

## Chao Yan, Ph.D.

Vanderbilt University Medical Center

#1400, 2525 West End Ave, Nashville, TN, 37203.

✉ [chao.yan.1@vumc.org](mailto:chao.yan.1@vumc.org)  [authoryanchao@gmail.com](mailto:authoryanchao@gmail.com)

### RESEARCH INTERESTS

---

AI for Medicine, Generative AI, Synthetic Health Data, Medical AI Fairness, Health Disparity, Representation Learning, Healthcare Workflow Modeling, Health Data Privacy.

### EXPERIENCE

---

#### Incoming Assistant Professor (Tenure Track)

10/2024 –

*Department of Biomedical Informatics  
Vanderbilt University Medical Center  
Nashville, USA*

#### Postdoctoral Fellow

02/2022 – 10/2024

*Department of Biomedical Informatics  
Vanderbilt University Medical Center  
Nashville, USA  
Advisor: Dr. Bradley Malin*

### EDUCATION

---

#### Ph.D. in Computer Science

08/2015 – 01/2022

*Health Information Privacy Laboratory  
Vanderbilt University  
Advisor: Dr. Bradley Malin*

*Nashville, USA*

#### M.S. in Computer Science

09/2012 - 06/2015

*University of Chinese Academy of Sciences  
Computer Network Information Center, Chinese Academy of Sciences*

*Beijing, China*

#### B.Eng. in Network Engineering

09/2008 - 06/2012

*Southwest University*

*Chongqing, China*

### INTERNSHIP EXPERIENCE

---

#### Research Intern

08/2019 – 11/2019

*School of Biomedical Informatics  
The University of Texas Health Science Center at Houston  
Advisor: Dr. Xiaoqian Jiang*

*Houston, USA*

*Research Project: Embedding Harmonization across multiple healthcare organizations*

### HONORS & AWARDS

---

- One of the best clinical research informatics papers published in 2020, selected by the International Medical Informatics Association 2021
- Honorable Mention of C.F. Chen Best Graduate Student Paper Award 2021
- Elected as member of *Student Editorial Board of JAMIA* 2020

▪ Distinguished Paper Award of AMIA 2019 Annual Symposium (Top 1%)	2019
▪ Second & Third Place of IEEE ICHI Data Analytics Challenge	2019
▪ Travel Grant of Vanderbilt University for Research Presentation at IEEE ICDE	2018
▪ Travel Grant of Vanderbilt University for Research Presentation at IEEE CIC	2018
▪ Excellent Graduates of Beijing (Top 1%)	2015
▪ National Scholarship for Graduate Students (Top 2%)	2014
▪ Excellent Graduates of Chongqing City (Top 1%)	2012
▪ National Scholarship (Top 2%)	2011
▪ National Scholarship (Top 2%)	2009

## PUBLICATIONS

---

### Accepted Journal Articles & Conference Papers (\* Equal Contribution)

[49] **Chao Yan**, Henry Ong, Monika Grabowska, Matthew Krantz, Wu-Chen Su, ..., Wei-Qi Wei, “Large Language Models Facilitate the Generation of Electronic Health Record Phenotyping Algorithms,” *Journal of the American Medical Informatics Association (JAMIA)*, 2024. In press [IF: 6.4]

[48] **Chao Yan\***, Monika Grabowska\*, Alyson Dickson, Bingshan Li, Zhexiong Wen, Dan Roden, Michael Stein, Peter Embí, Josh Peterson, QiPing Feng, Bradley Malin, Wei-Qi Wei, “Leveraging generative AI to prioritize drug repurposing candidates: Validating identified candidates for Alzheimer’s disease in real-world clinical datasets.” **npj Digital Medicine**. 2024. [IF: 15.4]

[47] **Chao Yan**, Ziqi Zhang, Steve Nyemba, Zhuohang Li, “Generating synthetic electronic health record data: A tutorial.” *Journal of Medical Internet Research (JMIR) AI*. 2024.

