# Zhiyu Wan

2525 West End Avenue, Suite 1422-1, Nashville, TN 37203

zhiyu.wan.1@vumc.org ♦ (615) 839-8083

## EDUCATION

**Vanderbilt University**                                                                                          Nashville, TN
- ***Ph.D. in Computer Science***, *Department. of Computer Science, School of Engineering*          Dec. 2020
  -Dissertation: <u>Privacy-Preserving Sharing of High-Dimensional Data based on Computational Game Theory</u>
  -Committee: Drs. Bradley A. Malin, Daniel Fabbri, Douglas H. Fisher, Yevgeniy Vorobeychik, Murat Kantarcioglu

**Xi'an Jiaotong University**                                                                                   Xi'an, China
- ***B.Eng. in Automation***, *School of Electronic and Information Engineering*                        June 2011

## RESEARCH EXPERIENCE

**Vanderbilt University Medical Center**                                                                      Nashville, TN
- ***Postdoctoral Research Fellow***, *Center for Genetic Privacy and Identity in Community Settings*   2021 – 2024
  (PIs: Drs. Ellen Wright Clayton and Bradley A. Malin)
  -Designed re-identification risk assessment models for genomic and health data and applied such models to the
   NIH's All of Us Research Program. (Supported by NIH grant RM1HG009034 and NSF grant CNS-2029651.)

**Vanderbilt University**                                                                                         Nashville, TN
- ***Research Assistant***, *Health Information Privacy Lab* (PI: Dr. Bradley A. Malin)                  2013 – 2020
  -Developed game theoretical models for privacy-preserving health & genomic data sharing. (Supported by NIH
   grants, R01HG006844 and R01LM009989.)

## TEACHING EXPERIENCE

**Vanderbilt University**                                                                                         Nashville, TN
- ***Certificate in College Teaching***, *Center for Teaching*                                       Aug. 2019 – May 2020
  -Completed seminars, practicums, and specializations in college teaching.
- ***Guest Lecturer***, *Department of Electrical Engineering and Computer Science*                   Spring 2018/19/20/22
  -Taught one lecture each semester for Data Privacy in Biomedicine.
- ***Teaching Assistant***, *Department of Electrical Engineering and Computer Science*               Aug. 2012 – May 2013
  -Graded exams and taught help-sessions for two courses: Discrete Structures and Introductory Programing.

## PEER-REVIEWED PUBLICATIONS

### Book Chapter

1. **Zhiyu Wan**, Yevgeniy Vorobeychik, Ellen Wright Clayton, Murat Kantarcioglu, and Bradley A. Malin. <u>Game theory for privacy-preserving sharing of genomic data</u>. In *Responsible Genomic Data Sharing: Challenges and Approaches, Xiaoqian Jiang and Haixu Tang, Eds.*, pp. 135–160, Academic Press, 2020.

### Journal

2. Abinitha Gourabathina*, **Zhiyu Wan**\*, J. Thomas Brown, Chao Yan, and Bradley A. Malin. <u>PanDa Game: Optimized Privacy-Preserving Publishing of Individual-Level Pandemic Data Based on a Game Theoretic Model</u>. *IEEE Transactions on NanoBioscience*, 22(4): 808–817, 2023. [IF:3.9]

3. Rajagopal Venkatesaramani, **Zhiyu Wan**, Bradley A. Malin, and Yevgeniy Vorobeychik. <u>Defending against membership inference attacks on Beacon services</u>. *ACM Transactions on Privacy and Security*, 26(3): 42, 2023. [IF:2.3]

4. Rajagopal Venkatesaramani, **Zhiyu Wan**, Bradley A. Malin, and Yevgeniy Vorobeychik. <u>Enabling Trade-offs in Privacy and Utility in Genomic Data Beacons and Summary Statistics</u>. *Genome Research*, 33(7): 1113–1123, 2023. [IF:9.438]

