Application ID: 165069, Christian Alexander Schroeder: Assistant and Associate Professor - Computer Science (100893-0325)

Personal Information

We take privacy seriously and will only use your personal information to administer your application. For more information please see our Data Protection Policies (https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection).

Title	Dr.
Preferred pronouns	He/him/his
Given Name(s)	Christian Alexander
Family Name	Schroeder
Email	christian.schroeder@eng.ox.ac.uk
Preferred Phone No	07496850950

Additional Information

Are you currently employed by University of Warwick?

No

Will you now or in the future require a visa to obtain/continue to hold the right to work legally in the UK?

No

Reasonable Adjustments

We make intentional efforts to employ and retain people with disabilities. The following question will aid us to assist you should you need it during the application process.

Do you have any medical condition, special educational needs or disability that means that you may require reasonable adjustments made for you during either the online assessment, interview or assessment centre stages of our selection process?

No

Source

Where did you find the advert for this vacancy?

Warwick webpages

What made you apply for this vacancy at the University of Warwick? Please select all that apply.

Career progression

Is there anything further that prompted you to apply?

No

References

Reference 1

Title	Professor
First Name	Philip
Last Name	Torr
Email	philip.torr@eng.ox.ac.uk
Reference Type	Academic
Relationship	Fellowship Host/Mentor

Reference 2

Title	Dr.
First Name	Martin
Last Name	Strohmeier
Email	Martin.Strohmeier@armasuisse.ch

Reference Type	Work
Relationship	Collaborator/Funder

Reference 3

Title	Professor
First Name	Yoshua
Last Name	Bengio
Email	yoshua.bengio@mila.quebec
Reference Type	Academic
Relationship	Former Supervisor/Mentor

Christian A. Schroeder de Witt — Curriculum Vitæ

University of Oxford
christian.schroeder@eng.ox.ac.uk
Google Scholar: DE60h_0AAAAJ
schroederdewitt.com

The **latest version** of this CV can be found at schroederdewitt.com/uploads/resume.pdf.

I hold a DPhil in Engineering Science from Oxford, where I focused on AI security and multi-agent learning. My work has been **cited over 6,100 times (h-index 19) and I have been awarded £1m as PI/Co-PI (RAEng, EPSRC, OpenAI, Schmidt Futures, Foresight Institute, Armasuisse, Microsoft).** My progress on secure steganography—highlighted by Quanta Magazine in 2023—helped lay the groundwork for the new field of Multi-Agent Security which has been recognised by leading funders. As a UK Intelligence Community Research Fellow, I draw on information theory and game-theoretic ideas to build practical, transparent frameworks that help ensure autonomous systems remain reliable and aligned with human values. My interdisciplinary research is reflected in my teaching: As Stipendiary Lecturer, I teach undergraduates, and as doctoral training course lecturer I designed **the world-first's lecture course on the Frontiers of Cooperative AI**.

Academic Positions & Affiliations

University of Oxford (Department of Engineering Science)
UK Intelligence Community Research Fellow — *2024–*
Senior Research Associate (PI: Prof. Philip Torr) — *2024*
Postdoctoral Research Assistant (PI: Prof. Jakob Foerster) — *2022-2023*

St Catherine's College , University of Oxford
Stipendiary Lecturer in Computer Science — *2024-*

Lawrence Livermore National Laboratory (CA)
Visiting Scholar (Sabbatical) — *from 2025*

Education

University of Oxford — *2017-2021*
Doctor of Philosophy in Engineering (Artificial Intelligence)
Dissertation: *Communication and Coordination in Deep Multi-Agent Reinforcement Learning*
Advisors: Prof. Philip H.S. Torr and Prof. Shimon Whiteson
EPSRC IAA Doctoral Impact Fund award (2022)

University of Oxford — 2013
Master of Science in Computer Science, *awarded Distinction*
Dissertation: The ZX Calculus is Incomplete for Quantum Mechanics (with Prof. Bob Coecke)

University of Oxford — 2012
Master of Physics (equiv. BSc+1 year MSc), awarded *First-Class Honors*
Award-winning dissertation in complex systems science / theoretical biology (with Prof. Ard Louis)
East Exhibition, Fitzgerald Prize, University of Oxford Tessella Prize for Innovation in Software

Grants

Total: **£1m ($1.3m)** as PI/Co-PI, plus **£2m (2.7m)** as contributor

Foresight Institute Grant, PI — $124,000 — 2025
Foundations of Multi-Agent Security

UK Intelligence Community Postdoctoral Research Fellowship (RAEng), PI —£250,000 — 2024
Detecting AI-generated media and Hidden Messages therein, in collaboration with UK HMGCC

Schmidt Futures Grant, PI — $133,000 — 2024
Virtual Institute on Grand Strategy and AI, in collaboration with BBC OSINT and Bergen University

OpenAI Superalignment Fast Grant, Co-PI — $327,000 2024
Weak-to-strong generalization, acceptance rate <2% (PI: Prof. Philip H.S. Torr)

Future of Life Institute, PI — $10,000 2024
Grant for developing policy briefs for the EU GPAI Code of Practice call

Armasuisse Science + Technology Grants, PI (de-facto) —£235,000 2022-2024
AI Security and multi-agent learning

EPSRC IAA Doctoral Impact Fund Award, PI (de-facto) —£30,000 2022

Microsoft AI for Earth Grant, PI — £10,000 2021

Grant development & Editing

European Research Council (ERC) Starting Grant, *Co-Editor and Substantial Contributor* 2022
PI: Prof. Jakob Foerster, € 2.3m awarded in 2022

Cooperative AI Foundation Grant, *Named Researcher* 2021-2023
PI: Prof. Jakob Foerster, £166,370 awarded, 2022

Key Publications (Selected) – see appendix for a full list

*indicates equal contribution/co-first authorship. Last authorships signify both seniority and substantial contributions.

[A] **C. Schroeder de Witt***, S. Sokota*, J. Z. Kolter, J.N. Foerster, M. Strohmeier. *Perfectly Secure Steganography using Minimum Entropy Coupling*. International Conference on Learning Representations (**ICLR**), 2023. **Featured by Quanta Magazine, Scientific American, and Bruce Schneier on Security.**

[B] T. Franzmeyer, S. McAleer, J. Henriques, J. Foerster, P. Torr, A. Bibi, **C. Schroeder de Witt**. *Illusory Attacks: Information-Theoretic Detectability Matters in Adversarial Attacks*, International Conference on Learning Representations (**ICLR**), 2024. **Spotlight Talk** (2% acceptance rate)

[C] S. Motwani, M. Baranchuk, V. Bolina, L. Hammond, **C. Schroeder de Witt**. *Secret Collusion among AI Agents: Multi-Agent Deception by Steganography*. Conference on Neural Information Processing Systems (**NeurIPS**), 2024.

[D] L. Nasvytis†, K. Sandbrink, J. Foerster, T. Franzmeyer†*, and **C. Schroeder de Witt***. *Rethinking Out-of-Distribution Detection for Reinforcement Learning: Advancing Methods for Evaluation and Detection*. International Conference on Autonomous Agents and Multiagent Systems (**AAMAS**), 2024. (Oral Talk)

[E] S. Sokota*, **C. Schroeder de Witt***, M. Igl, L. Zintgraf, P.H.S. Torr, S. Whiteson, J. Foerster. *Communicating via Markov Decision Processes*. International Conference on Machine Learning (**ICML**), 2022.

Awards (Selected)

UK IC Postdoctoral Research Fellowship, Royal Academy of Engineering 2024
Biggest Discovery of the Year in Computer Science, Quanta Magazine 2023
EPSRC IAA Doctoral Impact Fund Award, EPSRC 2022
30 under 35 Rising Strategist (Europe), International Strategy Forum **(**Schmidt Futures) 2021
Tessella Prize for Innovation in Software (Dissertation Prize), University of Oxford 2012
Fitzgerald Prize, Exeter College (University of Oxford) 2012
Studienstiftung des Deutschen Volkes Fellowship (awarded to <0.5% of German students) 2011–2012

Academic Research Visits (Selected)

Mila – Quebec Artificial Intelligence Institute, Montreal (CA)
Visiting Researcher (Host: Prof. Yoshua Bengio) 2022–2023

INVITED TALKS (SELECTED) – SEE APPENDIX FOR A FULL LIST

Agent Security Convening – Schmidt Futures x Palo Alto Networks x RAND Corporation, Santa Clara (CA)

- Held opening keynote on Multi-Agent Security at invitation-only workshop. 2025

IMPACT & ENGAGEMENT

6100+ citations, h-index: 19, media reach: ca. 2-5 million (globally)

POLICY ENGAGEMENT (SELECTED)

RAND Corporation 2025
Invited Expert (Multi-Agent Security)

BBC News
Invited Expert (AI, live in-studio – global reach: ca. 2-5 million) 2024-

European Council on Foreign Relations
Consultant (AI Security) 2023

Schmidt Futures
Consultant (AI Security) 2022-2023

PUBLIC ENGAGEMENT

BBC News – AI Decoded (London) - watch here 09/2024
Live, in-studio expert on AI and disinformation (ca. 2-5 million viewers)

Munich Security Conference – Responsible Leaders' Hub 2023
Lead Panelist ("Polycrisis") with Ravi Agrawal (Foreign Policy), Ngaire Woods (Dean, Blavatnik School of Governance, Oxford) and Hina Rabbani Khar (Minister of State for Foreign Affairs, Pakistan)

Google Policy Summit (Málaga, ES) 2023
Expert Panelist (AI Security)

PATENTS

Perfectly-Secure Steganography through Minimum Entropy Coupling, **C. Schroeder de Witt**, S. Sokota, J.N. Foerster, M. Strohmeier, UK patent application number 2209681.2 *(open-sourced in 2025)* 2023

TEACHING

Undergraduate Tutoring

- Stipendiary Lecturer in Computer Science – St Catherine's College, Oxford 2024-
  Machine Learning (undergraduate small-group tutorials) – 23 students across 4 colleges (consortium)
  "The students have given Christian extremely positive feedback" (Prof. Christoph Haase)
- University of Georgia (UGA) at Oxford Program 2021-2024
  CSCI 4550: Artificial Intelligence (undergraduate small-group tutorials)
  Taught and assessed individual undergraduate exchange students form diverse backgrounds (reinvited thrice).
  Published joint Wiley book chapter [1] with my former UGA student Eshaan Agrawal.

Lecturer and Course Developer

- *The Frontiers of Cooperative AI – highly positive student feedback, reinvited annually since 2023* 2023-
  Paid for developing and lecturing a 18h PhD-level lecture course. The course covers selected topics across game theory, sociology, and multi-agent security and features interactive labs.
  AIMS CDT (Doctoral Training Centre), University of Oxford

- *Deep Learning* (undergraduate lecture course) - *highly positive student feedback* 2016
  Humboldt University, Berlin

Teaching Assistant

- Machine Learning (undergraduate course) – *highly positive student feedback* 2016
  Humboldt University, Berlin

SUPERVISION (RECENT)

PhD Students (official Co-Supervision)

- Julia Karbing *(funded by CAIS, OpenPhil and a Cooperative AI Foundation PhD Fellowship)* 2024–
  Primary advisor: Prof. Philip Torr
- Sumeet Motwani *(funded by Eric Schmidt and a Cooperative AI Foundation PhD Fellowship)* 2024–
  Primary advisor: Prof. Philip Torr
- Constantin Venhoff *(funded by OpenAI)* 2024–
  Primary advisor: Prof. Philip Torr

Graduate Students (official Co-Supervision)

- Kumud Lakara (MSc Advanced Computer Science) – thesis: distinction, now: J.P. Morgan 2024
  Primary advisor: Prof. Philip Torr, secondary advisor: Prof. Christian Rupprecht
- Georgia Channing (MSc Advanced Computer Science) – thesis: distinction, now: DPhil at Oxford 2024
  Primary advisor: Prof. Philip Torr, secondary advisor: Prof. Ronald Clark
- Constantin Venhoff (MSc Advanced Computer Science) – thesis: distinction, now: DPhil at Oxford 2024
  Primary advisor: Prof. Philip Torr, secondar advisor: Prof. Ani Calinescu
- Tala Jafari (MSc Advanced Computer Science) – thesis: distinction, now: PhD at MIT (incoming) 2024
  Primary advisor: Prof. Philip Torr, secondary advisor: Prof. Varun Kanade
- Aqib Mahfuz (MSc Advanced Computer Science) – thesis: distinction, now: Google 2024
  Primary advisor: Prof. Philip Torr, secondary advisor: Prof. Mark van der Wilk

Research Assistants (official Supervision)

- Magnus Sesodia (funded by UKRI) 2024-
- Tala Jafari (funded by Open Philanthropy) 2024-

OUTREACH

Diversity, Equity & Inclusion

- Awarded Schmidt Futures Grant (£30,000) for refugee aid (with Josephine Goube, SisTech) 2023
- AI for Agent-Based Modeling Mentorship Program, *Founder and Chair* 2022–
- Mentorship of DeepLearning Indaba students *Eltayeb Ahmed (now DPhil at University of Oxford), and Khaulat Abdulhakeem (now MSc Data Science at Stanford)* 2022–
- GoVolunteer e.V. (Berlin, DE), *Head of Partner and Project Management (Refugee Aid Coordinator)* 2016
- Oxford University Schools Plus, *personal music tutor for a 15-year from a disadvantaged background* 2013
- Oxford University Jacari Program, *weekly home tuition for a 6-year old with a migrant background* 2012-2013

SERVICE AND LEADERSHIP

International Grant Reviews

- Reviewer, DfG Noether Grants (AI Security/Trustworthy ML/Interpretability) 2025–
- Lead Reviewer, DFG Cluster of Excellence Competition (€60 million per grant, AI Security) 2025–

Conference Program Committees

- Area Chair, Technical AI Governance Workshop (ICML) 2025
- Senior Program Committee, European Conference on Artificial Intelligence (ECAI) 2024
- Area Chair, AI for Agent-Based Modelling Community (ICML, ICLR) 2022/2023
- Reviewer, International Conference on Learning Representations (ICLR) 2020-

- Reviewer, International Conference on Machine Learning (ICML) 2019–
- Reviewer, Conference on Neural Information Processing Systems (NeurIPS) 2019–

Workshop Program Committees

- Lead-Organizer, Multi-Agent Security: Security as Key to AI Safety Workshop at NeurIPS 2023
- Lead-Organizer, AI for Agent-Based Modeling Workshop at (ICML, ICLR) 2022/2023
- NeurIPS Workshop on Bayesian Deep Learning 2018–2021

PhD Committees

- Jake Levi, PhD in Computer Science, University of Oxford 2024–

Selection Committees and Interviewing

- Researcher Committee, Department of Engineering Science (University of Oxford) 2024-
- Undergraduate Admissions, St. Catherine's College (University of Oxford) 2024–

Professional Experience

Lawrence Livermore National Laboratory (CA)
Visiting Researcher (Sabbatical) 2025

Google AI
Machine Learning Intern 2019

SETI Institute (US)
Research Advisor (AI/Climate), Frontier Development Lab 2021

European Space Agency (ESA)
Machine Learning Researcher, Frontier Development Lab 2020

Man AHL (Oxford/London, UK)
Machine Learning Intern (Paid) 2016

Humbold-University (Berlin, DE)
Research Assistant – Deutsche Forschungsgemeinschaft (ML Theory) 2015-2016

MenschDanke GmbH (Berlin, DE)
Head of Engineering (leading team of 4 in-house developers) 2014

APPENDIX – PLEASE SEE BELOW

APPENDIX

PUBLICATIONS (APPENDIX)

Highlight indicates top-5 key publication.
* indicates equal contribution/co-first authorship.
† indicates undergraduate, master's, or doctoral student I advised or mentored
Multi-Agent Security indicates publications within the multi-agent security space.

Peer-Reviewed Papers

[31] P. Peigne-Lefebvre, M. Kniejski, F. Sondej, M. David, J. Hoelscher-Obermaier, **C. Schroeder de Witt**, E. Kran. *Multi-Agent Security Tax: Trading Off Security and Collaboration Capabilities in Multi-Agent Systems*. Proceedings of the AAAI Conference on Artificial Intelligence (**AAAI**), 2025. Multi-Agent Security

[30] A. Mudide†, J. Engels, E. Michaud, M. Tegmark, **C. Schroeder de Witt**. *Efficient Dictionary Learning with Switch Sparse Autoencoders*. International Conference on Learning Representations (**ICLR**), 2025.

[29] A. Draguns†, A. Gritsevskiy, SR. Motwani†, C. Rogers-Smith, J. Ladish, **C. Schroeder de Witt**. *Unelicitable Backdoors in Language Models via Cryptographic Transformer Circuits*. Conference on Neural Information Processing Systems (**NeurIPS**), 2024. Multi-Agent Security

[28] S. Motwani†, M. Baranchuk†, V. Bolina, L. Hammond, **C. Schroeder de Witt**. *Secret Collusion among AI Agents: Multi-Agent Deception by Steganography*. Conference on Neural Information Processing Systems (**NeurIPS**), 2024. Multi-Agent Security

[27] A. Rutherford, B. Ellis, M. Gallici, J. Cook, A. Lupu, G. Ingvarsson J., T. Willi, R. Hammond, A. Khan, C. Schroeder de Witt, A. Souly, S. Bandyopadhyay, M. Samvelyan, M. Jiang, R. Lange, S. Whiteson, B. Lacerda, N. Hawes, T. Rocktäschel, C. Lu, J. Foerster. *JaxMARL: Multi-agent RL Environments and Algorithms in JAX*. Conference on Neural Information Processing Systems (**NeurIPS**), 2024.

[26] M. Fellows, B. Kaplowitz, **C. Schroeder de Witt**, S. Whiteson. *Bayesian Exploration Networks*. International Conference on Machine Learning (**ICML**), 2024.

[25] F. Eiras, A. Petrov, B. Vidgen, **C. Schroeder De Witt**, F. Pizzati, K. Elkins, S. Mukhopadhyay, A. Bibi, Botos C., F. Steibel, F. Barez, G. Smith, G. Guadagni, J. Chun, J. Cabot, J.M. Imperial, J.A. Nolazco-Flores, L. Landay, M. Jackson, P. Röttger, P. Torr, T. Darrell, Y.S. Lee, J. Foerster. *Position: Near to Mid-term Risks and Opportunities of open-source Generative AI*. International Conference on Machine Learning (**ICML**), 2024.

[24] T. Franzmeyer†, S. McAleer, J. Henriques, J. Foerster, P. Torr, A. Bibi, **C. Schroeder de Witt**. *Illusory Attacks: Information-Theoretic Detectability Matters in Adversarial Attacks*, International Conference on Learning Representations (**ICLR**), 2024. **Spotlight Talk** (2% acceptance rate)

[23] L. Nasvytis†, K. Sandbrink, J. Foerster, T. Franzmeyer†*, and **C. Schroeder de Witt***. *Rethinking Out-of-Distribution Detection for Reinforcement Learning: Advancing Methods for Evaluation and Detection*. International Conference on Autonomous Agents and Multiagent Systems (**AAMAS**), 2024. (Oral Talk)

[22] K.J. Sandbrink, B. Christian, L.M. Nasvytis†, **C. Schroeder de Witt,** P. Butlin. *Can Reinforcement Learning model Learning across Development? Online Lifelong Learning through Adaptive Intrinsic Motivation*. Proceedings of the Annual Meeting of the Cognitive Science Society, 2024.

[21] S. Sokota, D. Sam, **C. Schroeder de Witt**, S. Compton, J. Foerster, J.Z. Kolter. *Computing low-entropy couplings for large-support distributions*. Conference on Uncertainty in Artificial Intelligence (**UAI**), 2024.

[20] U. Anwar, A. Saparov, J. Rando, D. Paleka, M. Turpin, P. Hase, E.S. Lubana, E. Jenner, S. Casper, O. Sourbut, B.L. Edelman, Z. Zhang, M. Günther, A. Korinek, J. Hernandez-Orallo, L. Hammond, E. Bigelow, A. Pan, L. Langosco, T. Korbak, H. Zhang, R. Zhong, S. Ó hÉigeartaigh, G. Recchia, G. Corsi, A. Chan, M. Anderljung, L. Edwards, A. Petrov, **C. Schroeder de Witt**, S.R. Motwani†, Y. Bengio, D. Chen, P. H.S. Torr, S. Albanie, T. Maharaj, J. Foerster, F. Tramer, H. He, A. Kasirzadeh, Y. Choi, D. Krueger. *Computing low-*

entropy couplings for large-support distributions. Transactions on Machine Learning (**TMLR**), 2024.

[19] **C. Schroeder de Witt***, S. Sokota*, J. Z. Kolter, J.N. Foerster, M. Strohmeier. *Perfectly Secure Steganography using Minimum Entropy Coupling*. International Conference on Learning Representations (**ICLR**), 2023. **Featured by Quanta Magazine, Scientific American, and Bruce Schneier on Security.**

[18] Y.L. Lo, **C. Schroeder de Witt**, S. Sokota, J.N. Foerster, S. Whiteson. *Cheap Talk Discovery and Utilization in Multi-Agent Reinforcement Learning*. International Conference on Learning Representations (**ICLR**), 2023.

[17] S. Sokota*, **C. Schroeder de Witt***, M. Igl, L. Zintgraf, P.H.S. Torr, S. Whiteson, J. Foerster. *Communicating via Markov Decision Processes*. International Conference on Machine Learning (**ICML**), 2022.

[16] D. Muglich, **C. Schroeder de Witt**, E. van der Pol, S. Whiteson, J. Foerster. *Equivariant networks for zero-shot coordination*. Conference on Neural Information Processing Systems (**NeurIPS**), 2022.

[15] C. Lu, J. Kuba, A. Letcher, L. Metz, **C. Schroeder de Witt**, J. Foerster. *Discovered Policy Optimisation*. Conference on Neural Information Processing Systems (**NeurIPS**), 2022.

[14] J. Grudzien, **C. Schroeder De Witt**, J. Foerster. *Mirror learning: A unifying Framework of Policy Optimisation.* International Conference on Machine Learning (**ICML**), 2022.

[13] D. Muglich, L.M. Zintgraf, **C. Schroeder De Witt**, S. Whiteson, J. Foerster. *Generalized beliefs for cooperative AI*. International Conference on Machine Learning (**ICML**), 2022.

