

G. THOMAS HUDSON

thomas@gthudson.xyz || [Github](#) || [LinkedIn](#) || [ORCID](#)

Machine learning researcher working on fundamental multimodal models and applications within healthcare. Specialising in multimodal and health data, developing foundation models and intelligent systems that bridge the gap between advanced AI methods and real-world clinical impact.

RESEARCH EXPERIENCE

Research Fellow, Centre for AI Fundamentals, University of Manchester 2026-

- Developing AI fundamentals (bayesian inference, human-in-the-loop, foundation models) and their applications to healthcare with Professor Sami Kaski.

Senior Research Associate, Durham University 2025-2026

- Developed methods to monitor the quality of hospital discharge notes, enabling clinicians to more effectively communicate, improving patient outcomes.
- Worked with a multidisciplinary team alongside clinicians from Salford Royal Hospital, producing multiple papers and a spinout company creating real-world impact.
- Separately, I independently built and led my own team developing fundamental ML methods, leading to multiple papers including a novel multimodal approach presented at ICCV 2025.
- Mentored Masters and Doctoral students, guiding research directions which led to publications in top venues such as EMNLP.

Senior Data Scientist, Evergreen Life 2025-

- Working at the intersection of research and industry, building new models to analyse electronic health records.
- Developed methods to automatically triage patient requests in GP practices and detect skin cancer in patient-taken images.

Post Doctoral Research Associate, Durham University 2023-2025

- Worked closely with veterinarians from industry to use LLMs to understand discussions on veterinary message-boards.
- Developed methods to extract diseases from unstructured case discussions and match them to custom veterinary vocabularies.
- Built tools which enable longitudinal trends in diseases and treatments to be identified, and to facilitate more productive discussions between veterinary professionals with problematic cases.
- Required project management and communication skills to facilitate collaboration between an international multidisciplinary team.

Data Scientist, Caspian

2018-2023

- Researching how machine learning techniques can help solve NLP problems in the anti-money laundering (AML) sector, including information extraction from large unstructured documents.
- Role involved applying research techniques to industry, including developing methods for managing and monitoring models in production.

EDUCATION AND AWARDS

PhD in Computer Science, Durham University

2019-2024

- Developed methods for multitask learning and long-document understanding in NLP and vision, introducing task-reformulation and task-decomposition frameworks.

MEng in Computer Science, Durham University

2014-2018

- Published my thesis on Native Language Identification
- BCS Chartered Institute for IT Prize - Awarded for being the top performing student in Computer Science

TEACHING

- Developed teaching materials, gave lectures and led practical sessions for *COMP4167: Natural Language Processing* and MSc in Business Analytics (*Natural Language Analysis*) at Durham University.
- Developed and led multiple workshops for the N8 Excellence in Computationally Intensive Research centre, where I lectured and ran practical sessions on NLP with examples drawn from my research experience targeted at researchers in other disciplines.
- Led practical sessions in other topics including introductory programming, networking, algorithms and data structures.
- Currently working towards the Associate Fellow of the HEA (AFHEA)

GRANTS

The Medical Research Council (MRC) Impact Acceleration Accounts (IAAs) [grant number 2867143] (£64,576). “*Automated patient discharge letter analysis for more cost-effective, in-depth hospital audits and quality improvement projects*”, involved in the writing and reviewing.

[under review] Health and Social Care Delivery Research (HSDR), “*Improving Care Quality and Safety in the NHS Through Automated Monitoring of Clinical Communication*” (£1.2m), key role in the writing of the application.

PUBLICATIONS

Benedict Jones, James Harmsworth King, Matthew Watson, **G. Thomas Hudson**, and Noura Al Moubayed (2026), “*Bridging Survival Analysis and Machine Learning to Improve Healthy Life Expectancy Estimation using PHR Records*”, npj Digital Medicine [in press]

G. Thomas Hudson*, Chenghao Xiao*, Benedict Jones, James Harmsworth King, and Noura Al Moubayed, (*joint first authors) (2026), “*Generalizable multilingual medical text de-identification using generative instruction tuning*”, Nature Communications Medicine [in press]

G. Thomas Hudson, Benjamin David James, Matthew Watson, Mark Holland, Olivier Gaillemin, Noura Al-Moubayed, Darren Green (2026), “*Automating the quality monitoring of a hospital discharge summary improvement project utilising large language models*”, npj Digital Medicine [\[url\]](#)