5. Jia Guo, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, Myrna Wooders, **Zhiyu Wan**, Zhijun Yin, and Bradley Malin. <u>A Game Theoretic Approach to Balance Privacy Risks and Familial Benefits</u>. *Scientific Reports*, 13(1): 6932, 2023. [IF:4.996]

6.  Weiyi Xia, Melissa Basford, Robert Carroll, Ellen Wright Clayton, Paul Harris, Murat Kantacioglu, Yongtai Liu, Steve Nyemba, Yevgeniy Vorobeychik, **Zhiyu Wan**, and Bradley A. Malin. Managing Re-identification Risks While Providing Access to the All of Us Research Program. *Journal of the American Medical Informatics Association*, 30(5):907–914, 2023. [IF:7.942]

7.  Yongtai Liu, Zhijun Yin, Congning Ni, Chao Yan, **Zhiyu Wan**, and Bradley A. Malin. Examining Rural and Urban Sentiment Difference in COVID-19 Related Topics on Twitter: Word Embedding–Based Retrospective Study. *Journal of Medical Internet Research*, 23(1): e42985, 2023. [IF:7.076]

8.  Chao Yan*, Yao Yan*, **Zhiyu Wan**\*, Ziqi Zhang, Larsson Omberg, Justin Guinney, Sean D. Mooney, and Bradley A. Malin. A Multifaceted Benchmarking of Synthetic Electronic Health Record Generation Models. *Nature Communications*, 13(1): 7609, 2022. [IF:17.694]

9.  Yongtai Liu, Zhijun Yin, **Zhiyu Wan**, Chao Yan, Weiyi Xia, Congning Ni, Ellen Wright Clayton, Yevgeniy Vorobeychik, Murat Kantarcioglu, and Bradley A. Malin. Implicit Incentives Among Reddit Users to Prioritize Attention Over Privacy and Reveal Their Faces When Discussing Direct-to-Consumer Genetic Test Results: Topic and Attention Analysis. *JMIR Infodemiology*, 2(2): e35702, 2022.

10. **Zhiyu Wan**, James W. Hazel, Ellen Wright Clayton, Yevgeniy Vorobeychik, Murat Kantarcioglu, Bradley Malin. Sociotechnical safeguards for genomic data privacy. *Nature Reviews Genetics*, 23(7): 429–445, 2022. [IF:59.581]

11. J. Thomas Brown, Chao Yan, Weiyi Xia, Zhijun Yin, **Zhiyu Wan**, Aris Gkoulalas-Divanis, Murat Kantarcioglu, and Bradley A. Malin. Dynamically adjusting case-reporting policy to maximize privacy and utility in the face of a pandemic. *Journal of the American Medical Informatics Association*, 29(5): 853–863, 2022. [IF:7.942]

12. Congning Ni, **Zhiyu Wan**, Chao Yan, Yongtai Liu, Ellen Wright Clayton, Bradley Malin, and Zhijun Yin. The Public perception of the #GeneEditedBabies event across multiple social media platforms: Observational study. *Journal of Medical Internet Research*, 24(3): e31687, 2022. [IF:7.076]

13. **Zhiyu Wan**, Yevgeniy Vorobeychik, Weiyi Xia, Yongtai Liu, Myrna Wooders, Jia Guo, Zhijun Yin, Ellen Wright Clayton, Murat Kantarcioglu, and Bradley A. Malin. Using game theory to thwart multistage privacy intrusions when sharing data. *Science Advances*, 7(50): eabe9986, 2021. [IF:14.136]

14. Weiyi Xia, Yongtai Liu, **Zhiyu Wan**, Yevgeniy Vorobeychik, Murat Kantacioglu, Steve Nyemba, Ellen Wright Clayton, and Bradley A. Malin. Enabling realistic health data re-identification risk assessment through adversarial modeling. *Journal of the American Medical Informatics Association*, 28(4): 744–752, 2021. [IF:4.497]