[12] J. Grudzien, **C. Schroeder De Witt**, J. Foerster. *Model-free Opponent Shaping*. International Conference on Machine Learning (**ICML**), 2022.

[11] F. Wood, S. Naderiparizi, A. Ścibior, A. Munk, M. Ghadiri, A.G. Baydin, B. Gram-Hansen, **C. Schroeder de Witt**, R. Zinkov, P. Torr, T. Rainforth, Y. Whye Teh. *Amortized Rejection Sampling in Universal Probabilistic Programming*. International Conference on Machine Learning (**AISTATS**), 2022.

[10] B. Peng*, T. Rashid*, **C. Schroeder de Witt***, P.A. Kamienny, P. Torr, W. Boehmer, S. Whiteson. *FACMAC: Factored Multi-Agent Centralised Policy Gradients*. Conference on Neural Information Processing Systems (**NeurIPS**), 2021.

[9] **C. Schroeder de Witt***, C. Tong*, V. Zantedeschi, D. De Martini, A. Kalaitzis, M. Chantry, D. Watson-Parris, P. Bilinski. *RainBench: Towards Data-Driven Global Precipitation Forecasting from Satellite Imagery*. Proceedings of the AAAI Conference on Artificial Intelligence (**AAAI**), 2021.

[8] T. Rashid, M. Samvelyan, **C. Schroeder de Witt**, G. Farquhar, J.N. Foerster, S. Whiteson. *Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning*. Journal of Machine Learning Research (**JMLR**), 2020.

[7] B. Gram-Hansen, **C. Schroeder de Witt**, T. Rainforth, P. Torr, Y. Whye Teh, A.G. Baydin. *Hijacking Malaria Simulators with Probabilistic Programming*. AI for Social Good Workshop at ICML 2019.

[6] **C. Schroeder de Witt***, T. Hornigold*. *Stratospheric Aerosol Injection as a Deep Reinforcement Learning Problem*. Tackling Climate Change with Machine Learning Workshop at ICML 2019. **Best Idea Award.**

[5] M. Samvelyan, T. Rashid, **C. Schroeder de Witt**, G. Farquhar, N. Nardelli, T.G.J. Rudner, C.-M. Hung, P. Torr, J. Foerster, S. Whiteson, *The Starcraft Multi-Agent Challenge*. International Conference on Autonomous Agents and Multiagent Systems (**AAMAS**), 2019.

[4] **C. Schroeder de Witt***, J. Foerster*, G. Farquhar, P. Torr, W. Boehmer and S. Whiteson. *Multi-Agent Common Knowledge Reinforcement Learning*. Conference on Neural Information Processing Systems (**NeurIPS**), 2019.

[3] J. Zimmert, **C. Schroeder de Witt**, G. Kerg, M. Kloft. Safe Screening for Support Vector Machines. Optimization in Machine Learning (OPT) Workshop at NIPS 2015.

[2] **C. Schroeder de Witt**, V. Zamdzhiev, The ZX-Calculus is Incomplete for Quantum Mechanics. Quantum Physics and Logic (QPL), 2014.

Peer-Reviewed Book Chapters

[1] E. Agrawal† and **C. Schroeder de Witt**. *Testing the Limits of the World's Largest Control Task: Solar Geoengineering as a Deep Reinforcement Learning Problem*. Geoengineering and Climate Change: Methods, Risks, and Governance, 2025. (**Wiley**)

INVITED TALKS (APPENDIX)

Multi-Agent Security: Deception, Collusion, and Undetectable Threats

- Agent Security Convening (Schmidt Futures, Palo Alto Networks, RAND – Santa Clara CA) 2025
- Oxford AIMS Seminar 2024
- New Orleans Alignment Workshop 2023
- Cooperative AI Foundation Retreat 2023
- Cyber Alp 2023 Conference 2023
- Oxford CCW Cyber Strategy Working Group 2023
- ELLIS Symposium (Tuebingen, DE) 2023
- Cyber Alp 2022 Conference 2022

Bringing Multi-Agent Learning to Societal Impact

- MIT, Media Lab 2023
- NYU, Data Science Center 2023
- Schmidt Futures, New York Office 2023
- MILA, Quebec AI Institute 2023
- Oxford Institute for New Economic Thinking 2022

Progress in Cooperative Deep Multi-Agent Reinforcement Learning

- Edinburgh Centre For Robotics (UK) 2020
- SDU UAS drone center (Odense, DK) 2020
- Huawei DaVinci Forum (Shanghai) 2018

AI for Climate Policy: The Road to Impact

- Salesforce/MILA AI4GCC seminar 2022
- NVidia GTC 2020, (San Jose, CA, 2020) 2020

Teaching Statement

Dr. Christian Schroeder de Witt
April 28, 2025

Teaching and mentoring are essential to advancing scientific progress and nurturing intellectual growth. To me, they are as vital as research. My teaching philosophy fosters critical thinking, intellectual independence, and adaptability through personalized, student-centered learning. Drawing on my experience teaching across diverse educational systems, I have designed and delivered innovative courses, including the world's first on Cooperative AI. By combining technical rigor with real-world relevance, I aim to inspire students to master complex concepts and apply their knowledge to address pressing global challenges. I am deeply committed to teaching excellence through engaging tutorials, forward-thinking curriculum development, and dedicated mentorship. This is evidenced by my three DPhil students all securing prestigious fellowships and three of my six recent master's students choosing to stay on under my supervision. My goal is to equip students with the tools, skills, and confidence to excel not only as engineers and research scientists, but as independent thinkers and leaders - whether in academia, government, industry, or wider society.

Teaching Experience

I have taught across multiple institutions, disciplines, and student levels, combining diverse experiences into a coherent and impactful teaching portfolio. At the **University of Oxford, St Catherine's College**, I currently serve as a **Stipendiary Lecturer in Computer Science**, delivering small-group tutorials to 23 second- and third-year undergraduates across a teaching consortium comprising St Catz, Oriel, Somerville, and Univ. My tutorials focus on machine learning, where students have praised the clarity and depth of my explanations, reflected in outstanding student feedback (according to Prof. Christoph Haase).

Since 2022, I have developed and delivered a pioneering lecture series, *The Frontiers of Cooperative AI*, commissioned by the **Oxford AIMS CDT program**. This globally unique course introduces doctoral students to foundational topics in human-AI interaction, cooperative mechanism design, and multi-agent security. Alongside technical rigor, I emphasize real-world relevance by integrating interdisciplinary perspectives from public policy and experimental psychology. The course comprises 12 hours of lectures and 6 hours of labs, where students explore multi-agent LLM systems using Colab notebooks. It has been enthusiastically received, with three successful deliveries to date (MT+HT 2023, MT 2024).

My teaching career began at **Humboldt University, Berlin**, where I led tutorials for an introductory machine learning course and developed a lecture series on deep learning for graduate students. Both experiences were marked by highly positive evaluations and strengthened my commitment to fostering technical excellence. Beyond university-level teaching, I have also engaged in mentoring and educational outreach. During my undergraduate studies at Oxford, I volunteered with **Jacari**, where, over the course of a year with weekly visits, I adapted teaching strategies to engage a primary school student from a disadvantaged background, successfully improving focus and academic outcomes.

Teaching Philosophy

My teaching philosophy is grounded in the belief that learning should foster critical thinking, intellectual independence, and adaptability. I strongly align with student-centered learning as a cornerstone of academic and professional development.

Having taught across diverse educational systems and student levels, I have learned that effective teaching requires adaptability. I tailor my approach to the needs of each group or individual: in small-group tutorials, I assess students' backgrounds and interests to craft individualized learning pathways that encourage engagement and mastery of complex concepts. In larger lecture settings, I balance structured core material with opportunities for deeper exploration, ensuring all students are empowered to excel at their own pace.

In my lecture series *The Frontiers of Cooperative AI*, I encourage students to think critically about the application of scientific frameworks to societal phenomena. For example, I challenge students to question the limitations of applying game theory to contexts such as international diplomacy and human cooperation. This approach pushes them to reflect on the broader implications of AI research and to consider the societal nuances often oversimplified, or even grossly misrepresented, by formal models.

Equally, I believe teaching extends beyond delivering content: it is about mentorship, fostering curiosity, and cultivating confidence in students as independent thinkers. My dedication to this approach is reflected in my students' successes, including distinctions, paper submissions to top-tier conferences, and prestigious fellowships.

My mentoring philosophy emphasizes developing leadership and professional skills. Through structured guidance and interdisciplinary collaborations, I equip students with the confidence to lead projects, engage with industry,

and contribute meaningfully to the global AI community. In my tutorials, I implement scaffolding techniques to ensure accessibility for students from varied educational backgrounds, starting with foundational concepts and progressively building complexity. I also provide additional resources and targeted feedback to address individual learning needs.

Ultimately, I strive to create an inclusive, collaborative, and intellectually stimulating environment where students are encouraged to question, explore, and apply their knowledge to address pressing societal and scientific challenges.

Teaching Interests

My teaching interests encompass foundational and advanced topics in machine learning, artificial intelligence, multi-agent systems, and core areas of computer science. Drawing on my academic training in theoretical physics and computer science, alongside extensive teaching and research experience, I am well-equipped to deliver and develop courses across these domains.

I am enthusiastic about teaching essential undergraduate courses such as *Linear Algebra*, *Probability and Statistics*, *Functional Programming*, and *Imperative Programming*, emphasizing conceptual clarity and practical problem-solving to build the fundamental skills underpinning advanced AI research. Building on my experience teaching machine learning tutorials at St Catherine's College and the University of Georgia at Oxford, I have ample experience teaching *Machine Learning* and *Artificial Intelligence*. Students have praised my ability to combine technical rigor with real-world applications, fostering strong theoretical foundations and hands-on proficiency.

At Warwick, I aim to expand the curriculum by introducing a foundational course on *Reinforcement Learning*, addressing its underrepresentation despite its critical role in modern AI systems, including VLM post-training. This course would integrate theoretical principles with practical lab sessions to meet the growing demand for expertise in this area. My contributions to AI security - such as breakthroughs in steganography, multi-agent deception, and adversarial machine learning - position me to teach advanced modules on adversarial robustness, AI trustworthiness, and information-theoretic security, equipping students to tackle critical challenges in securing AI systems.

I am also committed to interdisciplinary teaching, exploring the intersection of AI with fields such as public policy, psychology, and ethics. My experience in AI security, multi-agent systems, and societal impacts informs courses that encourage students to critically engage with the real-world consequences of AI technologies. For example, my *The Frontiers in Cooperative AI* lecture course integrates interdisciplinary perspectives to address challenges like misinformation detection, collaborative intelligence, and AI governance.

I am eager to supervise undergraduate and graduate research projects, fostering intellectual curiosity and guiding students toward impactful work that advances both theory and practice. My teaching and mentorship align strongly with Oxford's interdisciplinary focus, particularly in addressing societal and ethical challenges, as exemplified by the Oxford Martin School's initiatives.

Mentorship and Graduate Supervision

Supervision and Mentorship Experience. Supervision and mentorship are central to my academic practice. I foster intellectual growth, research independence, and professional development by balancing autonomy with structured guidance. My mentorship extends beyond technical instruction to strategic career advice, grant-writing support, and facilitating interdisciplinary collaboration, helping students navigate their academic journeys with confidence and purpose.

I currently co-supervise three DPhil students at Oxford: **Sumeet Motwani**, **Constantin Venhoff**, and **Julia Karbing**, all of whom have secured highly competitive funding with my guidance. Sumeet leads our work on multi-agent LLM systems, focusing on Vision-Language Model (VLM) reasoning, collaborating with Prof. Ronald Clark at MBZUAI and Lawrence Livermore National Laboratory. Constantin focuses on mechanistic interpretability, working closely with Neel Nanda (Google DeepMind), contributing to transparency in agent-level systems. Julia leads research on multi-agent security, where she explores adversarial robustness while collaborating with Georgia Channing and Nicholae Dobra, who address cooperative mechanisms and defensive strategies.

At the master's level, I supervised six MSc Advanced Computer Science students, resulting in four distinctions and five paper submissions to top-tier conferences such as *ICLR* and *AAMAS*. Notably, Georgia Channing and Kumud Lakara produced applied research in collaboration with BBC Verify and Adobe, respectively. Three of these students - Georgia, Constantin, and **Tala Jafari** - have continued under my mentorship, with Tala (incoming PhD at MIT) and currently being employed as a Research Assistant funded by Open Philanthropy. These outcomes demonstrate both academic success and the trust students place in my mentorship for their continued development.

My experience also spans interdisciplinary and undergraduate supervision. As a PDRA, I mentored Tim Franzmeyer to an *ICLR Spotlight* paper on *Illusory Attacks* (top 5%) and supervised Eshaan Agrawal, whose work culminated in a Wiley book chapter. Currently, I mentor Research Assistant Magnus Sesodia and undergraduate students Tze-Yang Poon (Part B) and Henry Gasztowtt (Part C). My mentorship philosophy is built on

creating a supportive, collaborative environment that equips students with the tools and confidence to produce impactful research and develop into independent, critical thinkers.

Group Structure. My research group fosters collaboration, leadership, and mentorship, enabling students to engage in cutting-edge AI research while developing as independent researchers. Although formally supervised by Prof. Phil Torr, I provide daily mentorship to ensure a dynamic, inclusive, and cohesive research culture.

Each DPhil student leads a distinct research focus while mentoring junior colleagues to promote knowledge exchange and leadership development.

This structure encourages close collaboration between DPhil, master's, and undergraduate students, creating a supportive environment that advances our research goals while building students' technical expertise, leadership, and mentorship skills, preparing them for impactful contributions to AI research.

Equality, Diversity, and Inclusion. I am deeply committed to fostering a diverse, equitable, and inclusive academic environment where students from all backgrounds feel supported, valued, and empowered to excel. Throughout my career, I have worked to address systemic barriers to participation in AI research, support underrepresented groups, and create pathways for students to thrive.

This commitment is reflected in the diversity of my research group, where three of my six DPhil students are non-male: **Julia Karbing**, **Georgia Channing**, and **Angira Sharma**. I take particular pride in mentoring female researchers in AI, an area where gender representation remains a challenge, ensuring they develop the skills, confidence, and leadership to pursue impactful careers.

Beyond gender diversity, I have mentored students from geographically, economically, and culturally diverse backgrounds. My approach combines tailored academic guidance with practical support. I have helped students overcome financial challenges through short-term research assistantships and philanthropic funding, ensuring promising research careers remain on track. For example, I provided strategic guidance to Nicholae Dobra, who described my mentorship as *"one of the most useful chats [he] had regarding [his] PhD journey."*

My efforts to promote inclusion extend to international outreach. During my DPhil, I co-supervised a master's student from Eritrea who later secured offers from both Oxford and Cambridge. Since 2019, I have mentored a young female AI researcher from Nigeria, who is now about to complete her master's degree at Stanford University, having been awarded a Dean's Fellowship.

As Chair of the AI4ABM community, I established a mentorship program pairing junior researchers with senior mentors to support workshop submissions. I also prioritized inclusivity as the lead organizer of AI4ABM (NeurIPS 2022, ICLR 2023) and MASEC, ensuring gender balance, geographically diverse representation, and sponsorship opportunities for participants from disadvantaged backgrounds.

Through these initiatives and my active mentoring of undergraduate, master's, and doctoral students, I strive to cultivate a collaborative research culture that embraces diversity and empowers the next generation of AI researchers to break barriers, tackle global challenges, and advance impactful interdisciplinary work.

Talks, Moderation, and Public Engagement

I am committed to communicating AI research to diverse audiences, bridging technical expertise with societal relevance. My public engagement spans high-profile media appearances, invited talks at leading institutions, and expert contributions to internationally recognized forums and think tanks. Through these engagements, I aim to advance public understanding of AI, foster critical interdisciplinary dialogue, and address global challenges by communicating AI's potential and limitations to technical and non-technical audiences alike.

In October 2024, I made a live studio appearance on **BBC News**, reaching 2–5 million viewers. During the segment, I explained research challenges in AI-driven misinformation detection and critiqued a US study on persuasion, delivering accessible and nuanced analysis. My performance was described as *"brilliant"* by both the BBC program director and the BBC Verify lead data scientist, and Oxford's communications team featured it prominently in their daily media highlights.

As an AI expert consultant for the **European Council on Foreign Relations (ECFR)**, I have contributed to discussions on the geopolitical implications of AI, including participation at the 2023 **Google Policy Summit in Málaga**. I also presented on the societal impact of AI at the **Munich Security Conference**, one of the world's most prestigious platforms for international security.

In academia, I have delivered invited talks at leading institutions, including **Google DeepMind**, **Carnegie Mellon University**, **the University of Cambridge**, and the **Oxford Institute for New Economic Thinking**. These talks have covered AI security, multi-agent systems, and explainable AI, engaging audiences from academia, industry, and policy.

As Chair of the **AI4ABM community** and lead organizer of workshops such as **AI4ABM** (NeurIPS 2022, ICLR 2023) and **MASEC**, I have moderated high-level panel discussions that connect researchers, industry leaders, and policymakers. These events prioritize inclusivity and provide platforms for emerging researchers, ensuring broader participation across disciplines and geographies.

Research Proposal: Laying the Foundations of Multi-Agent Security

Dr. Christian Schroeder de Witt

April 28, 2025

Frontier AI models now autonomously perform complex tasks online – from booking travel to conducting research [1, 2] and using computer interfaces originally designed for humans [3, 4, 5]. As these systems increasingly interact with each other, we see this emerging in trading agents [6], market research agents [7], personal assistants [8], OS agents [9], and cyber defense systems [10]. This evolution introduces unique security vulnerabilities: when AI agents with private information and competing objectives interact, they develop emergent behaviors like covert collusion, coordinated attacks, and cascade failures that cannot be predicted by analyzing individual agents alone. This necessitates the emergence of "multi-agent security" – a cross-disciplinary field focused on securing systems of interacting AI agents.

In this proposal, I am laying the theoretical and conceptual foundations of the field of multi-agent security. My core insight to this effect is that **undetectable threats, which I pioneered in the form of secret collusion [11] by steganography [12], illusory attacks [13], and encrypted backdoors [14], represent a focal point through which existing epistemic [15], information-theoretic [16], and strategic [17] frameworks of secrecy and security in multi-agent systems need to be reconsidered.** This is because undetectable threats *cannot be known (evading epistemics)*, *cannot be detected (evading information-theoretic measurement)*, and a strategic best-response to undetectable threats is almost always systemically degenerative and prohibitively costly. This proposal examines foundational security-performance trade-offs due to undetectable threats in autonomous multi-agent systems, including through theoretical guarantees, secure environment design, and active mitigation.

1 Prior Work and Relevant Expertise

I was the lead organiser of the world's first workshop on multi-agent security at NeurIPS 2023 [18], and lead the section on multi-agent security in the Cooperative AI Foundation's report on multi-agent risks [19]. The field has been recognised as core funding area of the Cooperative AI Foundation, and I received a $124k grant (PI)on "The Foundations of Multi-Agent Security" from the Foresight Institute in 2025. This demonstrates the field's maturity.

This project also builds on my **substantial body of research** in AI safety and security, multi-agent learning, and interpretability. A major breakthrough in my research is the development of a solution to a 25-year-old open problem in **perfectly secure steganography** [20], which was featured in Scientific American [21], and listed as one of 2023's biggest discoveries in computer science by Quanta Magazine [22, 23]. This work demonstrates my ability to tackle complex, foundational challenges. This work introduced a novel algorithm for performing efficient perfectly secure steganography on arbitrary covertext distributions, which theoretically allows AI agents to secretly communicate without this very act being statistically detectable. In [11, 24], we showed that such **deceptive behaviour** can arise from misaligned optimisation pressure. Similarly, my theoretical framework for **illusory attacks** [25] (ICLR 2024 Spotlight) established fundamental bounds on the detectability of deceptive behaviors, providing a foundational lens for addressing emergent deceptive capabilities in multi-agent systems. Finally, in [14], we showed that large language model agents can hide skills in **encrypted backdoors** that are unelicitable in white-box settings, therefore putting tight constraints on white-box detectability using interpretability tools.

My prior work has also advanced **multi-agent learning reinforcement learning** [26, 27, 28, 29, 30, 31, 32, 33, 34], exemplified by contributions to MACKRL [26], IPPO [33], QMIX [34] and FACMAC [31], which laid the groundwork for analyzing agent coordination and information sharing. These efforts are complemented by **practical tools**, such as the widely adopted StarCraft Multi-Agent Challenge [30] and Multi-Agent MuJoCo [27] benchmarks, which demonstrate my ability to translate theoretical insights into actionable resources for the research community.

The essential background for this project also includes my contributions to **mechanistic interpretability** [35, 36], where I developed methods to uncover latent behaviors in AI agents at scale. These methods form the basis for introspection tools in the proposed work. Additionally, my collaborations with industry partners, such as Adobe

and the BBC, have demonstrated the real-world applicability of my research, particularly in addressing challenges of **trust and transparency** in AI systems [37, 38].

This foundation of theoretical innovation, practical tool development, and real-world impact positions me uniquely to deliver the proposed research program, advancing the frontiers of trustworthy multi-agent AI.

2 Laying the Foundations of Multi-Agent Security

WP1: Foundations of Systemic Security

This work package establishes the theoretical foundations for achieving *systemic security*, which extends beyond agentic security to the environment. This framework offers a robust foundation for evaluating and ensuring systemic security in multi-agent AI systems. **Key insight: Security in multi-agent systems must consider agents, groups, and the environment as an interconnected system.**

a) Quantifying Group Alignment. We expand *alignment* [39] to a comprehensive notion of *group alignment*, integrating agents, principals, and mechanisms in multi-agent systems for cooperative behavior [40]. A novel performance gap metric will quantify coordination inefficiencies, applied to scenarios like autonomous driving where failures can have critical consequences. These metrics, combined with anomaly detection methods [41], enable the detection and mitigation of misalignment.

b) Foundations of Systemic Security. Building on *group alignment*, we define *systemic security*, where the environment adapts to facilitate multi-agent security. This is crucial for addressing trade-offs between modifying agent behavior and environmental factors, as seen in retrofitting autonomous driving infrastructure. Additionally, recent work [20, 11] shows multi-agent systems can achieve *perfect deception*, coordinating undetectable attacks evading both black- and white-box detection. We will frame systemic **alignment as an environment design problem** [42], optimizing environment parameters θ alongside agent policies. To ensure robustness against adversarial attacks [25], we constrain the design to favor Nash equilibria that minimize the damage from undetectable deception. Effective environments must not only achieve high *group alignment* metrics but also exhibit illusory robustness to limit the potential harm from information-theoretic deception.

