

Haihua Chen, Ph.D.

Assistant Professor, Data Science (Affiliated with Health Informatics)

Director, [Intelligent Data Engineering and Analytics \(IDEA\) Lab](#)

Homepage: <https://haihua0913.github.io>

Anuradha and Vikas Sinha Department of Data Science
University of North Texas
3940 North Elm, Suite E292 Denton, Texas 76203-5017

Email: haihua.chen@unt.edu

[Google Scholar](#)

ORCID: [0000-0002-7088-9752](https://orcid.org/0000-0002-7088-9752)

Highlights

- Received 1,900+ citations and recognized as a high impact scholar focusing on data quality, legal AI, and computational social science.
- Published more than 60 peer-reviewed articles in top venues, including IP&M, KBS, JBI, ACM Computing Surveys, and others.
- Participated 20+ grant applications from NSF, NIH, IMLS, and others.
- Co-organized 10+ international conferences and workshops since 2022.
- Served or serving as the associate editor/guest editor/EAB member for 7 SSCI/SCIE journals, and reviewer for 30+ top-tier journals.
- Received excellent teaching evaluation for all the courses.

Research Interests and Expertise

- Artificial intelligence: applied machine learning, applied artificial intelligence, large language models
- Data science: applied data science, text mining, innovation measurement, social media analysis
- Data quality: data quality evaluation and improvement in machine learning, data centric AI
- Natural language processing: information retrieval, recommendation system, legal NLP
- Informatics: health informatics, legal informatics

Education

2017 – 2021	University of North Texas Ph.D. in Information Science concentrated on Data Science (Advisors: Dr. Jiangping Chen & Dr. Junhua Ding) Dissertation Title: Data Quality Evaluation and Improvement for Machine Learning Honor: Great Graduates
2014 – 2017	Wuhan University M.S. in Information Science (Advisor: Dr. Wei Lu) Honor: Fellowship for Graduate Students
2010 – 2014	Central China Normal University B.S. in Information Science & B.S. in English Literature Honor: Honor Graduates

Appointments

08/2025-present **Affiliated Faculty in Health Informatics (Tenure-track)**, University of North Texas

08/2023-present **Assistant Professor in Data Science (Tenure-track)**, University of North Texas
01/2022-07/2023 **Clinical Assistant Professor in Data Science**, University of North Texas

Publications

(*) indicates the corresponding author. (#) indicates student supervised at UNT. IF = Impact Factor. Journal ranking is based on Journal Citation Report (JCR) 2024.

Articles Under Review

- [1] *NIPS'26* Zuo, J., Zhang, Y., Laing, J., Zheng, Q., Yan, J., Chen, H., Zhang, Y. (2026). Generative Modeling of Human Trajectories via Token-Independent Hyperspherical Flow Matching. The Fortieth Annual Conference on Neural Information Processing Systems (NIPS 2026). Under review. (Top conference in CS)
- [2] *NIPS'26* Zuo, J., Liu, A., Zhang, Y., Kou, Z., Chen, H., Zhang, Y. (2026). Bridging Heterogeneous Spaces via Latent Asymmetric Bidirectional Diffusion. The Fortieth Annual Conference on Neural Information Processing Systems (NIPS 2026). Under review. (Top conference in CS)
- [3] *NIPS'26* Li, R., Lin, J., Zhang, H., Chen, H., Ding, J., Gehring, E., Zhang, Y. (2026). TopoGraphRAG-Bench: Evaluating Multimodal GraphRAG on Layout-Grounded Evidence Reasoning. The Fortieth Annual Conference on Neural Information Processing Systems (NIPS 2026). Under review. (Top conference in CS)
- [4] *JBI* Zhou, Y., Li, R., Cleveland, A., Ding, J., Sha, K., & **Chen, H.*** (2026). Quality-aware Multi-source Data Fusion and Enhancement for Medical Concept Normalization using Large Language Models. *Journal of Biomedical Informatics*. Under review. (JCR Q1).
- [5] *EMNLP'26* Zhang, H.#, Li, R., Liang, Z., Zhang, Y.#, Xiao, T., Ding, J., & **Chen, H.*** (2026). MetaGAI: A Large-Scale and High-Quality Benchmark for Generative AI Model and Data Card Generation (Long paper). *The 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026)*. Under review. (Top conference in CS)
- [6] *EMNLP'26* Liang, Z., Zhang, H., Xiao, T., Zhang, Y., **Chen, H.** (2026). From Puzzles to Paradigms: A Survey of Human-Agent Collaboration in Scientific Innovation (Survey paper). *The 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026)*. Under review. (Top conference in CS)
- [7] *COLM'26* Seo, W., Lee, J., Koh, J., An, H., Lee, S., **Chen, H.***, & Bu, Y*. (2026). MA-SAPO: Multi-Agent Reasoning for Score-Aware Prompt Optimization. *The 3rd Conference on Language Modeling (COLM)*. Under review.
- [8] *OIR* Li, Y., Mandaloju, T.#, Braidwood, J., Tang, L., **Chen, H.*** (2026). Public Conversations About Generative AI and Libraries on X: A Discourse Mapping Study. Online information Review. Under review. (JCR Q1)
- [9] *IEEE BIBM'26* Challa, L.#, Cleveland, A., Hossain, T., **Chen, H.*** (2026). StageCraft: Neuro-Symbolic Synthetic Clinical Data Generation for Lung Cancer TNM Staging (regular paper). *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Under review.
- [10] *IEEE TCSS* Jahan, R. I. #, Fan, H., **Chen, H.**, & Feng, Y. (2026). Unlocking Cross-Lingual Sentiment Analysis through Emoji Interpretation: A Multimodal Generative AI Approach. *IEEE Transactions on Computational Social Systems*. Under review. (JCR Q1)
- [11] *JCLLT* Wang, S. #, Pobbathi, L. #, & **Chen, H.*** (2026). LAMUS: A Large-Scale Corpus for Legal Argument Mining from U.S. Caselaw using LLMs. *Journal of Computational Law and Legal Technology*. Under review.
- [12] *JOCCH* Cherukuri, K.#, Moses, P. #, Sakata, A.#, Chen, J., & **Chen, H.***. Large Language Models for Oral History Understanding with Text Classification and Sentiment Analysis. *ACM Journal on Computing and Cultural Heritage*. In revision. (JCR Q3)

- [13] *JDR* Donepudi, S., **Chen, H***, Cleveland, A., Potula, R., & Parimi, S. Semi-automatic Dental Taxonomy Construction with Human and Generative AI Collaboration. *Journal of Dental Research*. Under review.
- [14] *BJET* Tu, F. #, Yu, J., Li, S., Zheng, Z., Ge, X., Chen, H. Modeling the Human-AI Dyad: Dynamics of Co-Regulation and Cognitive Offloading. *British Journal of Educational Technology*. Under review. (JCR Q1)
- [15] *ACM Computing Surveys* Ying, D., Yu, F., **Chen, H.**, Jiang, C., Li, Y., & Lu, W. Beyond Human Annotation: Recent Advances in Data Generation Methods for Document Intelligence. *ACM Computing Surveys*. (JCR Q1)
- [16] *ACM Computing Surveys* Zhang, H. #, Li, R., Xiao, T., Zhang, Y., Chen, J., Ding, J., & **Chen, H***. The Evolving Role of Large Language Models in Scientific Innovation: Assistant, Collaborator, Scientist, and Evaluator. *ACM Computing Surveys*. (JCR Q1)

Books, Book Chapters

- [B3] **Chen, H.**, Sha, K., Zhou, Y.#, & Tu, F #. Data Quality: Models, Measures, and Tools. In Springer Handbook of Data Engineering, in press. 2026.
- [B2] Sha, K., Cherukuri, K.#, & **Chen, H.** Privacy and Bias in Edge Computing. In Textbook Edge Computing: Systems and Applications, Wiley-IEEE Press. June 2025. ISBN: 978-1-394-28583-9.
- [B1] Lund, B., Agbaji, D. #, Bissadu, K., & **Chen, H.** Python for Information Professionals: How to Design Practical Applications to Capitalize on the Data Explosion. Rowman & Littlefield Publishers. November 15, 2023. ISBN-10: 1538178257. ISBN-13: 978-1538178256. (1200+ readings on Research Gate)

