

Akhil Pandey Akella

🏠 akhilpandey95.github.io 🎓 Akhil Pandey Akella 📧 akhilpandey95@allsci.com
🌐 [akhilpandey](https://akhilpandey.github.io) ✉ aakella@allsci.com

SUMMARY

My current research focus is bringing together the revolution of representational learning and language models to understand how it can play a dominant role in running long-horizon exploration tasks to accelerate scientific hypothesis discovery for life sciences. I build models, datasets, benchmarks and tools to understand how we can accelerate scientific discovery predominantly for furthering the causes of *AI4Science*.

EXPERIENCE

Applied Research Scientist (AERIS)

July 2025 - Present

AllSci Corp, Sunwater Institute

Chicago, IL

- Post-training and RL for *AERIS*, AllSci Enhanced Research Intelligence System.
- Post-training and data generation with frontier language models to break down scientific publications into hypotheses, research questions, and results.
- Build AI co-scientists to accelerate scientific discovery in life sciences.

Research Scientist (Kellogg Center for Science of Science & Innovation)

Aug 2024 - July 2025

Kellogg School of Management, Northwestern University

Evanston, IL

- Overseeing the AI, data analytics, data security, and data management efforts at CSSI for the \$ 20 Million National Science Foundation grant for Assessing and predicting technological outcomes.
- Built SciScinet-v2, the largest data lake for open science encompassing 250 million papers, 2.5 Billion citations, 100 Million authors, and 45 Million patent connections. More: <https://northwestern-cssi.github.io/sciscinet>
- Created the dataset to benchmarking (policyLLM) policy brief generation for language models and did strategic research into training and fine-tuning policy specific large language models that assist with demonstrating policy implications of scientific work.
- Established hybrid High Performance Computing CPU-GPU nodes for accelerating computational research at Center for Science of Science and Innovation, and Ryan Institute of Complexity.

Research Assistant (DATA Lab)

Sep 2017 - Aug 2024

Dept. of Comp.Sci, Northern Illinois University

Dekalb, IL

- Identifying Reproducible research using Human-in-the-Loop Machine Learning.
- Strategic research into models that address ways to comprehend, and estimate the Reproducibility crisis in AI.
- Create new datasets, 'reproducibility' metrics, and machine learning models that estimate confidence level in the reproducibility of a published work.
- Conducted comprehensive literature reviews and data analysis to support ongoing research projects.
- Presented findings at CIKM, JCDL, AACL-IJCNLP, Journal of Infometrics, enhancing the lab's visibility in the Science of Science, and applied machine learning community.

MCS Division, Givens Research Associate

May 2022 - Aug 2022

Argonne National Laboratory

Lemont, IL

- Played an instrumental role in the v1.0 rollout of DeepHyper/MetalgPy onto PyPI (<https://pypi.org/project/metalgpy/>).
- Built NAS pipelines for Graph Neural Architecture Search using Pytorch-Geometric and MetalgPy.
- Exponentially increased the search space for Conditional Hyperparameter Optimization on GNN's using DeepHyper and MetalgPy's symbolic program manipulation.
- Gave a session on Graph neural architecture search for molecular property prediction for ALCF DeepHyper Automated Machine Learning Workshop (<https://deephyper.github.io/events/workshop-anl-2022-summer>)

MCS Division, ALCF Summer Student

May 2020 - Aug 2020

*Argonne National Laboratory**Lemont, IL*

- Spearheaded a first of a kind comparative study of Uncertainty Quantification Approaches for Graph Neural Networks (GNN) on MoleculeNet.
- Built an architecture agnostic ML pipeline to estimate aleatoric and epistemic uncertainty in predictive models for classification and regression tasks.
- Empirically proved the effectiveness of Simultaneous Quantile Regression + Orthonormal Certificates as the best Uncertainty Quantification technique(<https://github.com/akhilpandey95/uncertainty>) for GNN's on Quantum Chemistry datasets.

