

## RESEARCH INTERESTS

---

- Language Modeling
- Causal Inference
- Theory of Optimization and Learning
- Data Distillation

## EDUCATION

---

- **Sharif University of Technology** Tehran, Iran  
*Bachelors in Electrical Engineering; GPA: 19.4 / 20* Sep 2019 - Jun 2023
- **Massachusetts Institute of Technology** Cambridge, MA  
*Institute for Data, Systems and Society (IDSS)* Sep 2023 -  
*Laboratory for Information and Decision Systems (LIDS)*  
*Ph.D. Program in Social Engineering Systems; GPA: 4.8/5*

## PUBLICATIONS

---

\* Indicates equal contribution.

- **M. JafariNodeh**, M. Ghadiri, A. Abadie, A. Jadbabaie, “Non-Asymptotic Continuous Average Treatment Effect Estimation.”, *Econometrica*, (under review)
- **M. JafariNodeh**, B. Khdoabandeh, A. Afazali, “Aligning visual contrastive learning models via preference optimization.”, *ICLR 2025* (under review)
- **M. JafariNodeh**, A. Ajorlou, A. Jadbabaie, “Belief Samples Are All You Need For Social Learning.”, *IEEE Conference on Decision and Control*, 2024. Paper
- G. Aminian, A. Bagheri\*, **M. JafariNodeh**\*, R. Karimian\*, MH. Yassaee, “Semi-supervised Learning under Self-training via  $f$ -Divergence.”, *ISIT*, 2024. Paper

## EXPERIENCE

---

- **MIT** Cambridge, MA  
*Research Assistant* Sep 2023 -
  - **Alignment of Language Models** : (Joint work with A. Beirami at Google DeepMind) Working on extension of current literature on alignment of language models to cases where penalizing function over the alignment dataset is more general divergence function such as  $\alpha$ -Rényi instead of current functions such as KL in order to avoid getting large log-likelihood ratios which can happen with small probability (heavy-tailed) which don't affect the KL considerably. This approach could make alignment of models much more reliable.
  - **Data Distillation**: Working on discovering methods that can recover the training dataset using the convergence parameter of an over-parameterized model while conjecturing requiring number of parameters  $p = \tilde{\Omega}(nd)$  and expecting a connection the stream of works on the universal law of robustness.
  - **Adaptive Methods for Language Models**: Studying the question ”Is the slow convergence of the linear classification head of the model the reason for overall slow convergence? and how heavy-tailed nature of training data of LLMs can drastically favor using adaptive optimization methods instead of SGD and how it changes the optimization landscape of the model and if this process could be explained in a multi-phase manner”
  - **Non-Asymptotic Continuous Average Treatment Effect Estimation**:
    - \* What We Did: We explored regression adjustment techniques (giving new non-trivial high probability upper bounds using independent leverage score sampling) for experiments with multiple and continuous treatments, offering unbiased estimators with variance tied to linear fit quality, and extended our analysis to include a finite set of dosages using non-parametric estimators.
    - \* What Was the Result: Provided non-asymptotic bias and variance bounds for both types of treatments, revealing a dependency on the continuous treatment effect curve's smoothness and analyzing the balance between dosage diversity and unit allocation.
    - \* Why It Is Important: Our work broadens the understanding of experimental design beyond traditional treatment/control setups, with implications for scientific, medical, and industrial research, especially in optimizing the inclusion and analysis of varied treatment intensities.

- Marco Mondelli's Lab**

Scientific Intern - Prof. Marco Mondelli

ISTA, Austria  
July 2022 - Sep 2022

  - Analyzing Behavior of SAM Optimization Techniques:**
    - What We Did: Investigated SAM's impact on shallow neural networks, modifying it to probe generalization in over-parameterized scenarios, simplifying analysis to derive a closed-form solution for weights.
    - What Was the Result: Our adapted SAM algorithm matched or outperformed the original, demonstrating a monotonic decrease in test loss with increased model complexity, indicative of mitigating double descent.
    - Why It Is Important: This study enhances understanding of optimization and model generalization, suggesting SAM's potential in self-tuning regularization and reducing computational costs in neural network training.

