AYUSH KUMAR SHAH

Research Scientist at Meta

shahayush@meta.com | Menlo Park, California | (585) 471-9866 | @ayushkumarshah
@ayushkumarshah | @ayush7 | shahayush.com | Ayush Kumar Shah

EDUCATION

PhD in Computing and Information Sciences, CGPA: 3.93/4 — *Aug 2020 – June 2025*
Rochester Institute of Technology (RIT) — *Rochester, NY, USA*
Area of focus: Developing AI models for visual parsing of graphical structures and notations from documents
Relevant Courses: Pattern Recognition, Computer Vision, Deep Learning Mathematics, NLP, Software Engineering.

Bachelors in Computer Engineering, CGPA: 3.96/4 — *Aug 2015 – Oct 2019*
Kathmandu University — *Kavre, Nepal*

PROFESSIONAL EXPERIENCE

Meta — Menlo Park, California
Research Scientist — *July 2025 – Present*

- Applied AI (AAI) org: developing data, evaluation, and tooling systems to support training and post-training of next-generation large-scale language and multimodal models.
- Media Foundation Video team (AI): designing transformer- and diffusion-based models for video enhancement (super-resolution, restoration, denoising), compression, quality optimization, and accessibility across the Family of Apps (Facebook, Instagram, WhatsApp, Messenger).
- Collaborating with the Meta Superintelligence Lab on large vision models (LVMs) for video generation, curation, and multimodal understanding.

Amazon - Alexa Speaker Understanding AI — Sunnyvale, California
Applied Scientist Intern — *May 2022 – Aug 2022*

- Designed and implemented AI models with semi-supervised learning that improved speaker identification accuracy in Alexa by 5% while reducing training time by 80% and cutting annotation costs by millions of dollars annually.
- Developed automated speech data annotation pipelines using clustering techniques, labeling over 10 million hours of speech data and accelerating dataset creation by 5x to enable large-scale training of next-generation voice models.

Fusemachines — Kathmandu, Nepal
Machine Learning Engineer — *June 2019 – Aug 2020*

- Optimized client's business decisions for chemical products that go unsold using boosting classifiers.
- Automated bank data extraction by building a 95% accurate handwritten text (English & Nepali) recognizer.
- Increased a subscription-based e-commerce client revenue by 6% building a recommendation system.
- Prepared Fusemachines AI Education Programs course materials for AI Democratization.

PUBLICATIONS

- **A. K. Shah**, et al., "Multimodal Search in Chemical Documents and Reactions", in Proceedings of the 48th International ACM **SIGIR** Conference on Research and Development in Information Retrieval, in SIGIR '25. ACM, July 2025, pp. 4030–4034, doi: 10.1145/3726302.3730152.
- **A. K. Shah**, et al., "ChemScraper: Leveraging PDF Graphics Instructions for Molecular Diagram Parsing," in Document Analysis and Recognition - **IJDAR 2024**, vol. 27, pp. 395-414, doi: 10.1007/s10032-024-00486-7.
- **A. K. Shah**, and R. Zanibbi, "Line-of-Sight with Graph Attention Parser (LGAP) for Math Formulas," in Document Analysis and Recognition - **ICDAR 2023**, doi: 10.1007/978-3-031-41734-4_25.
- B. M. Amador, M. Langsenkamp, A. Dey, **A. K. Shah**, and R. Zanibbi. "Searching the ACL Anthology with Math Formulas and Text" in Proceedings of the 46th International ACM **SIGIR** Conference on Research and Development in Information Retrieval, in SIGIR '23. ACM 2023, July 2023, pp. 3110–3114, doi: 10.1145/3539618.3591803
- **A. K. Shah**, A. Dey, and R. Zanibbi, "A Math Formula Extraction and Evaluation Framework for PDF Documents," in Document Analysis and Recognition - **ICDAR 2021**, doi: 10.1007/978-3-030-86331-9_2

RESEARCH EXPERIENCE

Document and Pattern Recognition Lab (DPRL), RIT — Rochester, New York
Graduate Research Assistant — *Aug 2020 – June 2025*

- Built **ChemScraper**, the first parser to extract molecular diagrams directly from born-digital PDF graphics (no OCR), enabling automated generation of large annotated datasets for training visual chemical parsers; **adopted by Pfizer R&D (Groton, CT) for internal document analysis**.
- Led **Multimodal Chemical Search (ReactionMiner)**, integrating text, molecular structures, and reactions into structured "reaction cards" for chemical literature exploration.
- Increased math formula recognition accuracy by 15% using a modified Graph Attention Network (GAT) with spatial pyramidal pooling (LGAP), and co-developed **LgEval**, a graph-based evaluation framework now widely used by the document recognition community.
- Designed **MathDeck**, a search system for the ACL Anthology combining text and formula queries via an intuitive "chip" interface for context-aware matching.
- Achieved a 6× training speedup via a custom dynamic-batching data loader and distributed parallel training, and containerized systems (Docker, FastAPI, Kubernetes) for real-world deployment.
- Mentored 6+ graduate and undergraduate researchers and served as DPRL's liaison across 3 institutions and 20+ collaborators (RIT, UIUC, NCSA); featured in the MMLI Researcher Spotlight (2025).

Research Interests: Pattern recognition, recognition of graphical structures, computer vision, video understanding and generation, speaker understanding, large language models, multi-modal deep learning, natural language processing

PEER REVIEW CONTRIBUTIONS

Program Committee Member, 20th International Conference on Document Analysis and Recognition (ICDAR) — reviewed 6 manuscripts — *2026*

Program Committee Member, 19th International Conference on Document Analysis and Recognition (ICDAR) — reviewed 6 manuscripts — *2025*

Journal Reviewer, Pattern Recognition (Elsevier) — reviewed 7 manuscripts — *2024 – 2026*

Journal Reviewer, International Journal on Document Analysis and Recognition (IJDAR) — reviewed 2 manuscripts — *2024 – 2025*

Program Committee Member, 17th International Conference on Document Analysis and Recognition (ICDAR) — reviewed 5 manuscripts — *2023*

HONORS AND AWARDS

RIT Ph.D. Assistantship. Full funding via NSF-supported research projects. — *2020 – 2025*

Kathmandu University Merit Scholarship (4x). Awarded $440 total for highest GPA in the Computer Engineering cohort across 4 of 7 semesters. — *2015 – 2019*

Fusemachines AI Scholarship. Selected from a nationwide pool for the Fuse.ai Artificial Intelligence Scholarship Program. — *Nov 2018*

American Society of Nepalese Engineers Merit Award. $200 award for top university entrance rank in Nepal. — *May 2016*

46$^{\text{th}}$ International Physics Olympiad (IPhO) Contestant. Selected among Nepal's top 5 to compete internationally with participants from 100+ countries. — *June 2015*

TEACHING EXPERIENCE

Graduate Teaching Assistant, RIT, Rochester, NY — *Aug 2022 – Dec 2022*
Course: *CSCI 335: Machine Learning*

Instructor, Samriddhi College, Kathmandu, Nepal — *Jan 2020 – June 2020*
Course: *Foundations in AI: Computer Science and Mathematics*

TECHNICAL SKILLS

Programming Languages	Python, R, MATLAB, C, C++, Java
Python Packages	PyTorch, TensorFlow, scikit-learn, OpenCV, NLTK, Pandas, NumPy, Matplotlib, FastAPI, BeautifulSoup, Regex, NetworkX, Jupyter
Database	MySQL, MongoDB
Miscellaneous	Git, GitHub, Bash, LaTeX, Jira, Linux, Arduino, Raspberry Pi