

Shashwat Gupta

✉ guptashashwatme+r@gmail.com / sg129@illinois.edu | ☎ +1 217-418-1825 / +91 8765694611 | 🎓 Google Scholar
🌐 linkedin.com/in/shashwat-gupta-uiuc | 🐙 github.com/ShashwatGPT | 🌐 shashwatgpt.github.io

Summary

CS graduate student and Microsoft Research India Fellow with **2+ years** across AI/ML research and production software engineering. Builds **agentic AI systems** and **distributed LLM infrastructure**: formulated multi-agent LLM communication as an RL problem for a **6.7% accuracy gain** (ACL 2026, US Patent Filed). Works down the stack into **systems and hardware**: co-designed a production **GPU/LLM cluster scheduler** at Microsoft Research (ILP with queuing-theoretic constraints, **6%** over baseline), and on **NCSA's DeltaAI** supercomputer root-caused Slurm, vLLM and CUDA failures on **GH200 Grace-Hopper** nodes; and shipped PySpark data pipelines serving **30M users / 60M features**. Proficient in **Python, C/C++, SQL, CUDA, PyTorch**/TensorFlow, data structures & algorithms, distributed-systems design, and cloud (Azure, Docker). Ships reliable, efficient, benchmarked software in cooperative teams, using AI-assisted development (GitHub Copilot, Claude Code).

Education

University of Illinois at Urbana-Champaign (UIUC), United States, Masters in Computer Science (MScS thesis based) 2026 to Present
Indian Institute of Technology (IIT) Kanpur, India, Bachelors (B.Tech.) in Computer Science (GPA: 9.3/10) 2020 to 2024
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, B.S. in Computer Science (Exchange Semester) (GPA: 5.5/6) Feb 2023 to June 2023

Interests

Research, Engineering, Operations (MLOps, DevOps), Agentic AI & LLM Systems, Applied Machine Learning & Data Science, Optimisation, Computer Vision, NLP, Large-Scale Data Pipelines, High-Performance / GPU Computing, Distributed Systems

Experience

Microsoft Research India (M365 Research)

Bangalore, India

Jan 2025 to Aug 2026

RESEARCH FELLOW (RESEARCH SOFTWARE DEVELOPMENT ENGINEER (RSDE) / SOFTWARE ENGINEERING (SWE))

- **OPTiMACS**: Formulated multi-agent LLM communication as an Extended MDP using an **RL** framework for message-structure selection, improving multi-hop QA accuracy by **6.7%** (5.1% over o3). **ACL 2026 Findings (A*)**, US Patent Filed.
- **BOFlow, Agentic Evolution**: Self-discoverable agentic workflows via **Bayesian optimisation**, **10% avg. accuracy gain** over SOTA baselines. **ACL 2027 (A*)**.
- **GPU Router-Scheduler Co-Design**: Co-designed a **distributed scheduling system** for cloud GPU/LLM clusters using **queuing-theoretic constraints**, pushed to production, achieving a **6% improvement** over the existing baseline.
- **AutoToolStub (FLASH)**: Built a parallelised test-case generator using **100-way concurrent BFS** and **provenance tracking** to auto-generate pre-verified tool stubs, benchmarking FLASH's incident-triaging agent across **60,000+ tickets**.

Bangalore, India

PwC US (Bangalore Advisory Centre)

Jul 2024 to Dec 2024

GENERATIVE-AI ASSOCIATE

- Built large-scale data pipelines using **PySpark** over a data fabric, scaling to serve **30M users / 60M features**.
- Built and shipped an **agentic workflow** for the Customer Links App from scratch to production in **1 month** under Scrum, using **RAG pipelines** and **LLM agents** to reduce **tokens used** and improve **context management**.

Bangalore, India

HP Tech Ventures

Jul 2024 to Aug 2024

BUSINESS ANALYTICS & VC

- Evaluated **50+** generative-AI/SaaS startups using an investment-ranking framework (**market sizing, tech-stack diligence**, unit economics, cap-table review), completing due-diligence for **3** of them.

Uttar Pradesh, India

HealthPredict (Pro-Bono Startup)

Sep 2023 to Dec 2025

FOUNDER

- **Agri-Health Vulnerability Engine**: Engineered a geospatial ML risk pipeline using **QGIS, Google Earth Engine, Sentinel imagery**, and **XGBoost** to route mobile health camps (**Recognised by Central Government**).
- **Saathi**: Built a **Flutter**-based interactive learning app for underprivileged children; led UI/UX design iterations and key feature development; inaugurated by the Dean and covered in local newspapers.
- **CRM for Hospitals**: Built a lightweight CRM to streamline patient outreach and care coordination, adopted across partner clinics.

Technical Skills

Programming Languages: Python, C/C++, SQL, CUDA

Data & ML: Pandas, NumPy, PyTorch, TensorFlow, Scikit-learn, PySpark

AI/Agentic Methods: RL post-training, Agent-based modeling, Bayesian Inference, Information Retrieval, Model Interpretability

Systems & Tools: Git/GitHub, Linux, Developer Tools, HPC (Slurm, NCSA DeltaAI), Azure, Docker

Misc: AI/ML, Machine Learning, Computer Science, Data Structures & Algorithms, Programming

Soft Skills: Problem-Solving, Accountability, Growth Mindset, Creativity, Reliability, Cooperative Teamwork

Languages: English (professional), Hindi (native), French (basic)

Publications

- **OPTiMACS: Learning Optimal Message Representations for Agentic Communication** (First-authored) | **ACL 2026 (Findings) (A*)** + US Patent Filed
- **BOFlow: Bayesian Optimisation over Latent Agentic Flows** (First-authored) | **ACL 2027 (A*)**, Accepted
- **EXACT-CT: Explainable Analysis for Crohn's and Tuberculosis using CT** (First-authored) | Under Review, *RSNA Radiology (A*)*
- **CISLR: Corpus for Indian Sign Language Recognition**, A. Joshi et al., **Shashwat Gupta** | *EMNLP 2022 (A*)*

Relevant Coursework

UIUC (MScS, Fall 2026): Distributed Systems (CS 425), Machine Learning Systems (CS 498), Advanced Topics in NLP (CS 546)

IIT Kanpur (B.Tech.): Machine Learning (CS771), Data Structures & Algorithms, Algorithms II, Operating Systems, Computer Networks, Computer Organisation, Compiler Design, Software Development & Operations (DevOps), Image Processing, Mathematics for CS I to III (Discrete Maths, Probability & Statistics, Linear Algebra)

EPFL (Exchange): Optimisation for Machine Learning (CS-439), Database Systems (CS-322)

Awards & Honours

- Dr. Prateek Mishra Memorial Scholarship: highest CPI, 1st year (10/10) | Academic Excellence Award (x2) | JEE Advanced AIR 603 (top 0.3%)