

Javid Dadashkarimi

Email : javid.dadashkarimi@yale.edu

LinkedIn: <https://www.linkedin.com/in/javid-dadashkarimi-0b26b3199/>

<https://github.com/dadashkarimi>

<https://dadashkarimi.github.io>

Mobile : +1-203-781-6115

EDUCATION

- **Yale University** New Haven, CT
PhD in Computer Science; GPA: 4 *Sep 2017 - Sep 2023*
Courses: Object-Oriented Programming, NeuroImaging Statistics, Applied Data Mining & Machine Learning, Optimization Techniques.
- **University of Tehran** Tehran, Iran
Master of Software Engineering; GPA: 3.7 *Sep 2012 - Sep 2015*
- **University of Tehran** Tehran, Iran
Bachelor of Software Engineering; GPA: 3.5 *Sep 2008 - Sep 2012*

SKILLS SUMMARY

- **Languages:** Java, JavaScript, NodeJS, C++, Python, PyTorch, TensorFlow, C, R, Scala, SQL, Numpy, Pandas, Unix scripting
- **Tools:** Kubernetes, Docker, Springboot, Apache Spark, Jupyter Notebook, GIT, MATLAB, XCode, Postgres, AWS, MongoDB

EXPERIENCE

- **Yale Institute of Network Sciences** New Haven, CT
Graduate Research Assistant *Sep 2017 - Current*
 - **Cross Atlas Remapping via Optimal Transport (CAROT):** As a key team member for the development of <https://carotproject.com>, I helped build a groundbreaking platform for brain connectomics studies. Leveraging my expertise in Python for back-end programming, I implemented the website using Node.js and Mongo DB. Our innovative platform democratizes science by enabling unprecedented sample sizes while preserving participant privacy, transforming data from one brain atlas to another without requiring raw data. We also provide open-source datasets, driving collaborative efforts and advancing our understanding of the human brain. Through this project, I developed skills in data science, artificial intelligence (AI), statistical analysis, computational neuroscience, machine learning, trustworthiness, privacy, and software engineering, and I am excited to apply these skills to new challenges.
 - **Brain Connectivity Viewer:** I contributed to the development of a web-based visualization toolkit for complex connectome data as part of the BioImage Suite Web project. The toolkit, developed using JavaScript, Node.js, Python, C++, and WebAssembly, enables easy visualization of complex connectome data from multiple sources and provides a range of visualizations across multiple levels of feature summarization, with significant implications for connectomics research. My works resulted in an online platform: BioImage Suit Web.
 - **Cross-lingual Information Retrieval and Neural Rankings:** I led the development of a cross-language document ranking tool that utilized large language models, even for resource-lean languages like Persian, Arabic, and Swahili. To overcome limited translation resources, we designed a two-step ranking algorithm that adjusted probabilities based on top-retrieved documents in the source language and relevant documents in the target language. Implementing learning-to-rank algorithms led to significant improvements, as reported in our paper. Our industry-standard solution is available on GitHub.
- **BlackBox AI** Remote
Machine Learning Engineer *Feb 2023 - Current*
 - **Generating Code by Retrieving the Docs:** As a Machine Learning Engineer, I specialize in creating tools that automate code generation using large language models. By incorporating information retrieval techniques, I am able to identify relevant documentation to produce robust and accurate functions, making developers more efficient and productive. My goal is to revolutionize software development and streamline workflows for businesses. I stay up-to-date with the latest academic publications in machine learning to continually improve algorithm design and performance.
- **PathIvy** Remote
Academic Research Associate *March 2023 - Current*
 - **Mentoring High School Students in AI:** As a mentor, I guide and prepare high school students for real-world work in machine learning and programming, with a particular focus on neuroscience and the brain. I impart critical skills in programming, math, statistics, and probability, and teach students how to apply these skills to real-world problems, including natural language processing, information retrieval, and computational neuroscience. Through this work, I am able to give back to the community, while also fostering the next generation of machine learning leaders.
- **University of Tehran** Tehran, Iran
Software Engineer *Jan 2012 - Feb 2013*
 - **Java Developer:** As a skilled software developer, I successfully designed and developed an automated framework for a commodity importing company that improved customer satisfaction, transparency, and privacy. By leveraging multiple technologies such as Apache Maven, MySQL, and JavaScript, I created a powerful platform that allowed customers to easily track the status of their applications and validate whether they met the country's standards. This project not only showcased my ability to work with cutting-edge technologies but also demonstrated my commitment to delivering real-world solutions that have a positive impact on end-users.

