Koyilbek Valiev

E-mail: valievkoyiljon112@gmail.com **Phone:** +(998)-95-495-5735
Github: github.com/koyilbek **Website:** valiev-koyiljon.github.io/Web

Summary

AI Engineer with 3+ years of experience building production-grade vision-language systems, autonomous AI agents, and multimodal ML pipelines. Proven expertise in video anomaly detection, time-series forecasting, cross-modal retrieval, and multi-agent orchestration. Published research paper in reinforcement learning for smart grid energy management. Passionate about advancing VLM/LLM-based intelligent systems and deploying them at scale.

Experiences

AI / ML Engineer *TelecomSoft, Tashkent, Uzbekistan*
Jan 2026 – Present

- Built and deployed a production-grade autonomous AI agent harness from scratch — architecting a 3-layer system with 20+ intent routing rules, 10+ procedural skill workflows, and multiple tool classes spanning billing, services, payments, and knowledge retrieval — delivering high-accuracy, human-style customer support conversations with multi-skill orchestration and external tool use at scale.
- Integrated ASR and TTS pipelines end-to-end, enabling natural voice-based agent interactions with support for diverse accents and dialects.
- Designed NLU modules capable of understanding varied chat styles, accents, and multilingual customer inputs.
- Deployed open-source models to production, built API-connected backend services, and developed web-based UI for real-time agent monitoring and interaction.
- Designed benchmarks and evaluation pipelines to assess AI agent workflows, measuring accuracy, latency, and end-to-end performance across scenarios.
- Built observability and tracing infrastructure to track agent decisions, tool calls, and conversation quality throughout the entire pipeline.

Multimodal AI Engineer (Co-op) *Pia Space, Seoul, South Korea*
March 2025 - mid June 2025

- **AI Video Anomaly Detection :** Collected and annotated diverse anomaly video datasets (e.g., violence, falls) using proprietary annotation tool. Optimized lightweight VLMs (<1B params) for video anomaly classification with dynamic sampling and prompt engineering in few-/zero-shot settings. Conducted both qualitative and quantitative analysis to evaluate and refine model performance.
- **Multimodal Retrieval for Video Anomaly Detection:** Co-developed MACS 3.0, a prompt-driven multimodal system for CCTV anomaly detection, achieving an overall model performance improvement of approximately 24% over previous versions. Built a benchmark dataset with manually captioned videos for video-text and video-video retrieval. Implemented and evaluated cross-modal retrieval models (video-to-text, text-to-video, video-to-video) to enhance scalable multimodal video understanding. Adapted a research method on multimodal explanation maps for vision-language models, generating visualizations that highlight key text prompt words and corresponding video regions to improve model interpretability and alignment.
- CTO-issued recommendation letter available: View Letter

AI Team Lead | AI Engineer (Co-op) *Recs Innovation Ltd, Naju, South Korea*
Feb 2024 – March 2025

- **Bidding Amount Prediction:** Developed a dynamic ML pipeline that analyzes incoming data distributions and selects the best-performing algorithm. Implemented data cleaning, anomaly detection, feature engineering, and clustering, achieving a 0.0017% prediction error and a 15× increase in bid win rate.
- **Solar Power Forecasting:** Built a 48-hour ahead forecasting system based on a hybrid neural network combining CNN, LSTM, GRU, and Transformer layers to capture spatial, temporal, and long-range dependencies. Leveraged forecasted weather data to reach an average 6% prediction error. The solution was integrated and deployed on the Sun-Q EMS platform for real-time energy optimization.
- **Photovoltaic Sensor Anomaly Detection:** Led development of an unsupervised AI solution using sensor data from solar plants. Employed models including LSTM Autoencoder, LSTM-VAE, TranAD, and VAE to detect anomalies, achieving over 70% F1-score on real-world data. The solution is deployed in the Sun-Q EMS platform for continuous monitoring and fault detection.
- **Documentation & Team Leadership:** Authored comprehensive technical documentation on system architecture and data workflows. Managed a small AI team using Notion and ClickUp for project planning, task delegation, and progress tracking, enhancing team efficiency and delivery.