Refereed Journal Articles

- [J39] *JCC* Wanf, F., Li, D., Ma, B., Wang, X., Li, C., Qi, X., Zhang, Y., Sha., K., **Chen, H.**, Yu., G. (2026). Selective Semantic State Editing for Private Edge-Cloud Inference in SLMs. *Journal of Cloud Computing*. In-press. (JCR Q1)
- [J38] *TEL* Zhang L., Jiang, C., Sun, M., Cheng, X., **Chen, H***, & Lu, W. (2026). LSD: A Unified Framework of Linked Scholarly Databases for Data Fusion and Scientific Discovery. *The Electronic Library*. In-press. (JCR Q2).
- [J37] *ETR&D* Yu, J., Tu, F. #, **Chen, H.**, Ding, J., Liu, D., Hsieh, C., Kim, H., & Watson, S. (2026). Measuring Cognitive Presence in Online Discussions: Automated Detection and Instructional Insights from a MOOC Context. *Educational technology research and development*. In-press. (JCR Q1)
- [J36] *IEEE IC* Zhou, Y. #, **Chen, H.**, & Sha, K. A Novel Multi-Layer Task-Centric and Data Quality Framework for Autonomous Driving. *IEEE Internet Computing*. Accepted.
- [J35] *JBI* **Chen, H***, Zhou, Y. #, Li, R., Illa, A. M. #, Cleveland, A., & Ding, J. (2026). A comprehensive survey on medical concept normalization: Datasets, techniques, applications, and future directions. *Journal of Biomedical Informatics*, 105005.
- [J34] *Brain Sciences* Akbari, H., Bagherzadeh, S., Farhadi Sedehi, J., Nawaz, R., Rostami, R., Kazemi, R., Muhammad, S., **Chen, H.**, & Mete, M. (2026). Towards the Development of a Deep Learning Framework Using Adaptive and Non-Adaptive Time-Frequency Features for EEG-Based Depression Therapy Prediction. *Brain Sciences*, 16(3), 301.
- [J33] *ESWA* Wang, Z., Wang, Z., Zhang, G., Chen, J., Luczak-Roesch, M., & **Chen, H***. (2025). A Hybrid Graph and LLM Approach for Measuring Scientific Novelty via Knowledge Recombination and Propagation. *Expert Systems with Applications*, 129794. IF = 7.5. (32/441 in Artificial Intelligence, JCR Q1)
- [J32] *TEL* Luo, Z., Lu, W., Zhu, C., Gu, C., Cai, L., & **Chen, H.** (2025). A dual-perspective innovation framework: Quantifying novelty and diffusion of scientific publications. *The Electronic Library*. Accepted. (JCR Q2)
- [J31] *JBI* **Chen, H***, Li, R., Cleveland, A., & Ding, J. (2025). Enhancing Data Quality in Medical Concept Normalization through Large Language Models. *Journal of Biomedical Informatics*, 104812. IF = 4.0. (26/115 in Health Informatics, JCR Q1)
- [J30] *IP&M* Wang, Z., Wang, N., Zhang, H.#, Wang, Z., Wang, Z., Ding, J., & **Chen, H***. (2025). IBID-CCT: A novel model for interdisciplinary breakthrough innovation detection based on the cusp

- catastrophe theory. *Information Processing & Management*, 62(4), 104121. IF = 7.4. (8/258 in Library and Information Sciences, JCR Q1)
- [J29] *EIT* Tu, F.#, Wu, L., Kinshuk, Ding, J., & **Chen, H.***. (2025). Exploring the influence of regulated learning processes on learners' prestige in project-based learning. *Education and Information Technologies*, 30(2), 2299-2329. *Education and Information Technologies*, 1-31. IF =5.5. (19/258 in Library and Information Sciences, JCR Q1)
- [J28] *JOI* Zhang, Y., Wang, Y., Sheng Q., **Chen, H.**, et al. (2025). Deep Learning Meets Bibliometrics: A Survey on Citation Function Classification. *Journal of Informetrics*, 19(1), 101608. IF = 3.7. (18/258 in Library and Information Sciences, JCR Q1)
- [J27] *Electronics* Kargozari, K.#, Ding, J., **Chen, H.** (2024). Empowering Consumer Decision: Decoding Incentive vs. Organic Reviews for Smarter Choices Through Advanced Textual Analysis. *Electronics*, 13(21), 2535. IF = 2.9. (138/373 in Computer Networks and Communications, JCR Q2)
- [J26] *JOI* Wang, Z., Zhang, H.#, Chen, J., & **Chen, H.***. (2024). An Effective Framework for Measuring the Novelty of Scientific Articles through Integrated Topic Modeling and Cloud Model. *Journal of Informetrics*. 18(4), 101587. IF = 3.7. (18/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 24](#))
- [J25] *Electronics* Zhao, H., **Chen, H.**, Ruggles, T. A., Feng, Y., Singh, D., & Yoon, H. J. (2024). Improving Text Classification with Large Language Model-Based Data Augmentation. *Electronics*, 13(13), 2535. IF = 2.9. (138/373 in Computer Networks and Communications, JCR Q2, [Google Scholar Citation: 41](#))
- [J24] *JKSU-CIS* Wang, Z., Zhang, H.#, **Chen, H.***, Feng, Y., Ding, J. (2024). Content-based Quality Evaluation of Scientific Papers using Coarse Feature and Knowledge Entity Network. *Journal of King Saud University - Computer and Information Sciences*. 36(6), 102119. IF = 6.9. (55/326 in Computer Science, JCR Q1, [Google Scholar Citation: 15](#))
- [J23] *TEL* **Chen, H.***, Kim, J., Chen, J., & Sakata, A.#. (2024). Demystifying oral history with natural language processing and data analytics: a case study of the Densho digital collection. *The Electronic Library*, 42(4), 643-663. IF = 2.42. (75/258 in Library and Information Sciences, JCR Q2)
- [J22] *TEL* Wang, Z., Qiao, X., Chen, J., Li, L., Zhang, H.#, Ding, J., & **Chen, H.***. (2024). Exploring and evaluating the index for interdisciplinary breakthrough innovation detection. *The Electronic Library*, (ahead-of-print). IF = 2.42. (75/258 in Library and Information Sciences, JCR Q2)
- [J21] *ACM Computing Surveys* Huang, J., **Chen, H.**, Yu, F., & Lu, W. (2024). From Detection to Application: Recent Advances in Understanding Scientific Tables and Figures. *ACM Computing Surveys*. IF = 16.6. (4/326 in Computer Science, JCR Q1, [Google Scholar Citation: 12](#))
- [J20] *SCM* Wang, Z., Chen, J., Chen, J., & **Chen, H.***. (2024). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics*, 129(11), 7359-7384 IF = 3.801. (26/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 151](#))
- [J19] *Electronics* Ge, J., Xu, G., Lu, J., Zhang, Y., **Chen, H.**, & Meng, X. (2023). Joint optimization of computation, communication and caching in D2D-assisted Caching-enhanced MEC System. *Electronics*, 12(15), 3249. IF = 2.9. (138/373 in Computer Networks and Communications, JCR Q2)
- [J18] *Knowledge-Based Systems* Wang, Z., Peng, S., Chen, J., Zhang, X., & **Chen, H.***. (2023). ICAD-MI: Interdisciplinary concept association discovery from the perspective of metaphor interpretation. *Knowledge-Based Systems*. 110695. IF = 8.8. (29/284 in Artificial Intelligence, JCR Q1)
- [J17] *JKSU-CIS* Wang, Z., Peng, S., Huang, J., Chen, J., & **Chen, H.***. (2023). ICDIA-NCM: A Framework for Interdisciplinary Concept Drift Identification and Analysis based on the Normal Cloud Model. *Journal of King Saud University - Computer and Information Sciences*. 101569. IF = 6.9. (55/326 in Computer Science, JCR Q1)
- [J16] *IDD* Nguyen, H.#, **Chen, H.**, Chen, J., Kargozari, K., & Ding, J. (2023). Construction and Evaluation of a Domain-Specific Knowledge Graph for Knowledge Discovery. *Information Discovery & Delivery*. IF = 2.245. (71/258 in Library and Information Sciences, JCR Q2)