[46] Jiayi Tong, Yishan Shen, Alice Xu, Xing He, Chongliang Luo, Mackenzie Edmondson, Dazheng Zhang, Yiwen Lu, **Chao Yan**, ..., David Asch, Yong Chen, “Evaluating site-of-care-related racial disparities in kidney graft failure using a novel federated learning framework.” *Journal of the American Medical Informatics Association (JAMIA)*, 2024. In press [IF: 6.4]

[45] **Chao Yan\***, Xinmeng Zhang\*, Yuyang Yang\*, Kaidi Kang, Martin Were, Peter Embí, Mayur Patel, Bradley Malin, Abel Kho, You Chen, “Differences in health professionals’ engagement with electronic health records based on inpatient race and ethnicity.” **JAMA Network Open**, 2023. [IF: 13.8]

[44] Abinitha Gourabathina\*, Zhiyu Wan\*, Thomas Brown, **Chao Yan**, Bradley Malin, “PanDa Game: Optimized privacy-preserving publishing of individual-level pandemic data based on a game theoretic model.” **IEEE Transactions on NanoBioscience**, 2023. [IF: 3.9]

[43] Zhuohang Li, **Chao Yan**, Xinmeng Zhang, Gharib Gharibi, Zhijun Yin, Xiaoqian Jiang, Bradley Malin, “Split learning for distributed collaborative training of deep learning models in health informatics.” In *Proceedings of 2023 American Medical Informatics Association Annual Symposium (AMIA’23)*, 2023.

[42] Yongtai Liu, Zhijun Yin, Congning Ni, **Chao Yan**, Zhiyu Wan, Bradley Malin, “Plandemic or Covidiot: Examining rural and urban sentiment difference in COVID-19 related topics on Twitter.” *Journal of Medical Internet Research (JMIR)*, 2023. [IF: 7.4]

[41] **Chao Yan\***, Yao Yan\*, Zhiyu Wan\*, Ziqi Zhang, Larsson Omberg, Justin Guinney, Sean Mooney, Bradley Malin, “A multifaceted benchmarking of synthetic electronic health record generation models.” **Nature Communications**. 2022. [IF: 17.4]

- [40] Abinitha Gourabathina\*, Zhiyu Wan\*, Thomas Brown, **Chao Yan**, Bradley Malin, “Privacy-preserving publishing of individual-level pandemic data based on a game theoretic model.” In Proceedings of 2022 IEEE International Conference on Bioinformatics and Biomedicine (**BIBM’22**), 2022.
- [39] Xinmeng Zhang\*, Yuying Zhao\*, **Chao Yan**, Tyler Derr, You Chen, “Inferring EHR utilization workflows through audit logs.” In Proceedings of 2022 American Medical Informatics Association Annual Symposium (**AMIA’22**), 2022.
- [38] Steve Nyemba, **Chao Yan**, Ziqi Zhang, Amol Rajmane, Pablo Meyer Rojas, Prithwish Chakraborty, Bradley Malin, “Distillation to enhance the portability of risk models across institutions with large patient claims databases.”. In Proceedings of 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining Workshop on Applied Data Science for Healthcare (**KDD DSHealth**), 2022.
- [37] Ziqi Zhang, **Chao Yan**, Bradley Malin, “Keeping synthetic patients on track: feedback mechanisms to mitigate performance drift in longitudinal health data simulation.” Journal of the American Medical Informatics Association (**JAMIA**), 2022. [IF: 6.4]
- [36] 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 Proceedings of 2022 International Conference on Privacy in Statistical Databases (**PSD’22**), 2020.
- [35] Ziqi Zhang, **Chao Yan**, Xinmeng Zhang, Steve Nyemba and Bradley Malin, “Forecasting the Clinical Status of a Patient Through Contrastive Learning.” Journal of the American Medical Informatics Association (**JAMIA**), 2022. [IF: 6.4]
- [34] Yongtai Liu, Zhiyu Wan, **Chao Yan**, Congning Ni, Yevgeniy Vorobeychik, Ellen Clayton, Murat Kantarcioglu, Zhijun Yin and Bradley Malin, “My DNA and ... My Face: How Information Sharing in Social Platforms Generates Attention and Privacy Concerns.” Journal of Medical Internet Research (**JMIR**) Infodemiology, 2022.
- [33] Xiaoge Zhang, Felix Chen, **Chao Yan**, and Indranil Bose. “Towards risk-aware machine learning systems: An overview.” Decision Support Systems (**DSS**), 2022. [IF: 7.5]
- [32] Thomas Brown, **Chao Yan**, Weiyi Xia, Zhijun Yin, Zhiyu Wan, Aris Gkoulalas-Divanis, Murat Kantarcioglu and Bradley Malin, “Dynamically adjusting case-reporting policy to maximize privacy and utility in the face of a pandemic.” Journal of the American Medical Informatics Association (**JAMIA**), 2022. [IF: 6.4]
- [31] Congning Ni, Zhiyu Wan, **Chao Yan**, Yongtai Liu, Ellen Clayton, Bradley Malin and Zhijun Yin, “Beyond the academic opinion on #GeneEditedBabies: Learning how the public’s stance about a controversial scientific event differs across online social platforms.” Journal of Medical Internet Research (**JMIR**), 2022. [IF: 7.4]
- [30] Ziqi Zhang, **Chao Yan** and Bradley Malin, “Membership inference attacks against synthetic healthcare data.” Journal of Biomedical Informatics (**JBIM**), 2022. [IF: 4.5]
- [29] Hannah Mannering, **Chao Yan**, Mhd Alrifai, Daniel France and You Chen, “Assessing collaboration structures in the neonatal intensive care unit before and during the COVID-19 pandemic: a network analysis.” Journal of Medical Internet Research (**JMIR**), 2021. [IF: 7.4]
- [28] Xinmeng Zhang\*, **Chao Yan\***, Bradley Malin, Mayur Patel and You Chen, “Predicting next day discharge via electronic health record access logs.” Journal of the American Medical Informatics Association (**JAMIA**), 2021. [IF: 6.4]

