and good understanding of DevOps practices and workflows.

Over the last couple of years I developed strong interest in such fields of AI/ML as language modelling, diffusion models, speech recognition, and computer vision. I am trying to stay tuned to new trends in AI/ML and new tools that facilitate adoption of AI technology in enterprises.

Also, I have over 8 years of experience as tech content creator in such fields as AI/ML, web development, and cloud-native applications (Kubernetes, Docker). I have developed technical documentation for tech companies and software projects. This experience allows me to effectively convey complex tech topics to my colleagues and wider audience.

WORK EXPERIENCE

ML Engineer / Data Scientist May 2022 - March 2024

Orichal Partners

- created data pipelines (ETL) to enable model training and fine-tuning.
- developed a robust pipeline to extract blockchain data from Graph nodes using GraphQL queries.
- performed statistical analysis on created datasets using various ML and statistical methods including classification, clustering, time series analysis, regression analysis, gradient boosting, etc.
- used LLMs (large language models) for downstream tasks such as sentiment analysis, summarization, and named entity recognition.
- fine-tuned language models with in-house data for the use in automatic question-answering applications and chatbots.

- integrated third-party AI/ML APIs, libraries and applications into in-house software and tools
- created visualizations to communicate data insights to stakeholders.
- developed dashboards and reports using tools like Tableau and Power BI.
- worked with cross-functional teams, including engineers, product managers, and business analysts to ensure seamless integration of developed statistical models

Senior Software Engineer (Blockchain, Smart contracts) July 2020 - May 2022

Orichal Partners

- Built core contracts for the DeFi risk tranching product including tranching, staking, oracle integration, voting escrow, etc.
- Developed and deployed subgraphs using the Graph protocol to index contract data and make it available to frontend applications.
- Managed contract deployments, operations, and upgrades on the Binance Smart Chain mainnet.
- Conducted peer review of junior developers' code
- Interviewed prospective candidates for the smart developer roles in the company.

GitCoin Project Developer March 2021 - June 2021

Freelance

Author of the GitCoin Grant Project "Deep NFTs."

This project seeks to expand the application of generative Machine Learning models to NFTs. Using ML we can transfer styles from existing NFTs and mint new NFTs with unique features and traits derived from the model's custom parameters and parent NFTs.

How it Works

• You have two NFTs like Hashmasks or CryptoPunks.

• Select one NFT as a content object and the other one as a style object or use two NFTs as style objects.

• Run pretrained ML model in your browser to transfer and/or combine styles between NFTs using custom settings such as stylization ratio.

• Mint a new NFT on the blockchain. Minted NFTs persist unique data about ML parameters and settings used to create them on blockchain. NFTs inherit rarity and unique properties from their parent NFTs.

Project Demo Website -- https://deepnft.vercel.app/

Senior Tech Writer (AI/ML, cloud-native technologies) November 2019 - July 2020

IOD

Key Topics

- Kubernetes and cloud-native technology
- AI/ML algorithms, ML libraries (Tensorflow, PyTorch)
- ML automation frameworks and ML pipelines for Kubernetes (Kubeflow, IBM's Fabric for Deep Learning (FfDL)
- Cloud ML services (AWS, Google Cloud)

Content Director April 2017 - November 2019

Qbox Inc.

Key Responsibilities

- managing content pipeline (blog creation, editing, scheduling, publishing, social media distribution).
- managing a team of content writers (assigning topics, technical and quality review of content, editing).

- writing documentation for the company's software products.
- producing content for Kubernetes training courses
- working closely with the marketing team to improve content creation strategies, SEO, trend spotting, social media marketing.
- delivering content to the video production team.
- writing e-books, blogs and tutorials about Kubernetes, container technologies, and Elasticsearch.

Tech Writer and Editor (AI/ML, data science) June 2015 - Feb. 2018

Data Science Foundation

Key Responsibilities

- making regular contributions to the Foundation's blog.
- covering recent trends in Machine Learning research and innovation
- producing educational content about Machine Learning, Deep Learning and data science
- covering AI/ML business use cases and applications
- editing tech articles by other contributors.

Senior Tech Writer May, 2013 - May, 2015

Boss Writers

Areas of specialization

- Web Programming Tutorials (JavaScript, React.js, Node.js, Ruby on Rails)
- Analysis of IT industry trends and news.
- Artificial Intelligence and Machine Learning tutorials, articles, and blogs.
- Data Science tutorials
- Cloud Computing blogs.

