

Stanford



Jason Fries

Research Engineer, Med/BMIR

Bio

CURRENT ROLE AT STANFORD

I'm currently working as a staff research scientist in the Shah Lab and research scientist at Snorkel AI. My interests fall in the intersection of computer science and medical informatics. My research interests include:

- Machine learning with limited labeled data, e.g., weak supervision, self-supervision, and few-shot learning.
- Multimodal learning, e.g., combining text, imaging, video and electronic health record data for improving clinical outcome prediction
- Human-in-the-loop machine learning systems.
- Knowledge graphs and their use in improving representation learning

PROJECTS

- Weakly supervised classification of rare aortic valve malformations using unlabeled cardiac MRI sequences - Stanford University (1/1/2018 - 7/1/2019)
- Snorkel: Rapid Training Data Creation with Weak Supervision - Stanford University (6/1/2016 - 12/1/2018)

SERVICE, VOLUNTEER, AND COMMUNITY WORK

- Co-organizer for Machine Learning for Health Workshop @ NeurIPS (12/2016 - 12/2018)
- Area Chair @ Machine Learning for Healthcare Conference (MLHC) (2019 - 2021)

Publications

PUBLICATIONS

- **Ontology-driven weak supervision for clinical entity classification in electronic health records.** *Nature communications*
Fries, J. A., Steinberg, E., Khattar, S., Fleming, S. L., Posada, J., Callahan, A., Shah, N. H.
2021; 12 (1): 2017
- **Weakly supervised classification of rare aortic valve malformations using unlabeled cardiac MRI sequences** *Nature Communications*
Fries, J. A., Varma, P., Chen, V. S., Xiao, K., Tejada, H., Saha, P., Dunnmon, J., Chubb, H., Maskatia, S., Fiterau, M., Delp, S., Ashley, E., Ré, et al
2019; 10
- **Snorkel: Rapid Training Data Creation with Weak Supervision** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., Re, C.
2017; 11 (3): 269–82
- **Exploring the Potential of Large Language Models in Neurology, Using Neurologic Localization as an Example.** *Neurology. Clinical practice*
Chiang, C., Fries, J. A.
2024; 14 (3): e200311