ACADEMIC PROJECTS

- **Trustworthiness in Biomedical Datasets:** As a biomedical data scientist, I led the development of three techniques to improve prediction performance while maintaining high feature similarities in biomedical datasets, highlighting the prevalence of “enhancement attacks” in machine learning research. Our work demonstrates the importance of robust data sharing and tracking and has important implications for the development of more reliable and trustworthy machine learning models in healthcare, as detailed in our preprint available at arXiv.
- **Convolutional Neural Networks:** As a teaching assistant with a deep understanding of image convolution operations and their relevance to convolutional neural networks (CNNs) in image processing. I designed and implemented a problem set on advanced edge detection techniques for image classification using CNNs, resulting in significant improvements in student understanding.
- **Autonomous Taxi Driver with Reinforcement Learning:** I applied reinforcement learning and Q-learning, a popular model-free reinforcement learning algorithm, using OpenAI Gym to train an agent to play a game, which involved generating the environment, collecting experiences, updating Q-values, and selecting actions. Through this hands-on experience, I gained a solid understanding of the practical application of reinforcement learning and Q-learning.
- **Variational Autoencoders and Hand Written Digits:** I created a Deep Learning model using Variational Autoencoders (VAEs) with Keras for image generation, as part of my intermediate machine learning course. I implemented VAEs and denoising autoencoders to generate and clean images, fine-tuning the model for high accuracy. This project gave me practical experience with Keras and a solid understanding of deep learning.
- **Graphical Lasso for Stock Market:** I have extensive experience analyzing multivariate data sets and working with Gaussian graphical models. Specifically, I have expertise in estimating graph structures using the graphical lasso and applying this technique to real-world equity data from Yahoo finance. I also used this approach to select a diverse range of ETFs representing various countries. Through this work, I developed strong skills in model building, data science, and their practical applications in economics.
- **Object-Oriented Programming and Block Chain in C++:** I demonstrated my proficiency in programming and problem-solving by completing a challenging programming course where I simulated a population of agents attempting to reach consensus on a choice value using Blockchain consensus algorithms. I gained practical experience in developing efficient code and applying algorithmic solutions to complex problems including implementing Blockchain.
- **Zero-shot Transfer Learning for Semantic Parsing:** In our project, we propose a method for zero-shot semantic parsing using feature transfer. Our approach learns the shared space between multiple domains based on domain label predictions. We show that cross-domain adversarial attacks can improve accuracy by identifying useful examples even from dissimilar domains.
- **Method and System for Information Retrieval:** Developed and co-invented the Perspective-Based Search System, which allows users to select and view search results based on specific perspectives related to their search topic. This patent-pending method provides users with greater clarity in inferring their own search intent without the need to submit complex or multiple queries. The patent for this invention is available online at US Patent.
- **English-Persian Machine Translator:** Curated bilingual texts in English and Farsi to improve the performance of the Faraazin machine translator through crowdsourcing. Leveraged C# backend software to train translation models using word co-occurrences in sentence pairs, and actively contributed to the development of the online platform at <http://faraazin.ir>.

PRESENTATIONS

- Stacking multiple optimal transport policies to map functional connectomes, **57th IEEE Annual Conference on Information Sciences and Systems**, Baltimore, March 22th, 2023.
- Cross atlas remapping via optimal transport, Evelina Fedorenko’s Language Lab at **McGovern Institute for Brain Research at MIT**, Boston, August 8th, 2023.
- Cross atlas remapping via optimal transport, at **Martinos Center at Massachusetts General Hospital at Harvard University**, August 15th, Boston, 2023.
- Data-driven mapping between functional connectomes using optimal transport, at Computational Radiology Laboratory at **Boston Children Hospital at Harvard University**, Jan 24th, Boston, 2023.
- Meet the Inference, Information and Decision Lab: Cross Atlas Remapping via Optimal Transport, **Yale Institute of Network Sciences**, New Haven, April 6th, 2022.