Freelance Javascript Developer March 2011 - Nov. 2014

Key tasks:

- Develop, maintain and support web applications using Javascript and related tools
- Analyze code for testing and debugging
- Create test transactions to identify, isolate, and resolve issues.

Areas of specialization

- jQuery Plugin Development
- MEAN (MongoDB, Express.js, Angular.js, Node.js) full-stack development.

PROJECTS

Fine-tuning Llama2 Language Model Mar 2024 - Mar 2024

Fine-tuned Llama2 language model on a 1000-sample subset of the Dolly 15k instruction dataset using Supervised Fine-Tuning (SFT) with QLoRA 4-bit precision. Repo.

Blockchain Data Extraction and Statistical Analysis May 2023 - Mar 2024

The objective of the project was to extract and analyze blockchain data from Graph nodes to uncover meaningful insights and trends. The data included blockchain transactions, dynamics of APR (annual percentage rate), investment amounts, etc.

During the project the following activities were performed:

- Developed a robust pipeline to extract blockchain data from Graph nodes using GraphQL queries.
- performed data wrangling and feature engineering
- applied various statistical methods and models to analyze the blockchain data and identify

key trends, patterns, and anomalies within the data. Methods applied included classification, regression, time series analysis, random forrests, etc.

The findings were conveyed to the product managers to improve the quality of existing offerings and develop new products based on the insights obtained from the data.

Statistical analysis of chess games played in "Titled Tuesday" events on chess.com Apr 2024 - Apr 2024

As part of this project, I created a dataset of chess games played in "Titled Tuesday" events on chess.com between July 2022 and December 2023 and executed various statistical analyses on it. The goal was to find interesting patterns and regularities in online chess tournaments.

Statistical tools applied were:

- Descriptive statistics (mean, number of wins, losses, rank distribution).
- Game accuracy distribution analysis between tournaments rounds to find differences in accuracy between initial and final rounds
- K-means clustering and k-nearest neighbors analysis.
- Linear regression to predict game results based on player accuracy.
- Statistical tests (Kolmogorov-Smirnov test, Shapiro-Wilk normality test).
- Kernel Density Estimation (KDE) and visualization
- Maximum Likelihood Estimation of mean and standard deviation assuming chess data normality
- Correlation analysis of score and accuracy (Spearman's rank corellation coefficient)

See more: Repo.

GitCoin Grant Project: Deep NFTs: Machine Learning for NFT space Mar 2021 - Jun 2021

This project sought to expand the application of generative Machine Learning models in blockchain (NFTs). Using ML we can transfer styles from existing NFTs and mint new NFTs with unique features and traits derived from the model's custom parameters and parent NFTs.

How it Works

• You have two NFTs like Hashmasks or CryptoPunks.

• Select one NFT as a content object and the other one as a style object or use two NFTs as style objects.

• Run pretrained ML model in your browser to transfer and/or combine styles between NFTs using custom settings such as stylization ratio.

• Mint a new NFT on the blockchain. Minted NFTs persist unique data about ML parameters and settings used to create them on blockchain. NFTs inherit rarity and unique properties from their parent NFTs.

Project Demo Website -- https://deepnft.vercel.app/

EDUCATION

MS in Finance 2008 - 2011

Donetsk National University of Economics and Trade, Ukraine

MS in International Relations September 2002 - June 2007

Donetsk National University, Donetsk, Ukraine

Front-end Development Course February 2016 - June 2016

Golt
Course Description:
The course provided basic training in HTML5, CSS, Javascript and OOP programming patterns.
Knowledge and Skills Acquired:
• Photoshop
• HTML5
• CSS
• Javascript Basics
• Javascript Frameworks and API (Google API, Angular)
• Introduction to OOP (Object Oriented Programming)

Machine Learning Course (Stanford University) March 2016 - September 2016
Coursera
Course Description:
The course provided basic introduction to Machine Learning algorithms and models
Knowledge and Skills Acquired:
• Machine Learning with Linear Regression
• Machine Learning with Logistic Regression (Classification)
• Neural Nets and Multilayer Perceptrons
• Backpropagation algorithms
• Unsupervised learning (Clustering algorithms)
• Introduction to Image Recognition AI

