Georgy Savva

georgy.savva@nyu.edu | georgysavva.github.io | github.com/georgysavva

Research Interests

World Models, Representation Learning, Digital Agents, Multimodal Learning

Education

New York University — New York, United States
Master of Science in Computer Science, Courant Institute of Mathematical Sciences. GPA: 3.8 — *2024 – Expected 2025*
Relevant courses: Computer Vision

Russian University of Cooperation — Moscow, Russia
Bachelor of Science in Computer Science — *2019 – 2021*

Research Experience

Research Assistant — Jan 2025 – Present
NYU Courant Institute of Mathematical Sciences, Advised by Prof. Saining Xie — *New York, United States*

- Developed the first multi-agent Minecraft World Model
- Investigated quality degradation in autoregressive video world models by applying Diffusion Forcing and Self Forcing

Research Intern — Jun 2024 – Nov 2024
NYU Courant Institute of Mathematical Sciences, Advised by Prof. Lerrel Pinto — *New York, United States*

- Applied online reinforcement learning to fix the human demonstration morphology gap for dexterous robotic manipulation
- Trained a behavior cloning policy using transformers on third-person human demonstration data to solve dexterous robotic hand tasks
- Set up an auto-resettable reinforcement learning environment for online robot training using the spacemouse device

Teaching Experience

Graduate Teaching Assistant — Sep 2025 – Dec 2025
NYU Courant Institute of Mathematical Sciences. Prof. Saining Xie — *New York, United States*

- Led weekly office hours for CSCI-GA.2271-001 Computer Vision. Website
- Helped students with their research projects and homework

Publications

1. *Solaris: A Multi-agent Video World Model*
   G. Savva*, O. Michel*, P. Waiwitlikhit*, D. Lu*, T. Meehan, D. Mishra, J. Lu, S. Poddar, S. Xie.
   (Expected Release Jan 2026)

2. *HuDOR: Bridging the Human to Robot Dexterity Gap through Object-Oriented Rewards*
   I. Guzey, Y. Dai, **G. Savva**, R. Bhirangi, L. Pinto.
   ICRA 2025. Website

Projects

Do Pre-Trained and Fine-Tuned World Models Generalize? — Jan 2025 – May 2025
Investigated the generalization and finetuning of two SOTA Minecraft World Models on three data distributions. Website

AppSim: A Learned World Model for an App API — Jan 2025 – May 2025
Used ChatGPT o3 in a zero-shot setting, achieving 74% accuracy, and compared its performance to a finetuned TinyLlama model. Website

Transformer-Based Diffusion for Game Generation — Oct 2024 – Dec 2024

Trained a transformer diffusion model to simulate DOOM trajectories. It achieves a PSNR of 32.21 in the teacher-forcing setting, producing an indistinguishable quality from the ground truth. Website

Advantage Actor-Critic with Optuna Jul 2024 – Sep 2024

Trained a reinforcement learning agent with Optuna for the HalfCheetah env, achieving a 24% better performance than the best publicly available policy. Website

Open-Source Library Scany May 2020 – Present

Created an open-source library to map data from a database into Go objects. The library has 1,500 stars on GitHub and is used by thousands of companies. GitHub

Industry Experience

Senior Software Engineer Oct 2023 – May 2024

Raylu *New York, United States*

- Led backend development of an LLM chatbot for healthcare and deployed it to production, serving first paying customers
- Developed an LLM-powered SaaS workflow automation application using the open source workflow engine Activepieces, allowing the company to release a new product in under 3 months

Technical Co-Founder Jul 2022 – Aug 2023

Scifind *Los Angeles, United States*

- Launched a troubleshooting platform for bioscientists into production, gaining 4,000 MAU in the first 2 months
- Led the development of the product with a team of 2 engineers using TypeScript, Next.js, and Node.js

Senior Software Engineer Jan 2021 – Jul 2022

IOTA *Berlin, Germany*

- Developed a new network layer of the blockchain node using libp2p, decreasing the number of peering errors by 70%
- Set up automatic deployment via Ansible and GitHub Actions, accelerating the development speed by 25%
- Introduced Go guidelines to the project and tools to ensure them, reducing the frequency of bugs by 50%

Senior Software Engineer Jan 2020 – Jan 2021

Elsa *San Francisco, United States*

- Implemented App Store and Google Play in-app subscriptions as microservices using FastAPI, which improved the correctness of billing by 50%
- Integrated payment gateway Instamojo, increasing the number of app purchases in Asia threefold

Software Engineer Mar 2017 – May 2019

Edwin *San Francisco, United States*

- Participated in building the English tutor chatbot from a prototype stage to being the top 1 bot on Facebook
- Implemented a YAML-file-based framework for building dialog systems, which allowed the company to launch a new product on Google Assistant
- Designed a distributed queue using PostgreSQL to handle messages from 800,000 users
- Developed the user knowledge graph service using Neo4j, improving the performance of read queries threefold

Software Engineer Jun 2016 – Mar 2017

Snaappy *Moscow, Russia*

- Rewrote the messaging service in Go, which increased the backend performance fivefold
- Migrated the chat storage from PostgreSQL to MongoDB, allowing horizontal scalability for the data layer
- Implemented a Websocket service for real-time updates using Go, driving user engagement by 30%

Skills

Languages: Python, SQL, Go, TypeScript
Machine Learning: Transformers, Diffusion, U-Net, CNN, LSTM, Reinforcement Learning, Distributed Training, TPU
Libraries: PyTorch, JAX, Gym, Optuna, Hydra, TensorFlow, Numpy, Pandas, FastAPI, React, Node.js
Tools: Docker, Kubernetes, PostgreSQL, AWS, GraphQL, Hasura

Certifications

Probability & Statistics for Machine Learning & Data Science Sep 2024
DeepLearning.AI, GPA: 93/100

YC Summer Startup School Aug 2022
Y Combinator

Mathematics for Machine Learning: Multivariate Calculus Dec 2021
Imperial College London, GPA: 96/100

Mathematics for Machine Learning: Linear Algebra Dec 2021
Imperial College London, GPA: 96/100