15. Weiyi Xia, **Zhiyu Wan**, Zhijun Yin, James Gaupp, Yongtai Liu, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A. Malin. It's all in the timing: calibrating temporal penalties for biomedical data sharing. *Journal of the American Medical Informatics Association*, 25(1): 25–31, 2018. [IF:4.270]

16. **Zhiyu Wan**, Yevgeniy Vorobeychik, Murat Kantarcioglu, and Bradley A. Malin. Controlling the signal: Practical privacy protection of genomic data sharing through Beacon services. *BMC Medical Genomics*, 10(Suppl 2): 39, 2017. [IF:2.848]

17. **Zhiyu Wan**, Yevgeniy Vorobeychik, Weiyi Xia, Ellen Wright Clayton, Murat Kantarcioglu, and Bradley A. Malin. Expanding access to large-scale genomic data while promoting privacy: a game theoretic approach. *The American Journal of Human Genetics*, 100(2): 316–322, 2017. ***Best papers in International Medical Informatics Association (IMIA) Yearbook of Medical Informatics***. [IF:10.794]

18. **Zhiyu Wan**, Yevgeniy Vorobeychik, Weiyi Xia, Ellen Wright Clayton, Murat Kantarcioglu, Ranjit Ganta, Raymond Heatherly, and Bradley A. Malin. A game theoretic framework for analyzing re-identification risk. *PLoS ONE*, 10(3): e0120592, 2015. [IF:3.234]

**Conference**

19. Rajagopal Venkatesaramani, **Zhiyu Wan**, Bradley A. Malin, and Yevgeniy Vorobeychik. Enabling Trade-offs in Privacy and Utility in Genomic Data Beacons and Summary Statistics. *Proceedings of the 27th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, Istanbul, Turkey, 2023.

20. Abinitha Gourabathina*, **Zhiyu Wan**\*, James Brown, Chao Yan, and Bradley Malin. Privacy-Preserving Publishing of Individual-Level Pandemic Data Based on a Game Theoretic Model. *Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 961–968, IEEE, 2022.

21. J. Thomas Brown, **Zhiyu Wan**, Aris Gkoulalas-Divanis, Murat Kantarcioglu, and Bradley Malin. Supporting

COVID-19 Disparity Investigations with Dynamically Adjusting Case Reporting Policies. *Proceedings of the 2022 American Medical Informatics Association Annual Fall Symposium (AMIA)*, Washington, DC, USA, 2022.

22. Xinmeng Zhang, **Zhiyu Wan**, Chao Yan, Thomas Brown, Weiyi Xia, Aris Gkoulalas-Divanis, Murat Kantarcioglu, and Bradley Malin. How Adversarial Assumptions Influence Re-identification Risk Measures: A COVID-19 Case Study. In *Privacy in Statistical Databases. PSD 2022. Lecture Notes in Computer Science, vol 13463, Domingo-Ferrer, J., Laurent, M.., Eds.*, pp. 361–374, Springer, 2022.

23. Yongtai Liu, Douglas Downey, **Zhiyu Wan**, Murat Kantarcioglu, Yevgeniy Vorobeychik, Bradley Malin. De-identifying socioeconomic data at the census tract level for medical research through constraint-based clustering. In *AMIA 2019 Annual Symposium Proceedings*, pp. 793–802, American Medical Informatics Association, 2021.

24. Yongtai Liu, Chao Yan, Zhijun Yin, **Zhiyu Wan**, Weiyi Xia, Murat Kantarcioglu, Yevgeniy Vorobeychik, Ellen Wright Clayton, and Bradley Malin. Biomedical research cohort membership disclosure on social media. In *AMIA 2019 Annual Symposium Proceedings*, pp. 607–616, American Medical Informatics Association, 2019. ***Distinguished Paper Awards***.

25. Yongtai Liu, **Zhiyu Wan**, Weiyi Xia, Murat Kantarcioglu, Yevgeniy Vorobeychik, Ellen Wright Clayton, Abel Kho, David Carrell, and Bradley Malin. Detecting the presence of an individual in phenotypic summary data. In *AMIA 2018 Annual Symposium Proceedings*, pp. 760–769, American Medical Informatics Association, 2018.

