Nirav Diwan

Website : https://nirav0999.github.io/ | **Email** : ndiwan2@illinois.edu | **Linkedin** : nirav-diwan-cse
Github : nirav0999 | **Twitter** : ocean_drifters | **Google Scholar** : ADOnuAAAAAJ
Address : 301 South Wright Street, Champaign, Illinois, United States of America (US), 61820

Education

- **University of Illinois Urbana-Champaign (UIUC)** — Aug 2022 – May 2024
  Master of Science (M.S) in Computer Science (CS) — *GPA: 4.0/4.0*
- **Indraprastha Institute of Information Technology-Delhi (IIITD) (Website)** — Aug 2017 – Jun 2021
  Bachelor's of Technology (B.Tech) in Computer Science Engineering (CSE) — *GPA: 8.96/10.00*

Publications

(C.1): Spotting Core Collusive Users in YouTube Blackmarket Network — AAAI-ICWSM, 2022

Hridoy Sankar Dutta*, **Nirav Diwan***, and Tanmoy Chakraborty

Publication | Code

(C.2): Fingerprinting Fine-tuned Language Models in the wild — ACL (Findings), 2021

Nirav Diwan, Tanmoy Chakraborty, Zubair Shafiq

Publication | Code

(J.1): RecipeDB: A Resource for Exploring Recipes — Database - OUP, 2020

Devansh Batra*, **Nirav Diwan*** ... Ganesh Bagler

Website | Publication | Code

(W.1): A Named Entity Based Approach to Modeling Recipes — DECOR Workshop@ICDE, 2020

Nirav Diwan, Devansh Batra and Ganesh Bagler

Publication | Code

(W.2): Nutritional Profile Estimation in Cooking Recipes — DECOR Workshop@ICDE, 2020

Jushaan Singh Kalra, Devansh Batra, **Nirav Diwan** and Ganesh Bagler.

Publication | Code

(P.1): Generating Adversarial Images using Diffusion Models — IEEE S&P (Oakland), 2024

Qingying Hao, **Nirav Diwan**, Ying Yuan, Gang Wang

(P.2): Robust Large Language model for Assembly Code — USENIX, 2024

Hadjer Benkraouda, **Nirav Diwan**, Gang Wang

* Equal Contribution, P - Planned Submission, C - Conference, J - Journal, W - Workshop

Awards

- **Dean's Thesis Appreciation Award:** Received a thesis appreciation award (awarded to approx. 1% students) from the Dean of Academic Affairs.
- **Dean's Award for Academic Excellence, IIITD:** Awarded to top-performing students ($CGPA > 9.0$) for the academic year (AY). Received it for the AY 2019-20, 2020-21.
- **OxML Summer School 2022 & 2023**: Selected to attend & present a poster at the Oxford Machine Learning Summer School 2022 and 2023.
- **EEML Summer School 2021 & 2022**: Selected to attend & present a poster at the Eastern European Machine Learning Summer School 2021 & 2022 on the work of synthetic text generated by fine-tuned language models.

- **Project Research & Innovation Showcase (P.R.A.I.S.E.)**: Awarded the 3^{rd} place at *Project Research & Innovation Showcase*, a research showcase held by IEEE Student branch for the Technical Festival - Elysium at IIITD. (Certificate).
- **Grants:** Awarded grant to attend AAAI sponsored *AIIDE* 2020 and *ICWSM* 2021 and 2022 - held virtually.
- **University Entrance Exams:** Secured All India Rank 1884 (top 0.15%) out of 1.5 million participants in *IIT-JEE Mains 2017* and 4186 out of 2,00,000 participants in *IIT-JEE Advanced 2017*.
- **NPTEL Course Award:** Awarded *Elite+Gold* certification for Cloud Computing and Distributed Systems by NPTEL. (Certificate).

