Haihua Chen, Ph.D.

Assistant Professor, Data Science (Joint Appointment with Health Informatics)

Director, Intelligent Information Access and Analytics Lab

Department of Information 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

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

Professional Experience

08/2023-present	**Assistant Professor (Tenure-track),** University of North Texas
01/2022-07/2023	**Clinical Assistant Professor,** University of North Texas
01/2019-01/2022	**Research Assistant & Teaching Fellow**, University of North Texas
05/2018-01/2022	**Lab Manager,** Intelligent Information Access Lab, University of North Texas
10/2020-08/2021	**Vice President,** Data Science Organization, University of North Texas
01/2018-01/2020	**Officer,** International Graduate Students Research Organization, University of North Texas
09/2017–01/2019	**Research Assistant & Teaching Assistant**, 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) 2023.

Articles In-progress

[1] *TBD* **Chen, H.,** Ding, J., A comprehensive survey on legal artificial intelligence: tasks, datasets, techniques, and future directions.

[2] *AITest 2024* Zhou, Y., Tu, F., Sha, K., Ding, J., & **Chen, H*.** A Survey on Data Quality for Machine Learning: Dimensions, Tools, and Future Directions.

[3] *ACM Computing Surveys* Tran, N.#, **Chen, H.,** & Cleveland, A. Disaster Informatics: Topics, Themes, and Future Direction. *ACM Computing Surveys.* (4/326 in Computer Science, JCR Q1)

[4] *JBI* **Chen, H*,** Illa, A. #, Li, R., & Ding., J. A Survey on Medical Concept Normalization: From Shallow to Deep Learning. *Journal of Biomedical Informatics.* (26/115 in Health Informatics, JCR Q1)

[5] *JOCCH* Sakata, A.# & **Chen, H***. JAIOH-Semantic: A High-quality Corpus for Understanding Japanese American Incarceration Oral History with NLP and Machine Learning. *Journal on Computing and Cultural Heritage.* (81/110 in Computer Science Interdisciplinary Applications, JCR Q3)

[6] *JBI* **Chen, H*,** Li, R., & Ding, J. Quality Evaluation of Data Augmentation with ChatGPT: Case Study on Medical Concept Normalization. *Journal of Biomedical Informatics.* (26/115 in Health Informatics, JCR Q1).

[7] *Information Fusion* Zhang L., **Chen, H***, & Lu, W*. LSD: A Unified Framework of Linked Scholarly Databases for Data Fusion and Scientific Discovery. *Information Fusion.* (5/376 in Information Systems, JCR Q1).

Articles Under Review

[1] *Electronics* Zhao, H., **Chen, H.,** Feng, Y., & Yoon, H. Optimizing Large Language Model-based Data Augmentation for Highly Imbalanced Text Classification. *Electronics.* Under review. IF = 2.9. (138/373 in Computer Networks and Communications, JCR Q2)

[2] *Electronics* Kargozari, K.#, Ding, J., **Chen, H.** Empowering Consumer Decision: Decoding Incentive vs. Organic Reviews for Smarter Choices Through Advanced Textual Analysis. Under review. IF = 2.9. (138/373 in Computer Networks and Communications, JCR Q2)

[3] *JOI* Wang, Z., Zhang, H., Chen, J., & **Chen, H*.** An Effective Framework for Measuring the Novelty of Scientific Articles through Integrated Topic Modeling and Cloud Model. *Journal of Informetrics.* Under review. IF = 3.7. (18/258 in Library and Information Sciences, JCR Q1)

[4] *EIT* Tu, F.#, Wu, L., Kinshuk, Ding, J., & **Chen, H***. Exploring the Influence of Regulated Learning Processes on Learners' Prestige in Project-Based Collaborative Learning. *Education and Information Technologies.* Under review. IF =5.5. (19/258 in Library and Information Sciences, JCR Q1)

