SUCHIR SALHAN

Cambridge Computer Science PhD Candidate · Writer & Journalist
Starred First BA (Cantab), MEng (Cantab)

https://www.suchirsalhan.com/ sas245@cam.ac.uk
07946290505 www.linkedin.com/in/ssalhan ResearchGate

I'm Suchir Salhan, a PhD Candidate in Computer Science at Gonville and Caius College, the University of Cambridge, specialising in Machine Learning and Natural Language Processing. My principal research interests focus on engineering more data-efficient "foundation models" (particularly Large Language Models and Vision-Language Models).

Education

Oct 2024 - Jul. 2028

PhD in Computer Science — Cambridge University

- Fully-funded PhD Studentship to research engineering Small-Scale and Interpretable Foundation Models to mitigate the risks associated with Transformer-based Foundation Models.

Oct 2020 - Jul. 2024

BA (Cantab) Class I* (Starred First), MEng (Cantab) Distinction — Computer Science and Linguistics Tripos

- **Part III Computer Science Tripos (Distinction):** Admitted to the integrated MEng/Part III of the Computer Science Tripos after a highly-competitive interview with Deputy Head of Computer Science, Prof Timothy Jones, and Russell Moore. Machine Learning Modules include **Advanced Topics in Machine Learning (Explainable AI, Imitation Learning)** and **Machine Learning for Language Processing**. Advanced Modules in Natural Language Processing, including **Natural Language Syntax and Parsing** and **Introduction to Computational Semantics**. Computer Security (Cybercrime). Achieved High Distinctions in several modules (80%+). Part III project, **Less is More: Augmenting Small-Scale Language Models with Syntactic Inductive Biases**, focussed on engineering **small-scale, compute-efficient Transformer-based Language Models**.
- **Bachelor of Arts (Class I with Distinction, "Starred First"):** **Class I with Distinction** aggregate mark in my undergraduate studies at Gonville & Caius College, University of Cambridge. High First Class in every Tripos Paper in Finals, including 75+ in Linguistic Theory, Syntax/Semantics, Typology, Morphology and Phonological Theory, Computational Linguistics (83, 98% in practical assessment) and Part IIB Dissertation Published.
- **Examiner's Feedback on Dissertation:** "This is an **outstanding dissertation** which presents a new approach to the topic of computational language modelling: incorporating theoretical principles from the work of Theresa Biberauer (from the model of 'Maximise Minimal Means') into the framework of modern-day 'Transformer' neural networks. The argumentation and discussion are **sophisticated and accomplished**, the point of view is confident. This dissertation could serve as the basis for a future graduate research project which would empirically test the predictions made in this work. The conjunction of linguistic theory and computational modelling is a relatively neglected area which has great potential for exploration, as proposed and demonstrated in this dissertation. There is evidence of extensive background reading and research, supplemented with thorough critical analyses in which the relevant literature is handled with dexterity and acumen. Presentation and referencing are immaculate. The addition of a case study also demonstrates strong practical skills, the ability to handle large datasets, and perform analyses at the intersection of computational method and linguistic theory."

Publications

Conference and Journal Publications

1. †**Salhan, S.A.**, Diehl-Martinez, Richard, Goriely, Zebulon & Buttery, Paula (2024) **Less is More: Pre-Training Cross-Lingual Small-Scale Language Models with Cognitively-Plausible Curriculum Learning Strategies**. In preparation for CoNLL 2024 BabyLM Challenge (Paper Track)

2. †**Salhan, S.A.** (2023) **On the Potential for Maximising Minimal Means in Transformer Language Models: A Dynamical Systems Perspective**. In Cambridge Occasional Papers in Linguistics. Volume 15, Article 3: pp. 55–110. ISSN: 2050-5949. Available online here.

3. **Salhan, Suchir**, Liu, Fangyu & Collier, Nigel. (2022/preprint) **Multimodal Language Modelling across Languages and Cultures: Grounding Strategies for Nouns and Verbs.** (Research Project, Language Technology Lab, Department of Theoretical and Applied Linguistics, University of Cambridge). [partially supported by a research award from Gonville & Caius College, Cambridge]

Seminar Presentations, Working Papers & Posters

4. **Salhan, S.A.** (2024), Diehl-Martinez, Richard, Goriely, Zebulon, Caines, Andrew & Buttery, Paula **LLMs "off-the-shelf" or Pretrain-from-Scratch? Recalibrating Biases and Improving Transparency using Small-Scale Language Models**. Learning and Human Intelligence Group, Department of Computer Science & Technology, University of Cambridge.

5. **Salhan, S.A.** (2023) **On the Potential for Maximising Minimal Means in Transformer Language Models**. Talk presented at SyntaxLab (14th February 2023), Department of Theoretical and Applied Linguistics, a departmental seminar organised by Dr Theresa Biberauer.

