Stanislau Beliasau

Green Card holder New York, 11229

🕡 bestasoff@gmail.com 💭 bestasoff in bestasoff 🕢 bestasoff

bestasoff.github.io +1-347-777-0545

Machine Learning Engineer with about 2 years of hands-on experience in machine learning projects (mainly in Computer Vision and NLP). Have experience with training and optimization of Large Language models. Have a strong background in CNNs, ViTs, GANs, and Diffusion models (like Stable diffusion, its applications, and downstream tasks).

Skills

Programming languages: Python, C++.

Frameworks: PyTorch, torch.distributed, torchlightning, huggingface, accelerate, FastAPI, Flask, deepspeed, scikit-learn, OpenCV, numpy, pandas, catboost, xgboost, coreml, bitsandbytes.

Machine Learning: Computer Vision, NLP, classical ML, distributed training, data-parallel training, model-parallel training, tensor parallelism, model pruning, model deployment, model distillation, model compression, memory footprint reduction. Tools: docker, docker-compose, git, Kubernetes, Terraform, Airflow, Prometheus, Postgres, Elasticsearch, Kibana, Logstash, gRPC, TensorRT, ONNX, Hadoop, Spark, Flink, Kafka.

Work Experience

🜀 📟 Gradient & Persona: AI Photo & Video mobile editors, New York, Remote

Computer Vision Engineer

- Make massive research on image generation, especially with Stable Diffusion model. Played a crucial role in developing the brand-new encoding method into its latent space.
- Run experiments with different Stable Diffusion downstream tasks like custom fine-tuning, introducing new modules, curating task-specific datasets, and implementing papers—accelerated image generation by 30%.
- Generate and curate custom datasets. Use CLIP, BLIP, StyleGAN, pix2pix models for processings. Resulted in obtaining datasets that helped to train new models.
- Train brand-new re-aging img2img filters both server and real-time.
- Train new versions of image-warping body-tune models both for images and videos and lightweight body segmentation models. Resulted in better postprocessing on inference.
- Making GANs work in real-time on mobile devices with HD quality, 60 fps and taking less than 2 MB.
- Deploy models both on IOS and server using torch.jit and coreml.

Yandex School of Data analysis, Remote

ML course tutor assistant

- Implement LLM (Large Language models) (GPT-2XL, OPT-6.7B) distributed training and inference pipelines: modelparallel, data-parallel, pipeline parallelism, memory offloading.
- Give lectures on machine learning including NLP and CV.
- Conduct seminars and check homework.

iTechArt, Minsk, Belarus

Machine Learning Engineer

- Designed and implemented an image classification service using the gRPC endpoint client/server architecture and the FastAPI framework.
- Utilized Uvicorn and Prometheus in conjunction with Docker and Supervisord to create a robust and scalable solution.
- Developed and implemented custom model architectures using C++, resulting in up to a 45% reduction in model latency.
- Generated synthetic datasets to supplement real data, leading to an increase in model accuracy of up to 10%.
- Successfully distilled the CNN model into a model that was 3 times smaller while maintaining nearly identical evaluation metrics.
- Created and curated custom datasets from unstructured client data using Pandas and SQL.
- Trained numerous time-series models for demand forecasting, reducing forecast MAPE by 20%.
- Constructed production pipelines with AirFlow to convert raw data into a feature vector, feed it into the model, and forecast the product demand.

9 Yandex, Minsk, Belarus

Software Engineer Intern

- Developed rule-based and NLP-based solutions for affiliations parsing.
- Developed a data annotation service using Flask framework. Wrapped it into Docker and deployed it to the server.

Education

Yandex School of Data analysis, Minsk, Belarus

Master's degree level Machine Learning developer academic program

• Relevant courses: Python, C++, Golang, Algorithms and Data Structures, Machine Learning, Computer Vision, NLP, Probability and Statistics, Reinforcement Learning, Efficient Deep Learning systems, Recommendation systems.

Feb. 2022 – Present.

Aug. 2022 - May. 2023

Sep. 2020 – June. 2022

Nov. 2021 - Aug. 2022

May. 2021 – Nov. 2021

Belarusian State University, Minsk, Belarus
Bachelor of Computer Science
Awarded a full scholarship and stipend by the government per entrance exam results.

Activities

Educational-Scientific Conference of Students on Recent Methods of ML and Data Analysis: Minsk, Belarus

"BERT-based biomedical Named Entity Recognition and Named Entity Normalization".

Annual Belarusian State University scientific conference: Minsk, Belarus

"BERT-based biomedical Named Entity Recognition and Named Entity Normalization".