

BIPASHA BANERJEE, Ph.D

CONTACT INFORMATION

OFFICE ADDRESS: 301, Newman Library, 560 Drillfield Dr, Blacksburg, VA, 24061

PHONE: (540) 231-6737

EMAIL: bipashabanerjee@vt.edu

EDUCATION

Ph.D. COMPUTER SCIENCE, VIRGINIA TECH, BLACKSBURG, VA 2024

Dissertation: Banerjee, B. (2024). Improving Access to ETD Elements Through Chapter Categorization and Summarization, Ph.D. Thesis, Virginia Tech, 2024. <https://hdl.handle.net/10919/120890>

M.S. COMPUTER SCIENCE, VIRGINIA TECH, BLACKSBURG, VA 2022

B.Tech. COMPUTER SCIENCE & ENGINEERING
WEST BENGAL UNIVERSITY OF TECHNOLOGY, KOLKATA, INDIA 2015

PROFESSIONAL EXPERIENCE

ASSISTANT PROFESSOR, August 2024 - Present
AI RESEARCH SCIENTIST, DIGITAL LIBRARIES University Libraries, Virginia Tech, Blacksburg, VA

GRADUATE RESEARCH ASSISTANT August 2019 – August 2024
University Libraries, Virginia Tech, Blacksburg, VA

GLOBAL PRODUCT AND TECHNOLOGY INTERN May – August 2019
ADP, Alpharetta, GA

GRADUATE TEACHING ASSISTANT January – May 2019
Computer Science, Virginia Tech, Blacksburg, VA

ASSISTANT SYSTEMS ENGINEER - FULL TIME April 2016 – November 2017
TATA Consultancy Services, Kolkata, India

RESEARCH INTERESTS

I am an Assistant Professor at Virginia Tech, where I currently work as an AI Research Scientist, Digital Libraries for the University Libraries. My research mainly focuses on digital libraries, natural language processing, machine learning, artificial intelligence, and scholarly big data.

RECENT PUBLICATIONS & TALKS

1. Klair, H., German, F., Banerjee, B., & Ingram, W. A. (Accepted). Evaluating Human-LLM Alignment in ETD Subject Classification. In 2025, The 29th International Conference on Theory and Practice of Digital Libraries. <https://tpdl2025.github.io/>
2. Ingram, W. A., Banerjee, B., & Fox, E. A. (2025). When LLMs Disagree: Diagnosing Relevance Filtering Bias and Retrieval Divergence in SDG Search. LLM4Eval at SIGIR 2025 DOI: <https://doi.org/10.48550/arXiv.2507.02139>

3. Miller, C., Banerjee, B. (2025). Preserving and Understanding the Past through Large Language Models. Generative AI in Libraries Conference. URL: <https://shslibraryguides.org/genailibraries/schedule>
4. Banerjee, B., Goyne, J. (2025). AI-driven Solutions for Digital Library Collections. Generative AI in Libraries Conference. URL: <https://shslibraryguides.org/genailibraries/schedule>
5. Banerjee, B., Goyne, J., & Ingram, W. A. (2024, December). Making History Readable. In 2024 IEEE International Conference on Big Data (BigData) (pp. 8620-8622). IEEE. DOI: [10.1109/BigData62323.2024.10826028](https://doi.org/10.1109/BigData62323.2024.10826028)
6. Banerjee, B., Ingram, W. A., & Fox, E. A. (2024, December). Automating Chapter-Level Classification for Electronic Theses and Dissertations. In 2024 IEEE International Conference on Big Data (BigData) (pp. 2400-2409). IEEE. DOI: [10.1109/BigData62323.2024.10825418](https://doi.org/10.1109/BigData62323.2024.10825418)
7. Ingram, W. A., Banerjee, B., & Fox, E. A. (2024, December). Agentic AI for Improving Precision in Identifying Contributions to Sustainable Development Goals. In 2024 IEEE International Conference on Big Data (BigData) (pp. 8677-8679). IEEE. DOI: [10.1109/BigData62323.2024.10825072](https://doi.org/10.1109/BigData62323.2024.10825072)
8. Ingram, W. A., Wu, J., Kahu, S. Y., Manzoor, J. A., Banerjee, B., Ahuja, A., ... & Fox, E. A. (2024). Building datasets to support information extraction and structure parsing from electronic theses and dissertations. International Journal on Digital Libraries, 25(2), 175-196. DOI: <https://doi.org/10.1007/s00799-024-00395-4>
9. Park, S. H., Banerjee, B., Ingram, W. A., & Fox, E. A. (2023). Case Study of Analyzing the Variety of ETD Layouts. URL: <https://ir.inflibnet.ac.in/server/api/core/bitstreams/27afc90c-1a82-470e-8baa-5c1a3769426b/content>
10. Chekuri, S., Chandrasekar, P., Banerjee, B., Park, S. H., Masrourisaadat, N., Ahuja, A., ... & Fox, E. A. (2023, June). Integrated digital library system for long documents and their elements. In 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 13-24). IEEE. DOI: [10.1109/JCDL57899.2023.00012](https://doi.org/10.1109/JCDL57899.2023.00012)
11. Banerjee, B., Ingram, W. A., Wu, J., & Fox, E. A. (2022, December). Applications of data analysis on scholarly long documents. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 2473-2481). IEEE. DOI: [10.1109/BigData55660.2022.10020935](https://doi.org/10.1109/BigData55660.2022.10020935)
12. Banerjee, B., Jude, P.M., Ingram, W.A., Luther K., Fox, E.A. (2022). Help Me Help You - A Mixed-Initiative Approach To Explore Book-length Documents. Workshop on Human-in-the-loop Data Curation at ACM CIKM 2022 <https://hilworkshops.github.io/hil-dc2022/>
13. Banerjee, B. (2022, June). Opening scholarly documents through text analytics. In Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (pp. 1-2). DOI: <https://doi.org/10.1145/3529372.3530948>
14. Uddin, S., Banerjee, B., Wu, J., Ingram, W. A., & Fox, E. A. (2021, December). Building A large collection of multi-domain electronic theses and dissertations. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 6043-6045). IEEE. DOI: [10.1109/BigData52589.2021.9672058](https://doi.org/10.1109/BigData52589.2021.9672058)
15. Banerjee, B., William A. Ingram, Jian Wu, and Edward A. Fox (2021), Applications of Mining ETDs. ETD 2021 Conference: Open scholarship in a post-pandemic world. DOI: <https://doi.org/10.26226/morressier.614c9b8c87a68d83cb5d59b2>
16. Aromando, J., Banerjee, B., Ingram, W. A., Jude, P., & Kahu, S. (2020). Classification and extraction of information from ETD documents. In Fox, E.A. course CS 6604: Digital Libraries, 2019. <http://hdl.handle.net/10919/96645>
17. Khawas, P., Banerjee, B., Kim, Y., Zhao, S., & Fan, Y. (2018). Big Data Text Summarization for Maryland Shooting Collection. In Fox, E.A. course CS 5984: Big Data Text Summarization 2018. <http://hdl.handle.net/10919/86407>
18. Ingram, W. A., Banerjee, B., & Fox, E. A. (2019). Summarizing ETDs with deep learning. Cadernos BAD, 1, 46-52. <https://opening-etds.github.io/files/2014-5500-1-PB.pdf>