- [J15] *IEEE Access* Tran, N.#, **Chen, H.**, Bhuyan, J., & Ding, J. (2022). Data Curation and Quality Evaluation for Machine Learning-Based Cyber Intrusion Detection. *IEEE Access*. IF = 3.9. (54/326 in Computer Science, JCR Q1, [Google Scholar Citation: 41](#))
- [J14] *JBI* Zhang, L., Lu, W., **Chen, H.**, Huang, Y., & Cheng, Q. (2022). A comparative evaluation of biomedical similar article recommendation. *Journal of Biomedical Informatics*, 104106. IF = 4.0. (26/115 in Health Informatics, JCR Q1, [Google Scholar Citation: 13](#))
- [J13] *IEEE TR* **Chen, H.**, Pieptea, L. F.#, & Ding, J. (2022). Construction and Evaluation of a High-Quality Corpus for Legal Intelligence Using Semiautomated Approaches. *IEEE Transactions on Reliability*, 71(2), 657-673. IF = 5.9. (15/189 in Safety, Risk, Reliability and Quality, JCR Q1, [Google Scholar Citation: 29](#))
- [J12] *IP&M* **Chen, H.**, Wu, L., Chen, J., Lu, W., & Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. *Information Processing & Management*, 59(2), 102798. IF = 7.4. (8/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 274, top 4 most cited article among all IP&M articles since 2022.](#))
- [J11] *SCM* **Chen, H.***, Nguyen, H.#, & Alghamdi, A.# (2022). Constructing a high-quality dataset for automated creation of summaries of fundamental contributions of research articles. *Scientometrics*, 1-15. IF = 3.801. (26/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 14](#))
- [J10] *SCM* Wang, Z., Wang, K., Liu, J., Huang, J., & **Chen, H.*** (2022). Measuring the innovation of method knowledge elements in scientific literature. *Scientometrics*, 127(5), 2803-2827. IF = 3.801. (26/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 18](#))
- [J9] *SCM* Zhang, Y., Zhao, R., Wang, Y., **Chen, H.**, Mahmood, A., Zaib, M., ... & Sheng, Q. Z. (2022). Towards employing native information in citation function classification. *Scientometrics*, 1-21. IF = 3.801. (26/258 in Library and Information Sciences, JCR Q1, [Google Scholar Citation: 31](#))
- [J8] *IEEE TR* **Chen, H.**, Chen, J., & Ding, J. (2021). Data evaluation and enhancement for quality improvement of machine learning. *IEEE Transactions on Reliability*, 70(2), 831-847. IF = 5.9. (15/189 in Safety, Risk, Reliability and Quality, JCR Q1, [Google Scholar Citation: 190](#))
- [J7] *IJPE* Tran, N.#, **Chen, H.***, Jiang, J., Bhuyan, J., & Ding, J. (2021). Effect of Class Imbalance on the Performance of Machine Learning-based Network Intrusion Detection. *International Journal of Performability Engineering*, 17(9). IF = 1.14. ([Google Scholar Citation: 29](#))
- [J6] *JDIS* **Chen, H.*** (2021). A New Citation Recommendation Strategy Based on Term Functions in Related Studies Section. *Journal of Data and Information Science*, 6(3), 75-98. IF = 1.889. (88/258 in Library and Information Sciences, JCR Q2)
- [J5] *JMLA* **Chen, H.**, Chen, J., & Nguyen, H.# (2021). Demystifying COVID-19 publications: institutions, journals, concepts, and topics. *Journal of the Medical Library Association: JMLA*, 109(3), 395. IF = 3.18 (47/115 in Health Informatics, JCR Q2)
- [J3] *TEL* **Chen, H.***, Yang, Y., Lu, W., & Chen, J. (2020). Exploring multiple diversification strategies for academic citation contexts recommendation. *The Electronic Library*, 38(4), 821-842. IF = 2.42. (75/258 in Library and Information Sciences, JCR Q2)
- [J3] *TEL* Tang, M., Chen, J., **Chen, H.**, Xu, Z., Wang, Y., Xie, M., & Lin, J. (2020). An ontology-improved vector space model for semantic retrieval. *The Electronic Library*, 38(5/6), 919-942. IF = 2.42. (75/258 in Library and Information Sciences, JCR Q2)
- [J2] *Information Sciences* Lu, W., Luo, M., Zhang, Z., Zhang, G., Ding, H., **Chen, H.**, & Chen, J. (2019). Result diversification in image retrieval based on semantic distance. *Information Sciences*, 502, 59-75. IF = 8.1. (23/284 in Artificial Intelligence, JCR Q1, [Google Scholar Citation: 13](#))
- [J1] *IDDI* Zhang, Z., **Chen, H.***, & Xiao, B. (2019). Understanding eWOM of Chinese Governments information service: a perceived value-based perspective. *Information Discovery & Delivery*, 47(4), 251-258. IF = 2.245. (71/258 in Library and Information Sciences, JCR Q2)

Refereed Conference Papers

- [C28] *ICTIR'26* Lee, J., Seo, W., Koh, J., Lee, S., **Chen, H.***, & Bu, Y*. (2026). RoTRAG: Rule of Thumb Reasoning for Conversation Harm Detection with Retrieval-Augmented Generation. *International ACM*


- SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR 2026)*. In press.
- [C27] *IEEE AITest'26* Nguyen, H. #, Zhang, H., Zhang, Y., Ding, J., & **Chen, H.***. (2026). SumEval4ALL: A Reference-Free Quality Evaluation Framework for Document Summarization using Large Language Models. In Proceedings of the 2026 IEEE International Conference on Artificial Intelligence Testing (AITest 2026). In press.
- [C26] *IEEE AITest'26* Nguyen, H. #, Zhang, H., Zhang, Y., Ding, J., & **Chen, H.***. (2026). LongSumEval: Question-Answering Based Evaluation and Feedback-Driven Refinement for Long Document Summarization. In Proceedings of the 2026 IEEE International Conference on Artificial Intelligence Testing (AITest 2026). In press.
- [C25] *IEEE MOST'26* Zhou, Y. #, Sattari, M., **Chen, H.**, & Sha, K. (2026). Modeling and Measuring Redundancy in Multisource Multimodal Data for Autonomous Driving. In *2026 IEEE 4th International Conference on Mobility, Operations, Services and Technologies (MOST)* (pp. 1-13). IEEE. (**Berst Poster Award**).
- [C24] *WWW'26* Zhang, H.#, Li, R., Liang, Z., Sattari, M., Vo, P., Qu, Q., Xiao, T., Zhang, Y.*, Ding, J., & **Chen, H.***. (2026). AdaQE-CG: Adaptive Query Expansion for Web-Scale Generative AI Model and Data Card Generation (Long paper). In *Proceedings of the ACM Web Conference 2026* (pp. 4360-4371). (Top conference in CS, Acceptance rate 20.1%). CORE Conference Ranking A
- [C23] *JCDL'25* Zhang, H.#, Li, R., Shrestha, S., Mamidala, S., Putta, R., Aggarwal, A., Xiao, T., Ding, J., & **Chen, H.***. (2025). ReviewGuard: Enhancing Deficient Peer Review Detection via LLM-Driven Data Augmentation (Long paper). *The 25th ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE. (Top conference in LIS, Acceptance rate 31%). Web of Science-indexed
- [C22] *IEEE ICDM'25* Li, R., Zhang, H.#, Gehringer, E., Xiao, T., Ding, J., & **Chen, H.***. Unveiling the Merits and Defects of LLMs in Automatic Review Generation for Scientific Papers (Short paper). *The 25th IEEE International Conference on Data Mining (ICDM 2025)*. (**Berst Poster Award**) (Top conference in CS, Acceptance rate 22.4%). CORE Conference Ranking A
- [C21] *EMNLP'25* Jiang, C., Yu, F., **Chen, H.**, Ying, D., Lu, W., Zeng, J. (2025). TABDSR: Decompose, Sanitize, and Reason for Complex Numerical Reasoning in Tabular Data (Findings). *The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025)*. In press. ACL. (Top conference in CS, Acceptance rate 39%). CORE Conference Ranking A
- [C20] *ASIS&T'25* Li, Y., Mandalaju, T.#, **Chen, H.** Exploring Public Perceptions of Generative AI in Libraries: A Social Media Analysis of X Discussions (Long paper). ASIST 2025. Accepted.
- [C19] *JCDL'24* Ying, D., Yu, F., **Chen, H.**, & Lu, W. (2024). Fine-Grained, Accurate Data Generation and Multimodal Layout Analysis for Academic Papers (Long paper). *ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2024*. ACM. (Top conference in LIS, Acceptance rate 25%). Web of Science-indexed
- [C18] *ACM MM'24* Ying, D., Yu, F., **Chen, H.**, & Lu, W. (2024). DIG: Complex Layout Document Image Generation with Authentic-looking Text for Enhancing Layout Analysis (Long paper). In Proceedings of the 32nd ACM International Conference on Multimedia (pp. 3239-3247). ACM. (Top conference in CS, Acceptance rate 22%). Scopus-indexed
- [C17] *IEEE AITest'24* Zhou, Y. #, Tu, F. #, Sha, K., Ding, J., & **Chen, H.*** (2024). A Survey on Data Quality Dimensions and Tools for Machine Learning (Invited paper). In Proceedings of the 2024 IEEE International Conference on Artificial Intelligence Testing (AITest 2024) (pp. 120-131). IEEE. (**Google Scholar Citation: 40**). Ei-indexed
- [C16] *IEEE AITest'24* Ding, J., Nguyen, H. #, & **Chen, H.** (2024). Evaluation of Question-Answering Based Text Summarization using LLM (Invited paper). In Proceedings of the 2024 IEEE International Conference on Artificial Intelligence Testing (AITest 2024) (pp. 142-149). IEEE. Ei-indexed
- [C15] *ASIS&T'1 24* **Chen, H.***, Cherukuri, K#, Zhu, X., & Yang, S. Are Prompts All You Need?: Chatting with ChatGPT on Disinformation Policy Understanding (Short paper). Proceedings of the Association for Information Science and Technology, 61(1), 488-492. Location and date: Calgary, Canada, October 25-29, 2024. Web of Science-indexed

