Sebastjan Cizel

in/sebastjan-cizel sebastjancizel sebastjancizel@gmail.com +44 7563 740128

Head of Engineering leading the model training, evaluation, and MLOps teams. Driving training and deployment of AI video compression models on edge devices. Experience with AI research and optimization for efficient inference in resource-constrained environments. Coauthor on 2 granted and 6 pending patents.

Work experience

Head of Engineering, Deep Render (October 2023 – Present)

- Technical lead and manager of 15 engineers responsible for productization of research models and internal hardware/software infrastructure.
- Built unified model inference library for major AI accelerators (Nvidia, Apple, Qualcomm, Intel) and CI/CD system reducing model port times from weeks to days.
- Spearheading the world's first integration of a neural codec into FFmpeg and VLC achieving realtime video encode and decode on consumer devices (demo).

Senior Research Scientist, Deep Render (February 2023 – October 2023)

- Led model quantization and pruning research achieving 2x improvements in memory footprint and runtime while retaining compression performance.
- Delivered client-facing applications and built model evaluation tools using Python and web stack (Vue, node).

Research Scientist, Deep Render (September 2021 – February 2023)

- Made foundational research contributions to video compression models with >10% efficiency gains over state-of-the-art.
- Lead maintainer of core model codebase (PyTorch, C++, CUDA) with 60k+ lines of contributions, focusing on generative CV models and optical flow.

Patents

- **Method and Data Processing System for Lossy Image or Video Encoding**
  US Patent 12,113,985. Issued October 8, 2024.
  Motion translations with flow-based processes for AI video compression.
- **Method and Data Processing System for Lossy Image or Video Encoding**
  US Patent 11,936,866. Issued March 19, 2024.
  Motion transformation handling in AI compression processes.
- **6 additional patent applications pending review** (Google Patents)
  Covering advanced neural compression techniques and optimization methods.

Education

2017 – 2021	**DPhil in Mathematics**, University of Oxford. Thesis title: *Supersymmetric Geometries in String Theory*
2016 – 2017	**MSc in Pure Mathematics** (with Distinction), Imperial College London Thesis title: *Glueing Construction of Calabi-Yau Metrics on Kummer Surfaces*
2013 – 2016	**BSc in Pure Mathematics**, Faculty of Mathematics and Physics, University of Ljubljana

Skills

Deep Learning	Highly proficient in applied generative AI research both in computer vision and applications of LLMs for internal business infrastructure projects.
Programming	**Python** (**advanced**) – 10+ years in software development, data analysis, and AI research. Expert in **PyTorch** with deep knowledge of model compilation stack (**FX, Inductor, Dynamo, Triton**). **C++** (**working knowledge**) – contributions to model inference libraries and entropy coders.
Other	AWS, Unix, git, Github actions, Weights and Biases, Docker, tmux, Vue, GraphQL, Kubernetes