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*
Research Group: Document and Pattern Recognition Lab (DPRL)
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*
Relevant Courses: AI, Data Structures & Algorithms, Software Engineering, Probability & Statistics, NLP, C, C++

PROFESSIONAL EXPERIENCE

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

- Media Foundation Video team (AI): designing and implementing AI algorithms for video processing, enhancement, and quality optimization across Family of Apps (FoA), including Facebook, Instagram, and WhatsApp.
- Collaborating with Meta Superintelligence Lab on video generation and curation using large vision models (LVMs).

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.

PUBLICATION

- **A. K. Shah**, A. Dey, L. Luo, B. M. Amador, P. Philippy, M. Zhong, S. Ouyang, D. M. Friday, D. Bianchi, N. Jackson, R. Zanibbi, and J. Han, "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**, B. M. Amador, A. Dey, M. Creekmore, B. Ocampo, S. Denmark, and R. Zanibbi, "ChemScraper: Leveraging PDF Graphics Instructions for Molecular Diagram Parsing," in Document Analysis and Recognition (Journal) - **IJDAR** 2024, vol. 27, Sep. 2024, 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, Cham: 2023, pp. 401–419, 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, Cham, 2021, pp. 19–34, 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*

- Developed a fast and accurate parser for molecular diagrams, including automated generation of annotated data for training visual chemical parsers, along with novel graph-based evaluation metrics and error analysis tools.
- Improved access to mathematical content by designing a search system for the ACL Anthology that integrates both textual and mathematical formula search, using context-aware word and formula matching.
- Increased math formula recognition accuracy by 15% through enhancements in attention mechanisms and contextual features, using a modified Graph Attention Network (GAT) combined with spatial pyramidal pooling.
- Achieved a 6× speedup in math formula recognition by implementing a custom data loader with dynamic batch sizing, fully utilizing GPU resources in a distributed parallel training framework.
- Contributed to the document recognition community by developing an open-source visualization tool to support the evaluation of graphical recognition results and enable detailed, context-aware error identification.

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

PEER REVIEW CONTRIBUTIONS

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

Journal Reviewer, Pattern Recognition (Elsevier) — reviewed 4 manuscripts — *2024 – 2025*

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*

46th 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

TALKS

Poster presentation on "ChemScraper: Pipeline for Parsing Raster and Vector Molecule Diagrams from PDFs" at the at Molecule Maker Lab Institute (MMLI) Symposium 2025 at **University of Illinois Urbana-Champaign (UIUC)**. *April 15, 2025*

Oral presentation on "ChemScraper: Leveraging PDF Graphics Instructions for Molecular Diagram Parsing" at the 18^{th} International Conference on Document Analysis and Recognition ICDAR 2024, Athens, Greece. *Sept 3, 2024*

Poster presentation on "ChemScraper: Extracting Molecule Diagrams from PDF Vector and Raster Images with CDXML and SMILES Output" at the Molecule Maker Lab Institute (MMLI) All-Institute Retreat at **University of Illinois Urbana-Champaign (UIUC)**. *Sept 12, 2023*

Poster presentation on the "Line-of-Sight with Graph-Attention and Task Interaction (LGATI) for math formula recognition" at the 17^{th} International Conference on Document Analysis and Recognition ICDAR 2023, San José, California. *Aug 22–23, 2023*

Poster presentation on "ChemScraper: Extracting Molecule Diagrams from PDF Vector Images with Page-Level CDXML (ChemDraw) and SMILES Output" in **NSF Annual Review Meeting** at **University of Illinois Urbana-Champaign (UIUC)**. *June 28, 2023*

Research Idea Ring (RIR) talk on "Line-of-sight with Graph Attention Parser (LGAP) for Math Formulas" at RIT. *April 17, 2023*

Poster presentation on "Reconstructing the Structure of Molecular Diagrams in PDF Documents using a CNN-Attention-Based Parsing Model" at the Molecule Maker Lab Institute (MMLI) All-Institute Retreat at **University of Illinois Urbana-Champaign (UIUC)**. *Sept 28, 2022*

Guest lecture on "Bayesian Decision Theory" for RIT's undergraduate course - Intro to Machine Learning (40 students). *Sept 5, 2022*

Research Idea Ring (RIR) talk on "A Fast and Interpretable Context-aware Parser for Isolated Formulas and Chemical Diagrams" at RIT. *April 7, 2022*

Poster presentation on the MathSeer extraction pipeline at the 16^{th} International Conference on Document Analysis and Recognition ICDAR 2021, Lausanne, Switzerland virtually. *Sept 9, 2021*

PROJECTS

Mulltimodal Chemical Search. *2025*
A system for searching chemical reactions, molecular structures, and text in scientific literature. It integrates text, SMILES, and reaction-based queries, linking extracted reaction details with molecular diagrams and textual descriptions. The interface provides structured reaction and molecule cards for easy navigation and retrieval, supporting chemists in literature exploration and data extraction.

ChemScraper. *2024*
A fast and accurate molecule diagram parser using characters and graphics extracted from born-digital (vector) PDF images—without the need for OCR, GPU, or vectorization. It uses these outputs to create training data for a new approach to visual parsing of molecule diagrams in raster images (i.e., pixel-based formats like PNGs) using a multi-task, segmentation-aware convolutional neural network (CNN).

MathDeck. *2022*
A system for searching PDF documents in a portion of the ACL Anthology, incorporating both formulas and text, displaying matched words and formulas in context. Its user-friendly interface includes formula 'chips' for easy formula creation, search, reuse and annotation. MathDeck supports both LaTeX and visual formula editing.

Math Formula Extraction. *2021*
A tool for detecting and recognizing math formulas from PDF documents using both PDF information (without OCR) and visual features through a multi-task learning framework based on ResNet-50 with line-of-sight graph-based attention.

Nepali Intelligent Character Recognition System. *2020*
A framework built using CRNN to predict the handwritten texts in both English and Nepali language in each field of a form.

Nepali Plagiarism Detector *2019*
An application that detects plagiarised Devanagari text files using a self-built rule-based stemming algorithm and Cosine similarity.

Guitar chord recognizer *2019*
An application that predicts the chord played by a guitar by extracting the Mel spectrograms of the sound and feeding it into a CNN.

Sarangi: Nepali lyrics emotions extraction *2018*
A framework that categorizes songs written in the Devanagari script into four different emotions using Naive Bayes.

AutoCar *2018*
A self-driving car that can detect lanes, stop sign, traffic light and avoid a collision, built using Canny edge detection, Hough transform, Haar cascade classifier, and Arduino programming.

MathMate – advanced mathematical calculator *2018*
An android app that solves different types of mathematical equations, numerical computations, and calculus problems showing involved steps.

ONLINE COURSES COMPLETED

Computer Vision Nanodegree	Udacity
Deep Learning Nanodegree	Udacity
Natural Language Processing Nanodegree	Udacity
Faster Python Code	Linkedin Learning
Python Design Patterns	Linkedin Learning
Full Stack Web Development with Flask	Linkedin Learning
Pandas foundations	Datacamp
Software Engineering for Data Scientists in Python	Datacamp
Multivariable calculus	Khan Academy
Statistics and Probability	Khan Academy
Linear Algebra by Gilbert Strang	MIT Lecture Series
Conda Essentials	Datacamp