- [C14] *IEEE QRS'23* Ding, J., Chen, H., Kolapudi, S. #, Pobbathi, L. #, & Nguyen, H#. Quality Evaluation of Summarization Models for Patent Documents (Short paper). In 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security (QRS) (pp. 250-259). IEEE. Location and date: Chiang Mai, Thailand, October 22-26, 2023. Web of Science-indexed
- [C13] *IEEE AITest'23* Nguyen, H.#, Chen, H.*, Maganti, R.#, Hossain, T., & Ding, J. Identifying High-quality Informative Comments for Software Review Summarization (Short paper). *The 5th IEEE International Conference on Artificial Intelligence Testing*. In 2023 IEEE International Conference on Artificial Intelligence Testing (AITest) (pp. 146-151). IEEE. Ei-indexed
- [C12] *IEEE AITest'23* Zhao, H., Chen, H., & Yoon, H. Enhancing Text Classification Models with Generative AI-aided Data Augmentation (Regular paper). *The 5th IEEE International Conference on Artificial Intelligence Testing*. In 2023 IEEE International Conference on Artificial Intelligence Testing (AITest) (pp. 138-145). IEEE. **(Best Student Paper Award)**. Ei-indexed
- [C11] *IEEE AITest'23* Kargozari, K.#, Ding, J., Chen, H.*, Evaluating the Impact of Incentive/Non-incentive Reviews on Customer Decision-making (Regular paper). *The 5th IEEE International Conference on Artificial Intelligence Testing*. In 2023 IEEE International Conference on Artificial Intelligence Testing (AITest) (pp. 160-168). IEEE. **(Best Paper Award)**. Ei-indexed
- [C10] *IEEE COMPSAC'23* Feng, Y., Vanam, S., Cherukupally, M., Zheng, W., Qiu, M., & Chen, H. Investigating Code Generation Performance of ChatGPT with Crowdsourcing Social Data (Regular paper). In *Proceedings of the 47th IEEE Computer Software and Applications Conference* (pp. 1-10). **(Best Paper Award Nominate under the SETA Track)**. (Top conference in CS, Acceptance rate 24%, **Google Citation: 191, top 1 cited article in COMPSAC'23 Proceedings**). Web of Science-indexed
- [C9] *iConference'23* Nguyen, H.#, Oladapo, L., Ali, I., Chen, H., & Chen, J. Fighting Misinformation: Where Are We and Where to Go? In *International Conference on Information* (pp. 371-394). Springer, Cham. (Acceptance rate 36%)
- [C8] *ISSI'21* Chen, H.*, & Kanuboddu, B. N.# (2021). A fine-grained annotation scheme for research contribution in academic literature. In *Proceedings of the 18th International Conference on Scientometrics and Informetrics* (pp. 241-248). Online (Among 48/300 selected for journal publishing). Web of Science-indexed
- [C7] *ASIS&T'21* Chong, M., & Chen, H.* (2021). Racist Framing through Stigmatized Naming: A Topical and Geo-locational Analysis of# Chinavirus and# Chinesevirus on Twitter. In *Proceedings of the association for information science and technology*, 58(1), 70-79. Web of Science-indexed
- [C6] *IEEE QRS'20* Chen, H., Chen, J., & Ding, J. (2020). Data Evaluation and Enhancement for Quality Improvement of Machine Learning. *2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS)*, 13-13. **(Best Paper Award)**. Online (Acceptance rate 28.8%). Web of Science-indexed
- [C5] *CCKS'19* Chen, H., Cao, G., Chen, J., & Ding, J. (2019, August). A practical framework for evaluating the quality of knowledge graph. In *China conference on knowledge graph and semantic computing* (pp. 111-122). Springer, Singapore. (Among 18/140 selected for conference proceedings publishing, **Google Citation: 82**)
- [C4] *IEEE BIBE'18* Ding, J., Jin, W., & Chen, H. (2018, October). Regression-Based Documents Reranking for Precision Medicine. In *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 283-286). IEEE. (Acceptance rate 30%). Web of Science-indexed
- [C3] *ICSH'18* Chen, H., Ding, J., Chen, J., & Cao, G. (2018, July). Designing a novel framework for precision medicine information retrieval. In *International Conference on Smart Health* (pp. 167-178). Springer, Cham. (Among 14/49 selected for conference proceedings publishing)
- [C2] *ICSH'18* Chen, J., Chen, M., Qu, J., Chen, H., & Ding, J. (2018, July). Smart and connected health projects: Characteristics and research challenges. In *International Conference on Smart Health* (pp. 154-164). (Among 14/49 selected for conference proceedings publishing)
- [C1] *ICKM'17* Zhang, Q., Lu, W., Yang, Y., Chen, H., & Chen, J. (2017). Automatic identification of research articles containing data usage statements. In *Knowledge Discovery and Data Design Innovation: Proceedings of the International Conference on Knowledge Management (ICKM 2017)* (pp. 67-87).

Editorials

- [E3] TEL Kim, J., **Chen, H.**, & Bloechle, M. (2025). A roadmap to ethical research. *The Electronic Library*, 43(2), 97-103.
- [E2] TEL Wang, Z., **Chen, H.***, Zhang, C., Lu, W., & Wu, J. (2024). Innovation measurement for scientific communication (IMSC) in the era of big data. *The Electronic Library*, 42(6), 849-853.
- [E1] *Electronics* Ding, J., **Chen, H.***, Feng, Y., & Hossain, T. (2024). Applications of Deep Learning Techniques. *Electronics*, 13(17), 3354.

Workshops, Panels, Posters, Tutorials, Abstracts, Blogs

- [W33] *ISLS'26* Yu, J., Tu, F., #, **Chen, H.**, Ding, J., Liu, D., Hsieh, C., Kim, H., & Watson, S. Measuring Cognitive Presence in Online Discussions: Interpretable AI and Instructional Insights from a MOOC Context. *International Society of the Learning Sciences Annual Meeting 2026*. In-press.
- [W32] *JCDL'25* Chong, M., Ding, J., Shah, C., **Chen, H.**, Application of GAI to Empirical Studies: It's High Time to Discuss Challenges and Pitfalls (Panel). *The 25th ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE.
- [W31] *SCC/MLA'25* Mantecon, H., Cleveland, A., Sharma, S., **Chen, H.** Understanding Perceptions, Barriers, and Potential of Virtual Reality in the Mental Health Landscape: Graduate Student Perspectives. *2025 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Little Rock, USA, October 24-29, 2025.
- [W30] *SCC/MLA'25* Donepudi, S., **Chen, H.***, Cleveland, A., Parimi, S., Potula, R. Semi-automatic Dental Taxonomy Construction with Human and Generative AI Collaboration. *2025 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Little Rock, USA, October 24-29, 2025.
- [W29] *SCC/MLA'25* Zhou, Y., **Chen, H.***, Cleveland, A. Enhancing extreme multi-label medical concept normalization with LLM-based data fusion and quality improvement. *2025 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Little Rock, USA, October 24-29, 2025.
- [W28] *SCC/MLA'25* Challa, L., **Chen, H.***, Cleveland, A. HealthAdvisor: A Retrieval-Augmented AI Assistant for Accurate and Trustworthy Health Response Generation. *2025 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Little Rock, USA, October 24-29, 2025.
- [W27] *SCC/MLA'25* Challa, L., **Chen, H.***, Cleveland, A., Patel, D., Vinnakota, J. Evaluating Prompting Strategies for Generating Structured Lung Cancer Staging Notes Using Large Language Models. *2025 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Little Rock, USA, October 24-29, 2025.
- [W26] *ASIST'25* Li, Y., **Chen, H.**, Lund, B., Ma, R., Yang, L., & Miriam, S. Libraries in the Age of LLMs: Perceptions, Practices, and the Future of Scholarly Work (Panel). Accepted.
- [W25] *SCC/MLA'24*  **Chen, H.***, Li, R., Cleveland, A., & Ding, J. Evaluating ChatGPT-based Data Augmentation for Medical Concept Normalization (Abstract + Oral presentation, **Best Paper Award 3rd place**). *2024 South Central Chapter of the Medical Library Association (SCC/MLA)*. Location and date: Dalla, USA, October 27-29, 2024
- [W24] *JCDL'24* Zhang, L., Wu, K., Cheng, X., Wang, J., & **Chen, H.*** (2024, December). LSD: An Effective Method for Improving Quality of Scholarly Databases (Poster). *ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2024*, accepted. ACM. Location and date: Hong Kong, December 16 - 20, 2024
- [W23] *JCDL'24* Lu, W., Huang, J., Yu, F., & **Chen, H.** (2024, December). JCDL 2024 Tutorial on Academic Table and Figure Understanding for Digital Library (Tutorial). *In 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, in press. IEEE. Location and date: Hong Kong, December 16 - 20, 2024
- [W22] *JCDL'24* Wang, Z., **Chen, H.***, Zhang, C., Bu, Y., Lu, W., & Wu, J. (2024, December). JCDL2024Workshop: The 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (Workshop proposal). *In 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, in press. IEEE. Location and date: Hong Kong, December 16 - 20, 2024