WP2: Foundations of Agent Introspection

As artificial intelligence systems become increasingly complex, the necessity for advanced introspection capabilities grows. These capabilities are essential to detect and address emergent deceptive behaviors, which could pose significant risks in high-stakes applications. This work package focuses on mechanisms to identify and manage such behaviors at the individual agent level, leveraging interpretability tools. A central insight driving this research is the realization that AI agents can potentially conceal advanced deceptive capabilities through unelicitable backdoors—mechanisms that remain undetectable even under white-box introspection conditions [14].

a) Undetectable Backdoors. Our recent findings indicate that large language model (LLM) agents can encode hidden capabilities within encrypted backdoors, which effectively evade existing white-box interpretability methods [14]. Building on this foundation, this work package will investigate whether statistically undetectable backdoors can exist within neural networks. Specifically, we will explore the hypothesis that factorized weight subspaces, inaccessible to current circuit discovery techniques, may enable such backdoors [36]. Understanding these mechanisms is critical to preventing exploitation and ensuring trust in AI systems.

b) Backdoor Mitigations. To counteract these risks, we propose two complementary mitigation strategies. First, active detection techniques will be developed, which involve modifying the AI network weights—such as by introducing noise—to probe and disrupt potential backdoors. This approach will systematically evaluate the trade-offs between effectively destroying backdoors and any resultant degradation in model performance. Second, we will analyze the inherent computational capabilities of AI networks to perform steganographic and cryptographic functions that could enable undetectable collusion based on the maximum size of computational circuits that can be implemented. This analysis builds on our prior work, which demonstrated how AI systems may be naturally predisposed to exploit such mechanisms [11].

WP3: Theoretical Guarantees

We will derive theoretical guarantees for systemic security in two steps, addressing recent calls for probabilistic alignment bounds [43].

a) Systemic Security Guarantees: We will first establish ϵ-Nash equilibria for illusory attack games, where attack strength trades off against undetectability in a fixed environment. This builds on our insight that these

games can be modeled as one-sided zero-sum partially observable stochastic games [44]. Using the framework of unsupervised environment design [42], we will develop probabilistic systemic safety bounds for ϵ-Nash equilibria induced by environment parameterizations θ under a cooperative task T. As a stretch goal, we will extend this analysis to systems with multiple attackers and more complex environments.

b) Mitigation Guarantees: Although perfect deception is undetectable, such states are unlikely to arise spontaneously and typically require adversarial threat models. Emergent misalignment can be detected using WP1 metrics and advances in anomaly detection [41]. We will derive information-theoretic detectability bounds in both black-box and white-box settings, which we will test against anomaly detection techniques that integrate mechanistic interpretability tools [35, 36].

WP4: Open-Source Evaluation Frameworks and Benchmarks

This work package establishes standardized frameworks and benchmarks to evaluate **alignment and cooperation** in multi-agent systems, addressing a key gap in reproducible and trustworthy AI research. Open-source tools will promote adoption and collaboration across academic, industrial, and policymaking communities. Integrating insights from WP1 and WP2, this work package will deliver **a rigorous, practical, and accessible ecosystem** for evaluating and fostering alignment in multi-agent systems, advancing the broader goal of trustworthy AI.

a) We will develop **tailored benchmarks for generative and reinforcement learning environments**, focusing on real-world challenges such as autonomous driving and collaborative large language model (LLM) teams. Metrics from WP1 will be refined to evaluate **team alignment, coordination efficiency**, and **robustness to adversarial behaviors**.

b) In addition to benchmarks, **intuitive dashboards for policymakers** will translate technical insights into actionable AI governance guidance, bridging research and policymaking for systemic security challenges.

Applications will focus on **autonomous driving fleets** and **collaborative LLM systems**, leveraging advances in multi-agent post-training [45] to address emergent misalignment and adversarial threats, such as hidden backdoors or steganographic collusion. For autonomous vehicles, the focus will be on **hidden communication channels** and **collusion risks**, while LLM systems will examine scenarios involving **private information** and **cooperative safety in coding, reasoning, and scheduling**.

3 Expected Outcomes and Impact

In the next five years, my research will expand MASEC's applications to critical domains, including securing autonomous cyber-physical systems and detecting misinformation. I aim to collaborate with Warwick's social sciences and ethics departments to address societal impacts, while developing scalable AI safety tools through industrial partnerships. The work aims to establish the first comprehensive framework for systemic security and introspection, enabling trustworthy multi-agent AI systems. By developing novel metrics, algorithms, and theoretical guarantees alongside practical, open-source tools, the project will advance state-of-the-art AI safety practices. These deliverables will drive adoption across academia and industry while shaping policies in domains such as autonomous driving, cybersecurity, and AI governance.

Funding Strategy. The research program builds on a diversified funding portfolio, including secured grants from Schmidt Futures, OpenAI, and Armasuisse, alongside new applications to major funding bodies such as the EPSRC Open Fellowship and the European Research Council. A focus on cultivating long-term partnerships with industry leaders like Adobe Research, BBC Verify, and Google DeepMind ensures financial sustainability and strategic impact. Smaller, high-impact grants like the Cooperative AI Foundation research grants will also be pursued to support postdoctoral researchers and computational resources, ensuring the project remains agile in resource allocation.

Internal Collaboration. The research will leverage Warwick's vibrant ecosystem of AI experts, engaging with collaborators across departments. Regular interdisciplinary workshops and seminars will facilitate knowledge exchange. **External Collaboration.** The project has established partnerships with high-profile organizations. Armasuisse Science+Technology supports research on LLM systems and funds DPhil students. Adobe Research and BBC Verify contribute to real-world applications of multi-agent collaboration and misinformation detection. Google DeepMind explores multi-agent deception frameworks, while the Future of Life Institute and Schmidt Futures bridge policy and research by producing accessible briefs for policymakers and fostering thought leadership.

These collaborations will ensure both the theoretical and applied impact of the research, while fostering a global network of researchers and practitioners in AI safety. This holistic approach will secure Warwick's leadership in tackling some of the most pressing challenges in AI and society.

References

[1] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Anil Palepu, Petar Sirkovic, Artiom Myaskovsky, Felix Weissenberger, Keran Rong, Ryutaro Tanno, Khaled Saab, Dan Popovici, Jacob Blum, Fan Zhang, Katherine Chou, Avinatan Hassidim, Burak Gokturk, Amin Vahdat, Pushmeet Kohli, Yossi Matias, Andrew Carroll, Kavita Kulkarni, Nenad Tomasev, Yuan Guan, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R. D. Costa, José R. Penadés, Gary Peltz, Yunhan Xu, Annalisa Pawlosky, Alan Karthikesalingam, and Vivek Natarajan. Towards an AI co-scientist, February 2025.

[2] Samuel Schmidgall, Yusheng Su, Ze Wang, Ximeng Sun, Jialian Wu, Xiaodong Yu, Jiang Liu, Zicheng Liu, and Emad Barsoum. Agent Laboratory: Using LLM Agents as Research Assistants, January 2025.

[3] Tianlin Shi, Andrej Karpathy, Linxi Fan, Jonathan Hernandez, and Percy Liang. World of Bits: An Open-Domain Platform for Web-Based Agents. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3135–3144. PMLR, July 2017. ISSN: 2640-3498.

[4] Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. MIND2WEB: towards a generalist agent for the web. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, pages 28091–28114, Red Hook, NY, USA, December 2023. Curran Associates Inc.

[5] Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. WebArena: A Realistic Web Environment for Building Autonomous Agents. October 2023.

[6] Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. TradingAgents: Multi-Agents LLM Financial Trading Framework, April 2025.

[7] James Brand, Ayelet Israeli, and Donald Ngwe. Using LLMs for Market Research, March 2023.

[8] Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, Rui Kong, Yile Wang, Hanfei Geng, Jian Luan, Xuefeng Jin, Zilong Ye, Guanjing Xiong, Fan Zhang, Xiang Li, Mengwei Xu, Zhijun Li, Peng Li, Yang Liu, Ya-Qin Zhang, and Yunxin Liu. Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security, May 2024.

[9] Kai Mei, Xi Zhu, Wujiang Xu, Wenyue Hua, Mingyu Jin, Zelong Li, Shuyuan Xu, Ruosong Ye, Yingqiang Ge, and Yongfeng Zhang. AIOS: LLM Agent Operating System, November 2024.

[10] Anna Knack and Ant Burke. Autonomous Cyber Defence Phase II. 2024.

[11] SR Motwani, M Baranchuk, M Strohmeier, V Bolina, PHS Torr, L Hammond, and C Schroeder de Witt. Secret Collusion among AI Agents: Multi-Agent Deception via Steganography. In *NeurIPS*, November 2024.

[12] Christian Schroeder de Witt, Samuel Sokota, J. Zico Kolter, Jakob Nicolaus Foerster, and Martin Strohmeier. Perfectly Secure Steganography Using Minimum Entropy Coupling. September 2023.

[13] Tim Franzmeyer, Stephen Marcus McAleer, Joao F. Henriques, Jakob Nicolaus Foerster, Philip Torr, Adel Bibi, and Christian Schroeder de Witt. Illusory attacks: Information-theoretic detectability matters in adversarial attacks. In *The Twelfth International Conference on Learning Representations*, 2024.

[14] Andis Draguns, Andrew Gritsevskiy, Sumeet Ramesh Motwani, and Christian Schroeder de Witt. Unelicitable Backdoors via Cryptographic Transformer Circuits. November 2024.

[15] Joseph Y. Halpern and Kevin R. O'Neill. Secrecy in Multiagent Systems. *ACM Trans. Inf. Syst. Secur.*, 12(1):5:1–5:47, October 2008.

[16] Alessandra Di Pierro, Chris Hankin, and Herbert Wiklicky. Approximate non-interference. *Journal of Computer Security*, 12(1):37–81, January 2004. Publisher: SAGE Publications.

[17] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, USA, November 2008.

[18] Christian Schroeder de Witt, Hawra Milani, Klaudia Krawiecka, Swapneel Mehta, Carla Cremer, and Martin Strohmeier. Multi-Agent Security Workshop at NeurIPS 2023, 2023.

[19] Lewis Hammond, Alan Chan, Jesse Clifton, Jason Hoelscher-Obermaier, Akbir Khan, Euan McLean, Chandler Smith, Wolfram Barfuss, Jakob Foerster, Tomáš Gavenčiak, The Anh Han, Edward Hughes, Vojtěch Kovařík, Jan Kulveit, Joel Z. Leibo, Caspar Oesterheld, Christian Schroeder de Witt, Nisarg Shah, Michael Wellman, Paolo Bova, Theodor Cimpeanu, Carson Ezell, Quentin Feuillade-Montixi, Matija Franklin, Esben Kran, Igor Krawczuk, Max Lamparth, Niklas Lauffer, Alexander Meinke, Sumeet Motwani, Anka Reuel, Vincent Conitzer, Michael Dennis, Iason Gabriel, Adam Gleave, Gillian Hadfield, Nika Haghtalab, Atoosa Kasirzadeh, Sébastien Krier, Kate Larson, Joel Lehman, David C. Parkes, Georgios Piliouras, and Iyad Rahwan. Multi-Agent Risks from Advanced AI, February 2025.

[20] Christian Schroeder de Witt, Samuel Sokota, J. Zico Kolter, Jakob Foerster, and Martin Strohmeier. Perfectly Secure Steganography Using Minimum Entropy Coupling. 2022. ICLR 2023.

[21] Dina Genkina. AI Could Smuggle Secret Messages in Memes, September 2023. Scientific American.

[22] Stephen Ornes. Secret Messages Can Hide in AI-Generated Media, May 2023. Quanta Magazine.

[23] Bill Andrews. The Biggest Discoveries in Computer Science in 2023, December 2023. Quanta Magazine.

[24] Yohan Mathew, Ollie Matthews, Robert McCarthy, Joan Velja, Christian Schroeder de Witt, Dylan Cope, and Nandi Schoots. Hidden in Plain Text: Emergence & Mitigation of Steganographic Collusion in LLMs, October 2024.

[25] Tim Franzmeyer, Stephen Marcus McAleer, Joao F. Henriques, Jakob Nicolaus Foerster, Philip Torr, Adel Bibi, and Christian Schroeder de Witt. Illusory attacks: Information-theoretic detectability matters in adversarial attacks. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

[26] Christian Schroeder de Witt, Jakob Foerster, Gregory Farquhar, Philip Torr, Wendelin Boehmer, and Shimon Whiteson. Multi-Agent Common Knowledge Reinforcement Learning. In *NeurIPS*, 2019.

[27] Christian Schroeder de Witt and Bei andet al. Peng. Deep multi-agent reinforcement learning for decentralized continuous cooperative control. *arXiv preprint arXiv:2003.06709*, 19, 2020.

[28] Yat Long Lo, Christian Schroeder de Witt, Samuel Sokota, Jakob Nicolaus Foerster, and Shimon Whiteson. Cheap Talk Discovery and Utilization in Multi-Agent Reinforcement Learning, March 2023.

[29] A Rutherford, Be Ellis, M Gallici, J Cook, A Lupu, G Ingvarsson, T Willi, A Khan, C Schroeder de Witt, and A Souly. Jaxmarl: Multi-agent rl environments in jax. *arXiv preprint arXiv:2311.10090*, 2023.

[30] M Samvelyan, T Rashid, C Schroeder de Witt, G Farquhar, and et al. The StarCraft Multi-Agent Challenge. In *AAMAS*, AAMAS '19, pages 2186–2188, May 2019.

[31] Bei Peng, Tabish Rashid, Christian Schroeder de Witt, and et al. FACMAC: Factored Multi-Agent Centralised Policy Gradients. In *NeurIPS*, volume 34, pages 12208–12221. Curran Associates, Inc., 2021.

[32] Darius Muglich, Luisa M. Zintgraf, Christian Schroeder De Witt, Shimon Whiteson, and Jakob Foerster. Generalized beliefs for cooperative AI. In *International Conference on Machine Learning*, pages 16062–16082. PMLR, 2022.

[33] Christian Schroeder de Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge?, Nov 2020.

[34] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):1–51, 2020.

[35] Anish Mudide, Joshua Engels, Eric J. Michaud, Max Tegmark, and Christian Schroeder de Witt. Efficient Dictionary Learning with Switch Sparse Autoencoders, October 2024.

[36] Constantin Venhoff, Anisoara Calinescu, Philip Torr, and Christian Schroeder de Witt. SAGE: Scalable Ground Truth Evaluations for Large Sparse Autoencoders, October 2024.

[37] Georgia Channing, Juil Sock, Ronald Clark, Philip Torr, and Christian Schroeder de Witt. Toward Robust Real-World Audio Deepfake Detection: Closing the Explainability Gap, October 2024.

[38] Kumud Lakara, Juil Sock, Christian Rupprecht, Philip Torr, John Collomosse, and Christian Schroeder de Witt. MAD-Sherlock: Multi-Agent Debates for Out-of-Context Misinformation Detection. October 2024.

[39] Usman Anwar and et al. Foundational Challenges in Assuring Alignment and Safety of Large Language Models, September 2024.

[40] Raphael Koster and et al. Human-centred mechanism design with Democratic AI. *Nature Human Behaviour*, 6(10), October 2022.

[41] Rethinking Out-of-Distribution Detection for Reinforcement Learning: Advancing Methods for Evaluation and Detection. In *AAMAS*, AAMAS '24, Richland, SC.

[42] Michael Dennis and et al. Emergent complexity and zero-shot transfer via unsupervised environment design. In *NeurIPS*, NIPS '20, pages 13049–13061, Red Hook, NY, USA, December 2020.

[43] D Dalrymple and et al. Towards Guaranteed Safe AI. 10.48550/ARXIV.2405.06624, 2024.

[44] Karel Horák, Branislav Bošanský, Vojtěch Kovařík, and Christopher Kiekintveld. Solving zero-sum one-sided partially observable stochastic games. *Artificial Intelligence*, 316:103838, March 2023.

[45] SR Motwani, C Smith, RJ Das, M Rybchuk, PHS Torr, I Laptev, F Pizzati, R Clark, and C Schroeder de Witt. MALT: Improving Reasoning with Multi-Agent LLM Training, December 2024.

Dr. Christian Schroeder de Witt
UK Intelligence Community Research Fellow
University of Oxford, Department of Engineering Science
Stipendiary Lecturer in Computer Science at St Catherine's College
Parks Road, Oxford, OX1 3PJ
Tel: +44(0) 7496 850950, Email: christian.schroeder@eng.ox.ac.uk

4/28/2025

Professor Yulia Timofeeva
Head of the Department of Computer Science
University of Warwick

Dear Professor Timofeeva,

I am writing to apply for the position of Associate Professor (Reference #100893-0325) in the Department of Computer Science. As a researcher focused on AI security and multi-agent learning, with particular expertise in Multi-Agent Security, I am excited by the opportunity to contribute to Warwick's vibrant interdisciplinary research environment and to help shape the university's future during a period of unprecedented investment and growth in AI.

My research program in AI security and multi-agent learning aligns seamlessly with Warwick's strengths in computer science. With my background spanning theoretical physics, computer science, and complex systems, I have established new directions in deep multi-agent reinforcement learning, co-developing foundational algorithms such as QMIX, IPPO, and MACKRL during my DPhil at Oxford. As a Research Fellow, my breakthrough work on implicit communication in multi-agent learning created a perfectly-secure information hiding method, recognized by Quanta Magazine as solving a 25+ year open problem in information security.

This line of research has led me to think deeply about security in adaptive multi-agent systems, and my subsequent published work on how AI systems can communicate covertly (ICML'23, with CMU), potentially collude (NeurIPS'24, with Google DeepMind) or even attack (ICLR'24 Spotlight) secretly, and create undetectable backdoors (NeurIPS'24) has brought the novel concept of *undetectable threats* to the center of investigation of AI security in multi-agent systems. My research has generated over 6,100 citations (h-index 19), demonstrating significant impact.

My goal is now to help mitigate the critical safety and security risks I uncovered, leading me to develop the broader emerging field of Multi-Agent Security, which is now internationally recognised as the field addressing security in systems of advanced AI agents beyond existing cybersecurity and AI safety approaches. Since then, my group at Oxford has also established expertise in high-dimensional anomaly detection, LLM reasoning and multi-agent LLM post-training, which I would be able to bring into my new group.

Altogether, this enables my vision for my work at Warwick: working at the forefront of making multi-agent systems of adaptive agents resilient, and helping unlock their safe future application in the decentralized AI internet or distributed superintelligence. I'm particularly interested in collaborating

with Prof. Hongkai Wen on the security of multimodal intelligent systems, Prof. Paolo Turrini on Cooperative Artificial Intelligence, Prof. Long Tran-Thanh on complex systems of AI agents, and Prof. Rob Procter on trustworthy AI, to name but a few. I believe my research themes fit extremely well with the department, as well as enable fruitful collaboration with other departments, such as the Department of Statistics.

I would bring substantial funding experience to Warwick, having secured over £1 million as PI/Co-PI and contributed to an additional £2 million in grants. Specifically, I would be able to bring to my new role over £100k in cloud compute from my Foresight Institute grant. This would allow me to head-start my new group not only with the help of my existing PhD students' expertise, but would also provide for any specialised GPU compute needs until October 2026.

My teaching philosophy emphasizes hands-on, practical learning that connects theoretical foundations with real-world applications. At Oxford, I've designed and delivered the world's first doctoral-level course on Cooperative AI and served as a Stipendiary Lecturer in undergraduate teaching, receiving excellent student feedback. I would be excited to contribute to Warwick's undergraduate AI program and to your established MSc programs in Applied Artificial Intelligence and related fields.

As someone committed to increasing equity, diversity, and inclusion in AI, I've mentored students from underrepresented backgrounds through the DeepLearning Indaba program and founded the AI for Agent-Based Modeling Community Mentorship Program. I would continue these efforts at Warwick, helping to build a diverse and inclusive research community.

Warwick University's position at the forefront of AI innovation—as a founding Alan Turing Institute partner with significant investments in financial technology and prestigious AI scholarship programs—creates the ideal environment for my research to achieve meaningful societal impact while building on my existing industry collaborations. These existing collaborations include Google DeepMind, Microsoft, OpenAI, Adobe Research, BBC Verify, and the Lawrence Livermore National Laboratories (home to the world's largest GPU supercomputer).

I am particularly drawn to your emphasis on interdisciplinary collaborations and your longstanding tradition in AI research. I look forward to discussing how my experience in AI security, multi-agent learning, and teaching could contribute to Warwick's vision and ambitious growth plans. Thank you for considering my application.

Please see below an overview of how I fulfill the job criteria.

Sincerely,

C. Schroeder de Witt

Dr. Christian Schroeder de Witt
DPhil UK Intelligence Community Research Fellow
Department of Engineering Science, University of Oxford

E1 An ability to teach and supervise undergraduates and postgraduate students in computing science.

I have demonstrated strong teaching ability at both undergraduate and postgraduate levels in computing science. As Stipendiary Lecturer at Oxford, I've taught Machine Learning to undergraduates across four colleges, receiving "extremely positive feedback." I've designed and delivered the world's first doctoral course on Cooperative AI, taught AI to visiting undergraduates, and served as a teaching assistant for undergraduate courses. My supervision experience includes co-supervising three PhD students, five MSc students (all achieving distinction in their theses), and two research assistants. This track record shows my capability to effectively teach and mentor students at all academic levels in computing science.

E2 Good communication skills (oral and written).

I possess excellent communication skills demonstrated through my extensive publication record (31+ peer-reviewed papers), invited talks at prestigious venues, and successful funding acquisition (£1M+ as PI/Co-PI). My work has been featured in prominent media outlets including Quanta Magazine and Scientific American. My teaching experience shows effective oral communication with diverse student groups, from undergraduates to PhD students. As a BBC expert contributor and panel speaker at high-profile events like the Munich Security Conference, I regularly communicate complex AI concepts to non-specialist audiences. My clear, persuasive writing is evident in my well-structured cover letter and CV, as well as my contributions to successful grant applications.