Matthew Watson, Thomas Winterbottom, **Thomas Hudson**, Benedict Jones, Hubert P. H. Shum, Amir Atapour-Abarghouei, Toby Breckon, James Harmsworth King, and Noura Al Moubayed (2026), “*Multimodal Models for Skin Cancer Classification using Clinical Free Text and Dermatoscopic Images*”, Nature Communications Medicine [\[url\]](#)

G. Thomas Hudson, Dean Slack, Thomas Winterbottom, Jamie Sterling, Chenghao Xiao, Junjie Shentu, and Noura Al Moubayed (2025). “*Everything is a video: Unifying modalities through next-frame prediction*”. International Conference on Computer Vision (ICCV 2025).[\[url\]](#)

Dean Slack, **G. Thomas Hudson**, Thomas Winterbottom, and Noura Al Moubayed (2025). “*Video Prediction of Dynamic Physical Simulations With Pixel-Space Spatiotemporal Transformers*”. In: IEEE Transactions on Neural Networks and Learning Systems.[\[url\]](#)

Chenghao Xiao, **G. Thomas Hudson**, and Noura Al Moubayed (2023). “*Towards more human-like language models based on contextualizer pretraining strategy*”. EMNLP/CoNLL 2023 BabyLM Challenge; Loose Track Winner. [\[url\]](#)

Chenghao Xiao, Yizhi Li, **G. Thomas Hudson**, Chenghua Lin, and Noura Al Moubayed (2023). “*Length is a Curse and a Blessing for Document-level Semantics*”. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023). [\[url\]](#)

Chenghao Xiao, Zihuiwen Ye, **G. Thomas Hudson**, Zhongtian Sun, Phil Blunsom, and Noura Al Moubayed (2023). “*Can Text Encoders be Deceived by Length Attack?*” In: The 14th International Conference on Learning Representations (ICLR2023). [\[url\]](#)

G. Thomas Hudson and Noura Al Moubayed (2022). “*MuLD: The Multitask Long Document Benchmark*”. In: Proceedings of the Language Resources and Evaluation Conference (LREC2022). [\[url\]](#)

G. Thomas Hudson and Noura Al Moubayed (2021). “*Ask me in your own words: paraphrasing for multitask question answering*”. In: PeerJ Computer Science 7, e759. [\[url\]](#)

G. Thomas Hudson and Sardar Jaf (2018). “*On the Development of a Large Scale Corpus for Native Language Identification*”. In: Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018). [\[url\]](#)

Under Review

G. Thomas Hudson, Benjamin David James, Matthew Watson, Mark Holland, Olivier Gaillemin, Noura Al-Moubayed, Darren Green, “*Predicting Transitional Care Breakdown from Discharge Requests Using Large Language Models*”, npj Digital Medicine

Matthew Watson, **G. Thomas Hudson**, Benedict Jones, James Harmsworth King, and Noura Al Moubayed, “*Multimodal Machine Learning Classification of Skin Lesions using Macroscopic Images with Uncertain Labels*”, IEEE transactions on Medical Imaging

Nour Mubarak, **G. Thomas Hudson**, and Noura Al Moubayed, “*A Comprehensive Analysis of Gender Bias in Arabic-English Multimodal Image Captioning*”, Language Resources and Evaluation (LRE)

In Preparation

G. Thomas Hudson, Dean Slack, Chenghao Xiao, “*Everything is Image In-painting: Unifying Modalities through Masked Image Completion*”

G. Thomas Hudson, Benjamin David James, Matthew Watson, Mark Holland, Olivier Gaillemin, Noura Al-Moubayed, Darren Green, “*Temporal Patterns in Clinician Discharge Requests: Insights from Large Language Model Extraction*”

Chenghao Xiao, **G. Thomas Hudson**, and Noura Al Moubayed. “*Rar-b: Reasoning as retrieval benchmark*” [\[url\]](#)

Chenghao Xiao, Zhuoxu Huang, Danlu Chen, **G. Thomas Hudson**, Yizhi Li, Haoran Duan, Chenghua Lin, Jie Fu, Jungong Han, Noura Al Moubayed. “*Pixel sentence representation learning*” [\[url\]](#)

Conference Presentations/Posters

G. Thomas Hudson, Benjamin David James, Matthew Watson, Mark Holland, Olivier Gaillemin, Noura Al-Moubayed, Darren Green, “*Clinician Language in Discharge Requests: Using Large Language Models to Predict Transitional Care Breakdown*”, HDR UK 2025

G. Thomas Hudson, Benjamin David James, Matthew Watson, Mark Holland, Olivier Gaillemin, Noura Al-Moubayed, Darren Green, “*Automating the quality monitoring of a discharge summary improvement project utilising large language models*”, SAM 2025