Research Assistant, CREATE Lab

Jun 2018 - Dec 2018

*Dept. of Education, Northern Illinois University**Dekalb, IL*

- Utilized NLP techniques such as Automatic Speech Recognition to build speech models that focused on using child speech data to identify the text.
- Developed a web application in NodeJS, and Angular JS to separate child speech data and transcribe into word documents.
- Built a software interface to interact with Furhat- an advanced social robot.

Technical Task Force member

Aug 2016 - Oct 2016

*Mozilla India**Hyderabad, India*

- webcompat/webcompat-reporter-extension – Co-maintained one of the largest software(<https://github.com/webcompat/webcompat-reporter-extensions>) repositories for Mozilla's webcompat project that focuses on checking web page compatibility across browsers.
- Built a JavaScript Linter that automatically recognizes the JS version and highlights the necessary changes before writing them to the disk.
- Built a software interface to interact with Furhat- an advanced social robot.

Summer of Code Student

Apr 2016 - Aug 2016

*Google**Hyderabad, India*

- sTeam web Interface – Built a UI framework in AngularJS that talks to services from a Pike Webserver.
- Weekly Blog posts – Summarized the work and published bi-weekly blogs to showcase the open-source work.
- Actively engaged with mentors from Google, and FOSSASIA Singapore.

EDUCATION**Doctor of Philosophy (Ph.D), Northern Illinois University**

2024, USA

*Effort of Reproducibility: Metrics, Methods, & Models to Navigate the Landscape of Reproducible Research***Master of Science (M.S), Northern Illinois University**

2019, USA

*Using Machine Learning models to discover promising research***Bachelors in Computer Science (B.S), GITAM University**

2016, India

PUBLICATIONS**Published:**

- **Akella, A.P.**, Siravuri, H.V. and Rohatgi, S., 2025. "Pre-review to Peer review: Pitfalls of Automating Reviews using Large Language Models". arXiv preprint arXiv:2512.22145.
- Wu, W., Furnas, A.C., Yang, E., Liu, G., **Akella, A.P.**, Song, X., Wang, D. and Liu, H. "Sci2Pol: Evaluating and Fine-tuning LLMs on Scientific-to-Policy Brief Generation". arXiv preprint arXiv:2509.21493.
- Ibrahim Al Azher, Seethi, V.D.R., **Akella, A. P.** and Alhoori, H., "LimTopic: LLM-based Topic Modeling and Text Summarization for Analyzing Scientific Articles limitations" In 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL).

- **Akella, A. P.**, Choudhury, S.R., Koop, D. and Alhoori, H., “Navigating the Landscape of Reproducible Research: A Predictive Modeling Approach”. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (pp. 24-33).
- **Akella, A. P.**, & Alhoori, H. “Influence of Reproducibility on Scientific Impact” Accepted as an extended abstract to 3rd International Conference on the Science of Science and Innovation (ICSSI).
- **Akella, A. P.**, Koop, D., & Alhoori, H. (2023, June). “Laying Foundations to Quantify the “Effort of Reproducibility”.” In 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 56-60). IEEE.
- **Akella, A. P.**, Alhoori, H., & Koop, D. (2022, November). “Reproducibility Signals in Science: A preliminary analysis.” In Proceedings of the first Workshop on Information Extraction from Scientific Publications (pp. 140-144).
- **Akella, A. P.** (2022). “A Brief Survey on Representation Learning based Graph Dimensionality Reduction Techniques.” arXiv preprint arXiv:2211.05594.
- **Akella, A. P.**, Alhoori, H., Kondamudi, P. R., Freeman, C., & Zhou, H. (2021). “Early indicators of scientific impact: Predicting citations with altmetrics.” *Journal of Informetrics*, 15(2), 101128.
- **Akella, A.P.**, “Using Machine Learning Models to Discover Promising Research”, MS thesis, Northern Illinois University, 2019.
- Siravuri, H. V., **Akella, A. P.**, Bailey, C., & Alhoori, H. (2018, May). Using social media and scholarly text to predict public understanding of science. In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (pp. 385-386).