- **Scalable Approach to Consumer Wearable Postmarket Surveillance: Development and Validation Study.** *JMIR medical informatics*
Yoo, R. M., Viggiano, B. T., Pundi, K. N., Fries, J. A., Zahedivash, A., Podchiyska, T., Din, N., Shah, N. H.
2024; 12: e51171
- **Characterizing the limitations of using diagnosis codes in the context of machine learning for healthcare.** *BMC medical informatics and decision making*
Guo, L. L., Morse, K. E., Aftandilian, C., Steinberg, E., Fries, J., Posada, J., Fleming, S. L., Lemmon, J., Jessa, K., Shah, N., Sung, L.
2024; 24 (1): 51
- **The Stanford Medicine data science ecosystem for clinical and translational research.** *JAMIA open*
Callahan, A., Ashley, E., Datta, S., Desai, P., Ferris, T. A., Fries, J. A., Halaas, M., Langlotz, C. P., Mackey, S., Posada, J. D., Pfeffer, M. A., Shah, N. H.
2023; 6 (3): ooad054
- **Self-supervised machine learning using adult inpatient data produces effective models for pediatric clinical prediction tasks.** *Journal of the American Medical Informatics Association : JAMIA*
Lemmon, J., Guo, L. L., Steinberg, E., Morse, K. E., Fleming, S. L., Aftandilian, C., Pfohl, S. R., Posada, J. D., Shah, N., Fries, J., Sung, L.
2023
- **The shaky foundations of large language models and foundation models for electronic health records.** *NPJ digital medicine*
Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., Pfeffer, M. A., Fries, J., Shah, N. H.
2023; 6 (1): 135
- **EHR foundation models improve robustness in the presence of temporal distribution shift.** *Scientific reports*
Guo, L. L., Steinberg, E., Fleming, S. L., Posada, J., Lemmon, J., Pfohl, S. R., Shah, N., Fries, J., Sung, L.
2023; 13 (1): 3767
- **Evaluation of Feature Selection Methods for Preserving Machine Learning Performance in the Presence of Temporal Dataset Shift in Clinical Medicine.** *Methods of information in medicine*
Lemmon, J., Guo, L. L., Posada, J., Pfohl, S. R., Fries, J., Fleming, S. L., Aftandilian, C., Shah, N., Sung, L.
2023
- **Efficient Diagnosis Assignment Using Unstructured Clinical Notes**
Blankemeier, L., Fries, J., Tinn, R., Preston, S., Shah, N., Chaudhari, A., Boyd-Graber, J., Okazaki, N., Rogers, A.
ASSOC COMPUTATIONAL LINGUISTICS-ACL.2023: 485-494
- **Investigating real-world consequences of biases in commonly used clinical calculators.** *The American journal of managed care*
Yoo, R. M., Dash, D., Lu, J. H., Jenkins, J. Z., Rabbani, N., Fries, J. A., Shah, N. H.
2023; 29 (1): e1-e7
- **A computational approach to measure the linguistic characteristics of psychotherapy timing, responsiveness, and consistency.** *Npj mental health research*
Miner, A. S., Fleming, S. L., Haque, A., Fries, J. A., Althoff, T., Wilfley, D. E., Agras, W. S., Milstein, A., Hancock, J., Asch, S. M., Stirman, S. W., Arnow, B. A., Shah, et al
2022; 1 (1): 19
- **Perspective Toward Machine Learning Implementation in Pediatric Medicine: Mixed Methods Study.** *JMIR medical informatics*
Alexander, N., Aftandilian, C., Guo, L. L., Plenert, E., Posada, J., Fries, J., Fleming, S., Johnson, A., Shah, N., Sung, L.
2022; 10 (11): e40039
- **Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine.** *Scientific reports*
Guo, L. L., Pfohl, S. R., Fries, J., Johnson, A. E., Posada, J., Aftandilian, C., Shah, N., Sung, L.
2022; 12 (1): 2726
- **Dataset Debt in Biomedical Language Modeling**
Fries, J., Seelam, N., Altay, G., Weber, L., Kang, M., Datta, D., Su, R., Garda, S., Wang, B., Ott, S., Samwald, M., Kusa, W., Assoc Computat Linguist
ASSOC COMPUTATIONAL LINGUISTICS-ACL.2022: 137-145
- **PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts**
Bach, S. H., Sanh, V., Yong, Z., Webson, A., Raffel, C., Nayak, N., Sharma, A., Kim, T., Bari, M., Fevry, T., Alyafeai, Z., Dey, M., Santilli, et al
ASSOC COMPUTATIONAL LINGUISTICS-ACL.2022: 93-104
- **Systematic Review of Approaches to Preserve Machine Learning Performance in the Presence of Temporal Dataset Shift in Clinical Medicine.** *Applied clinical informatics*

Guo, L. L., Pfohl, S. R., Fries, J., Posada, J., Fleming, S. L., Aftandilian, C., Shah, N., Sung, L.
2021; 12 (4): 808-815