HONORS AND AWARDS

- Best Paper Award, Graphs in Biomedical Imaging, MICCAI, Singapore, 2022.
- MICCAI Student Travel Award, 2022
- Brain Initiative Trainee Award, Virtual, 2020.
- Ranked 331th among batch of 200k students at university entrance exam, Tehran, Iran.
- Google Student Travel Award, 2016
- SIGIR Student Travel Award, 2016

TEACHING

- **Intermediate Machine Learning (S&DS 365/665)** by John Lafferty, Spring 2022.
- **Introduction to Machine Learning (CPSC 481/581)** by Andre Wibisono, Fall 2021.
- **Stochastic Processes (S&DS 351)** by Amin Karbasi, Spring 2020.

PUBLICATIONS

- **Stacking multiple optimal transport policies to map functional connectomes** Javid Dadashkarimi, Matthew Rosenblatt, Amin Karbasi, Dustin Scheinost. *IEEE Annual Conference on Information Sciences and Systems*, Baltimore, 2023.
- **Enhancement attacks in biomedical machine learning** Matthew Rosenblatt, Javid Dadashkarimi, Dustin Scheinost. *arXiv*, 2023.
- **Transforming connectomes to any parcellation via graph matching** Qinghao Liang, Javid Dadashkarimi, Wei Dai, Amin Karbasi, Joseph Chang, Harrison H. Zhou, and Dustin Scheinost, *International Workshop on Graphs in Biomedical Imaging*, — **Best Paper Award**, Singapore, 2022
- **Combining Multiple Atlases to Estimate Data-Driven Mappings Between Functional Connectomes Using Optimal Transport**, Javid Dadashkarimi, Amin Karbasi, Dustin Scheinost, *MICCAI*, — **Early Acceptance (top 13 %)**, Singapore, 2022 .
- **Machine learning and prediction in fetal, infant, and toddler neuroimaging: a review and primer**, Dustin Scheinost, Angeliki Pollatou, Alexander J Dufford, Rongtao Jiang, Michael C Farruggia, Matthew Rosenblatt, Hannah Peterson, Raimundo X Rodriguez, Javid Dadashkarimi, Qinghao Liang, Wei Dai, Maya L Foster, Chris C Camp, Link Tejavibulya, Brendan D Adkinson, Huili Sun, Jean Ye, Qi Cheng, Marisa N Spann, Max Rolison, Stephanie Noble, Margaret L Westwater, *Biological Psychiatry*, 2022, Impact Factor = 13.4
- **Altered brain dynamics across bipolar disorder and schizophrenia revealed by overlapping brain states** Jean Ye, Huili Sun, Siyuan Gao, Javid Dadashkarimi, Matthew Rosenblatt, Raimundo X Rodriguez, Saloni Mehta, Rongtao Jiang, Stephanie Noble, Margaret L Westwater, Dustin Scheinost, *medRxiv*, 2022.
- **Cross Atlas Remapping via Optimal Transport (CAROT): Creating connectomes for any atlas when raw data is not available** Javid Dadashkarimi, Amin Karbasi, Qinghao Liang, Matthew Rosenblatt, Stephanie Noble, Maya Foster, Raimundo Rodriguez, Brendan Adkinsone, Jean Yee, Huili Sun, Chris Camp, Michael Farruggia, Link Tejavibulya, Wei Dai, Rongtao Jiang, Angeliki Pollatou, Dustin Scheinost *bioRxiv*, 2022.
- **Predicting the future of neuroimaging predictive models in mental health** Link Tejavibulya, Max Rolison, Siyuan Gao, Qinghao Liang, Hannah Peterson, Javid Dadashkarimi, Michael C Farruggia, C Alice Hahn, Stephanie Noble, Sarah D Lichenstein, Angeliki Pollatou, Alexander J Dufford, Dustin Scheinost, *Molecular Psychiatry*, 2022, Impact Factor = 13.4.
- **Data-driven mapping between functional connectomes using optimal transport**, Javid Dadashkarimi, Amin Karbasi, Dustin Scheinost, *MICCAI*, 2021, — **Acceptance Rate = 29%**.
- **Functional connectivity during frustration: a preliminary study of predictive modeling of irritability in youth**, Dustin Scheinost, Javid Dadashkarimi, Emily S Finn, Caroline G Wambach, Caroline MacGillivray, Alexandra L Roule, Tara A Niendam, Daniel S Pine, Melissa A Brotman, Ellen Leibenluft, Wan-Ling Tseng, *Nature NPP*, 2021, Impact Factor = 8.29.
- **Transdiagnostic, connectome-based prediction of memory constructs across psychiatric disorders**, Daniel S Barron, Siyuan Gao, Javid Dadashkarimi, Abigail S Greene, Marisa N Spann, Stephanie Noble, Evelyn MR Lake, John H Krystal, R Todd Constable, Dustin Scheinost, *Cerebral Cortex*, 2021, Impact Factor = 4.8.
- **A hitchhiker's guide to working with large, open-source neuroimaging datasets**, Corey Horien, Stephanie Noble, Abigail S Greene, Kangjoo Lee, Daniel S Barron, Siyuan Gao, David O'Connor, Mehraveh Salehi, Javid Dadashkarimi, Xilin Shen, Evelyn MR Lake, R Todd Constable, Dustin Scheinost, *Nature Human Behaviour*, 2021, Impact Factor = 24.25.
- **Method and system for information retrieval**, Shayan Aliakbar Tabrizi, Azadeh Shakery, MohammadAli Tavallaei, Javid Dadashkarimi, *US Patent*, 2020.