Industrial Experience

- **Ema Inc.** June 2023 -
  Applied Science Intern
  - **Natural Language to SQL (NL2SQL)**:
    - Using **Large Language Models (LLMs)** to for the task of NL2SQL on single and multi-table settings.
    - Designed an **In-context learning** based approach using **Synthetic dataset** for NL2SQL task as a pilot study. [Code]
    - Designed systematic approaches for **Prompt Engineering** for NL2SQL Task that led to an increase in recall by **10**%.
- **Prodigal Technologies** Mumbai, India
  Natural Language Processing Engineer *Oct 2021 - May 2022*
  - **ProVoice**:
    - Designing frameworks for retrieving relevant information from conversations to enhance debt collection process.
    - Developing extractive and abstractive summaries of conversations.
    - **Keywords:** *Natural Language Processing, Deep Learning, Computational Finance*
- **Research and Innovation Labs, Tata Consultancy Services** Gurgaon, India
  Research Intern *May 2020 – Jul 2020*
  - **Optimising Stock Trading**:
    - Conducted a literature review for reinforcement and collaborative filtering techniques for portfolio optimisation.
    - Built custom stock trading environments and used neural reinforcement learning algorithms to obtain remarkably higher returns.
    - **Keywords:** *Machine Learning, Reinforcement Learning, Collaborative Filtering*

Research Experience

- **Illinois DASH Lab, UIUC** Champaign, USA
  Graduate Researcher *Aug 2022 - Present*
  - **Large Language Model for Assembly Code**:
    Guide(s) - Dr Gang Wang
    - Designed and trained a **Large Language Model (LLM)** for Assembly Code that outperforms models on downstream tasks by an average of 3% F1-Score.
    - Implemented a **pseudo-label based domain adaptation method** using **asymmetric tri-training loss**.
    - Improved the State of the Art performance by **3**% and **7**% on in-domain and out-of-domain datasets respectively.
    - **Keywords:** *Natural Language Processing, Domain Adaptation, ML for Code*
  - **Adversarial Image Generation using Diffusion Models**:
    Guide(s) - Dr Gang Wang
    - Developing methods to to generate adversarial images using **Guided Diffusion** and **Stable Diffusion** models.
    - Developed a **retrieval-based generative pipeline** for obfuscating text for evading online detectors.
    - Paper submitted to **IEEE Security & Privacy, Oakland (CORE A*) 2024**.**Keywords:** *Diffusion Model, Generative AI, Image Generation*
- **Laboratory for Computational Social Systems (LCS2), IIITD** Delhi, India
  Undergraduate Researcher *Aug 2019 - Jan 2022*

- **Fingerprinting Fine-tuned Language Models in the wild (B.Tech Thesis)**:

  Guide(s) - Dr Tanmoy Chakraborty, Dr Zubair Shafiq

  Designed a model for authorship attribution for fine-tuned language models. Accepted as a long research paper at Findings of ACL-IJCNLP 2021 conference. Our paper was also invited for a live Q&A session at Repl4NLP Workshop and for a poster presentation at Eastern European Machine Learning Summer School 2021.

  * Conducted an in-depth analysis between human-written and machine-generated text which highlighted the efficacy of synthetic text generated by language models.
  * Designed a convolutional model to identify the fine-tuned language model which generated the synthetic text. The model achieved a precision of 87% in a 108-class supervised classification setup.
  * **Keywords:** *Machine generated text, Natural Language Processing, Deep Learning, Computational Social Science, Security*

- **Weakening the Inner Strength: Spotting Core Collusive Users in YouTube Blackmarket Network**:
  Guide(s) - Dr Tanmoy Chakraborty

  Led an investigation into understanding the role and detection of influential users (core) in online collusive blackmarket services on Youtube. Project accepted as a long research paper at AAAI-ICWSM 2022 conference.

  * Developed a graph-based method *KORSE* by adapting the conventional core-periphery algorithm for collusive blackmarket network to detect *core* users.
  * Designed a deep learning framework using **BERT** - *NURSE* which leverages the natural language properties and metadata properties of users for the real-time detection of *core* users.
  * Designed experiments which highlighted the anomalous structure of the collusive Youtube blackmarket - a core surrounded by peripheral communities of collusive users
  * Conducted case-studies which showcased unusual properties of *core* users.
  * **Keywords:** *Natural Language Processing, Security & Privacy, Deep Learning, Computational Social Science*

- **Insect Ecology, Evolution, and Conservation Lab, IIITD** — Delhi, India
  Intern — *Aug 2019 – Apr 2020*