[27] **Chao Yan**, Cheng Gao, Ziqi Zhang, Wencong Chen, Bradley Malin, Wes Ely, Mayur Patel and You Chen, "Predicting Brain Function Status Changes in Critically Ill Patients via Machine Learning." Journal of the American Medical Informatics Association (**JAMIA**), 2021. [IF: 6.4]

[26] **Chao Yan**, Xinmeng Zhang, Cheng Gao, Erin Wilfon, Jonathan Casey, Dannel France, Yang Gong, Mayur Patel, Bradley Malin and You Chen, "The collaboration Structure in COVID-19 Critical Care: A Network Analysis." Journal of Medical Internet Research (**JMIR**) Human Factors, 2021. [IF: 2.7]

[25] Prithwish Chakraborty, ..., **Chao Yan**, et al., "Blending Knowledge in Deep Recurrent Networks for Adverse Event Prediction at Hospital Discharge." In Proceedings of 2021 AMIA Summits on Translational Science (**AMIA Summit'21**), 2021.

[24] Ziqi Zhang, **Chao Yan**, Thomas Lasko, Jimeng Sun and Bradley Malin, "SynTEG: A Framework for Temporal Structured Electronic Health Data Simulation." Journal of the American Medical Informatics Association (**JAMIA**), 2020. *[As Corresponding Author]* [IF: 6.4]

[23] **Chao Yan\***, Ziqi Zhang\*, Steve Nyemba and Bradley Malin, "Generating Electronic Health Records with Multiple Data Types and Constraints." In Proceedings of 2020 American Medical Informatics Association Annual Symposium (**AMIA'20**), 2020.

[22] You Chen, **Chao Yan** and Mayur Patel, "Network Analysis Subtleties in ICU Structures and Outcomes." American Journal of Respiratory and Critical Care Medicine (**AJRCCM**), 2020. [IF: 24.7]

[21] Xinmeng Zhang, **Chao Yan**, Cheng Gao, Bradley Malin and You Chen, "Predicting Missing Values in Medical Data via XGBoost Regression." Journal of Healthcare Informatics Research (**JHIR**), 2020. [IF: 5.9]

[20] Ziqi Zhang\*, **Chao Yan\***, Diego Mesa, Jimeng Sun and Bradley Malin, "Ensuring electronic medical record simulation through better training, modeling, and evaluation." Journal of the American Medical Informatics Association (**JAMIA**), 2020. *[As Corresponding Author]* [IF: 6.4]

***Named one of the best health informatics papers in 2020 by the International Medical Informatics Association.***

[19] **Chao Yan**, Haifeng Xu, Yevgeniy Vorobeychik, Bo Li, Daniel Fabbri and Bradley Malin, "To warn or not to warn: online signaling in audit games." In Proceedings of 36th IEEE International Conference on Data Engineering (**ICDE'20**), 2020.