- **A mass multivariate edge-wise approach for combining multiple connectomes to improve the detection of group differences**, Javid Dadashkarimi, Siyuan Gao, Erin Yeagle, Stephanie Noble, Dustin Scheinost, *International Workshop on Connectomics in Neuroimaging*, 2019, — **Best Poster Award**.
- **Dimension projection among languages based on pseudo-relevant documents for query translation**, Javid Dadashkarimi, Mahsa S Shahshahani, Amirhossein Tebbifakhr, Hesham Faili, Azadeh Shakery, *ECIR*, 2016, — **Acceptance Rate = 23%**.
- **An expectation-maximization algorithm for query translation based on pseudo-relevant documents**, Javid Dadashkarimi, Azadeh Shakery, Hesham Faili, Hamed Zamani, *Information Processing & Management*, 2016, Impact Factor = 6.2.
- **Learning to Weight Translations using Ordinal Linear Regression and Query-generated Training Data for Ad-hoc Retrieval with Long Queries**, Javid Dadashkarimi, Masoud Jalili Sabet, Azadeh Shakery, *COLING*, 2016 — **Acceptance Rate = 32%**.
- **Pseudo-relevance feedback based on matrix factorization**, Hamed Zamani, Javid Dadashkarimi, Azadeh Shakery, W Bruce Croft., *CIKM*, 2016, — **Acceptance Rate = 22%**.
- **SS4MCT: a statistical stemmer for morphologically complex texts**, Javid Dadashkarimi, Hossein Nasr Esfahani, Hesham Faili, Azadeh Shakery, *CLEF*, 2016, Sheffield, UK
- **Building a multi-domain comparable corpus using a learning to rank method**, Razieh Rahimi, Azadeh Shakery, Javid Dadashkarimi, Mozhdeh Arianneshad, Mostafa Dehghani, Hossein Nasr Esfahani, *Natural Language Engineering*, 2016, Impact Factor = 1.06.
- **Revisiting optimal rank aggregation: A dynamic programming approach**, Shayan A Tabrizi, Javid Dadashkarimi, Mostafa Dehghani, Hassan Nasr Esfahani, Azadeh Shakery, *ICTIR*, Indianapolis, US, 2015.