[5] *JKSU-CIS* Wang, Z., Zhang, H., **Chen, H*,** Feng, Y., Ding, J. Content-based Quality Evaluation of Scientific Papers using Coarse Feature and Knowledge Entity Network. *Journal of King Saud University - Computer and Information Sciences.* Under review. IF = 6.9. (55/326 in Computer Science, JCR Q1)

[6] *JOI* Zhang, Y., Wang, Y., Sheng Q., **Chen, H.,** et al. Deep Learning Meets Bibliometrics: A Survey on Citation Function Classification. *Journal of Informetrics.* Under review. IF = 3.7. (18/258 in Library and Information Sciences, JCR Q1)

[7] *JCMC* Chen, D., Wang, Z., & **Chen, H***. Investigating the Transmission Patterns between Emotional Resonance and Cognitive Biases during Public Emergency on Social Media. *Journal of Computer-Mediated Communication*. Under review. IF = 7.2. (5/96 in Communication, JCR Q1)

[8] *IP&M* Wang, Z., Wang, N., Wang, Z., Ding, J., & **Chen, H***. IBID-CCT: A Novel Model for Interdisciplinary Breakthrough Innovation Detection based on Cusp Catastrophe Theory. *Research Policy.* Under review. IF = 8.6. (8/258 in Library and Information Sciences, JCR Q1)

Books

[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.

Refereed Journal Articles

[J23] *TEL* — **Chen, H.***, Kim, J., Chen, J., & Sakata, A. "The Long Journey Home": Demystifying Oral History with Natural Language Processing and Data Analytics: A Case Study of Densho Digital Collection. *The Electronic Library*. Under review. 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*. In press. IF = 2.42. (75/258 in Library and Information Sciences, JCR Q2)

[J21] *ACM Computing Surveys* — Huang, J., **Chen, H.**, Yu, F., & Lu, W. 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)

[J20] *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)

[J19] *SCM* — Wang, Z., Chen, J., Chen, J., & **Chen, H*.** (2023). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics*, 1-26. IF = 3.801. (26/258 in Library and Information Sciences, JCR Q1)

[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)

[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 = 8.0. (26/115 in Health Informatics, JCR Q1)

[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)

[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 = 8.6. (8/258 in Library and Information Sciences, JCR Q1, received 120 citations)

[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)

[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)
[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)
[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, received 60 citations)
[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
[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, received 12 citations)
[J1]	*IDD*	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

[C15]	*ASIS&T'24*	**Chen, H.***, Cherukuri, K#., Zhu, X., & Yang, S. Are Prompts All You Need?: Chatting with ChatGPT on Disinformation Policy Understanding (Short paper). 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.
[C14]	*IEEE QRS 2023*	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
[C13]	*IEEE AITest 2023*	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. Location and date: Athens, Greece, July 17-20, 2023 (Acceptance rate 60%)
[C12]	*IEEE AITest 2023*	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).* Location and date: Athens, Greece, July 17-20, 2023 (Acceptance rate 40%, 1/16 selected as the best student paper award)

[C11]	*IEEE AITest 2023* 🏆	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).* Location and date: Athens, Greece, July 17-20, 2023 (Acceptance rate 40%, 1/16 selected as the best paper award)
[C10]	*IEEE COMPSAC 2023* 🏅	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).* Location and date: Torino, Italy, June 26-30, 2023 (Acceptance rate 24%, received 70 citations)
[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. Location and date: Barcelona, Spain, March 27-29, 2023 (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)
[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. Location and date: Salt Lake City, UT, USA, October 30 - November 2, 2021 (Acceptance rate 49%, 12 citations)
[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 Nominate).* Online (Acceptance rate 28.8%, and among 6/36 accepted regular papers selected for journal publishing)
[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. Location and date: Hangzhou, China, August 24-27, 2019 (Among 18/140 selected for conference proceedings publishing, received 36 citations)
[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. Location and date: Taichung, Taiwan, October 29-31, 2018 (Acceptance rate 30%)
[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. Location and date: Wuhan, China, July 1-3, 2018 (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). Springer, Cham. Location and date: Wuhan, China, July 1-3, 2018 (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). Location and date: Dallas, TX, USA, October 25 – 27, 2017

Workshops, Panels, Posters, Abstracts, Blogs

[W 19]	*ASIS&T'24*	Kim., J., **Chen, H.**, Yang, L., & Simic, J. Exploring the Application of Artificial Intelligence and Machine Learning in GLAM Collections. 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.