SKILLS

AI/ML

- **Programming Languages:** Python (NumPy, Pandas, scikit-learn, TensorFlow, PyTorch), Javascript, C++ (basic), Solidity.
- **Classical Machine Learning**: Linear models, dimensionality reduction methods (PCA, LDA), clustering algorithms, SVM (Support Vector Machines), gradient boosting, random forests, nearest neighbours algorithms.
- **Deep Learning (general):** Multilayer perceptrons, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAE), encoder-decoder architecture.
- **Model training**: hyper-parameter tuning, cross-validation, metrics and scoring, building full ML pipeline including training, hyperparameter optimization, fine-tuning, evaluation, and deployment.
- **Natural Language Processing (NLP):** transformer architecture, fine-tuning LLMs for downstream tasks such as Named Entity Recognition (NER), Sentiment Analysis, Topic Classification, Text Summarisation etc. Experienced in fine-tuning LLMs with customised enterprise-specific datasets for proprietary use. Experienced in prompt engineering and Chain of Thought (CoT) prompting.
- **Computer Vision:** Image Classification, Object Detection, Image Segmentation.
- **Diffusion Models**: DDPMs (Denoising Diffusion Probabilistic Models), conditioning diffusion models on text (text-to-image synthesis). Knowledge of Dalle(1,2) and Imagen.
- **Reinforcement Learning:** Q-Learning, Policy Gradient Methods.
- **Data-related skills:** data mining and data retrieval, data preprocessing (cleansing, dealing with missing values), data augmentation and modification, building and maintaining data

pipelines for online learning.
- **Math**: calculus, linear algebra, probability theory and statistics

Tools

- **AI/ML Frameworks:** TensorFlow, PyTorch, Keras, scikit-learn, NumPy, SciPy, Pandas.
- **Data Visualisation:** Matplotlib, Seaborn
- **Databases:** SQL, MongoDB, PostgreSQL, Elasticsearch
- **Cloud Platforms:** deploying and maintaining products on AWS, Google Cloud Platform (GCP), Azure, Heroku.
- **Version Control:** Git, GitHub.
- **IDEs:** Jupyter Notebook, PyCharm
- **Containerization:** Docker, Kubernetes
- **MLOps:** AWS SageMaker, KubeFlow, MLFlow
- **AI software ecosystem:** Hugging Face libraries(transformers, accelerate, PEFT, diffusers, text-generation-inference, TRL, optimum), WandB, LangChain
- **Cloud-Native Technologies and Tools**: Linux containers (Docker), Kubernetes, Prometheus, Fluentd, Traeffik, etc.

Blockchain technology

- Basic blockchain architecture (distributed ledgers, block finalization, state transition functions, cryptographic security, network fees and gas)
- consensus mechanisms (PoW, PoS, PoA, DPoS)
- Solidity programming language
- Architecture of Ethereum Virtual Machine (EVM)
- Smart contract development frameworks: Truffle and Hardhat
- Web3.js, ethers.js, OpenZeppelin contract libraries
- Basic understanding of the blockchain-related cryptography including hash functions (Keccak), public-key cryptography (ECDSA), etc.

Scientific Writing

- experience in writing papers for peer-reviewed journals
- knowledge of common citation styles (APA, MLA, Chicago, Oxford, Harvard);
- knowledge of scientific inquiry methods: qualitative (interviewing), quantitative (statistical analysis, sampling), case-study and content analysis.

Content management

- experienced in managing a team of content creators
- doing technical and quality reviews
- managing content marketing pipeline using JIRA and Trello
- interacting with engineering teams to produce relevant technical content including documentation and feature updates

PORTFOLIO

ML/Data Science Projects

- Fine-tuning Llama2 language model on a 1000-sample subset of the Dolly 15k instruction dataset using Supervised Fine-Tuning (SFT) with QLoRA 4-bit precision. Repo.
- statistical analysis of chess games played in "Titled Tuesday" events on chess.com between July 2022 and December 2023. Repo.
- Creating a dataset that includes chess games played in the Chess.com "Titled Tuesday" events in July 2022 - December 2023. Dataset.
- CNN (Convolutional Neural Network) implementation from scratch using NumPy. Repo.

Blockchain and Smart Contract Projects.

- a set of Solidity contracts to create NFT-backed fungible tokens. The goal is to make NFT-based fungible tokens that have the liquidity advantage of ERC20 while fully backed and convertible to NFT tokens. Repo.
- GitCoin Grant Project: Deep NFTs: Machine Learning for NFT space. Repo.

Blogs

My tech blogs on Medium -- https://medium.com/@kirill_86245
IOD articles -- https://iamondemand.com/blog/author/kirill-goltsman/

ML Models and Datasets

Hugging Face Profile -- https://huggingface.co/kirillgoltsman