

HUNAR BATRA

hunarbatra.com | i@hunarbatra.com | hunar.batra@cs.ox.ac.uk | hunar.batra@nyu.edu | [Linkedin](#) | [GitHub](#) | [Google Scholar](#)

EDUCATION

University of Oxford – DPhil Computer Science	[Oct 2023 - June 2026]
Multimodal learning and Interpretability Supervised by Prof. Ronald Clark	
University of Oxford – MSc Advanced Computer Science	[2021 - 2022]
Google Women in Computer Science Scholarship Awardee (Merit-based), Stanford MATS Scholar, GHC Scholar	
Dissertation: "Protein Language Representation Learning to predict SARS-CoV-2 mutational landscape", under Dr. Peter Minary [Overview]	
University of Delhi – BSc (Hons) Computer Science	[2017 - 2020]
8.42 CGPA, First Division Honours with Distinction	
Overall Rank 3 and Student of the Year Award (2020); Published 6 research papers	

RESEARCH EXPERIENCE

Research Consultant – Anthropic	[Sept 2023 – March 2024]
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 (Alignment Research Group)	[June 2023 – Present]
Applying mechanistic interpretability to decode intermediate encodings for LLaMa-2, & mitigating biases in LLMs via consistency training	
Research Scholar – Stanford Existential Risks Initiative	[Nov 2022 – Sept 2023]
Built tools for automating alignment research & simulating alignment researchers through human-AI collaboration & expert iteration	
Research Intern (EWADA) – Oxford Human Centred AI Group, University of Oxford	[Nov 2021 – Dec 2022]
Built decentralised apps using SOLID with privacy-preserving ML recommendations under Prof Jun Zhao, Nigel Shadbolt & Tim Berners Lee	
Research Lead – Oxford Rhodes AI Lab	[May – Oct 2022]
Leveraged GNNs to predict climate closures equation using symbolic regression in collab with CalTech (CLiMA), MIT & NASA JPL	
Language Modelling Research – Computational Biology Group, University of Oxford	[April – Oct 2022]
Applied 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 6 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

- **Visual Hierarchical Reasoning:** Improved acc of GPT-4V by 11% over MMMU by extracting segments + visual attributes [[Link](#)]
- **LLaMa-2-7B Tuned Lens:** Trained a tuned lens for LLaMa 2-7B to decoder outputs for intermediate layers [[Link](#)]
- **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++:** Self-learning agent with internet access + episodic memory for reasoning – built before internet access LLMs came out [[Link](#)]
- **MuFormer:** Inverted AlphaFold2 for inverse-folding with pLM inductive bias to generate mutational sequences [[Link](#)]
- **CoVBERT:** COVID-19 mutation prediction language model [[Link](#)]
- **GraphSAGE LSTM & BiLSTM Aggregators:** Merged in PyTorch Geometric Package [[Link](#)]
- **HunAI:** DialogPT DSTC telegram buddy bot
- **Real Time Voice Cloning:** Transfer Learning based voice cloning on the recorded input to generate TTS

AWARDS & ACHIEVEMENTS

- **Dan Kohn Scholarship**, KubeCon + CloudNative AI Conference EU 2024
- **Long Term Future Fund Grant**, Effective Ventures, 2023
- **Research Scholarship**, Stanford University, Machine Learning Alignment Theory Scholar, 2022
- **Google Women in Computer Science Generation Scholarship** EMEA, 2022

- **Grace Hopper Conference Scholarship**, Department of Computer Science, University of Oxford, 2022
- Deep Learning Theory Summer School Scholarship, Simons Institute for Theory of Computing, UC Berkeley, 2022
- **Rank 7**, G-Research Algorithmic Trading Oxbridge Challenge, 2021
- **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
- **Highest GPA** in Data Structure, Machine Learning, Computer Graphics, Android, Software Eng, System Programming, PHP, Microprocessor
- **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, Kubernetes, Google Cloud Platform, ROS