26. Fabian Prasser, James Gaupp, **Zhiyu Wan**, Weiyi Xia, Yevgeniy Vorobeychik, Murat Kantarcioglu, Klaus Kuhn, and Bradley Malin. An open source tool for game theoretic health data de-identification. In *AMIA 2017 Annual Symposium Proceedings*, pp. 1430–1439, American Medical Informatics Association, 2017.

27. Weiyi Xia, Murat Kantarcioglu, **Zhiyu Wan**, Raymond Heatherly, Yevgeniy Vorobeychik, and Bradley Malin. Process-driven data privacy. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1021–1030, ACM, 2015.

28. Xiaobo Ma, Jiahong Zhu, **Zhiyu Wan**, Jing Tao, Xiaohong Guan, and Qinghua Zheng. Honeynet-based collaborative defense using improved highly predictive blacklisting algorithm. In *2010 8th World Congress on Intelligent Control and Automation*, pp. 1283–1288, IEEE, 2010.

## PREPRINT

29. Leyao Wang\*, **Zhiyu Wan**\*, Congning Ni, Qingyuan Song, Yang Li, Ellen Wright Clayton, Bradley A. Malin, and Zhijun Yin. A Systematic Review of ChatGPT and Other Conversational Large Language Models in Healthcare. *medRxiv*:2024.04.26.24306390, 2024.

30. Chris Clifton, Bradley Malin, Anna Oganian, Ramesh Raskar, Vivek Sharma, Weiyi Xia, Jeremy Seeman, **Zhiyu Wan**, and Abhishek Singh. A Roadmap for Greater Public Use of Privacy-Sensitive Government Data: Workshop Report. *arXiv*:2208.01636, 2022.

## INVITED TALKS

- **Zhiyu Wan**. Harnessing AI Agents for Ethical and Responsive Real World Data Sharing in Medicine. *AI for Reliable and Equitable Real-world Evidence Generation in Medicine Workshop (in Conjunction with AIME2024) Rising Star Session on AI in Medicine*. Salt Lake City, UT, USA (July 9, 2024).
- **Zhiyu Wan**. Strategic Privacy-Preserving Sharing of High Dimensional Biomedical Data Enabled by Artificial Intelligence. *UTHealth School of Biomedical Informatics Spring 2023 Research Seminar*. Virtual (February 1, 2023).
- **Zhiyu Wan**. Game Theory For Shielding Personal Data Against Privacy Attacks On Anonymity. *2022 INFORMS Annual Meeting*. Indianapolis, IN, USA (October 16, 2022).

## POSTERS

- Yuhang Guo, **Zhiyu Wan**. Performance Evaluation of Multimodal Large Language Models (LLaVA and GPT-4-based ChatGPT) in Medical Image Classification Tasks. *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*. Orlando, FL, USA (June 4, 2024).
- **Zhiyu Wan**. Privacy-preserving collection and sharing of unbiased human voice data for automatic assessment of voice disorders. *17th Annual Vanderbilt Postdoctoral Association Symposium*. Nashville, TN, USA (Nov. 17, 2023).
- **Zhiyu Wan**. Privacy-preserving collection and sharing of unbiased human voice data for automatic assessment of

voice disorders. *Inaugural Bridge2AI Voice Symposium*. Washington, DC, USA (April 19, 2023). ***Best Poster***.

- Weiyi Xia, Yongtai Liu, **Zhiyu Wan**, Yevgeniy Vorobeychik, Murat Kantacioglu, Ellen W. Clayton, and Bradley A. Malin. A Scalable Tool for Realistic Health Data Re-identification Risk Assessment. *Proceedings of the 2022 American Medical Informatics Association Annual Fall Symposium (AMIA)*, Washington, DC, USA (Nov. 2022).
- **Zhiyu Wan**. Privacy-preserving collection and sharing of unbiased human voice data for automatic assessment of voice disorders and respiratory diseases: A pilot study. *Informatic Con 22*, Nashville, TN, US (September 28, 2022).
- **Zhiyu Wan**. A Game Theoretic Model for Privacy-Preserving Genomic Data Sharing. *9th Annual Oak Ridge Postdoctoral Association Research Symposium*. Virtual (July 29, 2021).