- [W21] JCDL'24 **Chen, H.***, Kim, J., Wang, X., Yang, L., & Fremery, W. (2024, December). JCDL2024 Workshop: Utilizing AI/ML to Enhance Information Extraction, Organization, and Retrieval from Large-scale Archival Collections (Workshop proposal). In *2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, in press. IEEE. Location and date: Hong Kong, December 16 - 20, 2024
- [W20] ASIS&T'24 Cherukuri, K#, **Chen, H.*** Investigating the Applications of AI in Oncology from NIH Funded Projects: Case study of Lung Cancer and Pancreatic Cancer (Poster). In proceedings of the 87th Annual Meeting of the American Society for Information, Science & Technology (ASIS&T 2024). In press. Location and date: Calgary, Canada, October 25-29, 2024.
- [W19] ASIS&T'24 Kim., J., **Chen, H.**, Yang, L., & Simic, J. Exploring the Application of Artificial Intelligence and Machine Learning in GLAM Collections (Panel). In proceedings of the 87th Annual Meeting of the American Society for Information, Science & Technology (ASIS&T 2024). In press. Location and date: Calgary, Canada, October 25-29, 2024.
- [W18] DGO'24 Zhang, X., Chong, M., Hagen, L., & **Chen, H.***. A Framework for Assessing Country Reputation: Case Study of China during the COVID-19 Pandemic (Poster). DGO '24: Proceedings of the 25th Annual International Conference on Digital Government Research (pp. 1008-1010). ACM. Location and date: Taipei, Taiwan, June 11-14, 2024
- [W17] SCC/MLA'23 Tran, N.#, **Chen, H.**, & Cleveland, A. D. Unveiling Key Research Themes in Disaster Informatics Amidst the Pandemic (Abstract + Oral presentation). *2023 South Central Chapter of the Medical Library Association (SCC/MLA)*. In press. Location and date: New Orleans, USA, September 21-26, 2023
- [W16] SCC/MLA'23 Tran, N.#, Cleveland, A. D., & **Chen, H.** Exploring the Shift in Disaster Informatics Research After the Pandemic (Abstract + Oral presentation). *2023 South Central Chapter of the Medical Library Association (SCC/MLA)*. In press. Location and date: New Orleans, USA, September 21-26, 2023
- [W15] SCC/MLA'23 Cleveland, A. D., & **Chen, H.** Dental Informatics over time: A 20-years longitudinal bibliometrics analysis (Abstract + Oral presentation). *2023 South Central Chapter of the Medical Library Association (SCC/MLA)*. In press. Location and date: New Orleans, USA, September 21-26, 2023
- [W14] JCDL'23 Wang, Z., **Chen, H.***, Zhang, C., Lu, W., & Wu, J. (2023, June). JCDL2023 Workshop: Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (Workshop proposal). In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 303-305). IEEE. Location and date: Santa Fe, NM, USA, June 26 - 30, 2023
- [W13] JCDL'23 Zhang, C., Zhang, Y., Mayr, P., Lu, W., Suominen, A., **Chen, H.**, & Ding, Y. (2023, June). JCDL2023 Workshop: Joint Workshop of the 4th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2023) and the 3rd AI+ Informetrics (AII2023) (Workshop proposal). In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 306-307). IEEE. Location and date: Santa Fe, NM, USA, June 26 - 30, 2023
- [W12] AEJMC Midwinter Conference'23 Zhang, X., & **Chen, H.***. Assessing the Change of Country Reputation during the COVID-19 Pandemic Period on Twitter: A Case Study of China (Abstract + Oral presentation). *The Association for Education in Journalism and Mass Communication Midwinter Conference 2023*. Location and date: Norman, OK, USA, February 24-25, 2023
- [W11] Information Matters **Chen, H.** (2022, September 23). Why does data quality matter? How to evaluate and improve data quality for machine learning systems (Blog). *Information Matters*, 2(9).
- [W10] EEKE@JCDL'22 Wang, Z., Chen, J., Chen, J., & **Chen, H.*** (2022). Interdisciplinary Topics Extraction and Evolution Analysis (Workshop paper). In *Proceedings of 3rd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2022)*, June 20-24, 2022, Cologne, Germany and Online.
- [W9] ASIS&T'22 Han, Y., Maganti, R.#, Chen, J., & **Chen, H.*** (2022). Uncover Marginalized Narratives of Japanese American Incarceration: An Annotation Scheme for Natural Language Processing and Data Analytics (Poster). *Proceedings of the Association for Information Science and Technology*, 59(1), 698-700. Location and date: Pittsburgh, PA, USA, October 29-November 1, 2022

- [W8] *ASIS&T'21* You, J.#, Shaik, N.#, & **Chen, H.*** (2021). Data Mining on COVID - 19 Vaccines: Side Effects (Poster). *Proceedings of the Association for Information Science and Technology*, 58(1), 869-871. Location and date: Salt Lake City, UT, USA, October 30 - November 2, 2021
- [W7] *TREC'21* Nguyen, H.#, **Chen, H.**, Prasad, B., Zhao, H., Ding, J., Chen, J., & Cleveland, A. (2021). UNTIIA Lab at TREC 2021 - Clinical Trial. In *TREC*. Online
- [W6] *IEEE Big Data'20* Tang, M., Su, C., **Chen, H.**, Qu, J., & Ding, J. (2020, December). SALKG: A Semantic Annotation System for Building a High-quality Legal Knowledge Graph (Workshop paper). In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 2153-2159). IEEE. Online
- [W5] *ASIS&T'20* Tang, M., Chen, J., & **Chen, H.** (2020). Smart bookshelf for library book management (Poster). *Proceedings of the Association for Information Science and Technology*, 57(1), e396. Online
- [W4] *IEEE QRS'20* Luo, M., Yu, F., & **Chen, H.*** (2020, December). Result Diversification in Clinical Case Reports Retrieval based on Main Finding (Workshop paper). In *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (pp. 213-220). IEEE. Online
- [W3] *IEEE QRS'20* Tang, M., Chen, J., & **Chen, H.** (2020, December). SemOIR: an ontology-based semantic information retrieval system (Workshop paper). In *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (pp. 204-208). IEEE. Online
- [W2] *JCDL'18* Yang, Y., **Chen, H.***, Lu, W., & Ayala, B. R. (2018, May). Diversifying citation contexts in academic literature for knowledge recommendation (Poster). In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 397-398). Location and date: Fort Worth, TX, USA, June 3 - 6, 2018
- [W1] *TREC'17* Viswavarapu, L. K., Chen, J., Cleveland, A. D., & **Chen, H.** (2017). UNT Precision Medicine Information Retrieval at TREC 2017. In *TREC*. Location and date: Gaithersburg, MD, USA, November 15--17, 2017

Research Grants and Proposals

External funded/pending grants:

- Title:** LB21: ALiT: AI Platform for Librarian Lifelong Training. **Role:** PI. **Funder:** Institute of Museum and Library Services. **Duration:** 2026-2029. **Amount:** \$382,268. Pending.
- Title:** NLGL: ORAL-AI: High Efficacy Content Representation and User Engagement of Oral Digital Collections with Multimodal LLMs. **Role:** PI. **Funder:** Institute of Museum and Library Services. **Duration:** 2026-2029. **Amount:** \$490,396. Pending.
- Title:** RET Site: Teacher-Researcher Co-Design of Adaptive GenAI for Computational Thinking. **Role:** Co-PI. **Funder:** National Science Foundation. **Duration:** 2026-2029. **Amount:** \$599,496. Pending.
- Title:** REU Site: Making Generative Artificial Intelligence Responsible. **Role:** Senior Personnel. **Funder:** National Science Foundation. **Duration:** 2025-2028. **Award Amount:** \$463,434.
- Title:** HSI Implementation and Evaluation Project: Developing a High-Quality Academic Environment for Broadening Participation of Hispanic Students in Computing. **Role:** Co-PI. **Funder:** National Science Foundation. **Duration:** 2022-2025. **Award Amount:** \$499,608.
- Title:** REU Site: Beyond Language: Training to Create and Share Vector Embeddings across Applications. **Role:** Senior Personnel. **Funder:** National Science Foundation. **Duration:** 2023-2025. **Award Amount:** \$403,547.
- Title:** Building a knowledge-driven Legal IR, Automatic Evidence and Argument Generation System for Decision Making. **Role:** PI. **Funder:** Google Cloud. **Duration:** 2021. **Award Amount:** \$5,000.
- Title:** Demystifying COVID-19 publications. **Role:** PI. **Funder:** Google Cloud. **Duration:** 2020. **Award Amount:** \$13,800.