6. **Salhan, S.A**. (2021). **UROP Project Report 2021: Providing Automatic Feedback on Argumentation Quality to Learners of English** (UROP Research Internship, Natural Language and Information Processing Group, Department of Computer Science and Technology, University of Cambridge). [supervised by Prof Paula Buttery, Dr Andrew Caines, Dr Russell Moore (NLIP Group); Dr Thiemo Wambsganss (University of St Gallen/EPFL) and funded by Cambridge Assessment]

Experience

2022 Language Technology Lab — Researcher

I undertook an extended research project with Prof Nigel Collier and Fangyu Liu (Google DeepMind) in the Language Technology Lab (University of Cambridge). We worked on a creative probing of state-of-the-art multimodal language models. This research has, in part, been supported by Gonville and Caius College.

2021 Department of Computer Science & Technology — Research Assistant

I supported my supervisor Dr Li Nguyen (University of Cambridge) as a research assistant in the drafting of her computational sociolinguistics paper on variation in overt/null pronouns by Vietnamese-English bilingual speakers.

2021 Department of Computer Science & Technology — Undergraduate Researcher

I worked on a project in collaboration with Thiemo Wambsganss on developing the back-end machine learning architecture for an application that supports the argumentation skills of English language learners.

Industry and Other Activities

2023- **Co-Founder and Editor-in-Chief of Per Capita Media** — Media & Journalism

Per Capita is a progressive national student publication founded in Cambridge University in collaboration with University of Art's London, covering News, Features and Culture. The publication is supported by Lady Stothard, Dr Ruth Scurr. President of Cambridge University Per Capita Media.

2024 **Founder and President of Gonville & Caius Media and Journalism Society** — Media & Journalism

2023-2024 **Head of Policy, Editor & Head of Events** — Policy

Editor of Cybersecurity Policy Paper for **policy thinktank**, The Wilberforce Society. Published in Nov 2023 and presented in St John's College, available online here. Organising **Freedom of Press Panel Event in St John's College** and a panel event with the Foreign Minister of Sri Lanka.

2024 **The Cambridge Student** — Media & Journalism

Deputy Editor— the student paper of Cambridge University.

2023 **The Cambridge Student** — Media & Journalism

News & Investigations Editor and Interviews writer— the student paper of Cambridge University.

2023 **BBC The One Show** — Media & Journalism

Invited by producers at The One Show to help produce a documentary about higher education post-COVID. Responsible for organising student contributors around the country. Also, invited by producers at BBC Radio 5Live to discuss the Marking and Asssessment Boycott.

2023 **Radio Presenter** — Media & Journalism

Qualified Radio Presenter at CamFM.

2022-23 **Varsity Newspaper** — Media & Journalism

Member of the Editorial Team (Investigations Editor, Senior Science Editor, News Correspondent, 'Chief Sub-editor') of Varsity— the student paper of Cambridge University.

2022 **Gonville & Caius College JCR** — Community

I was elected Secretary of Caius JCR in 2022-23

2021-2024 **University of Cambridge** — Societies

CamFM, Caius Boat Club (Novice), Member of Cambridge Law Society (CULS), Cambridge Investment Banking Society (CIBS), Cambridge Union, Sub-editor and Author of Polyglossia Magazine 2020-2021, Member of Gonville and Caius College Access Team (helping in open days, access events).

Awards & Scholarships

2024 Department of Computer Science & Technology

Fully-Funded PhD Studentship Award from the Department of Computer Science and Technology for research on mitigating risks of Large Language Models.

2023 Gonville & Caius College

Memorial Prize, one of two academic accolades awarded to Gonville & Caius graduating cohort 2023

2023 Gonville & Caius College

Examination Prize

2023 Gonville & Caius College

Re-election to Senior Scholarship

2022 Gonville & Caius College

Senior Scholarship

2022 Gonville & Caius College

Examination Prize for First Class Tripos Performance

2022 Gonville & Caius College

Research Award awarded by the Senior Tutor of Gonville and Caius (Dr Andrew Spencer), for my research project with Prof Nigel Collier.

2022 Computer Science Dept.

Funding of £175 for **Research Assistant** work

2021 Computer Science Dept.

Funding by Cambridge Assessment for a UROP in the Department of Computer Science & Technology, worth £3000. I was the only first-year student in the entire University to participate in the two/three month Natural Language Processing UROP program.

Skills

- **Programming Languages:** Python (proficient); basic experience in MATLAB.
- **Machine Learning Frameworks:** NumPy, SciPy, matplotlib, keras, scikit-learn, tensorflow
- **Natural Language Processing Frameworks/Libraries:** NLTK, SpaCy, HuggingFace libraries

Seminars

- **Computer Science:** Natural Language and Information Processing (NLIP) Seminars, Language Technology Lab (LTL) Seminars.
- **Cognitive Science and Linguistics:** Syntax Lab. Co-organised a student seminar and reading group for Phonological Theory, with Prof Bert Vaux. Audited Graduate Seminar in Optimality Theory.