- **Assessment of Extractability and Accuracy of Electronic Health Record Data for Joint Implant Registries.** *JAMA network open*
Giori, N. J., Radin, J., Callahan, A., Fries, J. A., Halilaj, E., Re, C., Delp, S. L., Shah, N. H., Harris, A. H.
2021; 4 (3): e211728
- **Trove: Ontology-driven weak supervision for medical entity classification.** *ArXiv*
Fries, J. A., Steinberg, E., Khattar, S., Fleming, S. L., Posada, J., Callahan, A., Shah, N. H.
2020
- **Estimating the efficacy of symptom-based screening for COVID-19.** *NPJ digital medicine*
Callahan, A., Steinberg, E., Fries, J. A., Gombar, S., Patel, B., Corbin, C. K., Shah, N. H.
2020; 3 (1): 95
- **Measure what matters: Counts of hospitalized patients are a better metric for health system capacity planning for a reopening.** *Journal of the American Medical Informatics Association : JAMIA*
Kashyap, S., Gombar, S., Yadlowsky, S., Callahan, A., Fries, J., Pinsky, B. A., Shah, N. H.
2020
- **Assessing the accuracy of automatic speech recognition for psychotherapy.** *NPJ digital medicine*
Miner, A. S., Haque, A., Fries, J. A., Fleming, S. L., Wilfley, D. E., Terence Wilson, G., Milstein, A., Jurafsky, D., Arnow, B. A., Stewart Agras, W., Fei-Fei, L., Shah, N. H.
2020; 3: 82
- **Assessing the accuracy of automatic speech recognition for psychotherapy** *NPJ DIGITAL MEDICINE*
Miner, A. S., Haque, A., Fries, J. A., Fleming, S. L., Wilfley, D. E., Wilson, G., Milstein, A., Jurafsky, D., Arnow, B. A., Agras, W., Li Fei-Fei, Shah, N. H.
2020; 3 (1)
- **Language models are an effective representation learning technique for electronic health record data.** *Journal of biomedical informatics*
Steinberg, E. n., Jung, K. n., Fries, J. A., Corbin, C. K., Pfohl, S. R., Shah, N. H.
2020: 103637
- **Assessing the accuracy of automatic speech recognition for psychotherapy.** *NPJ digital medicine*
Miner, A. S., Haque, A. n., Fries, J. A., Fleming, S. L., Wilfley, D. E., Terence Wilson, G. n., Milstein, A. n., Jurafsky, D. n., Arnow, B. A., Stewart Agras, W. n., Fei-Fei, L. n., Shah, N. H.
2020; 3 (1): 82
- **Snorkel: rapid training data creation with weak supervision.** *The VLDB journal : very large data bases : a publication of the VLDB Endowment*
Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., Re, C.
2020; 29 (2): 709–30
- **The accuracy vs. coverage trade-off in patient-facing diagnosis models.** *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*
Kannan, A., Fries, J. A., Kramer, E., Chen, J. J., Shah, N., Amatriain, X.
2020; 2020: 298–307
- **Cardiac Imaging of Aortic Valve Area from 34,287 UK Biobank Participants Reveal Novel Genetic Associations and Shared Genetic Comorbidity with Multiple Disease Phenotypes.** *Circulation. Genomic and precision medicine*
Córdova-Palomera, A. n., Tcheandjieu, C. n., Fries, J. n., Varma, P. n., Chen, V. S., Fiterau, M. n., Xiao, K. n., Tejada, H. n., Keavney, B. n., Cordell, H. J., Tanigawa, Y. n., Venkataraman, G. n., Rivas, et al
2020
- **Estimating the efficacy of symptom-based screening for COVID-19.** *NPJ digital medicine*
Callahan, A., Steinberg, E., Fries, J. A., Gombar, S., Patel, B., Corbin, C. K., Shah, N. H.
2020; 3: 95
- **Medical device surveillance with electronic health records.** *NPJ digital medicine*
Callahan, A. n., Fries, J. A., Ré, C. n., Huddleston, J. I., Giori, N. J., Delp, S. n., Shah, N. H.
2019; 2: 94

- **Multi-Resolution Weak Supervision for Sequential Data**

Sala, F., Varma, P., Fries, J., Fu, D. Y., Sagawa, S., Khattar, S., Ramamoorthy, A., Xiao, K., Fatahalian, K., Priest, J., Re, C., Wallach, H., Larochelle, et al
NEURAL INFORMATION PROCESSING SYSTEMS (NIPS).2019

- **ShortFuse: Biomedical Time Series Representations in the Presence of Structured Information.** *Proceedings of machine learning research*

Fiterau, M. n., Bhooshan, S. n., Fries, J. n., Bournhonesque, C. n., Hicks, J. n., Halilaj, E. n., Ré, C. n., Delp, S. n.
2017; 68: 59–74

- **Brundlefly at SemEval-2016 Task 12: Recurrent Neural Networks vs. Joint Inference for Clinical Temporal Information Extraction** *Jason Alan Fries*

Fries, J. A.
2016: 1274–79

PRESENTATIONS

- Workshop: Rapid Biomedical Knowledge Base Construction from Unstructured Data - Stanford University (10/21/2017 - October 25, 2017)
- Weakly Supervised Learning in Medicine - Stanford MedAI (4/15/2021)