## OTHER PRESENTATIONS

- **Zhiyu Wan**, Yuhang Guo, Shunxing Bao, Qian Wang and Bradley Malin. Evaluating Sex and Age Biases in ChatGPT-4 and LLaVA-1.6 for Skin Disease Identification Tasks Using Real-World Data. *2024 International Workshop on AI Applications in Public Health and Social Services (AI-PHSS 2024) (in Conjunction with AIME2024)*. Salt Lake City, UT, USA (July 12, 2024).
- Chao Yan, Yao Yan, Ziqi Zhang, **Zhiyu Wan**, Justin Guinney, Sean Mooney, and Bradley Malin. Assessing Machine Learning Based Generators for Synthetic Electronic Health Records: A Benchmarking. *Proceedings of the 2022 American Medical Informatics Association Annual Fall Symposium (AMIA)*, Washington, DC, USA, 2022.
- **Zhiyu Wan**. A Game Theoretic Model for Sharing Biomedical Data While Preventing Multistage Privacy Intrusions. *16th Annual Vanderbilt Postdoctoral Association Symposium*. Nashville, TN, USA (October 27, 2022).
- **Zhiyu Wan**. Game Theory for Genomic Data Privacy. *2nd event of the "Speak Easy" – A Postdoc Talk Series*. Nashville, TN, USA (July 21, 2022).
- **Zhiyu Wan**. Game Theory Can Expand Access to Genomic Data While Promoting Privacy. *AAAI 2017 Spring Symposium on AI for Social Good*. Stanford University. March 28, 2017.
- **Zhiyu Wan**. Practical Protection of Genomic Data Sharing Through Beacon Services. *iDASH Privacy and Security Workshop*, co-located with *AMIA 2016 Annual Symposium*. Chicago, IL. November 11, 2016.

## HONORS AND AWARDS

| | | |
|---|---|---|
| • Inaugural Bridge2AI Voice Symposium Poster Competition | **Winner** | Apr. 2023 |
| • Russell G. Hamilton Graduate Leadership Institute Travel Grant | **$1,000 USD** | Nov. 2019 |
| • 2019 American Medical Informatics Association Annual Fall Symposium | **Distinguished Paper** | Nov. 2019 |
| • 2018 IMIA Yearbook of Medical Informatics | **Best papers** | Aug. 2018 |
| • Vanderbilt University Graduate Student Travel Grant | **$500 USD** | Mar. 2017 |
| • 2016 iDASH Privacy and Security Competition Track 1 | **Winner** | Nov. 2016 |
| • iDASH Travel Award | **$1,250 USD** | Nov. 2016 |

## ACADEMIC SERVICE

- **Session Chair**:
  - IEEE International Conference on Bioinformatics and Biomedicine (BIBM '22/'21)

- **Committee Member**:
  - International Conference of AI in Medicine (AIME '24)
  - American Medical Informatics Association Clinical Informatics Conference (AMIA CIC '23)
  - International Medical Informatics Association – Yearbook – Edition 2022
  - Workshop on Artificial Intelligence for Social Good (AI4SG '21)
  - IEEE International Conference on Bioinformatics and Biomedicine (BIBM '23/'22/'21)
  - International Conference on Tools with Artificial Intelligence (ICTAI '22/'21/'20)

- **Reviewer**:
  Journal
    ACM Transactions on Privacy and Security (TOPS), Bioinformatics, BMC Medical Informatics and Decision Making, BMC Medical Research Methodology, Frontiers in Public Health, Heliyon, IEEE Transactions on

Dependable and Secure Computing, Journal of the American Medical Informatics Association, PLoS ONE, Methods of Information in Medicine, Nature Biotechnology, Transactions on Data Privacy