The poster is available in [here](#)
- Statistical Signal Processing lab**

Research Assistant

Sharif University, Tehran  
Oct 2021 - Sep 2023

  - A Framework for Recommender Systems:**
    - What We Did: Developed a method combining latent feature extraction from time series and user-item interactions with graph-based techniques to predict and prevent user churn, advancing to a personalized recommender system using graph signal processing.
    - What Was the Result: Achieved a sparse, high-performance network that outperforms current models in accuracy, utilizing frequency domain analysis for a fair and customizable recommender system.
    - Why It Is Important: Enhances subscription-based company strategies by preventing churn and personalizing recommendations, introducing a novel application of graph signal processing for improved service delivery.

## SKILLS SUMMARY

- Languages:** Java, C++, Python, C, SQL, R, Unix scripting
- Tools:** Kubernetes, Docker, Hadoop, GIT, XCode, Keras, Pytorch, Tensorflow, Jax
- Specialized Areas:** Large Language Models (LLM), Multi-Modality Models, Data Mining, NLP, Data Structures & Algorithms

## SERVICE

Reviewer for ISIT 2024, ICLR 2025

## SELECTED ADVANCED COURSES

Courses		
Mathematics	CS	Learning
Probability Theory	Analysis of Algorithms <sup>a</sup>	Foundations of Deep Learning
High Dimensional Probability		Theory of Machine Learning
Random Processes		Information Theoretical Statistical Learning
Real-Complex-Functional Analysis		Convex Optimization
Tensor Analysis & Advanced Optimization		
Foundations of Statistics		
Advanced Linear Algebra and Tensor Analysis		

<sup>a</sup>(Normal, Random, Approximate)

## ACADEMIC PROJECTS

- Fine-Tuning Large Language Models for Domain-Specific Text Generation:** Developed a pipeline for fine-tuning GPT-3 on domain-specific corpora to improve its performance on specialized text generation tasks. This involved curating a large dataset from domain-specific sources, applying transfer learning techniques, and conducting extensive evaluation using BLEU and ROUGE scores to assess the quality of generated content. The fine-tuned model achieved a 15% improvement in domain relevance compared to the base model.
- Inference Optimization in Large Language:** Designed and implemented an optimized inference pipeline for deploying large language models in real-time applications. This project focused on reducing latency and improving throughput by leveraging mixed-precision techniques and model quantization. The optimized pipeline was deployed in a simulated environment, demonstrating a 30% reduction in inference time while maintaining 98% of the original model's accuracy.
- Few-Shot Learning with LLMs for Content Moderation:** Explored few-shot learning techniques using GPT-3 to improve content moderation systems. The project involved designing prompt engineering strategies and testing the model's performance on identifying and categorizing harmful content with minimal labeled data. The approach demonstrated the potential to reduce the dependency on large labeled datasets while achieving high accuracy in content moderation tasks.
- Implement a multi-modal sentiment analysis and caption generator:** Develop from scratch a high-performance multi-modal model based on state-of-art RNN-based NLP tools and leveraging methods from Signal processing to perceive the voices and CNNs to process the video and maintain it's performance on Filimo movie subscription service. (Apr '22)

- **Sentiment Analysis Network using modified BERT**: developing and maintaining a new NLP framework to analyze and recognize sentiments of comments in the review sections of the company using tools like Python, Java, and Libraries such as Pytorch, Keras from structured and unstructured data-types and state-of-art algorithms to modify BERT (a highly used NLP-based tool) to increase the accuracy of the model and keeping the framework online to analyze user feedbacks at any time. (Dec '21)
- **Picture denoising Using RNN's**: Implemented an RNN-based system that denoises the noisy pictures by getting help from tools from signal processing such as wavelet transform and using RNNs on the output to output a more clear picture. (Feb '22)
- **A fair movie website**: Design a recommender system for a created social net that preserves fairness, in the sense that it can be personalized and each user may set his degree of personalization (By using graph signal processing tools in the users' graph). (Work is done in the second paper and implemented here) (Oct '22)

## PROFESSIONAL REFERENCES

---

1. Prof. Ali Jadbabaie  
Massachusetts Institute of Technology  
Email: jadbabai@mit.edu
2. Prof. Marco Mondelli  
Institute of Science and Technology Austria (ISTA)  
Email: marco.mondelli@ist.ac.at
3. Ahmad Beirami  
Senior Research Scientist  
Google Research  
Email: ahmad.beirami@gmail.com