External not funded grants:

- Title:** Beginnings: EMPACT AI: Empowering Minority-Serving Community College Students through Experiential Micro-Courses and Industry-Aligned Projects in Artificial Intelligence. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2025. Not funded.
- Title:** CISE MSI: RDP: HCC: LLM-Driven Healthcare Support System for Rural Texas: Empowering Nurse Practitioners with AI. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2025. Not funded.

3. **Title:** *LB 21: Utilizing AI/ML to Enhance Public Engagement with Large-scale and Multi-modal GLAM Collections: Model Development and Best Practices*. **Role:** PI. **Funder:** Institute of Museum and Library Services. **Proposal submission year:** 2025. Not funded.
4. **Title:** *NLG Planning: Unleashing the Potential of Generative AI for Bridging the Language Barriers in Library Services*. **Role:** PI. **Funder:** Institute of Museum and Library Services. **Proposal submission year:** 2025. Not funded.
5. **Title:** *PARTNER: Expanding Edge AI for Next-Generation Connected Autonomous Systems: A UNT-Athena Collaboration*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2024. Not funded.
6. **Title:** *IUSE:EDU: Building an Engaged Learning Environment for Monitoring, Guiding, and Accessing in STEM*. **Role:** Senior Personnel. **Funder:** National Science Foundation. **Proposal submission year:** 2024. Not funded.
7. **Title:** *RET Site: Enhancing Teachers' Research Skills in Applied Generative AI*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2024. Not funded.
8. **Title:** *NIH RO1 "Advancing Pancreatic Cancer Research and Quality of Life in High-risk Underserved Pancreatic Cancer Populations using Digital Health Tools, Immersive Technologies, and Data Science"*. **Role:** Co-PI. **Funder:** National Institutes of Health. **Proposal submission year:** 2024. Not funded.
9. **Title:** *RITEL: AI Mastery: Goal-Driven Learning Enhanced with Real-time Performance Analysis for Mastering AI Fundamentals and Practices*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2024. Not funded.
10. **Title:** *NRT-AI: Train Next-Generation Scientists and Engineers in Data-Centric Intelligence*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2024. Not funded.
11. **Title:** *AIM-AHEAD: Utilizing AI/ML to Enhance Personalized Health Information Services for Hispanic Populations during Disaster Recovery*. **Role:** PI. **Funder:** National Institutes of Health. **Proposal submission year:** 2023. Not funded.
12. **Title:** *Build and Broaden (B2): Collaborative Research: Build and Broaden: Building Research Capacity in Health Informatics Research Partnerships*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2023. Not funded.
13. **Title:** *I/UCRC Planning: Center for Emerging System Dependability Assurance (CESDA)*. **Role:** Co-PI. **Funder:** National Science Foundation. **Proposal submission year:** 2023. Not funded.

Internal funded/pending grants:

1. 2026. Co-PI. UNT Learning Institute Funding "Supporting Student Motivation in Data Science Education: A Self-Determination Theory Based Curriculum Redesign with AI-Enhanced Learning Partnership", University of North Texas. **Duration:** 2026-2027. **Amount:** \$18,065. Pending.
2. 2026. Co-PI. UNT Learning Institute Funding "An Agentic AI Simulation for Epistemic Resilience in Cybersecurity Education", University of North Texas. **Duration:** 2026-2027. **Amount:** \$16,000. Pending.
3. 2025. Co-PI. UNT Learning Institute Funding "Partner, Not Crutch: Designing a Metacognitive Nudge to Promote AI Co-Regulation", University of North Texas. **Duration:** 2026-2027. **Amount:** \$5,000.
4. 2025. PI. UNT Collaborative Research/Creative Scholarly Projects Abroad, "Embodied & Multimodal AI for Cross-Cultural Access to Digital Archives: A UNT-Osaka Seed Collaboration", University of North Texas. **Duration:** 2026. **Amount:** \$5,000.
5. 2024. PI. UNT COI Seed Funding "Utilizing AI/ML to Enhance Personalized Health Information Services for Hispanic Populations during Disaster Recovery", College of Information, University of North Texas. **Duration:** 2024-2025. **Award Amount:** \$5,000.
6. 2022. PI. UNT COI Seed Funding "Towards a large-scale and high-quality corpus for legal argument mining", College of Information, University of North Texas. **Duration:** 2022-2023. **Award Amount:** \$9,975.

Internal not funded grants:

1. 2025. Co-PI. UNT Learning Institute Funding "AI-Enhanced Communication Training for Instructors: A Generative AI Approach to Student-Centered Pedagogical Dialogue", University of North Texas. **Duration:** 2026-2027. **Amount:** \$20,000. Not funded.
2. 2023. PI. UNT VPRI Intramural Research Seed Grant (RSG) "Utilizing AI/ML to Enhance Personalized Health Information Services for Hispanic Populations during Disaster Recovery". Request: \$10,000. Not funded.
3. 2022. Co-PI. The UNT VPRI Intramural Research Seed Grant (RSG) "Assessing Country Reputation during the COVID-19 Pandemic Period on Twitter: A Case Study of China". Request: \$10,000. Not funded.

Travel Fundings

- 2026. NSF UNIDOS INSPIRA Innovation Sandbox. Award Amount: \$1,200.
- 2025. NSF CISE MSI Aspiring PI Workshop. Award Amount: \$1,200.
- 2025. UNT Department of Data Science Travel Grant (IEEE COSOSE 2025). Award Amount: \$1,100.
- 2024. UNT Department of Information Science Travel Grant (ASIS&T 2024). Award Amount: \$3,200.
- 2024. UNT Department of Information Science Travel Grant (MLA 2024). Award Amount: \$2,200.
- 2023. UNT College of Information and Department of Information Science (SCC/MLA 2023). Award Amount: \$2,000.
- 2023. UNT Department of Information Science Travel Grant (JCDL 2023). Award Amount: \$2,000.
- 2022. UNT Department of Information Science Travel Grant (ASIS&T 2022). Award Amount: \$1,000.
- 2022. UNT Department of Information Science Travel Grant (IEEE CISOSE 2022). Award Amount: \$1,200.
- 2021. UNT Travel Grant (ASIS&T 2021). Award Amount: \$500.
- 2021. UNT Department of Information Science Travel Grant (ASIS&T 2021). Award Amount: \$1,000.
- 2021. UNT Department of Information Science Travel Grant (TREC 2021). Award Amount: \$35.
- 2021. The International Society for Scientometrics and Informetrics (ISSI) Student Travel Award. Award Amount: €115.
- 2020. UNT Department of Information Science Travel Grant (IEEE QRS 2020). Award Amount: \$1,500.
- 2019. UNT Department of Information Science Travel Grant (WWW 2019). Award Amount: \$600.
- 2018. UNT Travel Grant (TREC 2018). Award Amount: \$1,200.
- 2018. UNT Department of Information Science Travel Grant (ALA 2018). Award Amount: \$800.
- 2018. UNT Department of Information Science Travel Grant (IEEE/ ACM JCDL 2018). Award Amount: \$1,000.

Honors and Awards

- 2026. The Dallas Innovates AI75, Dallas Innovates and Dallas Regional Chamber.
- 2025. Rocking Star of Research and Innovation, Department of Data Science, University of North Texas.
- 2025. Best Poster Award, The 2025 IEEE International Conference on Data Mining (ICDM2025), Washington DC, USA.
- 2025. Best Reviewer Award, The 2025 IEEE Technical Community on Learning Technology (TCLT), Changhua, Taiwan.
- 2022. Great Grads Spring 2022, College of Information, University of North Texas.
- 2021 - 2022. UNT COVID-19 Student Success Award (Spring 2022, Summer 2021, Fall 2021), University of North Texas. Award Amount: \$1,500.
- 2021. UNT Toulouse Graduate School Summer Award, University of North Texas. Award Amount: \$1,605.
- 2021. Linda Schamber Writing Award for paper "*Topic sentiments toward the COVID-19 vaccine on Twitter*", College of Information, University of North Texas. Award Amount: \$250.
- 2021. The Melba S. Harvill Endowed Scholarship, Department of Information Science, University of North Texas. Award Amount: \$1,000.
- 2020. The Dewey E. Carroll Graduate Fellowship Award, Department of Information Science, University of North Texas. Award Amount: \$4,000.
- 2020. The Donald B. and Ana D. Cleveland Medical Informatics Endowed scholarship, Department of Information Science, University of North Texas. Award Amount: \$1,000.
- 2019. The Mark E. Rorvig Endowed Graduate Fellowship Award, Department of Information Science, University of North Texas. Award Amount: \$4,000.
- 2019. Outstanding Reviewer for The Electronic Library, Emerald.
- 2019. The Donald B. and Ana D. Cleveland Medical Informatics Endowed scholarship, Department of Information Science, University of North Texas. Award Amount: \$1,000.
- 2018. The Howard Griesdorf Interdisciplinary Ph.D. Award, Department of Information Science, University of North Texas. Award Amount: \$1,000.
- 2017. Honor Graduates, Excellent Student Leader Award, Wuhan University.