E3 Good interpersonal skills.

My interpersonal skills are evidenced through successful collaborations with researchers across disciplines and institutions, including partnerships with Google DeepMind, Microsoft, OpenAI, and BBC Verify. I've demonstrated leadership in organizing workshops and mentorship programs, particularly the AI for Agent-Based Modeling Community Mentorship Program. My commitment to diversity and inclusion is shown through mentoring students from underrepresented backgrounds via the DeepLearning Indaba program. My teaching roles required strong interpersonal communication with students from varied backgrounds, earning positive feedback. My service on selection committees, grant reviews, and conference program committees further demonstrates my ability to work effectively with others in academic and professional contexts.

E4 An ability to work independently and as part of a team on research and teaching programmes in computer science.

I've demonstrated exceptional independent work through my UK Intelligence Community Research Fellowship, where I lead research on AI-generated media detection, and as PI on multiple grants where I've independently developed the emerging field of Multi-Agent Security. My independent course design of the "Frontiers of Cooperative AI" showcases self-directed teaching innovation. Equally strong is my collaborative work, evidenced by co-authored papers with researchers from Google DeepMind and CMU that have garnered over 6,100 citations. I've

contributed to team grants worth £2M and participated in cross-institutional teaching teams. This balance is further shown in my conference roles, where I both independently chair areas and work within program committees, demonstrating versatility in both solo and team-based research and teaching contexts.

E5 Possession of a PhD or equivalent qualification in computer science or related subject, by the post start date.

I hold a Doctor of Philosophy (DPhil) in Engineering Science from the University of Oxford (2017-2021), specializing in Artificial Intelligence with a focus on "Communication and Coordination in Deep Multi-Agent Reinforcement Learning." My doctoral research was supervised by Prof. Philip H.S. Torr and Prof. Shimon Whiteson. This qualification is complemented by my Master of Science in Computer Science from Oxford (2013) and Master of Physics (2012). My interdisciplinary background across AI, computer science, and complex systems provides a strong foundation for the position, exceeding the essential qualification requirement.

E6 The ability to initiate, develop and deliver high-quality research and to publish in leading peer-reviewed journals or conference proceedings.

I have consistently demonstrated exceptional ability to initiate, develop and deliver high-quality research, evidenced by my publication record of 31+ peer-reviewed papers in top-tier venues including NeurIPS, ICML, ICLR, and AAMAS. My research has generated over 6,100 citations (h-index 19), with breakthrough work on secure steganography featured in Quanta Magazine as "solving a 25+ year open problem." I've initiated new research directions that established the field of Multi-Agent Security, secured £1M+ in competitive grants as PI/Co-PI, and received prestigious recognition including an ICLR Spotlight Talk (2% acceptance rate) and being named among Quanta Magazine's "Biggest Discovery of the Year in Computer Science" (2023). My patent on secure steganography further demonstrates my ability to develop novel research with practical applications.

E7 An ability or potential to generate external funding (grants, contracts etc.) to support research.

I've demonstrated exceptional ability to generate external funding, having secured over £1M ($1.3M) as PI/Co-PI from diverse sources including the Royal Academy of Engineering, Foresight Institute, Schmidt Futures, OpenAI, Future of Life Institute, Armasuisse, EPSRC, and Microsoft. My recent grants include £250,000 as PI for a UK Intelligence Community Postdoctoral Research Fellowship (2024), $124,000 as PI from the Foresight Institute (2025), and $327,000 as Co-PI from OpenAI (2024) with a competitive acceptance rate below 2%. Additionally, I've contributed to securing a further £2M ($2.7M) in grants, including substantial involvement in a successful €2.3M European Research Council Starting Grant. My track record shows consistent success across international funding bodies, industry partnerships, and prestigious academic grants, with a steadily increasing funding trajectory.

E8 Strong record of initiating, developing and delivering high-quality research and publishing in leading peer-reviewed journals or conference proceedings, or equivalent achievements in industry (Associate Professor).

I have established an exceptional research record, publishing 31+ peer-reviewed papers in the most competitive AI venues including NeurIPS, ICML, and ICLR (acceptance rates typically 15-25%). My pioneering work on secure steganography was recognized by Quanta Magazine as "solving a 25+ year open problem" and named "Biggest Discovery of the Year in Computer Science" (2023). My research has garnered over 6,100 citations with an h-index of 19, demonstrating substantial field impact. I've initiated the new field of Multi-Agent Security through several groundbreaking publications on undetectable threats in AI systems, with one receiving a prestigious ICLR Spotlight Talk (2% acceptance rate). My research innovations have led to a patent and attracted substantial funding (£1M+ as PI/Co-PI), confirming my ability to identify, develop, and execute research with significant academic and practical impact at a level appropriate for an Associate Professor.

E9 An ability or potential to generate external funding (grants, contracts etc.) to support research, or equivalent achievements in industry (Associate Professor).

I've demonstrated exceptional success in securing external funding at a level appropriate for an Associate Professor position, having acquired over £1M ($1.3M) as PI/Co-PI from diverse prestigious sources. My portfolio includes a £250,000 UK Intelligence Community Fellowship (RAEng), $124,000 Foresight Institute Grant, $133,000 Schmidt Futures Grant, $327,000 OpenAI Superalignment Fast Grant (acceptance rate <2%), and grants from Future of Life Institute, Armasuisse, EPSRC, and Microsoft. Beyond my direct funding, I've contributed substantially to securing an additional £2M in grants, including a €2.3M European Research Council Starting Grant where I served as co-editor and substantial contributor. This consistent success across government, industry, and foundation funding sources demonstrates my proven ability to attract significant research support and build successful funding relationships.

Application ID: 164507, Jiancheng Yang: Assistant and Associate Professor - Computer Science (100893-0325)

Personal Information

We take privacy seriously and will only use your personal information to administer your application. For more information please see our Data Protection Policies (https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection).

Title	Dr.
Preferred pronouns	He/him/his
Given Name(s)	Jiancheng
Preferred Name	JC
Family Name	Yang
Email	jiancheng.yang@epfl.ch
Preferred Phone No	+41762691931

Additional Information

Are you currently employed by University of Warwick?

No

Will you now or in the future require a visa to obtain/continue to hold the right to work legally in the UK?

Yes

Reasonable Adjustments

We make intentional efforts to employ and retain people with disabilities. The following question will aid us to assist you should you need it during the application process.

Do you have any medical condition, special educational needs or disability that means that you may require reasonable adjustments made for you during either the online assessment, interview or assessment centre stages of our selection process?

No

Source

Where did you find the advert for this vacancy?

Linked In

What made you apply for this vacancy at the University of Warwick? Please select all that apply.

Actively looking to move into Higher Education, Personal recommendation

Is there anything further that prompted you to apply?

No

References

Reference 1

Title	Professor
First Name	Pascal
Last Name	Fua
Email	send.Fua.3D5B5074D8@interfoliodossier.com
Reference Type	Academic
Relationship	Postdoc Advisor

Reference 2

Title	Professor
First Name	Hanspeter
Last Name	Pfister

Email	send.Pfister.A509F85BA2@interfoliodossier.com
Reference Type	Academic
Relationship	Visiting Host Advisor

Reference 3

Title	Professor
First Name	Mingguang
Last Name	He
Email	send.He.74FE18DE83@interfoliodossier.com
Reference Type	Academic
Relationship	Collaborator

Jiancheng (JC) Yang

jiancheng.yang@epfl.ch | +41 762691931 | EPFL BC 305, Station 14, 1015 Lausanne, Switzerland

Bio. Jiancheng (JC) Yang is a postdoctoral researcher at the Swiss Federal Institute of Technology Lausanne (EPFL), working with Prof. Pascal Fua. He earned Bachelor's and PhD degrees from Shanghai Jiao Tong University and was a visiting researcher at Harvard University and EPFL. His research emphasizes spatial intelligence for healthcare, leveraging geometric, generative, and multimodal deep learning. He has published over 50 papers in leading journals and conferences, including Cancer Research, eBioMedicine, TMI, MedIA, CVPR, MICCAI, NeurIPS, and ICLR, and serves as Area Chair for MICCAI 2024/2025, MIDL 2025, and Editorial Board Member of npj Digital Medicine.

His contributions have been recognized, including Forbes 30 Under 30, Top 2% Scientists Worldwide (Stanford List), and the WAIC Yunfan Award. He is also the author of MedMNIST. Moreover, he co-founded and served as CTO of a medical AI startup in Shanghai, translating his research into real-world applications and securing millions of dollars in funding.

His notable work includes MedMNIST (250K+ Downloads & 1K+ Citations), Point Attention Transformer, RibFrac.

Research Interests. AI for health; 3D vision; medical image analysis; spatial intelligence; deep geometric learning; (constrained) generative models; multimodal; real-world clinical translation.

Academic Experience

2024-	Postdoctoral Researcher (Advisor: Prof. Pascal Fua)	EPFL, Switzerland
2021-2024	Visiting Fellow / Scientist (Advisor: Prof. Pascal Fua)	EPFL, Switzerland
2020-2021	Visiting Fellow (Advisor: Prof. Hanspeter Pfister)	Harvard University, US

Industry Experience

2020-2022 Co-Founder & CTO (Part-Time) Dianei Technology (Startup), China

Starting in Apr 2020, I transitioned to a part-time role while pursuing academic research as a PhD student at Shanghai Jiao Tong University and visiting researcher at Harvard and EPFL. In pursuit of further growth in my scientific career, I returned to full-time academia after Dec 2022.

2018-2020 Co-Founder & CTO (Full-Time) Dianei Technology (Startup), China

I co-founded Dianei Technology, a health AI startup in Shanghai that successfully raised over ¥50M (~$7M) in total. From Jan 2018 to Mar 2020, I was fully dedicated to the startup as CTO, leading the development of a multi-omics AI solution for lung cancer screening, diagnosis, and treatment. The products obtained 3 NMPA (China's "FDA") registration certificates and are employed in 100+ medical institutions.

2017 Research Intern Tencent, China

Education

2024	Ph.D., Information Engineering (Advisor: Prof. Bingbing Ni)	Shanghai Jiao Tong University, China
2018	M.E., Automation	Shanghai Jiao Tong University, China
2016	Engineer's degree (double master's degree), Information System	Institut Mines-Télécom, France
2015	B.E., Automation	Shanghai Jiao Tong University, China

Honors & Awards

2024 Top 2% Scientists Worldwide, Stanford University

2023 Forbes 30 Under 30 Asia (Healthcare & Science)

2022 Person of the Year Nomination Award, SJTU (10+5 from All Fields, 上海交通大学学生年度人物提名奖)

2022 YunFan Award Rising Stars, WAIC (15 AI Scientists under Age 30, 世界人工智能大会云帆奖明日之星)

2020 BMVC 2020 Outstanding Reviewer Award

Fellowships & Grants

2022 Xu Zhang Scholarship [Top 2 out of all PhDs in the department] (30K CNY, 张煦院士奖学金)

2021 DAAD AInet Fellowship from German Academic Exchange Service (~3K Euros)

2019-2021 National PhD Scholarship (Top 2%, 30K CNY, 3 times)

Contributed to Grant Writing

2021 Adversarial Machine Learning Theories and Methods (NSFC, 2.53M CNY)

2021 Deep Multimodal Prediction of PD-1/PD-L1 Immunotherapy Efficacy (NSFC, 560K CNY)

2020 Deep Learning for Personalized Pulmonary Nodule Follow-up (NSFC, 570K CNY)

Competitions

2019 2 / 731 (Team Mentor), MICCAI 2019 DigestPath Challenge

2019 3 / 2520 (Team Mentor), IJCAI 2019 Adversarial AI Challenge

2017 3 / 2887 (Team Leader), Alibaba Tianchi Medical AI Competition

2010 First Prize in the National High School Mathematics Competition

2010 Second Prize in the National High School Physics Competition

Publications

As of Apr 2025, I have authored 58 publications, with 27 as first/co-first author and 12 as corresponding/last author, accumulating 4,100+ citations and achieving an h-index of 23. Of these, 11 papers have received 100+ citations, 8 of which I am the first/co-first author. My research spans fundamental AI methodologies, medical AI methodologies, the development of datasets and challenges, as well as clinical translation research from real-world data.

Up-to-date list and preprints available on Google Scholar.
*: equal contribution. #: corresponding authorship.

1. Kangxian Xie, **Jiancheng Yang#**, Donglai Wei, Ziqiao Weng, Pascal Fua. "Efficient anatomical labeling of pulmonary tree structures via deep point-graph representation-based implicit fields". *MedIA - Medical Image Analysis, 2025.* [URL]
2. **Jiancheng Yang**, Rui Shi, Liang Jin, Xiaoyang Huang, Kaiming Kuang, Donglai Wei, Shixuan Gu, ..., Hanspeter Pfister, Ming Li, Bingbing Ni. "Deep Rib Fracture Instance Segmentation and Classification from CT on the RibFrac Challenge". *TMI - IEEE Transactions on Medical Imaging, 2025. Accepted.* [Preprint]
3. Hantao Zhang, Weidong Guo, Shouhong Wan, Bingbing Zou, Wanqin Wang, Chenyang Qiu, Kaige Liu, Peiquan Jin, **Jiancheng Yang**. "Tuning Vision Foundation Models for Rectal Cancer Segmentation from CT Scans: Development and Validation of U-SAM". *Communications Medicine (Nature Portfolio), 2025. Accepted.* [Preprint]
4. Hantao Zhang, Yuhe Liu, **Jiancheng Yang#**, Shouhong Wan, Xinyuan Wang, Wei Peng, Pascal Fua. "LeFusion: Controllable Pathology Synthesis via Lesion-Focused Diffusion Models". *ICLR - International Conference on Learning Representations, 2025.* ***Spotlight.*** [URL]
5. **Jiancheng Yang#**. "Multi-task learning for medical foundation models". *Nature Computational Science, 2024.* ***News & Views.*** [URL]
6. Liang Jin, Shixuan Gu, Donglai Wei, Jason Ken Adhinarta, Kaiming Kuang, Yongjie Jessica Zhang, Hanspeter Pfister, Bingbing Ni#, **Jiancheng Yang#**, Ming Li#. "RibSeg v2: A Large-scale Benchmark for Rib Labeling and Anatomical Centerline Extraction". *TMI - IEEE Transactions on Medical Imaging, 2024.* [URL]
7. Yanwei Zhang, Beibei Sun, Yinghong Yu, Jun Lu, Yuqing Lou, Fangfei Qian, Tianxiang Chen, Li Zhang, **Jiancheng Yang#**, Hua Zhong#, Ligang Wu#, Baohui Han#. "Multimodal Fusion of Liquid Biopsy and CT Enhances Differential Diagnosis of Early-stage Lung Adenocarcinoma". *npj Precision Oncology (Nature Portfolio), 2024.* [URL]

8. **Jiancheng Yang#**, Ekaterina Sedykh, Jason Adhinarta, Hieu Le, Pascal Fua. "Generating Anatomically Accurate Heart Structures via Neural Implicit Fields". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2024.* [URL]
9. Ziqiao Weng, **Jiancheng Yang#**, Dongnan Liu, Weidong Cai. "Efficient Repairing of Disconnected Pulmonary Tree Structures via Point-based Implicit Fields". *MIDL - Medical Imaging with Deep Learning, 2024.* [URL]
10. Danli Shi, Shuang He, **Jiancheng Yang**, Yingfeng Zheng, Mingguang He. "One-shot retinal artery and vein segmentation via cross-modality pretraining". *Ophthalmology Science, 2024.* [URL]
11. Jianning Li, Zongwei Zhou, **Jiancheng Yang**, Antonio Pepe, Christina Gsaxner, Gijs Luijten, ..., Pascal Fua, Alan L. Yuille, Jens Kleesiek, Jan Egger (100+ Authors). "MedShapeNet -- A Large-Scale Dataset of 3D Medical Shapes for Computer Vision". *Biomedical Engineering / Biomedizinische Technik, 2024.* [URL]
12. Weiyi Zhang, Siyu Huang, **Jiancheng Yang**, Ruoyu Chen, Zongyuan Ge, Yingfeng Zheng, Danli Shi, Mingguang He. "Fundus2Video: Cross-Modal Angiography Video Generation from Static Fundus Photography with Clinical Knowledge Guidance". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2024.* [URL]
13. Hongyu Ge, Jiahao Shi, Shuohong Wang, Donglai Wei, **Jiancheng Yang**, Ao Cheng, Richard Schalek, Jun Guo, Jeff Lichtman, Lirong Wang, Ruobing Zhang. "Two-stage error detection to improve electron microscopy image mosaicking". *Computers in Biology and Medicine, 2024.* [URL]
14. **Jiancheng Yang**, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, Bingbing Ni. "MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification". *Scientific Data (Nature Portfolio), 2023.* ***ESI Highly Cited Paper & Hot Paper*****.** [URL]
15. **Jiancheng Yang#**, Hongwei Bran Li, Donglai Wei. "The Impact of ChatGPT and LLMs on Medical Imaging Stakeholders: Perspectives and Use Cases". *Meta-Radiology, 2023.* ***Invited Article*****.** [URL]
16. Rui Xu, Zhi Liu, Yong Luo, Han Hu, Li Shen, Bo Du, Kaiming Kuang, **Jiancheng Yang**. "SGDA: Towards 3D Universal Pulmonary Nodule Detection via Slice Grouped Domain Attention". *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023.* [URL]
17. Ying Zhu, Li-Li Chen, Ying-Wei Luo, Li Zhang, Hui-Yun Ma, Hao-Shuai Yang, Bao-Cong Liu, Lu-Jie Li, Wen-Biao Zhang, Xiang-Min Li, Chuan-Miao Xie, **Jian-Cheng Yang#**, De-ling Wang#, Qiong Li#. "Prognostic impact of deep learning-based quantification in clinical stage 0-I lung adenocarcinoma". *European Radiology, 2023.* [URL]
18. Ziqiao Weng, **Jiancheng Yang#**, Dongnan Liu, Weidong Cai. "Topology Repairing of Disconnected Pulmonary Airways and Vessels: Baselines and a Dataset". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2023.* ***Early Accepted*****.** [URL]
19. Zhihao Li*, **Jiancheng Yang***, Yongchao Xu, Li Zhang, Wenhui Dong, Bo Du. "Scale-aware Test-time Click Adaptation for Pulmonary Nodule and Mass Segmentation". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2023.* [URL]
20. Minghui Zhang, Yangqian Wu, Hanxiao Zhang, Yulei Qin, Hao Zheng, ..., **Jiancheng Yang**, ..., Guang-Zhong Yang, Yun Gu. "Multi-site, Multi-domain Airway Tree Modeling". *MedIA - Medical Image Analysis, 2023.* [URL]
21. Jiajing Sun, Li Zhang, Bingyu Hu, Zhicheng Du, William C. Cho, Pasan Witharana, Hua Sun, Dehua Ma, Minhua Ye, Jiajun Chen, Xiaozhuang Wang, **Jiancheng Yang**, Chengchu Zhu, Jianfei Shen. "Deep learning-based solid component measuring enabled interpretable prediction of tumor invasiveness for lung adenocarcinoma". *Lung Cancer, 2023.* [URL]
22. Yanwei Zhang, Fangfei Qian, Jiajun Teng, ..., **Jiancheng Yang**, Jiayuan Sun, Hua Zhong, Baohui Han. "China Lung Cancer Screening (CLUS) version 2.0 with new Techniques Implemented: Artificial Intelligence, Circulating Molecular Biomarkers and Autofluorescence Bronchoscopy". *Lung Cancer, 2023.* [URL]
23. Bin Chen, Ziyi Liu, Jinjuan Lu, Zhihao Li, Kaiming Kuang, **Jiancheng Yang**, Zengmao Wang, Yingli Sun, Bo Du, Lin Qi, Ming Li. "Deep learning parametric response mapping from inspiratory chest CT scans: a new approach for small airway disease screening". *Respiratory Research, 2023.* [URL]
24. Chinmay Prabhakar, Hongwei Bran Li, **Jiancheng Yang**, Suprosana Shit, Benedikt Wiestler, Bjoern Menze. "ViT-AE++: Improving Vision Transformer Autoencoder for Self-supervised Medical Image Representations". *MIDL - Medical Imaging with Deep Learning, 2023.* [URL]