PATENTS

Co-inventor on: US Patent Application 18/585,685; filed February 23, 2024; published 8/29/2024; Document ID US 20240289356 A1: Structured Documentation Access for Electronic Documents: Edward A. Fox, William A. Ingram, Aman Ahuja, Satvik Chekuri, and Bipasha Banerjee. <https://patents.google.com/patent/US20240289356A1>

TEACHING EXPERIENCE

FACULTY MENTOR

Virginia Tech, Blacksburg, VA, 2024- Present

Mentoring graduate and undergraduate students on their research and curriculum work.

GRADUATE STUDENT MENTOR

Virginia Tech, Blacksburg, VA, 2019-2024

GRADUATE TEACHING ASSISTANT

Virginia Tech, Blacksburg, VA, JAN – MAY 2019

PROFESSIONAL SERVICE

- Co-chair, Conference Proceedings, IEEE/ACM Joint Conference on Digital Libraries, 2025
- Invited reviewer for Springer Nature Scientometrics Journal Article, 2025.
- Reviewer for IEEE/ACM Joint Conference on Digital Libraries 2022, 2023, 2024
- Co-organized a workshop titled “Hands-on digitizing texts with machine learning and AI”. A three-hour long workshop open to VT community on using machine learning and LLMs for optical character recognition

PROFESSIONAL MEMBERSHIPS

- Association for Computing Machinery (ACM) Member- #3768609
- Institute of Electrical and Electronics Engineers (IEEE) Member- #96669088

PROFESSIONAL ACTIVITIES

- Collaboration on a project geared toward using LLMs for University-wide analysis in partnership with University Libraries and the Office of Analytics & Institutional Effectiveness, Virginia Tech.
- Led breakout sessions at the workshop titled “Ensuring Scholarly Access to Government Archives and Records,” organized by Virginia Tech University Libraries in partnership with Virginia Tech Center for Humanities and the U.S. National Archives and Records Administration (funded by The Andrew W. Mellon Foundation)
- Digital Library Research Laboratory student representative for CS grad council at Virginia Tech (2020-2023).
- Association of Women in Computing Member, Virginia Tech Chapter.
- Volunteered at VT CS grad recruiting weekend and career fair.

AWARDS

- Recognized as an inventor by Virginia Tech for “identifying and disclosing a technology for FY 23” in 2024.
- Nominated for best student paper for “Integrated Digital Library System for Long Documents and their Elements” at IEEE/ACM JCDL 2023.
- Awarded a scholarship to attend the Grace Hopper Conference by the CS department at Virginia Tech in 2021 and the Association of Women in Computing (AWC, Virginia Tech Chapter) in 2022.
- Awarded free attendance at the International Conference in Pattern Recognition (ICPR), virtual 2020.
- “Student of the Year” awarded by the Institute of Engineering and Management for academics as well as contribution to the college in 2014.