

RESEARCH INTEREST

Transfer Learning, Representation Learning, AI-Human Alignment, Text Analytics

EDUCATION

- **Georgia Institute of Technology** Atlanta, GA, USA
Ph.D. in Computational Science and Engineering Aug 2018 - Present
- **Georgia Institute of Technology** Atlanta, GA, USA
M.S. in Computational Science and Engineering Aug 2018 - May 2022
- **Seoul National University** Seoul, South Korea
B.S. in Electrical and Computer Engineering, Minor in Computer Science & Engineering Mar 2011 - Feb 2018

PUBLICATIONS

- **Dongjin Choi, Andy Xiang, Ozgur Ozturk, Deep Shrestha, Barry Drake, Hamid Haidarian, Faizan Javed, Haesun Park**: “WellFactor: Patient Profiling using Integrative Embedding of Healthcare Data”, International Conference on Big Data, 2023. **Best Paper Award**
- **Dongjin Choi, Andy Xiang, Ozgur Ozturk, Deep Shrestha, Barry Drake, Hamid Haidarian, Faizan Javed, Haesun Park**: “Patient Clustering via Integrated Profiling of Clinical and Digital Data”, International Conference on Information and Knowledge Management (CIKM), 2023.
- **Dongjin Choi, Barry Drake, Haesun Park**: “Co-embedding Multi-type Data for Information Fusion and Visual Analytics”, International Conference on Information Fusion, 2023.
- **Dongjin Choi, Sara Evensen, Çağatay Demiralp, Estevam Hruschka**: “TagRuler: Interactive Tool for Span-Level Data Programming by Demonstration”, The Web Conference Demo., 2021.
- **Zijie J. Wang, Dongjin Choi, Shenyu Xu, Diyi Yang**: “Putting Humans in the Natural Language Processing Loop: A Survey”, EACL Workshop on Bridging HCI and NLP, 2021.
- **Sara Evensen, Chang Ge, Dongjin Choi, and Çağatay Demiralp**: “Data Programming by Demonstration: A Framework for Interactively Learning Labeling Functions”, Arxiv Preprint, 2020.
- **Hannah Kim, Dongjin Choi, Barry Drake, Alex Endert, and Haesun Park**: “TopicSifter: Interactive Search Space Reduction Through Targeted Topic Modeling”, IEEE VIS, 2019.
- **Dongjin Choi, Jun-gi Jang, and U Kang**: “S³CMTF: Fast, Accurate, and Scalable Method for Incomplete Coupled Matrix-Tensor Factorization”, PLOS ONE, 2019.
- **Dongjin Choi, and Lee Sael**: “SNeCT: Scalable Network Constrained Tucker Decomposition for Multi-Platform Data Profiling”, IEEE TCBB, 2019.
- **Jun-gi Jang, Dongjin Choi, and U Kang**: “Zoom-SVD: Fast and Memory Efficient Method for Time Ranged Singular Value Decomposition”, 27th ACM International Conference on Information and Knowledge Management (CIKM), 2018.
- **Jungwoo Lee, Dongjin Choi, and Lee Sael**: “CTD: Fast, Accurate, and Interpretable Method for Static and Dynamic Tensor Decompositions”, PLOS ONE, 2018.
- **Woojung Jin, Dongjin Choi, Youngjin Kim, and U Kang**: “Activity Prediction from Sensor Data using Convolutional Neural Networks and an Efficient Compression Method”, KIISE Journal, 2018.

INDUSTRY EXPERIENCE

- **Upstage** AI Research Intern
Mentor: Hwalsuk Lee May 2021 - Aug 2021
 - **Role**: Proposed and developed a system for data cleansing for OCR annotated data sets. Focused on improving the accuracy of outlier data samples.
- **Megagon Labs** Research Scientist Intern
Mentor: Çağatay Demiralp May 2020 - Aug 2020
 - **Role**: Proposed and developed a novel system for data programming by demonstration, focusing on span-level annotation. Focused on enhancing productivity and accuracy in data labeling.

- **NAVER Corp.** Data Science Intern
Mentor: Jaegul Choi *Apr 2018 - Jul 2018*
 - **Role:** Proposed and implemented an LSTM-based search model. The project involved designing and training deep learning models to enhance personalized search capabilities.

AWARDS AND HONORS

- **Best Paper Award:** International Conference on Big Data, for *WellFactor: Patient Profiling using Integrative Embedding of Healthcare Data* Dec 2023
- **Travel Grant Award:** International Conference on Big Data Dec 2023
- **Honorable Mention:** Humantech Paper Award, Samsung Feb 2018
- **Bronze Prize:** Humantech Paper Award, top 6 in the CS division, Samsung Feb 2017

TEACHING

- **Georgia Institute of Technology** Atlanta, GA
Graduate Teaching Assistant, Numerical Linear Algebra (CSE/MATH 6643) *Spring 2020*
 - **Tasks:** Grading, holding office hours, lecturing, proctoring.
 - **Instructor:** Haesun Park
- **Georgia Institute of Technology** Atlanta, GA
Graduate Teaching Assistant, Numerical Analysis I (CX/MATH 4640) *Fall 2019*
 - **Tasks:** Grading, holding office hours, lecturing, proctoring as head TA.
 - **Instructor:** Haesun Park

TECHNICAL SKILLS AND EXPERTISE

- **Large Language Models:**
 - **Transfer Learning/Fine Tuning:** RLHF, LoRA
 - **Utilizing External Sources:** Retrieval-Augmented Generation (RAG)
- **Machine Learning & AI:**
 - **Deep Learning Frameworks:** Tensorflow, Keras, PyTorch (Experienced in building and training large models)
 - **Boosted Trees:** XGBoost, LightGBM
 - **Model Explanation Methods:** Saliency maps, TCAV, LIME
 - **Other Libraries:** LangChain, pinecone, transformers, PyG
- **Web Development & Data Visualization:**
 - **Front-end Development:** React.js, streamlit
 - **Data Visualization:** D3.js, plotly (Proficient in creating interactive data visualizations)