Alexander Tsvyashchenko

Staff Research Engineer, Google DeepMind

Personal Data

Location:	Greater Zürich Area, Switzerland
Status:	Swiss Citizenship
Phone:	+41 76 738 71 30
E-mail, Matrix:	ndl@endl.ch
LinkedIn:	https://www.linkedin.com/in/ndlmaker
GitHub:	https://github.com/ndl
Website:	https://endl.ch

Summary

Research engineer with 25 years in R&D, the last 10 in machine learning at Google and the last 6 in ML research at Google Brain and Google DeepMind: sparsity, modularity and memory for frontier models, within internal "frontier challenges" projects on their core research questions. Co-author of the PaLM and T5X papers; inventor on a patent for extending multi-task networks to new modalities.

Before that: applied ML end to end for Google products and for certified autonomous flight, ad-fraud detection at Google scale, video and vision algorithms, and computational geometry for CAD and 3D printing.

Led a research team of six; tech lead for cross-industry botnet work; host and mentor of interns; teacher of ML inside and outside Google. Open-source contributor throughout.

Research Interests

Modularity and compositionality of neural networks; sparsity; memory for LLMs; continual learning; interpretability; large-scale training and inference systems.

Publications, Patents & Talks

Publications

Scaling Up Models and Data with t5x and seqio — 2023
Journal of Machine Learning Research 24(377):1–8

PaLM: Scaling Language Modeling with Pathways — 2023
Journal of Machine Learning Research 24(240):1–113

Patents

Extending multi-task neural network systems to new modalities — 2026
EP4713833A1, Google LLC; with Andrea Gesmundo; as Oleksandr Tsvyashchenko

A second patent application, on neural networks, is filed and pending.

Talks

"ML applications and typical errors" — 2023
AI Basics, an online course by Google Ukraine with the Ministry of Digital Transformation of Ukraine

"Applied Machine Learning" — 2018
Hack'n'Lead hackathon, Zürich

"Ad Fraud Botnets 101" — 2014
BotConf, Nancy

Professional Experience

ML research — 2023 – present

Staff Research Engineer — Apr 2023 – present
Google DeepMind, Zürich, Switzerland

- Research on sparsity, modularity and memory, within internal "frontier challenges" projects that tackle the core research questions for frontier AI models:
  - Lead of the inference sub-stream of the modular, composable and efficient LLM memory project.
  - Co-lead of the modularity in the music domain track.

  - Sparsity in text diffusion.
- Every project is a collaboration with other teams across Google DeepMind.
- Domains: LLM research: modularity, compositionality, sparsity, memory, continual learning, interpretability.
- Skills: Gemini and Gemma training, evaluation and serving stacks, XManager, TPUs, agentic R&D tooling (Antigravity), JAX, Grain, Python, C++.

Large-scale ML systems — 2020 – 2023

Staff Software Engineer — Nov 2020 – Apr 2023
Google / Brain, Zürich, Switzerland

- Research and development of large-scale ML systems:
  - ML Pathways, T5X and PaLM; co-author of the T5X and PaLM papers (see "Publications").
  - Multitask and transfer learning for text, vision and multimodal tasks, including applied research in YouTube; the patent on extending multi-task networks to new modalities comes from this work.
  - Sparsity and modularity research, continued at Google DeepMind.
- Host and mentor of interns through the Google internship program.
- Skills: JAX, Flax, T5X, XLA, TPUs, TensorFlow, Python, C++.

Applied ML, end to end — 2016 – 2020

Machine Learning Researcher / Autonomous Flight Specialist — Jul 2019 – Oct 2020
Daedalean AI, Zürich, Switzerland

- The research on the visual traffic detection system for uncooperative airborne obstacles: surveyed ML object detection and tracking, then led the development of several approaches for adding temporal information to the detection models.
- Co-designed, with the Simulation and Data teams, the pipeline that models and generates aircraft collision encounters at volume, the camera imaging simulation that narrows the gap to real data, and the traceable data flows that certification requires.
- Designed and implemented the training-to-production model conversion and inference stack for flight-compatible hardware, with proofs of concept for ML kernel verification under DO-178 and for FPGA inference on the Versatile Tensor Accelerator.
- Set up distributed training on Determined AI and optimised the training pipeline.
- Domains: Detect and Avoid; ML object detection and tracking; ML inference, compilers and optimisation; certification and formal verification; embedded devices, GPUs and FPGAs.
- Skills: TensorFlow / Keras, Determined AI, Unigine, GCP, OpenCV; TVM, ONNX / ONNX Runtime, LLVM / Clang, Z3, Nix, Bazel, Yocto, Xilinx Vivado.