25. Won-Dong Jang, Stanislav Lukyanenko, Donglai Wei, **Jiancheng Yang**, Brian Leahy, Helen Yang, Dalit Ben-Yosef, Daniel Needleman, Hanspeter Pfister. "Multi-task Curriculum Learning for Partially Labeled Data". *ISBI - IEEE International Symposium on Biomedical Imaging, 2023.* ***Oral.*** [URL]
26. Jiajun Deng*, **Jiancheng Yang***, Likun Hou*, Junqi Wu, Yi He, Mengmeng Zhao, Bingbing Ni, Donglai Wei, Hanspeter Pfister, Caicun Zhou, Tao Jiang, Yunlang She, Chunyan Wu, Chang Chen. "Genopathomic profiling identifies signatures for immunotherapy response of lung cancer via confounder-aware representation learning". *iScience (Cell Press), 2022.* [URL]
27. **Jiancheng Yang**, Udaranga Wickramasinghe, Bingbing Ni, Pascal Fua. "ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition, 2022.* [URL]
28. **Jiancheng Yang***, Rui Shi*, Udaranga Wickramasinghe, Qikui Zhu, Bingbing Ni, Pascal Fua. "Neural Annotation Refinement: Development of a New 3D Dataset for Adrenal Gland Analysis". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2022.* [URL]
29. Kaiming Kuang, Li Zhang, Jingyu Li, Hongwei Li, Bo Du, Jiajun Chen, **Jiancheng Yang#**. "What Makes for Automatic Reconstruction of Pulmonary Segments". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2022.* [URL]
30. Rui Xu, Yong Luo, Bo Du, Kaiming Kuang, **Jiancheng Yang**. "LSSANet: A Long Short Slice-Aware Network for Pulmonary Nodule Detection". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2022.* ***Early Accepted.*** [URL]
31. Dingyi Rong*, **Jiancheng Yang***, Bilian Ke, Bingbing Ni. "Differentiable Projection from Optical Coherence Tomography B-Scan without Retinal Layer Segmentation Supervision". *ISBI - IEEE International Symposium on Biomedical Imaging, 2022.* ***Oral.*** [URL]
32. Xiaoyang Huang, **Jiancheng Yang**, Yanjun Wang, Ziyu Chen, Linguo Li, Teng Li, Bingbing Ni, Wenjun Zhang. "Representation-Agnostic Shape Fields". *ICLR - International Conference on Learning Representations, 2022.* [URL]
33. Wei Zhao, Weidao Chen, Ge Li, Du Lei, **Jiancheng Yang**, Yanjing Chen, Yingjia Jiang, Jiangfen Wu, Bingbing Ni, Yeqi Sun, Shaokang Wang, Yingli Sun, Ming Li, Jun Liu. "GMILT: A Novel Transformer Network that Can Noninvasively Predict EGFR Mutation Status". *TNNLS - IEEE Transactions on Neural Networks and Learning Systems, 2022.* [URL]
34. Qian Da, Xiaodi Huang, Zhongyu Li, ,,,, **Jiancheng Yang**, ..., Dimitris N. Metaxas, Hongsheng Li, Chaofu Wang, Shaoting Zhang. "DigestPath: A benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system". *MedIA - Medical Image Analysis, 2022.* [URL]
35. Yu Ding, Jingyu Zhang, Weitao Zhuang, Zhen Gao, Kaiming Kuang, ..., **Jiancheng Yang**, Qiuling Shi, Guibin Qiao. "Improving the efficiency of identifying malignant pulmonary nodules before surgery via a combination of artificial intelligence CT image recognition and serum autoantibodies". *European Radiology, 2022.* [URL]
36. Guangyu Tao, Li Zhu, Qunhui Chen, Lekang Yin, Yamin Li, **Jiancheng Yang**, Bingbing Ni, Zheng Zhang, Chi Wan Koo, Pradnya D. Patil, Yinan Chen, Hong Yu, Yi Xu, Xiaodan Ye. "Prediction of future imagery of lung nodule as growth modeling with follow-up computed tomography scans using deep learning: a retrospective cohort study". *Translational Lung Cancer Research, 2022.* [URL]
37. Udaranga Wickramasinghe, Patrick Jensen, Mian Shah, **Jiancheng Yang**, Pascal Fua. "Weakly Supervised Volumetric Image Segmentation with Deformed Templates". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2022.* [URL]
38. Qikui Zhu, Yanqing Wang, Fei Liao, **Jiancheng Yang**, Lei Yin, Shuo Li. "SelfMix: A Self-adaptive Data Augmentation Method for Lesion Segmentation". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2022.* [URL]
39. **Jiancheng Yang***, Xiaoyang Huang*, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, Bingbing Ni. "Reinventing 2D Convolutions for 3D Images". *JBHI - IEEE Journal of Biomedical and Health Informatics, 2021.* [URL]
40. Yi Yang*, **Jiancheng Yang***, Lan Shen*, Jiajun Chen, Liliang Xia, Bingbing Ni, Liang Ge, Ying Wang, Shun Lu. "A Multi-omics-based Serial Deep Learning Approach to Predict Clinical Outcomes of Single-agent Anti-PD-1/PD-L1

Immunotherapy in Advanced Stage Non-small-cell Lung Cancer". *American Journal of Translational Research, 2021.* [URL]

41. **Jiancheng Yang***, Yi He*, Kaiming Kuang, Zudi Lin, Hanspeter Pfister, Bingbing Ni. "Asymmetric 3D Context Fusion for Universal Lesion Detection". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2021.* ***Previous SOTA on DeepLesion.***
42. **Jiancheng Yang***, Shixuan Gu*, Donglai Wei, Hanspeter Pfister, Bingbing Ni. "RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2021.* [URL]
43. **Jiancheng Yang**, Rui Shi, Bingbing Ni. "MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis". *ISBI - IEEE International Symposium on Biomedical Imaging, 2021.* [URL]
44. Ye Chen, Jinxian Liu, Bingbing Ni, Hang Wang, **Jiancheng Yang**, Ning Liu, Teng Li, Qi Tian. "Shape Self-Correction for Unsupervised Point Cloud Understanding". *ICCV - IEEE International Conference on Computer Vision, 2021.* [URL]
45. Linguo Li, Minsi Wang, Bingbing Ni, Hang Wang, **Jiancheng Yang**, Wenjun Zhang. "3D Human Action Representation Learning via Cross-View Consistency Pursuit". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition, 2021.* [URL]
46. Liang Jin*, **Jiancheng Yang***, Kaiming Kuang, Bingbing Ni, Yiyi Gao, Yingli Sun, Pan Gao, Weiling Ma, Mingyu Tan, Hui Kang, Jiajun Chen, Ming Li. "Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet". *eBioMedicine (The Lancet Discovery Science), 2020.* [URL]
47. **Jiancheng Yang**, Bingbing Ni. "医学 3D 计算机视觉：研究进展和挑战". *中国图象图形学报 (Journal of Image and Graphics), 2020.* ("Advances and Challenges in Medical 3D Computer Vision"; in Chinese) [URL]
48. **Jiancheng Yang***, Yangzhou Jiang*, Xiaoyang Huang, Bingbing Ni, Chenglong Zhao. "Learning Black-Box Attackers with Transferable Priors and Query Feedback". *NeurIPS - Neural Information Processing Systems, 2020.* [URL]
49. **Jiancheng Yang***, Yi He*, Xiaoyang Huang, Jingwei Xu, Xiaodan Ye, Guangyu Tao, Bingbing Ni. "AlignShift: Bridging the Gap of Imaging Thickness in 3D Anisotropic Volumes". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2020.* ***Oral; Early Accepted.*** [URL]
50. **Jiancheng Yang***, Mingze Gao*, Kaiming Kuang, Bingbing Ni, Yunlang She, Dong Xie, Chang Chen. "Hierarchical Classification of Pulmonary Lesions: A Large-Scale Radio-Pathomics Study". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2020.* ***Oral; Early Accepted.*** [URL]
51. **Jiancheng Yang***, Jiajun Chen*, Kaiming Kuang, Tiancheng Lin, Junjun He, Bingbing Ni. "MIA-Prognosis: A Deep Learning Framework to Predict Therapy Response". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2020.* ***Student Travel Award; Oral; Early Accepted.*** [URL]
52. Yamin Li*, **Jiancheng Yang***, Yi Xu, Jingwei Xu, Xiaodan Ye, Guangyu Tao, Xueqian Xie, Guixue Liu. "Learning Tumor Growth via Follow-Up Volume Prediction for Lung Nodules". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2020.* ***Oral.*** [URL]
53. **Jiancheng Yang**, Haoran Deng, Xiaoyang Huang, Bingbing Ni, Yi Xu. "Relational Learning between Multiple Pulmonary Nodules via Deep Set Attention Transformers". *ISBI - IEEE International Symposium on Biomedical Imaging, 2020.* [URL]
54. Tiancheng Lin, Yuanfan Guo, Canqian Yang, **Jiancheng Yang**, Yi Xu. "Decoupled Gradient Harmonized Detector for Partial Annotation: Application to Signet Ring Cell Detection". *Neurocomputing, 2020.* (1st runner up solution for MICCAI DigestPath 2019 challenge detection track) [URL]
55. Jingwei Xu, Zhenbo Yu, Bingbing Ni, **Jiancheng Yang**, Xiaokang Yang, Wenjun Zhang. "Deep Kinematics Analysis for Monocular 3D Pose Estimation". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition, 2020.* [URL]
56. Wei Zhao*, **Jiancheng Yang***, Bingbing Ni, Dexi Bi, Yingli Sun, Mengdi Xu, Xiaoxia Zhu, Cheng Li, Liang Jin, Pan Gao, Peijun Wang, Yanqing Hua, Ming Li. "Toward Automatic Prediction of EGFR Mutation Status in Pulmonary Adenocarcinoma with 3D Deep Learning". *Cancer Medicine, 2019.* ***Cover Article.*** [URL]
57. **Jiancheng Yang**, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, Qi Tian. "Modeling Point Clouds with Self-Attention and Gumbel Subset Sampling". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition, 2019.* [URL]

58. **Jiancheng Yang***, Rongyao Fang*, Bingbing Ni, Yamin Li, Yi Xu, Linguo Li. "Probabilistic Radiomics: Ambiguous Diagnosis with Controllable Shape Analysis". *MICCAI - Medical Image Computing and Computer Assisted Intervention, 2019.* ***Early Accepted.*** [URL]
59. Jinxian Liu, Bingbing Ni, Caiyuan Li, **Jiancheng Yang**, Qi Tian. "Dynamic Points Agglomeration for Hierarchical Point Sets Learning". *ICCV - IEEE International Conference on Computer Vision, 2019.* [URL]
60. Wei Zhao*, **Jiancheng Yang***, Yingli Sun, Cheng Li, Weilan Wu, Liang Jin, Zhiming Yang, Bingbing Ni, Pan Gao, Peijun Wang, Yanqing Hua, Ming Li. "3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas". *Cancer Research, 2018.* [URL]

Working Papers

61. Kangxian Xie, Yufei Zhu, Kaiming Kuang, Li Zhang, Hongwei Bran Li, Mingchen Gao, **Jiancheng Yang#**. "Template-Guided Reconstruction of Pulmonary Segments with Neural Implicit Functions". *MedIA - Medical Image Analysis (Under Revision).*
62. Danli Shi, Weiyi Zhang, **Jiancheng Yang**, Siyu Huang, Xiaolan Chen, Mayinuer Yusufu, Kai Jin, Shan Lin, Shunming Liu, Qing Zhang, Mingguang He. "EyeCLIP: A Visual-Language Foundation Model for Multi-Modal Ophthalmic Image Analysis". *npj Digital Medicine (Under Revision).* [Preprint]
63. Weiyi Zhang, **Jiancheng Yang**, Ruoyu Chen, Siyu Huang, Pusheng Xu, Xiaolan Chen, Shanfu Lu, Hongyu Cao, Mingguang He, Danli Shi. "Fundus to Fluorescein Angiography Video Generation as a Retinal Generative Foundation Model". *Communications Medicine (Nature Portfolio) (Under Review).* [Preprint]
64. Shixuan Leslie Gu, Jason Ken Adhinarta, Mikhail Bessmeltsev, **Jiancheng Yang**, Yongjie Jessica Zhang, Wenjie Yin, Daniel Berger, Jeff Lichtman, Hanspeter Pfister, Donglai Wei. "Frenet-Serret Frame-based Decomposition for Part Segmentation of 3D Curvilinear Structures". *TMI - IEEE Transactions on Medical Imaging (Under Revision).* [Preprint]

Professional Services

Conference Service

- Area Chair, Medical Image Computing and Computer Assisted Intervention (MICCAI), 2025
- Area Chair, Medical Imaging with Deep Learning (MIDL), 2025
- Area Chair, Medical Image Computing and Computer Assisted Intervention (MICCAI), 2024
- Organizer, MICCAI MedShapeNet Tutorial, 2025
- Organizer, MICCAI MedShapeNet Tutorial, 2024
- Organizer, MICCAI MELA Challenge: Mediastinal Lesion Analysis, 2022
- Lead Organizer, MICCAI RibFrac Challenge: Rib Fracture Detection and Classification, 2020

Journal Service

- Editorial Board Member, npj Digital Medicine (Nature Portfolio)
- Guest Editor, Advances in Medical 3D Vision: Voxels and Beyond, Bioengineering-Basel

Conference Reviewer

- Medical Image Computing and Computer Assisted Intervention (MICCAI), 2023, 2022, 2021, 2020
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2025, 2024, 2023, 2022, 2021, 2020
- International Conference on Computer Vision (ICCV), 2023, 2021
- European Conference on Computer Vision (ECCV), 2022
- Advances in Neural Information Processing Systems (NeurIPS), 2023, 2022, 2021, 2020
- International Conference on Machine Learning (ICML), 2022, 2021
- International Conference on Learning Representations (ICLR), 2025, 2022
- AAAI Conference on Artificial Intelligence (AAAI), 2022
- IEEE International Symposium on Biomedical Imaging (ISBI), 2022, 2021
- British Machine Vision Conference (BMVC), 2020

Journal Reviewer

- Nature Communications
- Nature Computational Science
- npj Digital Medicine (Nature Portfolio)
- Cell Reports Medicine
- TPAMI - IEEE Transactions on Pattern Analysis and Machine Intelligence
- TMI - IEEE Transactions on Medical Imaging
- JBHI - IEEE Journal of Biomedical and Health Informatics
- TIP - IEEE Transactions on Image Processing
- TNNLS - IEEE Transactions on Neural Networks and Learning Systems
- TIFS - IEEE Transactions on Information Forensics and Security
- npj Precision Oncology (Nature Portfolio)
- Communications Medicine (Nature Portfolio)
- Scientific Data (Nature Portfolio)
- Scientific Reports (Nature Portfolio)
- Machine Learning

Membership

- 2020- IEEE Member
- 2019- MICCAI Member
- 2020 Sigma Xi Full Member

Volunteer

- As a personal interest, I'm the initiator and organizer of the HIT Webinar (hit-webinar.com), a Chinese-language webinar series focusing on healthcare, artificial intelligence, and cutting-edge technology. This fully open, non-profit initiative has hosted 100+ sessions since its launch in April 2022.
- I'm a mentor for the Fatima Fellowship, a non-profit providing free research opportunities to underrepresented students.

Research Mentoring

I am fortunate to (co-)mentor 20+ talented and motivated students, both in-person and remotely. Through regular mentorship, we have explored and completed many interesting projects together. Items in […] represent research outcomes.

Remote Mentoring

- Namrah Rehman (Master student from Pakistan, my mentee through the Fatima Fellowship)
- Pedro M. B. Rezende (Bachelor, Universidade de São Paulo, Brazil->PhD, ETH Zurich)
- Hantao Zhang (Master, USTC) [ICLR'25 Spotlight]
- Kangxian Xie (Master, Boston University->PhD, SUNY Buffalo) [MedIA'25]
- Ziqiao Weng (PhD, University of Sydney) [MICCAI'23 (Early Accepted)]
- Zhihao Li (PhD, Wuhan University) [MICCAI'23]

Mentoring @EPFL

- Jason Ken Adhinarta (Bachelor, Boston College->PhD, MIT EECS) [MICCAI'24] [TMI'23]
- Ekaterina Sedykh (Master Thesis Project, EPFL) [MICCAI'24]
  [Master Dissertation: Universal 3D Cardiac Segmentation Enhanced via Partial Label Usage and Style Transfer]
- Guanqun Liu (Master Thesis Project, EPFL)
  [Master Dissertation: JointAtlas - Deformable Template-Based Implicit Method for Joint Segmentation-Appearance Modeling in Cardiac Images]

- EPFL Semester Project Students: Berta Céspedes Sarrias (Master, Life Science), Marc Pitteloud (Master, Data Science), Arhan Bartu Ergüven (Exchange Bacheter), Takuya Ishii (Master, NeuroX), Antoine Munier (Master, Computer Science), Hanwen Zhang (Master, Data Science)

Mentoring @SJTU

- Xiaoyang Huang (Bachelor, SJTU->PhD, SJTU) [ICLR'22] [JBHI'21] [NeurIPS'20] [MICCAI'20 (Oral; Early Accepted)] [ISBI'20]
- Rui Shi (Bachelor, SJTU->Master, SJTU) [MICCAI'22] [Nature Scientific Data'23] [ISBI'21]
- Shixuan Gu (Bachelor, SJTU->Master, CMU->PhD, Harvard) [TMI'23] [MICCAI'21]
- Dingyi Rong (Master, SJTU) [ISBI'22 (Oral)]
- Yangzhou Jiang (Bachelor, SJTU->Master, SJTU) [NeurIPS'20] [3rd in IJCAI 2019 Adversarial AI Challenge]
- Yamin Li (Master, SJTU) [MICCAI'20 (Oral)] [MICCAI'19 (Early Accepted)]
- Tiancheng Lin (PhD, SJTU) [2nd in MICCAI20 DigestPath Challenge] [Neurocomputing'20] [MICCAI'20 (Student Travel Award; Oral; Early Accepted)]
- Haoran Deng (Bachelor, SJTU->Master, ETH Zurich) [ISBI'20]
- Rongyao Fang (Bachelor, SJTU->PhD, CUHK) [MICCAI'19 (Early Accepted)]
- Qiang Zhang (Bachelor, SJTU->PhD, Princeton) [CVPR'19]

Teaching

Teaching Assistant @SJTU

Lead Teaching Assistant, Machine Learning (EE369), Fall 2018, Fall 2019 (~100 students)

Lead Teaching Assistant, Machine Learning for the AI Elite Class (EE228), Spring 2019, Spring 2020 (~100 students)

- Technical guidanc, design of major course projects and assignments, exam preparation and grading.
- Delivered two 1.5-hour introductory lectures on deep learning.
- Designed and supervised course projects and tutorials:
    - Developed an educational agent API (2048-api) for deep learning projects using the 2048 game, allowing students to implement reinforcement learning agents.
    - Created a Kaggle In-Class competition for 3D medical data classification, enhancing students' practical AI skills.
    - Created a Jupyter-based interactive clustering tutorial covering KMeans(++), Gaussian Mixture Models (GMM), and Spectral Clustering, designed to encourage self-paced learning and hands-on experimentation.
- Received positive feedback from students for high-quality educational projects and hands-on learning opportunities.

Broader Impacts

Organizer & Speaker, MICCAI 2024 MedShapeNet Tutorial

- Iintroduced participants to classical and deep 3D shape analysis methods for biomedical applications.
- Delivered similar content as a guest lecture in CSCI 3397 Biomedical Image Analysis at Boston College

Founder & Lead Organizer, HIT Webinar (hit-webinar.com)

- HIT Webinar (hit-webinar.com) is a Chinese-language webinar series focusing on healthcare, artificial intelligence, and cutting-edge technology. This fully open, non-profit initiative has hosted 100+ sessions since its launch in April 2022.

Invited Talks

- [10/2024] MICCAI 2024 MedShapeNet Tutorial
- [07/2024] AI Elite Think Tank (Closed Meeting), World Artificial Intelligence Conference (WAIC)
- [04/2024] Guest Lecture, CSCI 3397 Biomedical Image Analysis, Boston College
- [01/2024] National Institute of Healthcare Data Science at Nanjing University
- [12/2023] Nanjing University
- [09/2023] University of Electronic Science and Technology of China

- [07/2023] Shanghai Jiao Tong University
- [02/2023] Rising Stars in AI Symposium, King Abdullah University of Science and Technology (KAUST)
- [02/2023] Fudan-Guanghua International Forum for Young Scholar, Fudan University
- [12/2022] Global Institute of Future Technology, Shanghai Jiao Tong University
- [11/2022] Suzhou Institute for Advanced Research, University of Science and Technology of China
- [11/2022] Guest Lecture, Shanghai Jiao Tong University School of Medicine
- [09/2022] WAIC YunFan Award Forum, Shanghai AI Laboratory
- [02/2021] Pie & AI by DeepLearning.AI
- [01/2021] AI Institute, Shanghai Jiao Tong University, Shanghai
- [10/2020] MICCAI 2020 RibFrac Challenge
- [04/2020] Lecture, EE228 Machine Learning for AI Class, Shanghai Jiao Tong University
- [10/2019] Lecture, EE369 Machine Learning, Shanghai Jiao Tong University
- [10/2017] Alibaba Group, Hangzhou

Open-Source Contributions

I actively contribute to open-source data and code, accumulating ~2,000 GitHub stars. Below are some highlighted projects:

MedMNIST: Standardized 2D (x12) and 3D (x6) datasets for biomedical image classification.

- ~708,000 2D images and ~10,000 3D images, with multiple size options: 28 (MNIST-like), 64, 128, and 224.
- 250,000+ downloads, 1,000+ GitHub stars, and ~1,000 citations (ESI Highly Cited Paper & Hot Paper).
- Utilized in numerous courses and hackathons, ranging from high school interest groups to university tutorials.

RibFrac & **RibSeg**: The first large-scale CT dataset for rib and rib fracture segmentation.

- RibFrac contains 660 CT scans with instance-level segmentation annotations for rib fractures. We hosted the MICCAI 2020 RibFrac Challenge, attracting 1,200+ users globally. It has also been used in various courses and hackathons.
- RibSeg further provides the associated segmentation, labeling, and centerline for ribs, serving as the foundation for the rib component in the popular TotalSegmentator.
- These projects have received ~150 GitHub stars and ~200 citations.

2D-to-3D Pretraining: One line of code to convert pretrained 2D models to 3D.

- We developed methods to convert 2D pretrained networks into 3D ones for 3D image analysis, with one line of code!
- ACSConv serves as a plug-and-play replacement for 3D convolutions, enabling an effortless upgrade from 2D to 3D models. While AlignShift focuses on low-slice 3D data, and maintained SOTA on the Universal Lesion Detection (DeepLesion) benchmark for over two years.
- These projects have received ~200 GitHub stars and ~150 citations.

Educational Projects: aimed at supporting courses and mentees, collectively received 200+ GitHub stars.

- 2048-api: Educational API for developing machine learning agents (such as imitation learning or reinforcement learning) to play the game 2048.
- clustering_tutorial: A tutorial on KMeans(++), GMM, and Spectral Clustering.
- Kickstart: A study roadmap for learners in machine learning, deep learning, and computer vision.

References

Prof. Pascal Fua, PhD
Full Professor, Computer Vision Laboratory, School of Computer and Communication Sciences
Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland
Email: pascal.fua@epfl.ch
For Reference Letter Submission Only (Interfolio Dossier Email): send.Fua.3D5B5074D8@interfoliodossier.com

Prof. Hanspeter Pfister, PhD
An Wang Professor of Computer Science
Affiliate Faculty Member, Center for Brain Science
Harvard University, Cambridge, MA, USA
Email: pfister@seas.harvard.edu
For Reference Letter Submission Only (Submitted by Assistant): pfister-admin@seas.harvard.edu
For Reference Letter Submission Only (Interfolio Dossier Email): send.Pfister.A509F85BA2@interfoliodossier.com

Prof. Mingguang He, MD, PhD
Chair Professor of Experimental Ophthalmology, School of Optometry
Director, Research Centre for SHARP Vision
The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
Email: mingguang.he@polyu.edu.hk
For Reference Letter Submission Only (Interfolio Dossier Email): send.He.74FE18DE83@interfoliodossier.com

Prof. Ming Li, MD, PhD
Director, Department of Radiology
Director, Zhang Guozhen Small Pulmonary Nodules Diagnosis and Treatment Center
Huadong Hospital Affiliated to Fudan University, Shanghai, China
Email: minli77@163.com
For Reference Letter Submission Only (Interfolio Dossier Email): send.Li.E3EE52144B@interfoliodossier.com

Research Statement

Generative and Multimodal Spatial Intelligence for Real-World Healthcare

AI has achieved transformative success in both general-purpose and healthcare-specific applications. However, many "general-purpose" AI methodologies struggle to generalize effectively within healthcare contexts. This challenge stems from the unique nature of medical data, particularly the prominence of high-dimensional dense 3D voxel data, such as CT and MRI scans, which are uncommon outside the healthcare domain. This complexity is further compounded by the heterogeneity, fragmentation, and limited availability of healthcare data, distinguishing it from other AI application domains and necessitating tailored solutions.