[18] Liang Tong, Aron Laszka, **Chao Yan**, Ning Zhang and Yevgeniy Vorobeychik, "Finding needles in a moving haystack: Prioritizing alerts with adversarial reinforcement learning." In Proceedings of 34th AAAI Conference on Artificial Intelligence (**AAAI'20**), 2020.

[17] Yongtai Liu, **Chao Yan**, Zhijun Yin, Zhiyu Wan, Weiyi Xia, Murat Kantarcioglu, Yevgeniy Vorobeychik, Ellen Clayton and Bradley Malin, "Biomedical research cohort membership disclosure on social media." In Proceedings of 2019 American Medical Informatics Association Annual Symposium (**AMIA'19**), 2019.

***Distinguished Paper Award!***

[16] Cheng Gao, **Chao Yan**, Bradley Malin and You Chen, "Corpus size influences clinical concept embeddings." In Proceedings of 2019 American Medical Informatics Association Annual Symposium (**AMIA'19**), 2019.

[15] **Chao Yan**, Cheng Gao, Xinmeng Zhang, You Chen and Bradley Malin, "Deep imputation of temporal data." In Proceedings of 2019 IEEE International Conference on Healthcare Informatics (**ICHI'19**), 2019.

[14] Xinmeng Zhang, **Chao Yan**, Cheng Gao, Bradley Malin and You Chen, "XGBoost imputation for time series data." In Proceedings of 2019 IEEE International Conference on Healthcare Informatics (**ICHI'19**), 2019.

- [13] Cheng Gao, **Chao Yan**, Sarah Osmundson, Bradley Malin and You Chen, "A deep learning approach to predict Neonatal Encephalopathy from electronic health records." In Proceedings of 2019 IEEE International Conference on Healthcare Informatics (**ICHI'19**), 2019.
- [12] **Chao Yan**, Bo Li, Yevgeniy Vorobeychik, Aron Laszka, Daniel Fabbri and Bradley Malin, "Database audit workload prioritization via game theory." ACM Transactions on Privacy and Security (**ACM TOPS**), 2019. [IF: 2.3]
- [11] **Chao Yan**, Zhijun Yin, Stanley Xiang, You Chen, Yevgeniy Vorobeychik, Daniel Fabbri, Abel Kho, David Liebovitz and Bradley Malin, "Mining the best observational window to model social phenomena." In Proceedings of 2018 IEEE International Conference on Collaboration and Internet Computing (**CIC'18**), 2018.
- [10] Ruobing Zheng, Guoqiang Wu, **Chao Yan**, Renyu Zhang, Ze Luo and Baoping Yan, "Exploration in mapping kernel-based home range models from remote sensing imagery with conditional adversarial networks." **Remote Sensing**, 2018. [IF: 5.0]
- [9] Thomas Li, Cheng Gao, **Chao Yan**, Sarah Osmundson, Bradley Malin and You Chen, "Predicting neonatal encephalopathy from maternal data in electronic medical records." In Proceedings of 2018 AMIA Summits on Translational Science (**AMIA Summit'18**), 2018.  
**Finalist for Best Student Paper!**
- [8] **Chao Yan**, Bo Li, Yevgeniy Vorobeychik, Aron Laszka, Daniel Fabbri and Bradley Malin, "Get your workload in order: Game theoretic prioritization of database auditing." In Proceedings of 34th IEEE International Conference on Data Engineering (**ICDE'18**), 2018.
- [7] Monica Hedda, Bradley Malin, **Chao Yan** and Daniel Fabbri, "Evaluating the effectiveness of auditing rules for electronic health record systems." In Proceedings of 2017 American Medical Informatics Association Annual Symposium (**AMIA'17**), 2017.
- [6] Laszka, Aron, Yevgeniy Vorobeychik, Daniel Fabbri, **Chao Yan** and Bradley Malin, "A game-theoretic approach for alert prioritization." In Proceedings of 31th AAAI Conference on Artificial Intelligence on Artificial Intelligence for Cyber Security Workshop (**AICS'17**), 2017.
- [5] **Chao Yan**, You Chen, Bo Li, David Liebovitz and Bradley Malin, "Learning clinical workflows to identify subgroups of heart failure patients." In Proceedings of 2016 American Medical Informatics Association Annual Symposium (**AMIA'16**), 2016.
- [4] Cai Gao, **Chao Yan**, Andrew Adamatzky and Yong Deng, "A bio-inspired algorithm for route selection in wireless sensor networks." **IEEE Communications Letters**, 2014. [IF: 4.1]
- [3] **Chao Yan**, Cai Gao, Jianjun Yu, Yong Deng and Kai Nan, "The optimal path tour problem." International Journal of Unconventional Computing (**IJUC**), 2014.
- [2] Cai Gao, **Chao Yan**, Zili Zhang, Yong Hu, Sankaran Mahadevan and Yong Deng. "An amoeboid algorithm for solving linear transportation problem." **Physica A: Statistical Mechanics and Its Applications**, 2014. [IF: 3.3]
- [1] Cai Cao, **Chao Yan**, Daijun Wei, Yong Hu, Sankaran Mahadevan and Yong Deng, "A biologically inspired model for transshipment problem." arXiv preprint arXiv:1401.2181.

