HUNAR BATRA

i@hunarbatra.com | hunar.batra@nyu.edu | hunarbatra.com | Linkedin | GitHub | Google Scholar

EDUCATION

University of Oxford – *DPhil Computer Science* **[Expected Graduation- June 2026]**
Multimodal learning and Interpretability | Supervised by Prof. Ronald Clark

University of Oxford – *MSc Advanced Computer Science* **[2021 - 2022]**
Dissertation: "Protein Language Representation Learning to predict SARS-CoV-2 mutational landscape", under Dr. Peter Minary [Overview]

University of Delhi – *BSc (Hons) Computer Science* – 8.42 CGPA, First Division Honours **[2017 - 2020]**

RESEARCH EXPERIENCE

***Research Consultant* – Anthropic** **[Sept 2023 – Present]**
Improving chain of thought transparency in LLMs by mitigating issues of ignored reasoning, sycophancy & biased reasoning via SFT

***Visiting Researcher* – New York University Center for Data Science** **[June 2023 – Present]**
Applying mechanistic interpretability to decode truthful representations from large language models and training LMs for consistency

***Research Scholar* – Stanford Existential Risks Initiative & NYU Center for Data Science** **[Nov 2022 – Sept 2023]**
Built tools for automating alignment research & simulating alignment researchers through human-AI collaboration & expert iteration

***Privacy Preserving ML Researcher (EWADA)* – Oxford Human Centred AI Group, University of Oxford** **[Nov 2021 – Dec 2022]**
Building decentralised web apps with PPML based recommendations for SOLID under Prof. **Tim Berners Lee**

***Research Lead* – Oxford Rhodes AI Lab** **[May – Oct 2022]**
Leveraging GNNs to predict climate closures equation using symbolic regression in collaboration with **CalTech** (CLiMA), **MIT** & **NASA JPL**

***Language Modelling Research* – Computational Biology Group, University of Oxford** **[April – Oct 2022]**
Applying language modelling to predict COVID-19 mutations using transformer-based models & AlphaFold2 under Prof. Peter Minary

***Chatbot Development Research* – University of Oxford** **[Feb – Aug 2022]**
Developed a Question-Answering language model for the Philosophy Dept to help convey their research work over website/messenger

***NLP Student Researcher* – Department of Computer Science, University of Delhi** **[March 2020 – July 2021]**
Researched & developed multiple projects- GPT-3 use-case model extractor, Ensemble ML Fake News detection, GPT-2 Title Generation. COVID-19 News Summariser using transformers, Medical QA bot. Co-authored and published papers in IEEE & Springer Singapore

***Computer Vision Student Researcher* – AI Research Lab, University of Delhi** **[June – Sept 2019]**
Built a Computer Vision based Assistive System for Autonomous Vehicles. Compiled Darknet with OpenCV for real-time predictions

WORK EXPERIENCE

***Mobile Robotics Engineer* – Swift Robotics** **[Aug 2020 – Sept 2021]**
Developed Flask REST API to livestream video processed with Computer Vision techniques (OpenCV, image stitching- KNNs)
Built a React Native application which interacts with ROS melodic nodes to control robot's navigation & visualised LiDAR odometry

***Co-Founder* – HushTech Solutions** **[June 2019 – July 2021]**
Self-taught NLP engineer; developed omni-channel messenger chatbots & RPA solutions for businesses such as a DIET classifier email bot

***Machine Learning Engineer* – Omdena** *(One of the 28 Global AI experts selected)* **[March – June 2020]**
Applied statistical models: LDA topic modelling, VAR, ARIMA & EDA over COVID-19 policies. Results showcased at UN AI Summit

***Mobile Application Development Intern* – Impute Inc.** **[March – June 2019]**
Developed & extensively trained a contextual conversation QA agent for Fluent8 iOS app. Deployed webhooks on Firebase Cloud Function

***Chatbot Development Intern* – Inverted Sense** **[Dec 2018 – March 2019]**
Built chatbots using Twilio & developed an in-built shopping cart with up-selling resulting in higher lead conversions & ROAS

PROJECTS | Github : github.com/hunarbatra

- **Model written sycophancy evals**: Evals generation using expert oversight guided multiversal dynamics exploration [Link]
- **Scaffold**: Simulates alignment researchers comments on posts/drafts [Link]
- **Alignment Forum Summarisation tool**: Iterative MCTS with tuned expert agent to steer & generate summaries [Slides/Code]
- **GPT++**: Autonomous self-learning agent with tools access
- **MuFormer**: Inverted AlphaFold2 for inverse-folding with pLM inductive bias to generate mutational sequences
- **CoVBERT:** COVID-19 mutation prediction language model [Link]
- **GraphSAGE LSTM & BiLSTM Aggregators**: Merged in PyTorch Geometric Package [Link]
- **HunAI:** DialoGPT DSTC telegram buddy bot

AWARDS & ACHIEVEMENTS

- **Google Women in Computer Science** Generation Scholarship EMEA, 2022
- **Grace Hopper Conference Scholar**, Department of Computer Science, University of Oxford, 2022
- Deep Learning Theory Summer School Scholarship, Simons Institute for Theory of Computing, UC Berkeley, 2022
- **Student of the Year & Rank 3,** Department of Computer Science, University of Delhi, 2020
- The Mars Generation **24 under 24** Award for Leaders & Innovators in STEM, 2019
- **National Finalist, Smart India Hackathon** Software Edition, (out of 5,000 teams) in India's largest hackathon by MHRD Govt. of India, 2019
- **National Winner**, Summer with Google (out of 20,000 participants), 2018

SKILLS

Python, C++, C, Javascript, SQL, App Dev (Native, React Native), Web Dev (HTML, CSS, React.js, TypeScript, Node.js, Flask)
PyTorch, PyTorch Geometric [Merged PR], TensorFlow, JAX, Langchain, LLaMA-Index, OpenCV, Kubernetes, Google Cloud Platform, ROS

RESEARCH PUBLICATIONS (150+ Citations) - Google Scholar | Detailed *CV* can be found here