  Guide(s) - Dr Jainendra Shukla, Dr Swapna Purandare

  - **Perceive the Pollinator**:
    * Used Transfer Learning and Domain Adaptation techniques to design a classifier to label a database of pollinators indigenous to the Indian Subcontinent.
    * Achieved an F1-Score of 91.5 % on a 5-class classification process.
    * **Keywords:***Computer Vision, Deep Learning, Computational Ecology*

- **Complex Systems Lab, IIITD** — Delhi, India
  Intern — *May 2019 – Jul 2019*

  Guide(s) - Dr Ganesh Bagler

  - **RecipeDB (Website)**:
    Co-founded - RecipeDB, a website providing nutritional profiles of over 118,000 recipes along with the dietary profiles (DietRX) and flavour profiles (FlavorDB) of 2,000 ingredients. Website has over (10k+ users).
    * Developed a **Named Entity Recognition (NER)** model to extract ingredient information that achieved an **F1-Score 95%**.
    * Scaled the model to process data from upto **10 data sources** with over **118,000+ recipes** processed.
    * Conducted an exploratory data analysis of the recipe data (available on the website) on their diversity from across the world.
    * **Keywords:** *Natural Language Processing, Machine Learning, Computational Gastronomy*

- **Centre for Social Sciences and Humanities (CSH), CNRS** — Delhi, India
  Intern — *May 2019 – Jul 2019*

  Guide(s) - Dr Jean Thomas Martelli

  - **Populism in Indian Prime Minister's speeches**:
    * Contributed to the creation of speech dataset of Indian Prime Ministers.
    * Performed a quantitative analysis of all Indian Prime Minister's speeches, helping to develop a relative scale of populism of the leaders.
    * **Keywords:** *Natural Language Processing, Statistics, Computational Social Sciences*

Teaching Experience

- **University of Illinois Urbana Champaign (UIUC)** — Illinois, USA
  Teaching Assistant — *Aug 2022 –*
  - **CS124 Introduction to Computer Science**: Instructor(s) - Dr Geoffrey Challen
    - Teaching Assistant for the course of 1300 undergraduate students.
    - Responsibilities include making student Quizzes, proctoring quizzes and conducting office hours.
    - Teaching Assistant for 2 consecutive semesters.

- **Indraprastha Institute of Information Technology, Delhi (IIITD)** — Delhi, India
  Teaching Assistant — *Aug 2020 – Dec 2020*
  - **CSE543 Machine Learning**: Instructor(s) - Dr Tanmoy Chakraborty
    - Teaching Assistant for Machine Learning Course of 180 students comprised of 4^{th} year Undergraduates and Graduate Students.
    - Conducted tutorials, answered student doubts. evaluated quiz sheets, and created the mid-semester examination.

Selected Projects

- **Pseudo-Label Domain Adaptation (Code)**: Fall 2022, UIUC
  - Implemented the research paper Pseudo-Label Guided Unsupervised Domain Adaptation of Contextual Embeddings
  - Implemented Asymmetric Tri-training loss for Unsupervised Domain Adaptation.
  - Achieved an F1-Score of 75% on the target domain for a task of sentence classification.
  - *Language(s)* - Python, *Libraries* - pytorch, transformers *Deep Learning*

- **Site Energy Prediction**: Fall 2022, UIUC
  - Achieved 8^{th} best score in the Kaggle competition - Women in Data Science (WiDS) Datathon.
  - Problem Statement - Predicting the energy usage of buildings of different kinds - apartment, factory, warehouse etc.
  - Performed a feature-wise analysis for predicting the energy usage.
  - Trained and tested several machine learning models - *Language(s)* - Python, *Libraries* - pytorch, transformers *Deep Learning*

- **RustID:Identification of Rust Disease in Wheat (Challenge Link) (Report) (Medium Blog)**: Spring 2020, IIITD
  - Implemented State of the Art Deep Learning architectures including EfficientNet, DenseNet, ResNet, VGG16, VGG19 alongside augmentation techniques (mixup, cutmix).
  - Obtained a rank of 39/803 participants in a team of 3 students.

  Language(s) - Python *Libraries* - fastai, tensorflow, pytorch *Deep Learning*

- **Career Advisory System (Code)**: Fall 2020, IIITD
  - Implemented a career advisory system for final year undergraduate students using logic programming in Prolog.
  - Developed a terminal interface which was able to accurately respond to natural language answer

  Language(s) - Prolog *Logic Programming, Artificial Intelligence*

- **Argument Reasoning Task (Challenge link)** : Fall 2019, IIITD
  - Designed a embedding framework using Bi-LSTM, RNN and BERT for the challenge of argument reasoning.
  - Obtained a higher F1-score than the task's baseline by 2 %.