RESEARCH PUBLICATIONS | [Google Scholar](#) | 200+ Citations

1. Medbot: Conversational Artificial Intelligence powered Chatbot for delivering Telehealth after COVID-19, **IEEE** 5th International Conference on Communications and Electronic Systems (ICCES 2020); Urmil Bharti, Deepali Bajaj, **Hunar Batra** et al., IEEE Xplore [[Link](#)]
2. Serverless Deployment of a Voice-Bot for Visually Impaired, International Conference on Applied Soft Computing & Communication Networks (ACN 2020); Deepali Bajaj, Urmil Bharti, **Hunar Batra** et al., Book Chapter - **Springer** Singapore [[Link](#)]
3. CoVShorts: News Summarization application based on Deep NLP transformers for SARS-CoV-2, **IEEE** 9th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2021); **Hunar Batra**, Akansha J, et al. - IEEE [[Link](#)]
4. MUCE - A Multilingual Use Case Model Extractor using GPT-3; Deepali Bajaj, **Hunar Batra** et. al, International Journal of Information Technology (IJIT 2022), **Springer** [[Link](#)]
5. CovFakeBot: a machine learning based chatbot using ensemble learning technique for COVID-19 fake news detection; **Hunar Batra** et. al, International Journal of Artificial Intelligence and Soft Computing 2022 [[Link](#)]
6. TiGen - Title Generator based on Deep NLP Transformer Model for Scholarly Literature; **Hunar Batra**, Eshika G et. al, 3rd International Conference on Communication, Networks and Computing 2022, Springer [[Link](#)]
7. Protein Language Models are self-supervised mutagenesis learners [MSc Dissertation]; **Hunar Batra**, Peter Minary; To be published (2022)
8. Solidflix: A decentralised movie social sharing app with privacy-preserving recommendations; **Hunar Batra**, Jun Zhao et. al; Work in progress (2022)
9. Bias-Augmented Consistency Training Reduces Biased Reasoning in Chain-of-Thought; James Chua, Edward Rees, **Hunar Batra**, Sam Bowman, Julian Michael Ethan Perez, Miles Turpin; Under review for COLM 2024 [[Link](#)]

INVITED TALKS & WORKSHOPS

Oxford AI Mini-Conference, Feb 2024 – Large Language Models and AGI Panel

Oxford Women in Computer Science Panel, Nov 2023

Stanford Existential Risks Initiatives MATS Symposium, Feb 2023 – Accelerating Alignment Research via Human-AI Expert Iteration [[Slides](#)]

SolidWorld 2022 – Solidflix: A decentralised movie social sharing app with privacy-preserving recommendation [[Video](#)] [[Slides](#)]

GirlsWhoML 2022, University of Oxford – Introduction to Machine Learning (Linear Regression and Logistic Regression) [[Slides](#)]

ICML 2022, Oxford Women in Computer Science Virtual Social – Highlighting Women Researchers in Machine Learning

Oxford Computer Science Conference 2022 – Protein Language Modelling to generate de novo SARS-CoV-2 mutations [[Slides](#)]

Oxbridge Women in Computer Science Conference 2022 – Protein Language Modelling to generate de novo SARS-CoV-2 mutations [[Slides](#)]

ML Reading Group, University of Oxford (Feb and March 2022) – Presented state-of-the-art work on LLMs

ICRITO 2021, IEEE – Paper Presentation at 9th International Conference on Reliability, Infocom Technologies and Optimization, IEEE

ICCES'20, IEEE – Paper Presentation at 5th International Conference on Communication and Electronic Systems, IEEE

ACN'20, Springer – Paper Presentation at 5th International Conference on Applied Soft Computing and Communication Networks, Springer

Ryerson University (The DMZ, Think Outside the Valley 2020) – Process Automation with Chatbots [[Video](#)]

HackOn Hackathon 2020 – Ok Google! Let's build an action for Google Assistant [[Video](#)]

SRCC, University of Delhi 2019 – Chatbot Development for Marketing (youngest invited speaker)

Google DevFest New Delhi 2019 – Project showcase, Google Developer Students Club

WHRC 2018 – Ideation Paper Presentation on 'Ingestible Robots' at 15th WONCA World Rural Health Conference 2018

POSITIONS OF RESPONSIBILITY

Reviewer – ACM Transactions on Intelligent Systems and Technology 2024	[March 2024]
DPhil Academic Representative – Department of Computer Science, University of Oxford	[Oct 2024 - Present]
IT Officer – Oxford Women in Computer Science	[Oct 2024 - Present]
Reviewer – ICML 2022, AI4ABM workshop	[May 2022]
Student Entrepreneur – Oxford University Innovation and Oxford Science Enterprises	[June - July 2022]
Summer Fellow – Global Leadership Initiative, Oxford Character Project	[June - July 2022]
IT Officer – Oxford Women in Business	[April - Oct 2022]
IT Officer – Oxford Women in Computer Science	[Oct 2021 - Oct 2022]
MSc Academic Representative – Department of Computer Science, University of Oxford	[Oct 2021 - Oct 2022]
Solutions Challenge Lead – Google Developer Student Club Oxford	[Oct 2021 - Jan 2022]
Exam Marker – Mathematical Institute, University of Oxford	[Oct 2021 - Nov 2021]
Lead – Google Developer Student Club (One of the few students selected globally by Google)	[Jan 2019 - Aug 2020]
Mentor – Google Code-in at TensorFlow	[Nov 2019 - Jan 2020]