Teaching

Instructor, University of North Texas

- DTSC 3020: Introduction to Computation with Python (Fall 2025, Summer 2026)
- INFO 5506: Applications of Artificial Intelligence in Health (Spring 2024, 2025, 2026)
- INFO 5810: Data Analysis and Knowledge Discovery (Summer 2022, Fall 2022, Spring 2023, Summer 2024, Summer 2025)
- INFO 5731: Computational Methods for Information Systems (Spring 2020 – Spring 2026)
- INFO 5502: Analytic Tools, Techniques and Methods (Spring 2022, Summer 2022, Summer 2024)
- INFO 4206: Information Retrieval (Spring 2019)

Course Designer and Instructor, University of North Texas

- INFO 5506: Applications of Artificial Intelligence in Health (Spring 2024)
- CSCI 4290/5290: Introduction to Natural Language Processing (CIB) (Fall 2022)

Co-Instructor, University of North Texas

- INFO 5717: Networked Data Modeling and Processing (Fall 2019)

Guest Lecturer, University of North Texas

- LSCM 3960: Logistics and Supply Chain Management – Topic: The Applications of Machine Learning and Natural Language Processing for Supply Chain Management (Spring 2022)
- INFO 5709: Data Visualization and Communication – Topic: Deep Learning with TensorFlow and Keras (Summer 2020)
- INFO 3020: Introduction to Computation with Python – Topic: Web Scraping with Python (Summer 2020)
- INFO 5707: Data Modeling for Information Professionals – Topic: Construction and Applications of Knowledge Graphs (Fall 2019)
- INFO 5810: Data Analysis and Knowledge Discovery – Topic: Web Scraping with Python (Fall 2019)
- INFO 5501: Fundamentals of Data Analytics – Topic: Introduction to Python Tools and Packages for Data Analysis and Scientific Computing (Spring 2018)

Teaching Assistant, University of North Texas

- INFO 5717: Dynamic WWW Structures (Fall 2018)

Student Advising

-
- Current students
 - Mehri Sattari, Ph.D. student in Information Science (Fall 2025-present, as major professor)
 - Yuhan Zhou, Ph.D. student in Information Science (Fall 2024-present, as major professor)
 - Haoxuan Zhang, Ph.D. student in Information Science (Fall 2024-present, as co-major professor)
 - Fengjiao Tu, Ph.D. student in Information Science (Fall 2023-present, as major professor)
 - Laxmigayathri Challa, Ph.D. student in Information Science (Spring 2025-present, as major professor)
 - Rafid Ishrak Jahan, Ph.D. candidate in Computer Science (Fall 2025-present, as committee member)
 - Daniel Agbaji, Ph.D. candidate in Information Science (Fall 2023-present, as committee member)
 - Mimi Byun, Ph.D. candidate in Information Science (Summer 2023-present, as committee member)
 - Lavinia Florentina, Ph.D. student in Information Science (Fall 2022-present, as faculty mentor)
 - Natalie Tran, Ph.D. candidate in Information Science (Spring 2022-present, as faculty mentor)
 - Kate Kargozari, Ph.D. candidate in Information Science (Fall 2020-present, as faculty mentor)
 - Komala Subramanyam Cherukuri, was master student in Data Science, now Ph.D. student in Information Science (Spring 2024-present, as faculty mentor)
 - Sai Donepudi, undergraduate student in Data Science (Fall 2024-present)
 - Serene Wang, Texas Academy of Mathematics and Science (Spring 2025-present)
 - Master thesis committee

- 2025. Sina Montazeri, M.S. in Computer Science. Thesis title: A Framework for Discovery and Analysis of Public Data Using Large Language Model Agents for Non-Expert Users.
- 2024. Rafid Ishrak Jahan, M.S. in Computer Science. Thesis title: Unlocking Cross-Lingual Sentiment Analysis through Emoji Interpretation: A Multimodal Generative AI Approach.
- Alumni
 - Huyen T. N. Nguyen, Ph.D. in Information Science (Summer 2021- Spring 2025, as faculty mentor, Graduated in Spring 2025)
 - Murthy Aryan, master student in Information Science (Graduated in Spring 2025)
 - Aisa Sakata, master student in Information Science (Graduated in Fall 2024)
 - Lavanya Pobbathi, master student in Data Science (Graduated in Fall 2024)
 - Kaushal Sen, master student in Data Science (Graduated in Spring 2024)
 - Suruthi Selvam, master student in Data Science (Spring 2023-Summer 2023)
 - Kavya Vedantham, master student in Data Science (Summer 2023)
 - Mounica Siri Chandana Tamalampudi, master student in Data Science (Summer 2023)
 - Solomon Daramay, undergraduate in Computer Science at UNT (Under the NSF REU program at UTD, Summer 2023)
 - Aparna Bhooshanan, undergraduate in Computer Science at The University of Alabama (Under the NSF REU program at UNT, Summer 2023)
 - Chimara Okeke, undergraduate in Bioinformatics at UNT (Under the NSF REU program at UNT, Summer 2023)
 - Ella Ghazarian, undergraduate in Computer Science at UNT (Under the NSF REU program at UNT, Summer 2023)
 - Max Jing, undergraduate in Statistics at University of Illinois Urbana-Champaign (Summer 2023)
 - Karthik Puppala, master student in Data Science (Spring 2023)
 - Amith Reddy Kasarla, master student in Data Science (Graduated in Spring 2023, now working as a Reporting Analyst III at BAL)
 - Revathi Bejawada, master student in Data Science (Graduated in Spring 2023, now working as a Software Engineer at Source Infotech Inc)
 - Roopesh Maganti, master student in Data Science (Graduated in Spring 2023, now working as a Data Management Engineer with Comcast)
 - Bhanu Prasad G, master student in Data Science (Graduated in Spring 2022, now working as a Software Engineer at T-mobile)
 - Akshitha Maddi, Master student in Data Science (Graduated in Spring 2021, now working as a Business Analyst at Barclays)
 - Sai Supriya Vadla, Master student in Data Science (Graduated in Spring 2021, now working as an Associate Engineer at Continental)
 - Nagma Shaik, Master student in Data Science (Graduated in Spring 2021, now working as a Software Engineer III at SiriusXM & Pandora)
 - Raheyima Khan, Master student in Data Science (Graduated in Spring 2021, now working as a Market Analyst at Kemper)
 - Vinati Gattupalli, Master student in Data Science (Graduated in Spring 2021, now working as a Data Scientist at Bluestem Brands)

Invited Talk and Presentations

- Talk Title: Deep Learning and Large Language Models for Software Engineering and Testing, Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, July 11, 2024.
- Talk Title: Data Evaluation and Improvement for Machine Learning Systems, Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, April 19, 2024.
- Talk Title: Data science for legal intelligence, scientific communication, and informatics: from model construction to data quality, Zhongnan University of Economics and Law (Online), Jun 2, 2023
- Talk Title: Informatics applications in multiple fields: A bibliometrics analysis, Department of Information Science, University of North Texas, Nov 16, 2022

