# Jaydeep Jitendra Borkar

jaijborkar@gmail.com
http://jaydeepborkar.github.io/
https://github.com/jaydeepborkar

## Education & Research Experience

**Northeastern University** 2021 - present
*PhD in Computer Sciences*
Advisor: Prof. David A. Smith 2023 - present

**Meta Superintelligence Labs** June '25 - Nov '25
*Visiting Researcher*, New York
Research on LLM privacy and memorization.
Mentors: Diego Garcia-Olano, Zheng Xu, Karan Chadha, Niloofar Mireshghallah.

**MIT-IBM Watson AI Lab** 2020 - 2021
*External Research Student*
Advisor: Dr. Pin-Yu Chen
Research on adversarial machine learning.

**University of Pune** 2016 - 2020
*Bachelor's degree in Computer Engineering*

**CIFAR Deep Learning + Reinforcement Learning Summer School** Aug 2020
Hosted by Mila (25% acceptance rate)
Among 300 students selected across 45 countries for the summer school

**Research Interests**: Privacy and safety in language models, training data extraction (memorization) in LLMs, synthetic data.

## Papers & Research Projects

**Memorization Dynamics in Knowledge Distillation for Language Models**
**Jaydeep Borkar**, Karan Chadha, Niloofar Mireshghallah, Yuchen Zhang, Irina-Elina Veliche, Archi Mitra, David A. Smith, Zheng Xu, Diego Garcia-Olano
*under review*

**Privacy Ripple Effects from Adding or Removing Personal Information in Language Model Training**
**Jaydeep Borkar**, Matthew Jagielski, Katherine Lee, Niloofar Mireshghallah, David A. Smith, Christopher A. Choquette-Choo
*Association for Computational Linguistics (ACL)* 2025

**Recite, Reconstruct, Recollect: Memorization in LMs as a Multifaceted Phenomenon**
USVSN Sai Prashanth, Alvin Deng, Kyle O'Brien, Jyothir S V, Mohammad Aflah Khan, **Jaydeep Borkar**, Christopher A. Choquette-Choo, Jacob Ray Fuehne, Stella Biderman, Tracy Ke, Katherine Lee, Naomi Saphra
*International Conference on Learning Representations (ICLR)* 2025

**What can we learn from Data Leakage and Unlearning for Law?**
**Jaydeep Borkar**
*Generative AI and Law (GenLaw) workshop, ICML* 2023
https://genlaw.github.io/CameraReady/12.pdf

**Mind the gap: Analyzing lacunae with transformer-based transcription**
**Jaydeep Borkar** and David A. Smith
*Workshop on Computational Paleography, ICDAR* 2024
https://arxiv.org/abs/2407.00250

**Extracting Training Data from Pre-trained and Fine-tuned GPT-2**
*CS 7150 Deep Learning class project*
Showed that fine-tuned models can memorize and leak both fine-tuning and pre-training data.
Project report: `https://jaydeepborkar.github.io/7150_project_report.pdf`

**Simple Transparent Adversarial Examples**
**Jaydeep Borkar** and Pin-Yu Chen
*Workshop on Security and Safety in Machine Learning Systems, ICLR* 2021
https://aisecure-workshop.github.io/aml-iclr2021/papers/48.pdf

## Talks

**Google ML Privacy Seminar**: talk on Privacy Ripple Effects.
**Northeastern CS 5100**: invited guest lecture on privacy and security in LLMs.

## Organizing

**Trustworthy ML Initiative**
Co-organizer of the Trustworthy ML Initiative along with Prof. Hima Lakkaraju (Harvard), Sara Hooker (Cohere for AI), Dr. Sarah Tan (Salesforce AI), Dr. Subho Majumdar (Vijil), Chhavi Yadav (UC San Diego), Dr. Chirag Agarwal (Harvard), Prof. Haohan Wang (UIUC), and Marta Lemanczyk (Hasso-Plattner-Institut).

## Courses

Machine Learning CS 6140, Natural Language Processing CS 6120, Deep Learning CS 7150, Machine Learning Security and Privacy CY 7790, Theory and Methods in Human-Computer Interaction CS 7340, AI as an Archival Science CS 7180.

## Teaching Experience

| | |
|---|---|
| Natural Language Processing CS 6120 - TA | Summer 2024 |
| Foundations of AI CS 5100- TA | Spring 2024 |
| Foundations of Data Science DS 3000 - TA | Fall 2023 |
| Product Development for Large Language Models CS 7180 - TA | Summer 2023 |
| Introduction to Computer Science Research CS 3950 and CS 4950 - TA | Spring 2023 |
| Introduction to Machine Learning and Data Mining DA 5030 - TA | Summer and Fall 2022 |