Conference
- AAAI conference on Artificial Intelligence (AAAI '19/'18/'17/'16)
- IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE '19)
- American Medical Informatics Association 2019 Annual Symposium (AMIA '24/'23/'22/'21/'20/'19/'18/'17)
- IEEE International Conference on Big Data (BigData '16/'15)
- ACM Conference on Data and Application Security and Privacy (CODASPY '23/'22/'21/'20/'19)
- Conference on Data and Applications Security and Privacy (DBSec '20)
- IEEE International Symposium on High Assurance Systems Engineering (HASE '15)
- IEEE International Conference on Data Mining (ICDM '19/'18/'17)
- International Conference on Learning Representations (ICLR '22)
- International Conference on Machine Learning (ICML '21)
- International Joint Conference on Artificial Intelligence (IJCAI '19)
- European Conference on Computational Biology (ECCB '23)
- Conference on Neural Information Processing Systems (NeurIPS '21)
- Pacific Symposium on Biocomputing (PSB '21)
- Privacy in statistical databases (PSD '20)
- Transdisciplinary Artificial Intelligence (TransAI '20) …

## LEADERSHIP AND MEMBERSHIPS

| | | | |
|---|---|---|---|
| • Vanderbilt Postdoctoral Association (VPA) | **Vice Chair** | Aug. 2021 – May 2023 |
| • Vanderbilt University Chinese Students and Scholars Association | **Vice President** | May 2015 – May 2018 |
| • Sigma Xi, The Scientific Research Honor Society | **Full Member** | April 2024 – Present |
| • Association for the Advancement of Artificial Intelligence (AAAI) | **Member** | Mar. 2017 – Dec. 2022 |
| • American Society of Human Genetics (ASHG) | **Member** | Sep. 2016 – Dec. 2022 |
| • American Medical Informatics Association (AMIA) | **Member** | Nov. 2016 – Present |
| • Institute of Electrical and Electronics Engineers (IEEE) | **Member** | Jan. 2015 – Present |
| • Association for Computing Machinery (ACM) | **Member** | Jan. 2014 – Present |
| • Institute for Operations Research & Management Sciences (INFORMS) | **Member** | Sep. 2022 – Dec. 2022 |

## COURSE PROJECTS

| | | |
|---|---|---|
| **Big Data (A)** | *Freelancer Recommender System based on Mahout* | Spring 2014 |
| **Machine Learning (A)** | *Community Detection in Social Networks* | Spring 2014 |
| **Data Privacy (A-)** | *More Accurate Estimation of Quasi-Identifier Anonymity* | Spring 2013 |
| **Computer Networks (A)** | *Instant Messenger with Multimedia Messages* | Spring 2013 |
| **Distributed Systems (A)** | *Security and Privacy Preservation in Healthcare Data Flows* | Fall 2012 |
| **Network Security (A)** | *Botnet Attacks on Cyber-Physical System (CPS)* | Fall 2012 |
| **Model-Integrated Computing (A+)** | *Metamodeling of a Honeynet-based Collaborative Defense System with a Demonstration on Botnet Defense* | Fall 2012 |

## OTHER RELEVANT COURSES

- **Vanderbilt University**: Advanced Artificial Intelligence (B+), Cloud Computing (A), Algorithms (B+).

## TECHNICAL SKILLS

| | |
|---|---|
| • Programming Languages: | Python, MATLAB, Java, C/C++, Delphi, UML, BASIC, MASM. |
| • Machine Learning: | Pytorch, TensorFlow, Keras, Pandas, SciPy, Matplotlib, Numpy. |
| • Big Data: | Mahout, Scala, Spark, Hive, Apache Cassandra. |
| • Cloud Computing: | Hadoop, Hbase, Amazon Web Services. |
| • Network Security Tools: | VMware, GnuPG, OpenSSL, Snort, Netcat, Nessus, Wireshark, Sniffer. |