[W 18]	*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. 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 2023*	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 2023*	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 2023*	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 2023*	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. *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 2023*	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). *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 2023*	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. *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 grants under review or in preparation:

- PI. NEH "Utilizing AI/ML to Enhance Public Engagement with Historical Records: Model Development, Best Practices, and Educational Training", National Endowment for the Humanities. In preparation.
- Co-PI (PI: Dr. Kewei Sha). NSF Expanding AI Innovation through Capacity Building and Partnerships (ExpandAI) "PARTNER: Expanding Edge AI for Next Generation Connected Autonomous Systems", National Science Foundation. In preparation.
- Co-PI (PI: Dr. Junhua Ding). NSF Research Experience for Teachers (RET) Site *"Machine Learning and Data Analytics with Large Language Models"*. Request: $599,544. Pending.
- Co-PI (PI: Dr. Junhua Ding). NSF RITEL "AI Mastery: Goal-Driven Learning Enhanced with Real-time Performance Analysis for Mastering AI Fundamentals and Practices", National Science Foundation. Request: $899,757. Pending.
- Senior Personal (PI: Dr. Yunhe Feng). NSF REU Site "Aligning Artificial Intelligence with Human Values", National Science Foundation. Request: $456,540. Pending.

External funded grants:

- 2023-2025. **Senior Personnel** (PI: Dr. Ting Xiao). NSF REU Site *"Beyond Language: Training to Create and Share Vector Embeddings across Applications"*, National Science Foundation. Amount: $403,547.
- 2022-2025. **Co-PI** (PI: Dr. Junhua Ding). NSF HSI Implementation and Evaluation Project *"Developing a High-Quality Academic Environment for Broadening Participation of Hispanic Students in Computing"*, National Science Foundation. Award Amount: $499,608.
- 2021. PI. Google Cloud Research Credits for project *"Building a knowledge-driven Legal IR, Automatic Evidence and Argument Generation System for Decision Making"*, Google. Award Amount: $5,000.
- 2021. PI. Google Cloud Education Grant - google cloud platform credits for course *"INFO 5731: Computational Methods for Information Systems"*. Google. Award Amount: $1,550.
- 2020. PI. Google Cloud COVID-19 Research Credits Program - google cloud platform credits for project *"Demystifying COVID-19 publications"*. Award Amount: $13,800.
- 2020. PI. Google Cloud Education Grant - google cloud platform credits for course *"INFO 5731: Computational Methods for Information Systems"*. Award Amount: $2,800.
- 2019. PI. Google Cloud Education Grant - google cloud platform credits for course *"INFO 4206: Information Retrieval"*. Award Amount: $5,000.

External not funded grants:

- 2024. Co-PI (PI: Dr. Junhua Ding). NSF Research Traineeship (NRT) Program *"NRT-AI: Train Next-Generation Scientists and Engineers in Data-Centric Intelligence"*. Request: $3,000,000. Not funded.

- 2023. PI. NIH AIM-AHEAD "Utilizing AI/ML to Enhance Personalized Health Information Services for Hispanic Populations during Disaster Recovery". Amount requested: $99,991. Not funded.
- 2023. Co-PI (PI: Dr. Sharad Sharma). NSF Build and Broaden (B2) *"Collaborative Research: Build and Broaden: Building Research Capacity in Health Informatics Research Partnerships"*, National Science Foundation. Not funded.
- 2023. Co-PI (PI: Dr. Sharad Sharma). NSF I/UCRC Planning Grant *"Center for Emerging System Dependability Assurance (CESDA)"*, National Science Foundation. Not funded.