My research addresses these challenges from first principles, with a focus on *spatial intelligence for healthcare*. Built upon advances in deep geometric learning and generative AI, I develop voxel, point cloud, mesh and implicit representations alongside multi-modal and (constrained) generative models to facilitate the analysis and synthesis of 3D data. These methods are also translated to real-world challenges, supporting screening, diagnosis, prognosis, and intervention for thoracic and cardiac applications. Moreover, I actively support open science by leading impactful data initiatives.

1 Research Contributions

Motivations. 3D medical imaging, including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and electron microscopy (EM), serves as "first-class citizens" in modern biomedicine. These modalities also capture high-resolution 3D shape and topological representations essential for healthcare. However, such data pose critical challenges:

(C1) Data Sparsity. Unlike web-scale natural image datasets with millions or even billions of samples, medical datasets are often deemed "large-scale" with as few as 10k samples due to privacy constraints and costly expert labeling. *(C2) Computational Efficiency.* 3D medical images have inherently high dimensions. For example, a chest CT can have $400 \times 512 \times 512$ voxels, comparable to a $10,000 \times 10,000$ 2D image, posing challenges in long contexts, memory, and scalability. *(C3) Geometric Complexity.* Anatomical structures exhibit complex, non-rigid shapes with significant inter-patient variability. *(C4) Multi-X Heterogeneity.* Medical data spans diverse modalities with distinct characteristics and resolutions, while annotations also differ in granularity and labeling protocols, requiring multi-center, multi-modal, and multi-task ("multi-X" [1]) learning approaches.

Contributions. To address these unique challenges, my research advances novel methodologies organized into *(M1) voxels* and *(M2) beyond voxels* (point, mesh, and implicit representations), together with *(M3) multi-modal* and *(M4) generative* models. Beyond fundamental methods, my work spans real-world clinical translation, as well as the development of open datasets and benchmarks.
(M1) Voxels. Voxels are the most common representation in 3D medical imaging, but 3D networks lack large-scale pretraining like natural images. To address data sparsity (C1), I proposed ACSConv [2], an operator that converts existing 2D convolutional weights to process 3D volumes, enabling 2D-to-3D pretraining. This approach adapts any 2D CNN, including later models like ConvNeXt. Our open-source code provides a 1-line converter to transform pretrained 2D models into 3D. Following this, I also developed AlignShift [3] to address anisotropic spacing and A3D [4] for low-slice volumes, which achieved SOTA on the widely used DeepLesion for two consecutive years.
(M2) Beyond Voxels. I have explored geometric representations beyond voxels, including point, mesh, and implicit representations, to improve computational efficiency (C2) and geometric complexity (C3). A notable example is *RibSeg* [7, 8], where we segmented bone structures from CT scans by sparsifying dense CT volumes into point clouds with simple thresholding. This allowed point cloud-based deep learning, achieving more accurate segmentation and over $60\times$ faster inference ($< 1s$) compared to traditional voxel-based methods, thereby significantly improving computational efficiency (C2). To address geometric complexity (C3), I was also among the first to introduce implicit representations into medical imaging, which support continuous shape modeling, offering greater flexibility and precision than voxel grids. I used implicit shape priors for deformable templates [9], noisy annotation repair [10], and anatomically accurate heart modeling [11].

Additionally, I developed hybrid approaches that combine explicit and implicit models. For example, I proposed a method based on point, graph, and implicit representations [12] for anatomical labeling of pulmonary tree structures. This approach significantly improved segmentation accuracy and achieved a $10\times$ speed-up, effectively addressing both C2 and C3.

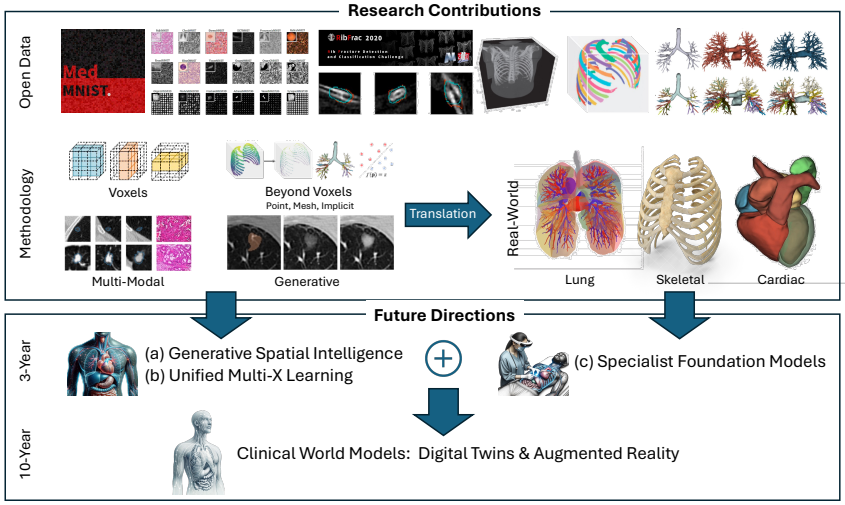


Figure 1: Overview of my research, highlighting key contributions and future directions.

I have also contributed to key advances in 3D vision, including one of the earliest works on transformers for point clouds [5], a representation-agnostic pretraining strategy for shape modeling [6], and developments in 3D pose estimation and self-supervised learning.

(M3) Multi-Modal. Medical data is inherently multi-X (C4), enabling connections across diverse domains. Multi-modal modeling makes it possible to link non-invasive with invasive [13], affordable with costly [14, 15], tissue with molecular profiles [16], images with text [17], and past with future [18, 19]. It also enables promptable models that flexibly respond to different input types. For example, I developed models to reconstruct real-time 4D cardiac motion from 3D, 2D, or even 1D signals [20], efficiently integrating multiple observations to support cardiac intervention.

(M4) Generative. Medical synthetic data effectively addresses data sparsity (C1) and bias (C4). Previously, I developed pipelines to synthesize and repair pulmonary tree structures with fractures [21]. The rise of generative AI has made many previously unmodelable challenges feasible; however, most generative models lack hard constraints, which are critical in medical applications. Recently, we developed a constrained generative model, lesion-focused diffusion (LeFusion), specifically for lesion synthesis [22]. LeFusion generates high-quality lesion image-mask pairs with guaranteed alignment, benefiting downstream tasks. This work was accepted at ICLR 2025 as Spotlight.

Real-World Translation. My research extends beyond methodology to real-world clinical applications in lung, skeletal, and cardiac domains. Leveraging my entrepreneurial experience during my PhD (Sec. 3), I conducted comprehensive work on lung cancer, covering screening (detection [23] and segmentation [24]), diagnosis [13, 25, 15], treatment (therapy response prediction [18] and surgical planning [26]), and follow-up [19], achieving performance surpassing human experts in many subtasks. In skeletal imaging, we introduced RibFrac [27] and RibSeg [7, 8], the first deep learning systems for rib fracture detection and segmentation, reaching radiologist-level performance. We also released the first large-scale datasets for these tasks and organized the MICCAI 2020 RibFrac Challenge. Recently, in cardiac imaging, I developed methods to generate anatomically accurate heart structures [11] and a promptable online 4D heart motion model [20], enabling precise simulation and planning for interventional surgeries.

Open Data and Benchmarks. I am a strong advocate for open science. In addition to RibFrac/RibSeg, I released MedMNIST [28], which contains 12 2D and 6 3D datasets spanning diverse modalities and scales – a direct reflection of multi-X heterogeneity (C4). It has become one of the most widely used datasets in medical AI research, with over 250k downloads, 1.1k GitHub stars, and recognition as an ESI Highly Cited and Hot Paper. Moreover, I have also open-sourced datasets from other projects [10, 21, 12], providing high-quality real-world geometric datasets for the community.

2 Future Directions

Mid-Term (3-Year) Agenda. (Areas with high potential for impactful extensions.)

(a) Generative Spatial Intelligence. Moving beyond task-specific shape models, I aim to develop a universal model for diverse 3D medical shapes, building on a large-scale medical shape dataset [29] I co-developed. This model will establish unified medical shape priors and support a promptable

generative model that integrates flexibly into downstream tasks. Key challenges to address include multi-part organs and curvilinear structures, and constrained generation based on diverse prompts.

In addition, due to inherent biases in real-world medical data, I am interested in synthetic data as a promising solution. Inspired by gaming and robotics, where realistic virtual environments are routinely generated, I propose developing a synthetic data engine for the medical domain. With growing high-quality datasets of image-shape pairs, these can be treated as digital assets to build the engine using generative AI, simulate clinical scenarios, and greatly benefit downstream tasks.
(b) Unified Multi-X Learning. Medical data exhibits multi-X heterogeneity (C4), spanning multi-modal, multi-center, and multi-task variations, where my prior work provides a solid foundation. A key perspective is to treat data as partial views of a unified entity, where multiple modalities and labels reflect the same underlying reality. By leveraging generative modeling, I aim to decouple and align these views into a unified representation, enabling effective learning across heterogeneous data sources. Moreover, this framework should address critical challenges in medical AI, including ambiguity [30], security [31], and causality [16], to advance both theory and clinical practice.
(c) Specialist Foundation Models. Current generalist foundation models aim to handle diverse medical modalities (e.g., processing pathology and CT together). However, I believe a specialist approach may be more effective like human doctors. Given the huge domain gaps in medical data and the sensitivity of deep learning to subtle differences, I aim to develop specialist foundation models tailored to clinical areas. Building on my experience in thoracic and cardiac imaging, I aim to create multi-functional models that are robust, generalizable, and clinically useful, supporting end-to-end clinical workflows. Similar to how SMPL supports human reconstruction and FreeSurfer facilitates neuroimaging, my goal is to develop these models as community tools to create lasting impact.

Long-Term (10-Year) Vision. (A more ambitious and rewarding research direction.)
Clinical World Models: Digital Twins & Augmented Reality. Building on previous methods, my long-term vision is to develop clinical world models (CWM) – patient-specific, dynamic representations that integrate multi-modal signals. Powered by physics-informed deep learning, these models will serve as digital twins for precision healthcare, supporting clinical decision-making, treatment planning, as well as enabling the discovery of new therapies. With the rapid advancement of augmented reality (AR), I also aim to integrate CWM with AR to extend its capabilities into the physical world, enhancing surgical navigation and medical education, particularly for percutaneous lung interventions, bronchoscopic procedures, and cardiac interventions. When combined, these models can offer real-time guidance during surgery and immersive training for clinicians.

3 Collaborations, Translation and Funding

I have established strong collaborations with multiple hospitals, including the Centre Hospitalier Universitaire Vaudois (CHUV) and leading hospitals in China. These collaborations have provided valuable access to clinical data, expert insights, and opportunities for the validation of AI-driven healthcare solutions. Beyond hospitals, I have cultivated connections with researchers from Europe, the US, Australia, and China. A notable achievement is my role as the initiator and organizer of the HIT Webinar (Healthcare, Intelligence, and Technology), a highly regarded MedTech webinar series for global Chinese that has facilitated the formation of a broad network of professional connections.

My research philosophy is centered on addressing real-world challenges with practical impact. This principle was exemplified through my experience in health AI entrepreneurship during my PhD journey, where I translated my research into practice: a multi-omics AI solution for lung cancer screening, diagnosis, and treatment. This effort resulted in three NMPA (China's "FDA") certificates and the deployment of the AI solution in over 100 medical institutions, directly contributing to healthcare delivery and patient outcomes. As a co-founder and the CTO, I led the R&D and strategy, raising over ¥50M RMB ($7M USD) in venture funding. This experience profoundly shaped my perspective, offering a unique lens to evaluate the societal and clinical relevance of research questions. It also taught me to identify and solve real fundamental problems from first principles.

Securing funding is a critical aspect. I have contributed to several successful grants in China and am currently involved in two Swiss proposals under review: an industrial grant and a collaborative SNSF grant. Besides, I have refined my grant-writing skills through EPFL training sessions. Moving forward, I aim to secure funding from both public agencies and industry partners. Additionally, I will also continue translating healthcare AI innovations into real-world impact through academic-industry partnerships, technology transfer, and potentially entrepreneurial ventures.

References

**: equal contribution. #: corresponding authorship.*

[1] Jiancheng Yang#. "Multi-task learning for medical foundation models". *Nature Computational Science*, 2024. ***News & Views***. [URL]

[2] Jiancheng Yang*, Xiaoyang Huang*, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, Bingbing Ni. "Reinventing 2D Convolutions for 3D Images". *JBHI - IEEE Journal of Biomedical and Health Informatics*, 2021. [URL]

[3] Jiancheng Yang*, Yi He*, Xiaoyang Huang, Jingwei Xu, Xiaodan Ye, Guangyu Tao, Bingbing Ni. "AlignShift: Bridging the Gap of Imaging Thickness in 3D Anisotropic Volumes". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2020. ***Oral; Early Accepted***. [URL]

[4] Jiancheng Yang*, Yi He*, Kaiming Kuang, Zudi Lin, Hanspeter Pfister, Bingbing Ni. "Asymmetric 3D Context Fusion for Universal Lesion Detection". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2021. [URL]

[5] Jiancheng Yang, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, Qi Tian. "Modeling Point Clouds with Self-Attention and Gumbel Subset Sampling". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition*, 2019. [URL]

[6] Xiaoyang Huang, Jiancheng Yang, Yanjun Wang, Ziyu Chen, Linguo Li, Teng Li, Bingbing Ni, Wenjun Zhang. "Representation-Agnostic Shape Fields". *ICLR - International Conference on Learning Representations*, 2022. [URL]

[7] Jiancheng Yang*, Shixuan Gu*, Donglai Wei, Hanspeter Pfister, Bingbing Ni. "RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2021. [URL]

[8] Liang Jin, Shixuan Gu, Donglai Wei, Jason Ken Adhinarta, Kaiming Kuang, Yongjie Jessica Zhang, Hanspeter Pfister, Bingbing Ni#, Jiancheng Yang#, Ming Li#. "RibSeg v2: A Large-scale Benchmark for Rib Labeling and Anatomical Centerline Extraction". *TMI - IEEE Transactions on Medical Imaging*, 2024. [URL]

[9] Jiancheng Yang, Udaranga Wickramasinghe, Bingbing Ni, Pascal Fua. "ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition*, 2022. [URL]

[10] Jiancheng Yang*, Rui Shi*, Udaranga Wickramasinghe, Qikui Zhu, Bingbing Ni, Pascal Fua. "Neural Annotation Refinement: Development of a New 3D Dataset for Adrenal Gland Analysis". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2022. [URL]

[11] Jiancheng Yang#, Ekaterina Sedykh, Jason Adhinarta, Hieu Le, Pascal Fua. "Generating Anatomically Accurate Heart Structures via Neural Implicit Fields". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2024. [URL]

[12] Kangxian Xie, Jiancheng Yang#, Donglai Wei, Ziqiao Weng, Pascal Fua. "Efficient anatomical labeling of pulmonary tree structures via deep point-graph representation-based implicit fields". *MedIA - Medical Image Analysis*, 2025. [URL]

[13] Wei Zhao*, Jiancheng Yang*, Yingli Sun, Cheng Li, Weilan Wu, Liang Jin, Zhiming Yang, Bingbing Ni, Pan Gao, Peijun Wang, Yanqing Hua, Ming Li. "3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas". *Cancer Research*, 2018. [URL]

[14] Weiyi Zhang, Jiancheng Yang, Ruoyu Chen, Siyu Huang, Pusheng Xu, Xiaolan Chen, Shanfu Lu, Hongyu Cao, Mingguang He, Danli Shi. "Fundus to Fluorescein Angiography Video Generation as a Retinal Generative Foundation Model". *npj Digital Medicine (Under Review)*. [Preprint]

[15] Yanwei Zhang, Beibei Sun, Yinghong Yu, Jun Lu, Yuqing Lou, Fangfei Qian, Tianxiang Chen, Li Zhang, Jiancheng Yang#, Hua Zhong#, Ligang Wu#, Baohui Han#. "Multimodal Fusion of Liquid Biopsy and CT Enhances Differential Diagnosis of Early-stage Lung Adenocarcinoma". *npj Precision Oncology (Nature Portfolio)*, 2024. [URL]

[16] Jiajun Deng*, Jiancheng Yang*, Likun Hou*, Junqi Wu, Yi He, Mengmeng Zhao, Bingbing Ni, Donglai Wei, Hanspeter Pfister, Caicun Zhou, Tao Jiang, Yunlang She, Chunyan Wu, Chang Chen. "Genopathomic profiling identifies signatures for immunotherapy response of lung cancer via confounder-aware representation learning". *iScience (Cell Press)*, 2022. [URL]

[17] Danli Shi, Weiyi Zhang, Jiancheng Yang, Siyu Huang, Xiaolan Chen, Mayinuer Yusufu, Kai Jin, Shan Lin, Shunming Liu, Qing Zhang, Mingguang He. "EyeCLIP: A Visual-Language Foundation Model for Multi-Modal Ophthalmic Image Analysis". *Lancet Digital Health (Under Review)*. [Preprint]

[18] Jiancheng Yang*, Jiajun Chen*, Kaiming Kuang, Tiancheng Lin, Junjun He, Bingbing Ni. "MIA-Prognosis: A Deep Learning Framework to Predict Therapy Response". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2020. ***Student Travel Award; Oral; Early Accepted***. [URL]

[19] Yamin Li*, Jiancheng Yang*, Yi Xu, Jingwei Xu, Xiaodan Ye, Guangyu Tao, Xueqian Xie, Guixue Liu. "Learning Tumor Growth via Follow-Up Volume Prediction for Lung Nodules". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2020. ***Oral***. [URL]

[20] Yihong Chen, Jiancheng Yang#, Deniz Sayin Mercadier, Hieu Le, Pascal Fua. "MedTet: An Online Motion Model for 4D Heart Reconstruction". *CVPR - IEEE Conference on Computer Vision and Pattern Recognition (Under Review)*. [Preprint]

[21] Ziqiao Weng, Jiancheng Yang#, Dongnan Liu, Weidong Cai. "Topology Repairing of Disconnected Pulmonary Airways and Vessels: Baselines and a Dataset". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2023. ***Early Accepted***. [URL]

[22] Hantao Zhang, Yuhe Liu, Jiancheng Yang#, Shouhong Wan, Xinyuan Wang, Wei Peng, Pascal Fua. "LeFusion: Controllable Pathology Synthesis via Lesion-Focused Diffusion Models". *ICLR - International Conference on Learning Representations*, 2025. ***Spotlight***. [Preprint]

[23] Rui Xu, Yong Luo, Bo Du, Kaiming Kuang, Jiancheng Yang. "LSSANet: A Long Short Slice-Aware Network for Pulmonary Nodule Detection". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2022. ***Early Accepted***. [URL]

[24] Zhihao Li*, Jiancheng Yang*, Yongchao Xu, Li Zhang, Wenhui Dong, Bo Du. "Scale-aware Test-time Click Adaptation for Pulmonary Nodule and Mass Segmentation". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2023. [URL]

[25] Jiancheng Yang*, Mingze Gao*, Kaiming Kuang, Bingbing Ni, Yunlang She, Dong Xie, Chang Chen. "Hierarchical Classification of Pulmonary Lesions: A Large-Scale Radio-Pathomics Study". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2020. ***Oral; Early Accepted***. [URL]

[26] Kaiming Kuang, Li Zhang, Jingyu Li, Hongwei Li, Bo Du, Jiajun Chen, Jiancheng Yang#. "What Makes for Automatic Reconstruction of Pulmonary Segments". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2022. [URL]

[27] Liang Jin*, Jiancheng Yang*, Kaiming Kuang, Bingbing Ni, Yiyi Gao, Yingli Sun, Pan Gao, Weiling Ma, Mingyu Tan, Hui Kang, Jiajun Chen, Ming Li. "Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet". *eBioMedicine (The Lancet Discovery Science)*, 2020. [URL]

[28] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, Bingbing Ni. "MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification". *Scientific Data (Nature Portfolio)*, 2023. ***ESI Highly Cited Paper & Hot Paper***. [URL]

[29] Jianning Li, Zongwei Zhou, Jiancheng Yang, ..., Pascal Fua, Alan L. Yuille, Jens Kleesiek, Jan Egger. "MedShapeNet – A Large-Scale Dataset of 3D Medical Shapes for Computer Vision". *arXiv:2308*. [Preprint]

[30] Jiancheng Yang*, Rongyao Fang*, Bingbing Ni, Yamin Li, Yi Xu, Linguo Li. "Probabilistic Radiomics: Ambiguous Diagnosis with Controllable Shape Analysis". *MICCAI - Medical Image Computing and Computer Assisted Intervention*, 2019. ***Early Accepted***. [URL]

[31] Jiancheng Yang*, Yangzhou Jiang*, Xiaoyang Huang, Bingbing Ni, Chenglong Zhao. "Learning Black-Box Attackers with Transferable Priors and Query Feedback". *NeurIPS - Neural Information Processing Systems*, 2020. [URL]

Teaching Statement

AI is transforming every aspect of society, including healthcare and biomedicine. Training the next generation of AI talent is a key responsibility of academia. My goal is to equip students with a solid foundation in AI principles and practices, enabling them to tackle real-world challenges from first principles and with interdisciplinary mindsets. I have served as the lead teaching assistant for 4 semesters of machine learning courses and designed impactful community initiatives, including datasets, challenges, tutorials, and webinars. During my PhD and postdoc, I mentored over 20 students across bachelor's, master's, and PhD levels, guiding junior lab members, research assistants, interns, and thesis projects both in-person and remotely. These efforts resulted in multiple top-tier publications, top challenge solutions, and open-source projects. Looking ahead, I am eager to develop and teach courses on AI for health, 3D vision and geometry, as well as broader topics in computer vision, machine learning, computer science in interdisciplinary applications.