## Book Chapters

- [A] Xiaoge Zhang and **Chao Yan**, "Physarum-inspired solutions to network optimization problems." In Shortest Path Solvers. From Software to Wetware, pp. 329-363. Springer, Cham, 2018.

## Manuscripts in Revision, Review or Submission

[A] **Chao Yan**, Haifeng Xu, Yevgeniy Vorobeychik, Bo Li, Murat Kantarcioglu, Bradley Malin, “Robust Bayesian Signaling Games for Database Access Auditing.” [in submission]

[B] Jiayi Tong, Jenna Reys, Chongliang Luo, Martijn Schuemie, **Chao Yan**, ..., David Asch, Yong Chen, “A new end-to-end data aggregation approach for comparing hospital performance without sharing patient-level data.” [in revision at Lancet Digital Health]

[C] Christopher Guardo, Srushti Gandireddy, **Chao Yan**, ..., Wei-Qi Wei, “Generic and Survey Data Improves Performance of Machine Learning Model for Long COVID.” [under review]

[D] Xinmeng Zhang, Kaidi Kang, **Chao Yan**, ..., You Chen, “Enhanced Patient Portal Engagement Associated with Improved Weight Loss Outcomes in Post-Bariatric Surgery Patients.” [in revision at JMIR]

[E] Yanwei Li, **Chao Yan**, Juan Zhao, Henry Ong, Wei-Qi Wei, “Evaluating phenotypic classification performance using synthesized online content.” [under review]

[F] Yubo Feng, **Chao Yan**, Xinmeng Zhang, Xiong-Fei Pan, Hui Cai, Xiao-Ou Shu, Qiuyin Cai, Wei Zheng, Danxia Yu, You Chen, “Predicting the risk of type 2 diabetes through machine learning of blood metabolites.” [in submission]

## Projects in Progress with a Leading Role

[A] Large language model powered drug repurposing and phenotyping

[B] Association study between EHR engagement disparities and patients’ outcomes

[C] Generating and evaluating synthetic HIV observational datasets

[D] Real time brain dysfunction modeling in COVID ICUs using daily observations from ~70 institutions

[E] High-throughput phenotyping based on graph neural networks

[F] Benchmarking of structural synthetic EHR generation models using multi-site data

[G] EHR access logs usage in promoting outcome predictions (e.g. real-time discharge prediction)

