SANZHAR UMBET

+49-157-372-34-656 | sanzhar.umbet@alumni.nu.edu.kz

WORK EXPERIENCE

Data Scientist - Bereke Bank June 2023 – Current
Major bank in Kazakhstan *Almaty, KZ*
- Engineered features for **AutoML**, leading to a **15% increase** in service personalization.
- Spearheaded the development of an anonymization library using NER models, improving **F1 score by 20%**.
- **Benchmarked Whisper and M4T models** across diverse speech tasks.
- Led a project using clustering algorithms for **automatic labeling** of text data.
- Optimized LLM inference with **contextual caching**, cutting API costs by **20%**.
- Led the development of an **Insight Service** to provide real-time analytics on customer interactions.

Core Contributor - Horde Research Dev. 2024 – Current
Leading Kazakh Open-Source AI Research Group *Remote*
- Developed a **synthetic dataset** for an online LLM arena, improving model benchmarking capabilities.
- Created **the largest Kazakh SFT dataset**, expanding the accessibility of LLMs in the Kazakh language.
- **Benchmarked various LLMs** to evaluate performance and guide model improvements.
- Trained and fine-tuned **dialogue models** to enhance contextual understanding.

Research Assistant - Nazarbayev University Aug. 2024 – Nov. 2024
National Laboratory of Astana *Astana, KZ*
- Developed a language identification model using Wav2Vec2 XLSR, **achieving 92.9% F1** score.
- Built an **inverse normalization model for Kazakh**, improving punctuation and capitalization by 25%.
- Optimized training pipelines with data augmentation, improving overall model robustness.

Research Assistant - Brock University Dec. 2023 – Feb. 2024
Brock NLP Lab – Prominent NLP Research Group *St. Catharines, CA*
- Led research on open-source models, analyzing the correlation between internal and external confidence levels.
- **Investigated model hallucination** by comparing how models express confidence versus internal logits.
- Tested OpenAI and other models, identifying discrepancies in confidence that affect reliability.
- Contributed to NLP studies on model behaviour, focusing on hallucinations. ACL main conference 2024.

Research Assistant - Nazarbayev University May 2022 – May 2023
Astrophysics Research Group, Physics Department *Astana, KZ*
- Applied Stokes parameters using astropy to visualize magnetic fields, enhancing filament detection **by 80%**.
- Developed a **identification algorithm in 3D space**, improving hub filament detection accuracy by 15%.
- Utilized the Spectral-Cube package to generate **3D visualizations**, reducing false positives cases by nearly 80%.
- Introduced Line Integral Convolution for improved filament visibility in magnetic field studies.

Research Assistant - Nazarbayev University Nov. 2022 – May. 2023
Autonomous Systems Research, Robotics Department *Astana, KZ*
- Led the enhancement of reinforcement learning implementations on PMSM motors, improving control and performance with added functionalities.
- Analyzed and interpreted RL experiments, driving developments in motor control and optimization.
- Translated Matlab/Simulink to Python, significantly boosting workflow efficiency and clarity.

EDUCATION

Nazarbayev University Astana, KZ
Courses: Complex Analysis, Matching Theory, Big Data Analytics

TECHNICAL SKILLS

Languages: Python, C++, R, Stata
Tools: Git, VS Code, PyCharm, Eclipse
Libraries: PyTorch, NumPy, Matplotlib, sklearn, pandas