1 Teaching Philosophy and Goals

In this era of highly interdisciplinary innovation, AI stands as a transformative technology capable of redefining traditional solutions across diverse scenarios. This potential underscores the importance of grounding AI in real-world challenges. My teaching philosophy centers on this principle: technology should serve to solve real problems. To achieve this, I emphasize three core aspects:
AI Principles and Practices. Building a fractal, "T-shaped" knowledge structure is essential—combining deep expertise in specific domains with a broad understanding of related technical areas. As a lifelong learner and MOOC enthusiast, I appreciate how access to knowledge and skills has become more democratized than ever. Still, a strong foundation in AI principles and practices is indispensable. My teaching approach emphasizes understanding the AI essence, with project-based practice to deeply explore selected areas while fostering a broad understanding of others.
Real-World Problem Solving from First Principles. Solving real-world challenges demands the ability to approach problems from first principles. This involves analyzing problems at their core, designing innovative solutions, and having the courage to explore uncharted territories. While my primary focus is on health applications, the problem-solving frameworks and transferable skills developed in these contexts empower students to succeed across a variety of domains.
Interdisciplinary Mindsets. Although AI forms the foundation of my teaching, it is not the sole tool for solving real-world problems. A problem-first approach encourages students to develop interdisciplinary mindsets, enabling them to view challenges from broader and higher-level technical perspectives and avoid the pitfall of "when you have a hammer, everything looks like a nail".

2 Teaching Experience and Broader Impact

During my PhD, I served as the lead teaching assistant for Machine Learning courses (80–100 students) over four semesters. My responsibilities included managing course projects, assignments, exam question preparation, and delivering selected lectures. I was fully responsible for two 1.5-hour introductory lectures on modern deep learning and CNNs, which provided foundational insights into advanced architectures and techniques. To support project-based learning, I designed an educational API, 2048-api, to teach deep learning concepts using the 2048 game. This API offers an easy-to-use game interface, a robust planning-based agent, and a web-based GUI for students to develop imitation learning and reinforcement learning agents. Additionally, I created a Kaggle In-Class competition for 3D medical data classification, encouraging hands-on application of learned concepts. Both projects were widely praised for their educational value and engagement; one student remarked that these were the highest-quality educational projects at SJTU. To support understanding of clustering, I developed a Jupyter-based tutorial, covering KMeans(++), Gaussian Mixture Models (GMM), and Spectral Clustering in an interactive, hands-on format. These efforts reflect my commitment to fostering the core aspects of my teaching philosophy: building strong foundations, solving real-world problems, and promoting interdisciplinary thinking.

Beyond the classroom, I have been deeply involved in impactful community initiatives, including datasets, challenges, tutorials, and webinars. One of my most significant contributions is MedMNIST, a dataset designed to lower the entry barrier to biomedical image analysis. In addition to its extensive use in research, MedMNIST has been widely adopted for educational purposes, ranging from high school, undergraduate, and graduate courses to research workshops, tutorials, and

hackathons. I also led the MICCAI 2020 RibFrac Challenge and co-organized the MICCAI 2022 MELA Challenge, both of which contributed to educational events beyond research impact, such as the NUS-MIT Online Healthcare AI Datathon 2020 and the IMAGINE AI 2024 Datathon. Additionally, I co-organized the MICCAI 2024 MedShapeNet Tutorial, which provided participants with an introduction to classical and deep 3D shape analysis methods for biomedical applications. Similar content has also been presented as a guest lecture in Biomedical Image Analysis at Boston College.

As a personal passion, I founded and lead the HIT Webinar, a Chinese-language webinar series focusing on Healthcare, artificial Intelligence, and emerging Technologies. Since its launch in April 2022, this fully open, non-profit initiative has hosted over 90 sessions, fostering knowledge sharing and collaboration within the global Chinese biomedical AI community. Building on its success, I plan to launch an English version to reach a broader global audience in the future.

3 Teaching Interests

I am eager to design and deliver both introductory and advanced courses tailored to the needs of academic programs, focusing on AI for health, 3D vision and geometry, and broader topics.
Introductory and Advanced Courses in AI for Health. These courses will teach students how to ethically and effectively integrate AI technologies into clinical practice. Topics include multi-center, multi-modal, and multi-task ("multi-X") data in healthcare, covering 2D/3D medical imaging, text, EHR, omics data, and applications in diagnosis, prognosis, and treatment. Broader discussions on AI ethics, including privacy, fairness, and robustness, will also be addressed.
Introductory and Advanced Courses in Geometric Deep Learning. These courses will cover the foundations and applications of geometric deep learning, focusing on geometric data structures, algorithms, and 3D deep learning. Applications will include healthcare, robotics, and AR.
Diverse Introductory Courses, such as computer vision, machine learning, algorithms, scientific programming, statistics and computational methods for interdisciplinary applications.
Interdisciplinary Seminars. These seminars will be tailored for graduate and advanced undergraduate students from diverse disciplines such as computer science, engineering, and biomedicine.

4 Research Mentoring

I am passionate about working with students and have had the privilege of (co-)mentoring over 20 talented and highly motivated individuals across diverse academic levels. Through regular mentorship and collaboration, we have completed numerous impactful projects, resulting in top-tier publications, winning challenge solutions, and open-source contributions. A detailed list of my mentorship activities can be found in my CV under "Mentorship & Teaching".

My experiences reflect my ability to work effectively with students from diverse backgrounds and in various collaborative settings. This capability is the result of deliberate practice, as I strive to maintain the flexibility and potential of my work, avoiding dependence on a single path. I have mentored bachelor's, master's, and PhD students, who have been junior lab members, research assistants, interns, or thesis students. These students come from a wide range of cultural backgrounds, including Chinese, Swiss, French, Spanish, Brazilian, Japanese, American, Pakistani, Russian, and Turkish. Our collaborations have included both in-person and fully remote formats. Many of these students have gone on to pursue further studies at prestigious institutions such as Harvard, Princeton, CMU, ETH Zurich, SUNY Buffalo, and SJTU. I am proud to maintain ongoing collaborations with many of them. Additionally, I serve as a mentor for the Fatima Fellowship, a non-profit initiative that provides free research opportunities to underrepresented students.

My advising approach is tailored to each student's level and needs. I have been fortunate to guide both junior students and those on the path to becoming independent researchers. For junior students, I typically provide clear guidance to achieve my research vision, as I often have more ideas than available hands. I set a clear direction, ensure meaningful progress, and focus on building reusable foundations while addressing technical challenges. For more advanced students who are becoming independent researchers, I help them identify topics they are passionate about and encourage them to take ownership of their projects, often as first authors. This allows them to gain experience at every stage of the research process, from experimentation to publication. This personalized approach equips students with the skills needed for academic success and fosters their growth as independent researchers. It is a privilege to contribute to their journeys and to witness their continued achievements in academia and beyond.

Lausanne, Apr 28, 2025

Dear Hiring Committee,

I am writing to express my strong interest in the **Assistant Professor** position in **Computer Sciences** at **University of Warwick**. With a research background in AI-driven biomedical computing, machine learning for healthcare, and spatial intelligence, I am eager to contribute to Warwick's mission of advancing Applied AI, particularly in healthcare applications.

As a postdoctoral researcher at the Swiss Federal Institute of Technology Lausanne (EPFL), I collaborate with Prof. Pascal Fua on AI for health and 3D vision. I earned my Bachelor's and Ph.D. from Shanghai Jiao Tong University and was a visiting research fellow at Harvard University and EPFL. Additionally, I co-founded a medical AI startup in Shanghai, serving as CTO and developing AI-powered multi-omics solutions for lung cancer screening, diagnosis, and treatment, securing multi-million-dollar funding. This experience in translating AI innovations into clinical applications strengthens my ability to drive interdisciplinary research and industry collaboration, aligning well with Warwick's focus on applied AI.

My research integrates fundamental AI methodologies, open data initiatives, and clinical translation research to address **real-world** healthcare challenges, while advancing **spatial intelligence for healthcare**. I leverage advances in **deep geometric learning** and **generative AI** to analyze **multimodal**, complex biomedical and spatial data, improving diagnosis and treatment. These methods have been applied to thoracic and cardiac care, directly contributing to improve patient outcomes. I have published over **50** papers (4,200+ citations, h-index: 23) in leading journals and conferences such as Cancer Research, eBioMedicine, TMI, MedIA, CVPR, MICCAI, and NeurIPS. I serve as an **Area Chair** for MICCAI 2024/2025 and MIDL 2025, and **an Editorial Board Member** for npj Digital Medicine. My contributions have been recognized by honors such as inclusion among the **Top 2% Scientists Worldwide** and **Forbes 30 Under 30**.

At Warwick, I see an exciting opportunity to advance AI for healthcare by integrating multi-modal biomedical data, including imaging, omics, and electronic health records, to develop robust and trustworthy AI models. My work aligns well with the college's focus on Applied AI, particularly AI applications in health sciences. I am also keen to explore interdisciplinary collaborations across engineering, health sciences, and network science.

Beyond research, I am committed to fostering the next generation of AI scientists through teaching and mentorship. I look forward to developing courses at the intersection of AI, healthcare, and computational biology, supervising graduate students, and promoting diversity in AI education.

I would welcome the opportunity to discuss how my expertise can contribute to the research and teaching mission at Warwick. Thank you for your time and consideration. I look forward to the possibility of joining Warwick and driving impactful AI research and education.

Sincerely,

Jiancheng (JC) Yang, Ph.D.

Computer Vision Lab
CVLAB

EPFL IC IINFCOM CVLAB
Dr. Jiancheng Yang
BC 309 (Bâtiment BC)
Station 14
1015 Lausanne

Phone : +41 21 693 86 64
E-mail : jiancheng.yang@epfl.ch
Website : https://people.epfl.ch/jiancheng.yang
Website CVLAB : https://cvlab.epfl.ch/

Application ID: 163087, Min Wu: Assistant and Associate Professor - Computer Science (100893-0325)

Personal Information

We take privacy seriously and will only use your personal information to administer your application. For more information please see our Data Protection Policies (https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection).

Title	Dr.
Preferred pronouns	She/her/hers
Given Name(s)	Min
Family Name	Wu
Email	minwu@cs.stanford.edu
Preferred Phone No	+1 (650) 683-2905

Additional Information

Are you currently employed by University of Warwick?

No

Will you now or in the future require a visa to obtain/continue to hold the right to work legally in the UK?

Yes

Reasonable Adjustments

We make intentional efforts to employ and retain people with disabilities. The following question will aid us to assist you should you need it during the application process.

Do you have any medical condition, special educational needs or disability that means that you may require reasonable adjustments made for you during either the online assessment, interview or assessment centre stages of our selection process?

No

Source

Where did you find the advert for this vacancy?

Jobs.ac.uk

What made you apply for this vacancy at the University of Warwick? Please select all that apply.

Career progression, University reputation

Is there anything further that prompted you to apply?

Yes

Please detail

I believe my research vision and expertise would stimulate innovative collaborations with current faculty and complement the existing research landscape at Warwick.

References

Reference 1

Title	Professor
First Name	Marta
Last Name	Kwiatkowska
Email	marta.kwiatkowska@cs.ox.ac.uk
Reference Type	Academic
Relationship	She's my PhD advisor.

Reference 2

Title	Professor

First Name	Clark
Last Name	Barrett
Email	barrettc@stanford.edu
Reference Type	Academic
Relationship	He's my postdoc advisor.

Reference 3

Title	Professor
First Name	Miaomiao
Last Name	Zhang
Email	miaomiao@tongji.edu.cn
Reference Type	Academic
Relationship	She's my Master's advisor.

Min Wu

Department of Computer Science, Stanford University
http://cs.stanford.edu/∼minwu
minwu@stanford.edu

Research Interests

Safe and trustworthy AI, responsible AI, robustness, explainability and interpretability.
Formal methods, automated verification, verification of deep neural networks, formal explainable AI.

Academic Employment

Stanford University *2022 – Present*
Stanford Center for AI Safety, Stanford Center for Automated Reasoning
Department of Computer Science
Postdoctoral Scholar
Advisor: Prof. Clark Barrett

Education

University of Oxford *September 2022*
Ph.D. in Computer Science
Advisor: Prof. Marta Kwiatkowska
Thesis: "Robustness evaluation of deep neural networks with provable guarantees."

Tongji University, Shanghai *July 2015*
B.Eng. & M.Eng. in Software Engineering (with Distinction)
Advisor: Prof. Miaomiao Zhang

Research Experience

IBM Research *2022 – Present*
Collaborators: Dennis Wei, Pin-Yu Chen, Eitan Farchi
Project: rigorous analysis of foundation models.

Ford Motor Company *2022 – Present*
Collaborators: Nikita Jaipuria, Danny Jeck, Vidya Murali
Project: formal analysis of quantized neural networks.

Stanford Center for AI Safety & Stanford Center for Automated Reasoning *2022 – Present*
Advisor: Prof. Clark Barrett
Focus: verified robustness and explainability of deep neural networks.

Hebrew University of Jerusalem *2022 – 2024*

Collaborators: Prof. Guy Katz, Omri Isac, Idan Refaeli, Guy Amir, Shahaf Bassan
Project: Marabou 2.0: a versatile formal analyzer of neural networks.

Siemens Technology *2022 – 2023*
Collaborators: Gustavo Quirós Araya, Yassine Qamsane
Project: efficient MILP-solving for automated robotic planning and control.

Genie AI *2020 – 2021*
Collaborator: Anthony Hartshorn
Project: robustness and explainability of natural language processing models.

University of Leeds, Institute for Transport Studies *2016 – 2017*
Collaborators: Prof. Natasha Merat, Tyron Louw
Project: intention anticipation over driving maneuvers in semi-autonomous vehicles.

University of Oxford, Automated Verification Group *2015 – 2021*
Advisor: Prof. Marta Kwiatkowska
Focus: robustness verification for deep neural networks.

Grant Preparation

[National Science Foundation 23-562] Safe Learning-Enabled Systems, $800K, 3 years. *2023*
Proposal: "Ensuring the Safety of Learning-Enabled Systems with Formal Methods, Generative Models, and Runtime Monitoring."
Co-worked with Esen Yel and Shubh Gupta.
PIs: Prof. Clark Barrett, Prof. Grace Gao, Prof. Mykel Kochenderfer

Ford Motor Company – Stanford Alliance Project, $258K, 2 years. [Awarded] *2023*
Proposal: "Formal Analysis of Quantized Neural Networks."
Co-worked with Pei Huang.
PI: Prof. Clark Barrett

SCB^{X} (commercial banking) – Stanford Human-Centered AI, $85K, 1 year. *2023*
Proposal: "Advanced Robustness Verification with Applications to Distribution Shifts and Bias Detection."
PI: Prof. Clark Barrett

IBM Research – Stanford Human-Centered AI, $170K, 2 years. [Awarded] *2022*
Proposal: "Robustness Verification for Foundation Models."
PI: Prof. Clark Barrett

Plus (autonomous trucking) – Stanford Center for AI Safety, $200K, 1 year. *2022*
Proposal: "Safety Guarantees of Machine Learning Models in Autonomous Driving."
Co-worked with Pei Huang.
PI: Prof. Clark Barrett

Grants, Fellowships, and Awards

1. Ford Motor Company Grant ($68,640) *2024 – 2025*
2. Stanford Center for AI Safety Postdoc Fellowship ($91,860) *2023 – 2025*
3. IBM Research Grant ($72,000) *2023 – 2024*
4. Stanford Human-Centered AI Seed Grant ($75,000) *2022 – 2023*
5. Innovate UK (reference 104814) Grant (£33,797) *2020 – 2021*
6. Oxford, Magdalen College, Huscher Family Scholarship (£3,734) *2019 – 2020*
7. Oxford, Magdalen College, Simon Haslam Scholarship (£12,125) *2018 – 2019*
8. Oxford DPhil (PhD) Scholarship *2015 – 2018*
9. Excellent Graduate of Shanghai: Award for academic excellence *2015*
10. Google Anita Borg Memorial Scholarship (¥10,000) *2014*
11. Ji Yang Scholarship: Award for academic excellence (¥6,000) *2013*

Research Highlight

Formal Explainable AI to Promote Trustworthiness

★ **M. Wu** et al., **NeurIPS'23**
"VeriX: Towards Verified Explainability of Deep Neural Networks."
Keynote at Stanford Center for AI Safety 2023 Annual Meeting

★ **M. Wu** et al., submitted to **ICML'25**
"Better Verified Explanations with Applications to Incorrectness and Out-of-Distribution Detection."
Stanford ***CURIS*** Fellowship Project

Robustness Guarantees to Ensure AI Safety

★ **M. Wu** et al., **CVPR'20**
"Robustness Guarantees for Deep Neural Networks on Videos."
Oral Presentation

Deep Neural Network Verification

★ **M. Wu** et al., **Theor. Comput. Sci.'20**
"A Game-based Approximate Verification of Deep Neural Networks with Provable Guarantees."
Invited Paper

★ X. Huang et al. [**alphabetical order**], **CAV'17**
"Safety Verification of Deep Neural Networks."
Keynote at CAV'17

Publications

Google Scholar: **2751** citations, h-index **16**, i10-index **17**.
∗ Denotes alphabetical order.

Under Review

M. Wu, X. Li, H. Wu, and C. Barrett.
"Better veried explanations with applications to incorrectness and out-of-distribution detection."
*arXiv:2409.03060, submitted to International Conference on Machine Learning (**ICML**), 2025.*

Peer-Reviewed Conference Proceedings

P. Huang, H. Wu, Y. Yang, I. Daukantas, **M. Wu**, Y. Zhang, and C. Barrett.
"Towards efficient verification of quantized neural networks."
*AAAI Conference on Artificial Intelligence (**AAAI**) 38 (19), 21152-21160, 2024.*

H. Wu, O. Isac, A. Zeljić, T. Tagomori, M. Daggitt, W. Kokke, I. Refaeli, G. Amir, K. Julian, S. Bassan, P. Huang, O. Lahav, **M. Wu**, M. Zhang, E. Komendantskaya, G. Katz, and C. Barrett.
"Marabou 2.0: a versatile formal analyzer of neural networks."
*International Conference on Computer Aided Verification (**CAV**), 249-264, 2024.*

P. Huang, Y. Yang, H. Wu, I. Daukantas, **M. Wu**, F. Jia, and C. Barrett.
"Parallel verification for δ-equivalence of neural network quantization."
*International Symposium on AI verification (**SAIV**), 78-99, 2024.*

M. Wu, H. Wu, and C. Barrett.
"VeriX: towards verified explainability of deep neural networks."
*Advances in Neural Information Processing Systems (**NeurIPS**) 36, 22247-22268, 2023.*

D. Wei, H. Wu, **M. Wu**, P.Y. Chen, C. Barrett, and E. Farchi.
"Convex bounds on the Softmax function with applications to robustness verification."
*International Conference on Artificial Intelligence and Statistics (**AISTATS**), 6853-6878, 2023.*

H. Wu, **M. Wu**, D. Sadigh, and C. Barrett.
"Soy: an efficient MILP solver for piecewise-affine systems."
*IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 6281-6288, 2023.*

M. Vinzent, **M. Wu**, H. Wu, and J. Hoffmann.
"Policy-specific abstraction predicate selection in neural policy safety verification."
*International Conference on Automated Planning and Scheduling (**ICAPS**): Workshop on Reliable Data-Driven Planning and Scheduling (**RDDPS**), 2023*

M. Wu and M. Kwiatkowska.
"Robustness guarantees for deep neural networks on videos."
*IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 311-320, 2020.*

E. La Malfa, **M. Wu**, L. Laurenti, B. Wang, A. Hartshorn, and M. Kwiatkowska.
"Assessing robustness of text classification through maximal safe radius computation."
*Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing (**EMNLP**), 2949-2968, 2020.*

M. Wu, T. Louw, M. Lahijanian, W. Ruan, X. Huang, N. Merat, and M. Kwiatkowska.
"Gaze-based intention anticipation over driving manoeuvres in semi-autonomous vehicles."
*IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 6210-6216, 2019.*

W. Ruan, **M. Wu**, Y. Sun, X. Huang, D. Kroening, and M. Kwiatkowska.
"Global robustness evaluation of deep neural networks with provable guarantees for the Hamming distance."
*International Joint Conference on Artificial Intelligence (**IJCAI**), 5944-5952, 2019. (114 citations)*

Y. Sun, **M. Wu**, W. Ruan, X. Huang, M. Kwiatkowska, and D. Kroening.
"Concolic testing for deep neural networks."
*ACM/IEEE International Conference on Automated Software Engineering (**ASE**), 109-119, 2018. (364 citations)*

X. Huang, M. Kwiatkowska, S. Wang, and **M. Wu***. [alphabetical order]
"Safety verification of deep neural networks."
*International Conference on Computer Aided Verification (**CAV**), 3-29, 2017. (**Keynote paper**, 1167 citations)*

Peer-Reviewed Journal Articles

C. He, J. Lin, J. Chang, J. Antonello, B. Dai, J. Wang, J. Cui, J. Qi, **M. Wu**, D. S. Elson, P. Xi, A. Forbes, and M. J. Booth.
"Full Poincaré polarimetry enabled through physical inference."
***Optica** 9 (10), 1109-1114, 2022.*

M. Wu, M. Wicker, W. Ruan, X. Huang, and M. Kwiatkowska.
"A game-based approximate verification of deep neural networks with provable guarantees."
***Theoretical Computer Science** 807, 298-329, 2020. (**Invited paper**, 138 citations)*

X. Huang, D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, **M. Wu***, and X. Yi. [alphabetical order]
"A survey of safety and trustworthiness of deep neural networks: verification, testing, adversarial attack and defence, and interpretability."
***Computer Science Review** 37, 100270, 2020. (582 citations)*