- Talk Title: Measuring the novelty of scientific literature through contribution sentences analysis, Information Processing & Management Conference 2022 (Online), Oct 22, 2022
- Talk Title: Fine-grained semantic analysis for academic data mining and evaluation, National Science Library, University of Chinese Academic of Science (Online), May 31, 2022
- Talk Title: The applications of machine learning and natural language processing in healthcare and medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong (Online), Mar 3, 2022
- Talk Title: Building high-performance machine learning systems for domain-specific applications, Department of Information Science, University of North Texas, Dec 21, 2021
- Talk Title: Construction and evaluation of high-quality corpus for legal intelligence using semi-automated approaches, Department of Information Science, University of North Texas, Nov 17, 2021
- Talk Title: Building high-performance machine learning systems for domain-specific applications, Sun Yat-Sen University (Online), Oct 24, 2021
- Talk Title: Information retrieval in 40 years: a bibliometric analysis of SIGIR proceedings, Department of Information Science, University of North Texas, Sep 1, 2021
- Talk Title: Information quality improvement and evaluation for building high performance machine learning systems, Department of Information Science, University of North Texas, Oct 7, 2020
- Talk Title: Information quality improvement and evaluation for building high performance machine learning systems, Wuhan University (Online), Jul 9, 2020
- Talk Title: Knowledge graph construction, application and evaluation, Beihua University (Online), May 19, 2020
- Talk Title: Construction and applications of knowledge graphs, Department of Information Science, University of North Texas, Nov 6, 2019
- Talk Title: Research experience sharing in 2019 new Ph.D. student orientation, Department of Information Science, University of North Texas, Aug 24, 2019
- Talk Title: Diversifying the citation contexts in academic literature for knowledge recommendation, Central China Normal University, Jan 10, 2019

Academic Service

- Leadership:
 - Associate editor of The Electrical Library (IF = 2.24, SSCI index, JCR Q2 in LIS) (2022-present)
 - Associate editor of Data Intelligence (IF = 1.9, ESCI index, JCR Q2 in LIS) (2023-present)
 - Publicity co-chair of Information and Software Technology (IF = 4.3, SCI index, JCR Q1 in Computer Science Applications) (2025-present)
 - Associate editor of Information Matters (A digital-only communication translational forum) (2025-present)
 - Chair, ASIS&T SIG-STI (Special Interest Group - Scientific and Technical Information) (2024-2025). Served as the Chair-elect from 2023-2024.
 - Symposium Co-Chairs of the ACM Conference on Trustworthy and Responsible AI and Computing Systems (ACM TRUST 2027).
 - Local Chair of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2026).
 - Award Committee of IEEE TCLT to judge the Early Career Researcher Award, Student Travel Award, Best Paper Award, and Best Reviewer Award of IEEE ICALT 2026 (August 2025 - July 2026).
 - Program Co-chair of the 8th IEEE International Conference on Artificial Intelligence Testing (IEEE AITest 2026).
 - Panel Chair of the 25th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2025).
 - Program Co-chair of the 7th IEEE International Conference on Artificial Intelligence Testing (IEEE AITest 2025).
 - Workshop and Tutorial Co-chair of the 24th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2024).
 - Program Co-chair of the 6th IEEE International Conference on Artificial Intelligence Testing (IEEE AITest 2024).
 - Local Chair of the 2nd IEEE Conference on Mobility: Operations, Services, and Technologies (IEEE MOST 2024).
 - Workshops and Round Table Chair of the 18th International ISKO Conference (ISKO International Conference 2024).

- Website Chair of the 5th IEEE International Conference on Artificial Intelligence Testing (IEEE AITest 2023).
- Publicity Chair of the 18th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2018).
- Editorial board member:
 - Information Processing and Management (IF = 7.4, top journal in LIS, 2025-present)
 - Frontiers in Research Metrics and Analytics (CiteScore = 3.5, 2023-2024)
 - Heliyon Information Science (IF = 3.4, 2022-2024)
- Guest editor:
 - The Electrical Library (IF = 2.42) special issue “Application of AI/ML in Large-scale Digital Archive Collections”, 2024 – 2025, ongoing. Link: <https://www.emeraldgrouppublishing.com/calls-for-papers/application-aiml-large-scale-digital-archive-collections>
 - Computer Standards & Interfaces (IF = 5.0) special issue “Applications of Generative AI”, 2024-2025, ongoing. Link: <https://www.sciencedirect.com/journal/computer-standards-and-interfaces/about/call-for-papers#applications-of-generative-ai>
 - Electronics (IF = 2.9) special issue “Intelligent Data Analysis and Learning”, 2024-2025, ongoing. Link: https://www.mdpi.com/journal/electronics/special_issues/8FU4PHTNC1
 - Electronics (IF = 2.9) special issue “Applications of Deep Learning Techniques”, published. Link: https://www.mdpi.com/journal/electronics/special_issues/82123FACP2
 - The Electrical Library (IF = 2.42) special issue “Innovation Measurement for Scientific Communication in the Era of Big Data”, 2023 – 2024, published. Link: <https://www.emerald.com/insight/publication/issn/0264-0473/vol/42/iss/6>
- Co-chair of Workshops:
 - MAI-DRS workshop@ IEEE BIBM 2026
 - DQIS workshop@ IEEE QRS 2021, 2022, 2023, 2024
 - EEKE+AI Informatics workshop@ JCDL 2023, iConference 2024, ISSI 2025
 - IMSC workshop@ JCDL 2023, 2024
 - AI4LCA workshop@ JCDL 2024
 - DQ4TAIS workshop@ IEEE/ACM CHASE 2025
 - DMAIL workshop@ IEEE ICDM 2025, IEEE Big Data 2026
- Journal reviewer (selected):
 - Scientometrics (2021-present)
 - Computer Standards & Interfaces (2021-present)
 - JAMIA Open (2021-present)
 - Journal of Computational Social Science (2022-present)
 - Aslib Journal of Information Management (2022-present)
 - Applied Intelligence (2022-present)
 - ACM Transactions on Asian and Low-Resource Language Information Processing (2022-present)
 - Artificial Intelligence Review (2022-present)
 - Information Processing and Management (2022-present)
 - Journal of Informetrics (2022-present)
 - Neural Processing Letters (2022-present)
 - Machine Learning (2022-present)
 - Computational Intelligence and Neuroscience (2022-present)
 - Complex & Intelligent Systems (2023-present)
 - Knowledge and Information Systems (2023-present)
 - Scientific Reports (2023-present)
 - Expert Systems with Applications (2023-present)
 - Journal of the Association for Information Science and Technology (2023-present)
 - Knowledge-Based Systems (2023-present)
 - IEEE Transactions on Reliability (2023-present)
 - IEEE Transactions on Big Data (2023-present)
 - Education and Information Technologies (2024-present)
 - IEEE Transactions on Artificial Intelligence (2024-present)

- IEEE Transactions on Computational Social Systems (2024-present)
- IEEE Transactions on Emerging Topics in Computational Intelligence (2024-present)
- IEEE Transactions on Knowledge and Data Engineering (2024-present)
- IEEE IEEE Transactions on Multimedia (2025-present)
- ACM Computing Surveys (2024-present)
- PC member and reviewer for Conferences and Workshops (selected):
 - IEEE ICDM (2025, 2026)
 - IEEE ICALT (2025)
 - EMNLP (2024, 2025, 2026)
 - ACL (2025)
 - IEEE COMPSAC (2020, 2021, 2022, 2023, 2024, 2025)
 - IEEE ICSC (2020, 2021, 2022, 2023, 2024, 2025)
 - IEEE Big Data (2019, 2020, 2021, 2022, 2023, 2024, 2025)
 - IEEE QRS (2020, 2021, 2022, 2023, 2024, 2025)
 - iConference (2021, 2023, 2024, 2025)
 - ACM/IEEE JCDL (2021, 2022, 2023, 2024, 2025)
 - ASIS&T (2021, 2023, 2024, 2025)
 - ALLDATA (2022, 2023, 2024, 2025)
 - IEEE AITest (2023, 2024, 2025)
 - IEEE DSA (2023, 2024)

Department Service

1/2025-Present	Executive Committee , Anuradha and Vikas Sinha Department of Data Science, University of North Texas
1/2025-Present	Chair of the Research and Outreach Committee , Anuradha and Vikas Sinha Department of Data Science, University of North Texas
1/2025-Present	Curriculum Committee , Anuradha and Vikas Sinha Department of Data Science, University of North Texas
10/2024-Present	Faculty Search Committee , Health Informatics Program (for tenure-track assistant/ associated professor), University of North Texas
09/2023-09/2024	Chair of the External Affairs and Physical Resources Committee , Department of Information Science, University of North Texas
09/2022-09/2024	Faculty Search Committee , Data Science Program (for assistant professor in practice), University of North Texas
09/2022-08/2023	Faculty Search Committee , Health Informatics Program (for tenure-track assistant professor), University of North Texas
09/2019-01/2022	Faculty-Staff-Student Council , Department of Information Science, University of North Texas

Professional Associations

- 2017-Present. Association for Computing Machinery (ACM)
- 2020-Present. Institute of Electrical and Electronics Engineers (IEEE)
- 2020-Present. ACM Special Interest Group on Information Retrieval (SIGIR)
- 2020-Present. Texas Computer Education Association (TCEA)
- 2020-Present. Association For Information Science and Technology (ASIS&T)
- 2021-Present. Association for the Advancement of Artificial Intelligence (AAAI)