AI Engineer (Capstone Project, Industry Partner: NetVision) *Woosong University, Daejon, South Korea*
Sep 2023 - Dec 2023

- Coordinated team of 4 engineers to develop a real-time trash-bag optimization system for NetVision.
- Developed YOLOv8 to detect, classify, segment, track & size-estimate bags in video.
- Participated with this project in Woosong University's 2023 Capstone Competition with permission from NetVision, winning 1st place award.

Computer Vision Intern *Sequus PTY LTD, Australia*
(Remote)
Jun 2023 - Aug 2023.

- Automated conversion of bounding-box outputs into YOLO, Pascal VOC, and custom formats.
- Annotated hundreds of architectural drawings (LabelImg) to improve model accuracy.

Publications

[1] Valiev, K., Ikromov, S., & Kim, Y. (2026). ***A Study of Reinforcement Learning Framework for Energy Management in Smart Grids: Integrating Market Trading, Load Forecasting, and Vertical Agents.*** Journal of Korean Institute of Communications and Information Sciences (J-KICS), Vol. 51, No. 2, pp. 324–347, Feb. 2026. DOI: 10.7840/kics.2026.51.2.324 | **View Paper**

Patents

[1] Valiev, K. **ESS**를 활용한 에너지 비용 및 판매 가격 계산 방법 및 장치 *(Method and Device for Calculating Energy Cost and Sales Price Using ESS)*
Recs Innovation Ltd, South Korea | Issued by KIPO, 2024 | **Patent Certificate (PDF)**

Projects: Refer portfolio koyilbek for other projects

- **Uzbek-English Pretrained Language Model (140M Parameters)**: Collected a diverse Uzbek-English dataset and trained a Byte Pair Encoding (BPE) tokenizer from scratch, resulting in a vocabulary size of ~62,016 tokens. Built a decoder-only modified Transformer architecture with 140.7M parameters and a 1024-token context window. Developed the full AI training pipeline and pre-trained the model using open-source multilingual datasets on 2× NVIDIA A100 GPUs for 16.5 hours. Reached ~3.5% cross-entropy loss during pretraining; model currently supports next-token generation. Planning to fine-tune the model for instruction-following and downstream task alignment. Demo includes both short and long prompt generation samples: **View Project**

- **LLMs from Scratch – GPT-2 Implementation & Fine-Tuning:** Completed the full implementation of a GPT-2 model from scratch based on "*Build LLMs from Scratch" book* by Sebastian Raschka. Developed essential components including text preprocessing, positional encodings, multi-head self-attention, and the GPT-2 architecture and pretraining the model using causal language modeling (CLM) on unlabeled data. Fine-tuned the pretrained model for spam classification, and further applied instruction tuning to enable natural language prompt-following. Gained hands-on experience in building, training, and adapting large language models for both generative and classification tasks. More details in: **View Project**

Education

Woosong University
Bachelor's in AI & Big Data | Sep *2021 – Aug 2025*

Key courses: Data Visualization, Artificial Intelligence, Machine Learning Theory & Lab, Computer Vision, Statistics, Linear Algebra, Discrete Mathematics, Capstone Project

Skills

Python, PyTorch, Hugging Face Transformers, LangChain, LlamaIndex, FastAPI, OpenCV, CuPy, TensorRT Ollama, Unsloth, MLflow, Weights & Biases, Optuna, Git, NumPy, Pandas, Scikit-learn, Matplotlib, Docker, Gradio, Sentencepiece, KT Cloud, Public Speaking, Team Player

Awards

- **2nd Place,** HumbleBeeAI Hackathon: Data Science Challenge (Aug 2025)
- **1st Place,** Woosong University Capstone Competition among 200 teams (*Dec 2023)*
- **President's Award,** Woosong University (*Dec 2023)*
- **100% Merit Scholarship,** Woosong University (*Sep 2023 – Dec 2023)*
- **1st Place,** IoT Learning Concert (*Sep 2023 – Dec 2023)*
- **2nd Place,** Machine Learning Lab Learning Concert (*Sep 2023 – Dec 2023)*