Selected working papers:

- **Akella, A. P.**, “Probing the Influence of Pretrained Scientific Language models”.(Under-revision).
- **Akella, A. P.**, “Large Language models for Reproducible Science of Science”. (Under-revision).

RESEARCH PRESENTATIONS

- Invited talk: **Data and software for AI**, “Leadership Computing Facility Seminar”, **Argonne National Laboratory**, *Courtesy: Dr. Huiho Zheng, May 2025.*
- Invited talk: **Direct Policy Optimization training with Large language models**, CS Lecture, **University of North Texas**, *Courtesy: Dr. Sagnik Roy Choudhury, April 2025*
- Invited talk: **Is reasoning the path forward for scientific text generation?**, CSCI Lecture, **Northern Illinois University**, *Courtesy: Dr. Hamed Alhoori, April 2025.*
- Invited talk: **Large Language Models for Computational Social Science**, MORS Lecture, **Kellogg School of Management, Northwestern University**, *Courtesy: Dr. Dashun Wang, March 2025.*
- Hands-on Workshop Speaker: **Graph neural architecture search for molecular property prediction**, **Argonne National Laboratory**, *Argonne Leadership Computing Facility Hands-on Training, July, 2022*
- Invited Seminar: **Symbolic programming for Neural architecture search on Graphs using Metalgpy**, **Argonne National Laboratory**, *Courtesy: Dr Prasanna Balaprakash, June 2022*
- Invited Seminar: **A comparative study of uncertainty quantification approaches for graph neural networks**, **Argonne National Laboratory**, *Courtesy: Dr Prasanna Balaprakash, August 2020.*
- Invited talk: **Understanding Backpropagation in LSTMs**, Department of Statistics, **Northern Illinois University**, *Courtesy: Dr Haiming Zhou, August 2018.*
- Invited Seminar: **Javascript & IoT, FOSSASIA Singapore**, *Courtesy: Hong Phuc Dang, March 2016.*

HONORS & AWARDS

- **National Science Foundation 2024 Travel grant for Conference on Information and Knowledge Management, CIKM'24** — October 2024
- **International Conference on Science of Science & Innovation (ICSSI) Best poster award** — July 2024
- **Northern Illinois University CS Department Travel Award** - June 2024
- **Special Interest Group on Information Retrieval (SIGIR) Student Travel Grant** — June 2023
- **Google Cloud Academic Research Grant**, Award: 274000118(2023-2024), Award: 331845891(2024-2025)
- **Lambda Labs Research Compute Grant** (2025-2026) .

SKILLS

Languages: C/C++, Python, Java, JavaScript/TypeScript, HTML/CSS, \LaTeX

Tools: Git/GitHub, Unix Shell, Webpack, VS Code, IntelliJ CLion/PyCharm/IDEA, Atom

PEER REVIEW AND SERVICE

Organizer/Co-Chair/Chair

- Resources Co-Chair ACM/IEEE Joint Conference on Digital Libraries, (JCDL'25) .
- Outreach and Publicity Chair ACM/IEEE Joint Conference on Digital Libraries, (JCDL'25) .

Program Committee Member

- Society for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining (SDM'25)
- Workshop on Information Extraction for Scientific Publications @ AACL-IJCNLP 2023

Reviewer

- Scientometrics

Professional Memberships

- ACM

REFERENCES

Dr. Dashun Wang

Professor and Kellogg Chair of Technology

Kellogg School of Management,

Northwestern University

Email: dashun.wang@kellogg.northwestern.edu

Website: dashunwang.com

Dr. Hamed Alhoori

Associate Professor - Computer Science

Northern Illinois University

Email: alhoori@niu.edu

Website: alhoori.github.io

Dr. Sagnik Roy Choudhury

Assistant Professor - Computer Science

University of North Texas

Email: Sagnik.Raychoudhury@unt.edu

Website: sagnik.github.io

Dr. Tirthankar Ghosal

Scientist (HPC and AI)

Oak Ridge National Laboratory

Email: ghosal@ornl.gov

Website: ornl.gov/tirthankar-ghosal