Internal funds:
- 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. Award Amount: $5,000.
- 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.
- 2022. Co-PI (PI: Dr. Xiaoqun Zhang). 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.
- 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. Award Amount: $9,975.
- 2020. PI. UNT Graduate Student Research Award. Award Amount: $800.

Travel Fundings

- 2023. UNT College of Information and Department of Information Science (SCC 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.
- 2017. UNT Department of Information Science Travel Grant (ICKM 2017). Award Amount: $300.

Honors and Awards

- 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

- INFO 5506: Applications of Artificial Intelligence in Health (Spring 2024)
- INFO 5810: Data Analysis and Knowledge Discovery (Summer 2022, Fall 2022, Spring 2023, Summer 2024)
- INFO 5731: Computational Methods for Information Systems (Spring 2020 – Spring 2024)
- 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)
- CSCE 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
  - Fengjiao Tu, Ph.D. student in Information Science (Fall 2023-Present, as major professor)
  - Nishith Reddy Mannuru, Ph.D. student in Information Science (Spring 2024-Present, as committee member)
  - Mimi Byun, Ph.D. student 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. student in Information Science (Spring 2022-Present, as faculty mentor)
  - Huyen T. N. Nguyen, Ph.D. student in Information Science (Summer 2021-Present, as faculty mentor)
  - Kate Kargozari, Ph.D. student in Information Science (Fall 2020-Present, as faculty mentor)
  - Komala Subramanyam Cherukuri, master student in Data Science (Spring 2024-Present)
  - Murthy Aryan, master student in Information Science (Fall 2023-Present)
  - Aisa Sakata, master student in Information Science (Fall 2023-Present)
  - Lavanya Pobbathi, master student in Data Science (Spring 2023-Present)

- o Kaushal Sen, master student in Data Science (Fall 2022-Present)
- o Suruthi Selvam, master student in Data Science (Spring 2023-Summer 2023)
- o Kavya Vedantham, master student in Data Science (Summer 2023)
- o Mounica Siri Chandana Tamalampudi, master student in Data Science (Summer 2023)
- o Solomon Daramay, undergraduate in Computer Science at UNT (Under the NSF REU program at UTD, Summer 2023)
- o Aparna Bhooshanan, undergraduate in Computer Science at The University of Alabama (Under the NSF REU program at UNT, Summer 2023)
- o Chimara Okeke, undergraduate in Bioinformatics at UNT (Under the NSF REU program at UNT, Summer 2023)
- o Ella Ghazarian, undergraduate in Computer Science at UNT (Under the NSF REU program at UNT, Summer 2023)
- o Max Jing, undergraduate in Statistics at University of Illinois Urbana-Champaign (Summer 2023)
- o Karthik Puppala, master student in Data Science (Spring 2023)
- Alumni
  - o Amith Reddy Kasarla, master student in Data Science (Graduated in Spring 2023, now working as a Reporting Analyst III at BAL)
  - o Revathi Bejawada, master student in Data Science (Graduated in Spring 2023, now working as a Software Engineer at Source Infotech Inc)
  - o Roopesh Maganti, master student in Data Science (Graduated in Spring 2023, now working as a Data Management Engineer with Comcast)
  - o Bhanu Prasad G, master student in Data Science (Graduated in Spring 2022, now working as a Software Engineer at T-mobile)
  - o Akshitha Maddi, Master student in Data Science (Graduated in Spring 2021, now working as a Business Analyst at Barclays)
  - o Sai Supriya Vadla, Master student in Data Science (Graduated in Spring 2021, now working as an Associate Engineer at Continental)
  - o Nagma Shaik, Master student in Data Science (Graduated in Spring 2021, now working as a Software Engineer III at SiriusXM & Pandora)
  - o Raheyma Khan, Master student in Data Science (Graduated in Spring 2021, now working as a Market Analyst at Kemper)
  - o 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: 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