Machine Learning Consultant — Oct 2019 – Oct 2020
FAIRTIQ, Zürich, Switzerland
Advised the research department on sequence models and time series, feature engineering and imbalanced datasets.

Staff Software Engineer — May 2016 – Jun 2019
Google / Google AI, Zürich, Switzerland

- Owned the full lifecycle of ML projects for ARCore (Tango), YouTube, Abuse and Ads: scoping with the product team, data analysis and collection, model research, evaluation and productionisation.
- Taught the product teams ML along the way, and ML classes inside and outside Google.
- Skills: TensorFlow, TFX; multiple Google technologies, languages and products.

Large-scale data, abuse and fraud — 2011 – 2016

Senior Software Engineer — Aug 2011 – May 2016
Google / Ad Traffic Quality, Zürich, Switzerland
Traffic analysis, signals, metrics and filters against abusive and fraudulent ad traffic; tech lead of the ad-fraud botnets effort, including information sharing and collaboration with the wider industry (see "Talks").

Video and computer vision — 2007 – 2011

Chief Scientist — Feb 2010 – Aug 2011
Deebmedia, Amsterdam, Netherlands (remote)
Owned the whole video-processing and algorithmic stack; R&D of robust online algorithms for background reconstruction in complex scenes and for highly accurate object tracking.

Algorithm Consultant — Sep 2007 – Nov 2007
Atoms Optical Measuring, Locarno, Switzerland (remote)
Researched and prototyped a highly accurate image-segmentation algorithm, one of the key components behind micron-level measurement accuracy.

Computational geometry, CAD and simulation — 2000 – 2011

Senior Software Engineer / Project Manager — Nov 2005 – Jul 2011
Automated Industrial Machinery, Inc, Chicago, IL (remote)
End-to-end ownership of four R&D projects for the wire-bending industry: a photogrammetric measurement system, the BenderCad CAD system, real-time bending-machine simulation and CAD wire import; managed three sub-contractors.

Senior Researcher / Team Leader — Aug 2000 – Jul 2007
Materialise, Kyiv, Ukraine
Researched and implemented the computational-geometry and rapid-prototyping algorithms at the core of Materialise DigitalCAD and of the Magics, Mimics, Simplant and 3-matic products; led a research team of six and introduced automatic testing, autobuild and code-quality processes.

Open source — 1998 – present

Author and contributor — 1998 – present
Multiple open-source projects
Stand-alone projects and patches to existing ones, from a Linux kernel Wi-Fi driver to XMPP archiving in Erlang; the recent ones are at github.com/ndl, the older ones under projects.

Skills

Current

ML:	Gemini and Gemma training, evaluation and serving stacks; JAX, Flax, Grain, XLA, SPMD / MPMD, TensorFlow.
Platforms:	TPUs, GPUs; XManager, GCP; Apache Beam (Google Flume).
Languages:	Python, C++, C.
Tooling:	Agentic R&D tooling (Antigravity, Claude Code); Nix, Docker, Bazel, Git; LaTeX. Own server (mail, web, IM, storage, calendaring) administered since 2007.
Leadership:	Leading research workstreams and teams, cross-team collaboration, mentoring interns, teaching ML classes, navigating organisational complexity, stakeholder communication, full lifecycle of ML projects.

Earlier

Used in the past; can be picked back up in days if the need arises.

Domains:	Autonomous flight, certification and formal verification, computer vision, embedded devices and FPGAs, image and video processing, advertising and traffic quality, CAD / CAM, computational geometry, rapid prototyping, visual simulation, instant messaging, networking.
Languages:	Zig, C#, OpenCL, Kotlin, Erlang, VB.NET, Ruby, Perl, SQL, Java, Haskell, Lua, Tcl, Pascal, x86 assembler, Lisp, Prolog.
Inference & compilers:	TVM, ONNX / ONNX Runtime, LLVM / Clang, Z3.
Libraries & frameworks:	OpenCV, Pandas, scikit-learn, Matplotlib, Dash / Plotly; Unigine, Yocto, Xilinx Vivado; .NET (WPF, LINQ), OpenGL, OpenMP, Qt, wxWidgets, Ruby on Rails; Boost, CGAL, ITK, VTK, OpenCASCADE, OpenMesh; RDBMS (MySQL, PostgreSQL, SQLite); Linux (desktop and embedded), Android (AOSP), Windows, Mac OS X, FreeBSD.

Education

1997 – 2003 — Master of Science in Computer Science with distinction, Cybernetics Faculty, Department for Theoretical Cybernetics, *Taras Shevchenko National University of Kyiv*, Kyiv, Ukraine

Languages

Russian, Ukrainian:	Mother tongues
English:	Fluent
German:	Intermediate (TELC B1 Certificate, 2017)
Swiss German:	Beginner