## MEDIA ABOUT RESEARCH

- 
- |  |             |
|--|-------------|
| ▪ <b>Project: Generative AI-powered drug repurposing for Alzheimer’s disease</b> | <b>2024</b> |
| <i>Reported by Vanderbilt University Medical Center Reporter</i>                 |             |
| ▪ <b>Project: Difference in health professionals’ engagement in HER</b>          |             |
| <i>Reported by Discovery in Medicine, VUMC</i>                                   |             |
| <i>Reported by Vanderbilt University Medical Center Reporter</i>                 |             |
| <i>Reported by POLITICO MAGAZINE, Medium, Medscape, MDedge</i>                   |             |
|  | <b>2023</b> |
| ▪ <b>Project: ICU acute brain dysfunction prediction</b>                         |             |
| <i>Reported by Vanderbilt University Medical Center Reporter</i>                 |             |
|  | <b>2021</b> |
| ▪ <b>Project: ICU structure investigation in COVID pandemic</b>                  |             |
| <i>Reported by Vanderbilt University Medical Center Reporter</i>                 |             |
|  | <b>2021</b> |
| ▪ <b>Project: Health data simulation</b>   |             |

<i>Cover story of Vanderbilt University Medical Center Breakthrough Page</i>	<b>2019</b>
<i>Reported by MedicalXpress</i>	<b>2019</b>
▪ <b>Project: Privacy Disclosure in Biomedical Research Cohort</b>	
<i>Reported by Vanderbilt University Medical Center Reporter</i>	<b>2019</b>

## CONFERENCE PRESENTATION

---

### Oral Presentations:

American Medical Informatics Association (AMIA) Annual Symposium: 2016, 2019, 2021, 2022, 2023  
 American Medical Informatics Association (AMIA) Informatics Summit: 2024  
 IEEE International Conference in Data Engineering (ICDE): 2018, 2020  
 IEEE International Conference on Collaboration and Internet Computing (CIC): 2018

## PROFESSIONAL SERVICE

---

### Invited Talks

- Ensuring electronic medical record simulation through better training, modeling, and evaluation. [JAMIA Journal Club Webinar](#), Jan 6, 2020.
- To warn or not to warn: online signaling in audit games. [AI for Social Good Workshop](#) at Harvard University, July 20, 2020

### Program Committee

- [Workshop on Data Privacy and Data Analysis in Healthcare Systems with IEEE ICHI](#). 2024.
- [Workshop on Security and Privacy of Machine Learning with ICML](#). 2019.
- [Workshop on Adversarial Machine Learning in Real-World Computer Vision Systems with IEEE CVPR](#). 2019.

### Session Moderation

- [JAMIA Journal Club Webinar Series Organized by AMIA](#). July, August 2021.

### Judges

- The selection of the annual best papers invited by the International Medical Informatics Association (IMIA). 2022, 2023, 2024.

### Journal Reviewing Activities

- Annals of Internal Medicine. 2023,2024.
- JAMA Network Open. 2023,2024.
- Npj Digital Medicine. 2022,2023,2024.
- Nature Machine Intelligence. 2020.
- Nature Communications. 2022,2023.
- Journal of the American Medical Informatics Association. 2016, 2019, 2020, 2021, 2022, 2023,2024. [Student Editorial Board 2021-2022]
- Journal of Medical Internet Research. 2021, 2022, 2023.
- BMC Medical Research Methodology. 2019.
- BMC Medical Informatics and Decision Making. 2018,2022.
- Journal of Biomedical Informatics. 2018, 2019, 2022, 2023.
- Journal of Data Science and Its Applications. 2020.

### Conference Reviewing Activities

- Neural Information Processing Systems (NeurIPS). 2020, 2021.
- AAAI Conference on Artificial Intelligence (AAAI). 2017, 2019, 2020.

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021, 2022.
- IEEE International Conference on Data Mining (ICDM). 2016, 2017, 2018, 2019.
- International Joint Conference on Artificial Intelligence (IJCAI). 2019,2023.
- Network and Distributed System Security Symposium (NDSS). 2018.
- ACM Conference on Data and Application Security and Privacy (CODASPY). 2019, 2020.
- IEEE International Conference on Big Data. 2017, 2019.
- AMIA Annual Symposium. 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023.
- AMIA Informatics Summit. 2019, 2020, 2021.
- Machine Learning for Healthcare Conference (ML4HC). 2018.
- IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2019.
- ISMB/ECCB, 2019.
- IEEE International Conference on Intelligence and Security Informatics (ISI). 2017.
- Australasian Data Mining Conference. 2017.

(Last updated on April, 2023)