  Language(s) - Python, *Libraries* - tensorflow, pytorch. *Natural Language Processing*

- **Movie Recommender via Collaborating Filtering (Code)**: Spring 2020, IIITD
  - Designed **User Based Recommender, Item Based Recommender and Matrix Decomposition Methods** to recommend movies based on Ratings provided by users.
  - Developed a website (link) using **Flask** and gives recommendations based on initial ratings provided by the user.

  Language(s) - Python, HTML, CSS, JS. *Web Development, Multi-threading*

- **SuperStore (Code)** : Spring 2019, IIITD
  - Designed a GUI application by applying **OOPS**.
  - Created Multi-User Login Capabilities for admin and customer using Multi-threading.

  Language(s) - Java *JavaFX, Multi-threading*

- **Page Swapping**: Spring 2019, IIITD
  - Implemented Page Swapping and developed priority queue in xv6.

  Language(s) - C *Operating Systems*

- **LINUX Shell**: Spring 2019, IIITD
  - Created a LINUX Shell with features such as pipelining and I/O redirection.

  Language(s) - C++ *Operating Systems*

Technical Skills

- **Languages (Programming + Scripting)** - Python , Java, C++, C, HTML, JavaScript, Prolog, Haskell
- **Databases** - MongoDB, SQL
- **AI & Maths** - Numpy, Pytorch, tensorflow, scikit-learn, fastai, networkX, dgl, OpenAI-Gym, nltk, TXM
- **Data Management and Text Scraping** - Twitter API, Youtube API, Selenium, beautifulsoup, pandas
- **Website Development** - flask, Django
- **Others** - Git, Linux, LaTeX

Courses Undertaken

- **Algorithms -** Data Structures and Algorithms, Algorithm Design and Analysis, Advanced Programming, Introduction to Haskell Programming
- **Data Science -** Trustworthy Machine Learning*, Introduction to Data Mining*, Deep Learning, Machine Learning, Fundamentals of Databases, Mining Large Networks, Big Data Analytics, Artificial Intelligence, Natural Language Processing
- **Systems & Security -** Foundations of Computer Security, Cloud Computing and Distributed Systems, Computer Networks, Operating Systems, System Management, Computer Organisation
- **Mathematics -** Probability and Statistics, Multi-variable Calculus, Discrete Mathematics, Linear Algebra, Introduction to Mathematical Logic
- **Other -** Social and Political Philosophy, Anthropology of Social Media, Environmental Science, Technical Communication, Portfolio Management, Introduction to Quantitative Biology, Body Language Studies

*Ongoing

Co-curricular Acitvities

- **Student Mentorship Programme, IIITD:** Mentored 7 Freshman B.Tech students for their academic and social growth in college for the academic year 2020-21.
- **Shanti Sahyog Volunteer, IIITD:** Volunteered at the NGO Shanti Sahyog. Conducted a workshop of HTML, CSS and Javascript for underprivileged female students of XII grade. Compiled and maintained a database of Shanti Sahyog's Inter-Continental Wide Connections. Completed 100 hours of volunteer work.
- **Talks:**
  - **SecureML Seminar:** A reading group on Adversarial Machine Learning. Presented two papers in the cross-domains of security and machine learning.
  - **LCS2, IIITD:** Routinely presented research papers at the weekly group meetings.
  - **Computational Gastronomy Symposium Season 3, IIITD:** Presented RecipeDB and organised the symposium alongside the research team and Dr. Ganesh Bagler. Attended by nation - wide food specialists, anthropologists and chefs.

- **Writing:** Contributed to two chapters for the book *Social Network Analytics* by Dr Tanmoy Chakraborty, explaining various complex concepts of graphs.
- **Conference Volunteer:** ICWSM'22, ACL'21, ICML'21, ICLR'21.
- **Reviewer:** Selected as a mentee for *ICLR Student-Reviewer Mentorship Program 2022* and as a shadow PC reviewer for *ACM COMPASS Shadow PC Program* 2021.