Professional Service

- Leadership:
    - Co-editor of The Electrical Library (2022-present)
    - 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:
    - Frontiers in Research Metrics and Analytics (2023-present)
    - Heliyon Information Science (2022-present)
    - American Journal of Information Science and Technology (2020-present)
- Guest editor:
    - Computer Standards & Interfaces (IF = 5.0) special issue "Applications of Generative AI", 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 "Applications of Deep Learning Techniques", ongoing. 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, ongoing. Link: https://www.emeraldgrouppublishing.com/calls-for-papers/innovation-measurement-scientific-communication-imsc-era-big-data
    - Information Discovery & Delivery (IF = 2.245) special issue "Information and Data Quality for Intelligent Systems", 2022-2023, finished.
- Co-chair of Workshops: DQIS workshop@ IEEE QRS 2021, 2022, 2023, 2024, EEKE+AI Informatics workshop@ JCDL 2023, iConference 2024, IMSC workshop@ JCDL 2023
- Journal reviewer:
    - The Electrical Library (2017-present)
    - IEEE Access (2018-present)
    - Information Discovery and Delivery (2018-present)
    - Scientometrics (2021-present)
    - Computer Standards & Interfaces (2021-present)
    - JAMIA Open (2021-present)

- o Journal of Computational Social Science (2022-present)
  - o Aslib Journal of Information Management (2022-present)
  - o Applied Intelligence (2022-present)
  - o ACM Transactions on Asian and Low-Resource Language Information Processing (2022-present)
  - o Artificial Intelligence Review (2022-present)
  - o Information Processing and Management (2022-present)
  - o Journal of Informetrics (2022-present)
  - o Information Resources Management Journal (2022 - present)
  - o Frontiers in Public Health (2022-present)
  - o Neural Processing Letters (2022-present)
  - o Machine Learning (2022-present)
  - o Sensors (2022-present)
  - o Computational Intelligence and Neuroscience (2022-present).
  - o PeerJ Computer Science (2022-present).
  - o Mathematics (2022-present).
  - o Computer Animation and Virtual Worlds (2023-present).
  - o Complex & Intelligent Systems (2023-present).
  - o Knowledge and Information Systems (2023-present).
  - o Scientific Reports (2023-present).
  - o Expert Systems with Applications (2023-present).
  - o PLOS One (2023-present).
  - o Journal of the Association for Information Science and Technology (2023-present).
  - o Knowledge-Based Systems (2023-present).
  - o IEEE Transactions on Reliability (2023-present).
  - o IEEE Transactions on Big Data (2023-present).
  - o Education and Information Technologies (2024-present).
  - o IEEE Transactions on Artificial Intelligence (2024-present).
  - o IEEE Transactions on Computational Social Systems (2024-present).
  - o IEEE Transactions on Emerging Topics in Computational Intelligence (2024-present).
  - o IEEE Transactions on Knowledge and Data Engineering (2024-present).
- PC member and reviewer for Conferences and Workshops:
  - o IEEE COMPSAC (2020, 2021, 2022, 2023, 2024)
  - o IEEE ICSC (2020, 2021, 2022, 2023, 2024)
  - o IEEE Big Data (2019, 2020, 2021, 2022, 2023, 2024)
  - o IEEE QRS (2020, 2021, 2022, 2023, 2024)
  - o IEEE ETCEA (2022)
  - o iConference (2021, 2023, 2024)
  - o JCDL (2021, 2022, 2023)
  - o ASIS&T (2021, 2023, 2024)
  - o ALLDATA (2022, 2023)
  - o IEEE AITest (2023)
  - o IEEE DSA (2023)

Department Service

09/2023-present	**Chair of the External Affairs and Physical Resources Committee,** Department of Information Science, University of North Texas
09/2022-present	**Faculty Search Committee,** Data Science Program, University of North Texas
09/2022-08/2023	**Faculty Search Committee,** Health Informatics Program, 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)