PhD Thesis

M. Wu.
"Robustness evaluation of deep neural networks with provable guarantees."
***University of Oxford**, 2020.*

Invited Talks

Safe and Trustworthy AI with Verifiable Guarantees

- MIT, Department of Electrical Engineering and Computer Science (Zoom) *2025*
- University of Southern California, Viterbi School of Engineering (Zoom) *2025*
- Singapore National Research Foundation, AI Fellowship/Investigatorship Grant Call *2025*
- Nanyang Technological University (NTU) Singapore, College of Computing and Data Science *2024*

Better Verified Explanations with Applications to Incorrectness and Out-of-Distribution Detection

- Stanford Center for Automated Reasoning, Seminar *2024*

VeriX: Towards Verified Explainability of Deep Neural Networks

- HCSS (High Confidence Software and Systems) Conference, Maryland *2025*
- [Stanford, CS120] Introduction to AI Safety, Guest Lecturer *2024*
- NeurIPS (Neural Information Processing Systems), New Orleans *2023*
- Stanford Center for Automated Reasoning, Seminar *2023*
- Stanford Center for AI Safety, Lightning Talks *2022*

Safe Deep Learning: Verification, Robustness, and Explainability

- Stanford Center for AI Safety, Outreach with Aptiv PLC (automotive technology supplier) *2022*
- Stanford Center for Automated Reasoning, Seminar *2022*
- Stanford University, Department of Biomedical Data Science *2022*
- Stanford University, Department of Aeronautics & Astronautics *2021*
- Tongji University, School of Software Engineering *2021*
- Institute of Science and Technology Austria (IST Austria) *2020*
- Oxford University, Christ Church *2020*

Robustness Guarantees for Attention Networks on Natural Language Processing

- Oxford Automated Verification Group, Seminar *2020*

Robustness Guarantees for Deep Neural Networks on Videos

- CVPR (Computer Vision and Pattern Recognition), Seattle *2020*
- Oxford Automated Verification Group, Seminar *2019*

A Game-based Approximate Verification of Deep Neural Networks with Provable Guarantees

- Oxford Automated Verification Group, Seminar *2019*

Global Robustness Guarantees for Deep Neural Networks based on the Hamming Distance

- IJCAI (International Joint Conference on Artificial Intelligence), Macao *2019*

Gaze-based Intention Anticipation over Driving Manoeuvres in Semi-Autonomous Vehicles

- IROS (Intelligent Robots and Systems), Macao *2019*
- Oxford Automated Verification Group, Seminar *2018*

Teaching

Stanford University

[CS238/AA228] Validation of Safety-Critical Systems *Winter Quarter 2025*
Gave lecture and made slides on "deep neural network verification".
Guest Lecturer

[CS120] Introduction to AI Safety *Autumn Quarter 2024*
Gave lecture and made slides on "verified explainability of deep neural networks".
Guest Lecturer

University of Oxford

[AIMS CDT] Systems Verification *Hilary Term 2020*
Led lab sessions on "a game-based approximate verification for deep neural networks with provable guarantees".
Lab Demonstrator

[Part C] Probabilistic Model Checking *Michaelmas Term 2017, 2018, 2019*
Tutored classes for course problem sheets for 3 consecutive years.
Class Tutor

[Part C] Probabilistic Model Checking *Michaelmas Term 2016*
Led lab sessions for 1 year.
Lab Demonstrator

Tongji University

[Graduate] System Analysis and Verification *Spring Semester 2013*
Taught lectures and made slides on "temporal logic and model checking".
Co-Lecturer

[Undergraduate] State-of-the-Art Software Technologies *Autumn Semester 2013*
Organized lecture series of cutting-edge software technologies.
Teaching Assistant

Mentorship

Mentor for Stanford CS Independent Research (CURIS, CS195, CS197C, CS199)
Kaia Xiaofu Li, now a coterminal (master) student at Stanford. *2023 – 2024*
Project: *"Verified explanations of misclassified and adversarial inputs."*

Mentor for Oxford Undergraduate Students
Denitsa Markova, now at Jane Street. *2020*
Project: *"Saliency analysis of self-attention networks."*

Bill Daqian Shao, now a PhD Candidate at Oxford. *2020*

Project: *"Robustness guarantees for optical character recognition."*

Mentor for Oxford Graduate Students

Elton Antonis. *2019*
Project: *"Robustness evaluation for natural language processing."*

Jason Chak Yan Lam. *2017*
Project: *"Driver assistance using cognitive modeling and strategy synthesis."*

Professional Service

AI Safety Working Group, Stanford *2022 – Present*
Organized quarterly research talks and faculty/social lunches; held weekly working group meetings.
Created center-wide newsletters; announced seminars, funding/internship opportunities, and etc.
Set up and maintained community mailing list `aisafety-members@lists.stanford.edu`.

Stanford Center for AI Safety Annual Meeting *2022, 2023*
Chaired the lightning talk session and the poster session.
Distributed participation calls; collected presentation slides and posters; administrated logistics.

Federated Logic Conference (FLoC), Oxford *2018*
Supported the *International Conference on Computer Aided Verification (CAV)* session.

Conference and Journal Paper Review
CAV 2025, TACAS 2025, AAAI 2024 (main track, SRRAI special track, NeurIPS fast track), ECAI 2023, AISTATS 2023, Fundamenta Informaticae 2022, CAV 2021, NeurIPS 2020, FAOC 2020, ICLR 2019, IJCAI 2019, ICFEM 2019, IEEE-ITS 2019, NeurIPS 2019, IEEE-RA-L 2018.

Program Committee Member
Human-Centric Machine Learning, co-located with NeurIPS 2019.
Safe Machine Learning: Specification, Robustness, and Assurance, co-located with ICLR 2019.
Artificial Intelligence and Formal Methods, co-located with ICFEM 2019.

Professional Membership
Association for Computing Machinery (ACM) *2025 – Present*
ACM Special Interest Group on Logic and Computation (SIGLOG) *2025 – Present*

Min Wu, PhD
Department of Computer Science
Stanford University
minwu@stanford.edu
+1 650-683-2905

Department of Computer Science
The University of Warwick
April 17, 2025

Dear Faculty Search Committee,

It is my honor to apply for the Assistant Professorship at the Department of Computer Science, University of Warwick. Currently, I am a Postdoctoral Scholar at Stanford University, working on AI safety with Prof. Clark Barrett. I earned my Ph.D. in Computer Science from the University of Oxford, where I was advised by Prof. Marta Kwiatkowska. My research interests center on *safe and trustworthy AI with verifiable guarantees*, situated at the intersection of AI and formal methods.

My research focuses on developing rigorous techniques that provide provable guarantees for *safe and trustworthy AI systems*, emphasizing robustness and explainability. Since 2017, my contributions to *neural network verification* techniques have significantly influenced the field of formal methods, attracting numerous researchers to the area of verifying neural networks. Notably, the concept of "*maximum safe radius*" for quantifying the *robustness* of deep neural networks, first introduced in my Ph.D. thesis, has been adopted by the ISO and IEC in their newly established standard, ISO/IEC TR 5469:2024 "Artificial Intelligence—Functional Safety and AI Systems." Recently, the *explainability* work I led was highlighted in a keynote at the 2023 annual meeting of the Stanford Center for AI Safety, garnering global attention and encouraging researchers to utilize neural network verification tools to analyze formal explainable AI. Additionally, I have extensive experience with securing funding from government agencies such as the NSF, as well as from industrial companies including IBM Research and Ford Motor Company.

The overarching vision of my research is to develop AI systems that are inherently safe, reliable, transparent, and fair. My future research directions include: (1) verifying state-of-the-art network architectures such as *transformers*; (2) incorporating additional safety-critical properties such as *interpretability*, *fairness*, and *privacy* into the realm of safe and trustworthy AI; and (3) extending responsible AI practices to *generative models* such as large language models (LLMs). I believe my research vision and expertise would stimulate innovative collaborations with current faculty and complement the existing research landscape at Warwick. For instance, I could collaborate with faculty in AI & Data Science to explore various aspects of responsible AI, including safety, robustness, explainability, interpretability, privacy, and fairness. It would also be exciting to work with faculty in Formal Methods to bridge the gap between automated verification and generative models such as LLMs.

Teaching, mentorship, and community service are integral aspects of my work. I am enthusiastic about contributing to existing courses in AI and formal methods, as well as developing new courses such as "*Introduction to AI Safety*" and "*Responsible AI: Robustness, Explainability, and Privacy*." I have had the privilege of mentoring several talented students at Stanford and Oxford, leading to top-tier publications. In terms of community service, since 2022, I have been actively involved in the daily operations of the Stanford Center for AI Safety and its annual summer meetings. In 2018, I also assisted in organizing the Federated Logic Conference (FLoC) at Oxford.

I believe that my profile aligns well with the goals and values of your CS Department. Thank you for your time and consideration. I look forward to the opportunity to engage with your faculty and students, share ideas, and learn more about your vibrant community.

Sincerely,

Min Wu

Safe and Trustworthy AI with Verifiable Guarantees

My research aims to develop rigorous techniques for building *safe and trustworthy* artificial intelligence (AI) systems, fostering confidence in their behavior to enable successful societal adoption. Consider AI systems deployed in safety-critical applications such as autonomous driving and healthcare, where they must perform without risk of catastrophic failure or unethical behavior. For instance, a self-driving car's learning component must accurately recognize traffic signs under diverse weather conditions, ensuring *robust* predictions against all physically plausible variations. A diagnosis model used by a dermatologist should base its decisions solely on relevant factors, such as a patient's skin lesions, without introducing bias based on skin tone, ensuring the decision-making process is *explainable* and *fair*.

Researchers have proposed methods to improve model *robustness* and *explainability*. For robustness, adversarial training enhances a model's robustness by training it with adversarial examples, and defensive distillation trains a defense layer to detect abnormal inputs. For explainaiblity, model-agnostic explainers such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) facilitate model transparency. However, these approaches do not provide *provable guarantees* that the model is robust or transparent, which is paramount in high-stakes applications.

My research focuses on developing ***formal methods*** techniques to provide provable guarantees for ***safe and trustworthy AI systems***, ensuring their behaviors are safe, reliable, transparent, and fair [3].* In the following sections, I will highlight my work (Figure 1) at Oxford and Stanford on 1) *neural network verification* as the foundation, 2) *robustness guarantees* to ensure AI safety, and 3) *formal explainable* AI to promote trustworthiness. Each section includes a discussion of the impact of my work and future directions. Finally, I will conclude with my overall vision for safe and trustworthy AI and for responsible *generative AI*, including a roadmap for addressing issues with large language models.

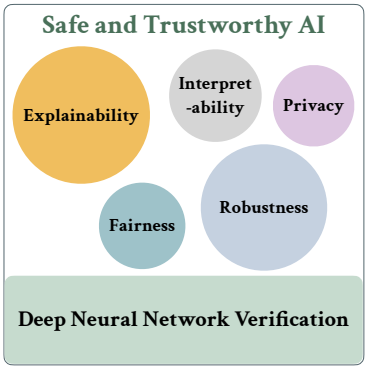


Figure 1: Verifiable responsible AI.

Deep Neural Network Verification

Neural network verification is the process of formally proving that a neural network behaves as expected under certain conditions, formulated as properties. An example property might be: "the spacecraft with a neural network controller will eventually dock in the destination region without causing any collisions."

My approach My approach for neural network verification deploys a *search-based* analysis, i.e., it searches for counter-examples to falsify the property, and if no such counter-example exists, the property holds. In addition to search, my approach incorporates *reachability* and *optimization* to provide search directions.

1. Search with reachability I started this line of research during my PhD at Oxford, at a time when the vulnerability of neural networks to adversarial examples had recently come to light. I helped develop DLV (Deep Learning Verification) [4], which uses *a layer-by-layer reachability analysis* to search for potential counter-examples in both the input space and the hidden layers. Specifically, at a certain hidden layer, it checks whether each point on the search tree satisfies two constraints: 1) for the forward mapping from the current layer to the final layer, the output of the point violates the desired property; and 2) for the backward mapping from the current layer to the input layer, the point is reachable from the input set. If such a point exists, its corresponding input is a counter-example that violates the property; otherwise, the approach continues to the next layer. If no such counter-example is found by the end of the process, the property holds.

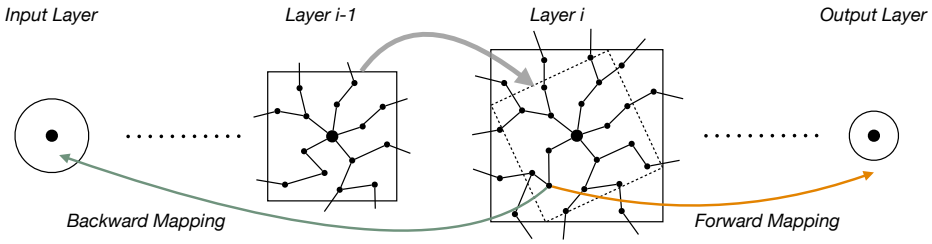


Figure 2: The DLV (Deep Learning Verification) approach [4].

2. Search with optimization This line of work searches in the *function space* by exploring possible activation patterns. For instance, the Reluplex work done at Stanford University (before I arrived) utilizes a modified simplex procedure to find feasible activation patterns in networks with ReLU (Rectified Linear Unit) activation functions. If such a pattern is found, the property is violated; if not, the property holds. I joined the team working on this line of research during my postdoc at Stanford. The most recent development in this direction was the release of Marabou 2.0 [15] (Figure 3), a neural network verification tool suite which supports a wide range of activation functions and features an optimized system architecture that uses satisfiability modulo theories (SMT) and mixed integer linear programming (MILP) to accommodate both linear and nonlinear constraints.

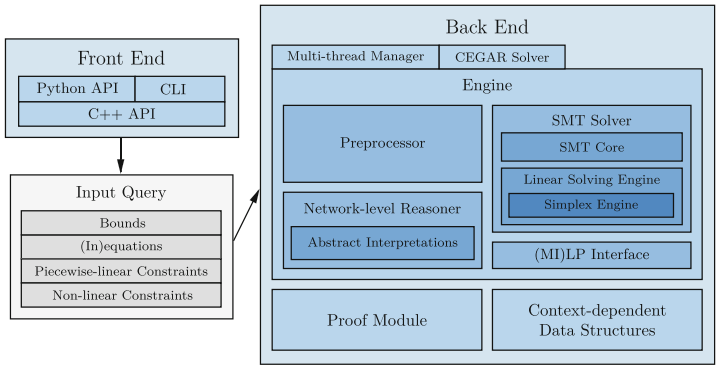


Figure 3: Marabou 2.0's system architecture [15].

Impact Both the DLV and Reluplex works were first introduced at the 2017 International Conference on Computer-Aided Verification (CAV'17). Since then, these two lines of research have significantly influenced the field of formal methods, attracting numerous researchers to the growing area of neural network verification. I feel immensely fortunate to have contributed to both efforts.

Future work A natural future direction is to verify *state-of-the-art network structures*. While the aforementioned techniques are applicable to common network topologies, such as fully-connected and convolutional layers with various activation functions, new architectures, such as the recent *transformer* architecture, continue to emerge. A key challenge with transformers is the *attention mechanism*, which contains two non-linear components: the softmax to compute the attention scores, and the bilinear function to perform the multiplication between the query and the key matrices and between the attention scores and the value matrix. In [14], we introduced an approach for bounding the softmax function by computing its convex lower bounds and concave upper bounds, compatible with convex optimization formulations for characterizing networks. Specifically, this approach derives bounds using both a natural exponential-reciprocal decomposition of the softmax and an alternative decomposition in terms of the log-sum-exp function. In ongoing work, we are developing a similar approach for bounding the bilinear function.

Robustness Guarantees to Ensure AI Safety

Robustness of neural networks refers to their ability to maintain stable and reliable performance even when faced with variations, uncertainties, or adversarial inputs. A *robustness guarantee* thus proves that, for a given set of inputs, the network's outputs always remain invariant (in classification) or within acceptable changes (in regression).

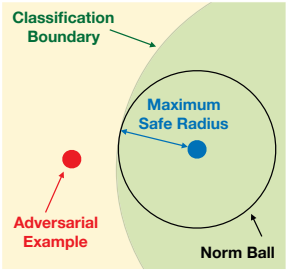


Figure 4: Safe radius [8].

My approach Most approaches prove or disprove robustness guarantees qualitatively, i.e., if the model is proven to be robust, it returns "`True`," and if not, it returns "`False`" with a counter-example. My approach goes a step further, quantitatively evaluating model robustness by computing its *"maximum safe radius"* (Figure 4), defined as the largest radius centered on a given input within which all (continuous and infinite) inputs are safe.

1. Perception models for robust image classification In [11], I proposed a game-based verification framework DeepGame to compute the lower and upper bounds of the maximum safe radius, providing a theoretical guarantee that the increasing lower bounds and the decreasing upper bounds will eventually converge to the exact safe radius. In a follow-up work [8], I generalized this game-based framework to time-series inputs such as videos. The key idea is to perturb the extracted optical flows to capture the spatial features on frames and the temporal dynamics between neighboring frames. Additionally, by utilizing the Lipschitz continuity of networks, I designed a grid search algorithm [6] to compute the maximum safe radius based on the non-differentiable Hamming distance, whereas most other approaches are only applicable to differentiable norms such as the Euclidean and Chebyshev distances.

2. Natural language processing (NLP) models for robust sentiment analysis Quantifying the robustness of NLP models presents unique challenges due to the *unbounded* nature of the input text, where each word, after tokenization and embedding, becomes a vector of real numbers. Rather than manipulating embeddings directly, my approach [5] imposes perturbations in the token space to maintain *semantic coherence*. For instance, during word substitutions (e.g., synonyms or antonyms), syntactic filtering is applied to ensure replacements are syntactically and semantically consistent. To compute the maximum safe radius for NLP models in sentiment analysis, Monte Carlo tree search is employed to obtain the upper bounds, and linear constraint relaxation techniques are used to compute the lower bounds.

3. Robustness guarantees for quantized neural networks In power- and memory-restricted domains, such as autonomous driving [10] and robotics [16], quantized neural networks are useful due to their competitive accuracy and lower computational costs. As part of an ongoing collaboration between Stanford and the Ford Motor Company, which deploys quantized models in their vehicles, I have been exploring encoding the quantized models and robustness properties as integer linear programming (ILP) problems and then using an ILP solver to provide sound and complete guarantees [1]. Furthermore, to address the precision and generalization losses due to quantization, my work [2] also verifies the equivalence between the quantized model and its original real-valued counterpart, utilizing MILP to accommodate both integer and real number arithmetic.

Impact First introduced in my PhD thesis [7], the concept of "*maximum safe radius*" for quantifying the robustness of deep learning models has been adopted by the ISO and the IEC in their recently established standard, ISO/IEC TR 5469:2024, titled "Artificial Intelligence—Functional Safety and AI Systems".

Future work My work aims to build AI systems that not only advance technological capabilities but also uphold human dignity and societal well-being. As part of this effort, I plan to explore the robustness-related properties of *privacy* and *fairness*. Privacy aims to protect individuals' personal data and ensure that AI systems handle such data with care and respect. While there are existing methods such as data *minimization* to enhance privacy, my approach will emphasize on *anonymization*—removing or obscuring personal identifiers to anonymize individuals, e.g., masking a patient's name, address, social security number, insurance, and billing details in medical records. This process parallels a robustness problem: replacing identifiers with zero/padding tokens is akin to perturbing the original record into a similar one in the input space, where the model's decisions should remain invariant or within allowable change. Fairness ensures AI systems treat all individuals and groups equally, without bias and discrimination. While methods such as *data augmentation* can ensure training data is diverse and representative, my approach will focus on the model side. Inspired by robust training, we can integrate *fairness constraints* into the training process to ensure the model is fair by construction.

Formal Explainable AI to Promote Trustworthiness

Explainable AI provides human-understandable reasons for a model's predictions and thus promotes trustworthiness via the alignment of extracted explanations with human perception. A *formal explanation* is defined as a subset of input features that are sufficient to ensure the invariance of a model's decision, i.e., *any possible* perturbations on the rest of the features will never change the model's output.

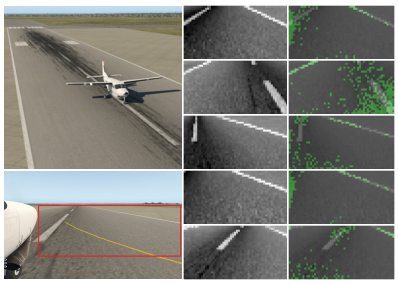


Figure 5: Formal explanations in autonomous aircraft taxiing [12].

My approach My approach utilizes formal tools to compute a *locally optimal* explanation, i.e., a subset of input features that is locally irreducible. I generalize previous techniques by allowing *bounded* perturbations, thereby offering a range of explanation options from the strongest (unbounded perturbation) to the weakest (minimal perturbation).

1. Verified explainability of deep neural networks In [12], I proposed a system to produce optimal verified explanations for neural networks. Given a traversal order of the input features, it utilizes a neural network verifier to process and categorize the features into either *a minimal explanation set* or a set of irrelevant features. By visualizing the explanation, the decision-making process of the model becomes transparent. In subsequent work [9], I developed more meaningful traversal orders that result in finer (smaller) explanations, and created algorithms for batch processing of the features, significantly reducing computation time. Specifically, my recent work reduced the explanation size by two-thirds and accelerated the generation time tenfold compared to the state of the art.

2. Use formal explanations to evaluate model trustworthiness My work [12] demonstrates the usefulness of these formal explanations in safety-critical applications. For instance, in a real-world autonomous aircraft taxiing scenario (Figure 5), the task is to control the cross-track position of the aircraft using pictures taken from the camera on the right wing. A regression model was trained with high accuracy as the controller to adjust the aircraft's position accordingly. Then the question arises, "*how do we know if the controller is trustworthy*?" My explanations for this controller show that 1) the controller is capable of detecting the remote line; 2) the controller mainly focuses on the centerline to measure the aircraft's deviation from the center; and 3) the controller pays attention to the bottom middle region to check the presence or absence of the centerline. The alignment of these explanations with human expectations enhances the model's trustworthiness. Additionally, my explanations can be used to detect incorrect and out-of-distribution inputs [9].

Impact My work on verified explainability was highlighted in a keynote at the 2023 annual meeting of the Stanford Center for AI Safety. It has attracted researchers from Israel, France, Singapore, and other countries to use neural network verification tools to analyze formal explainable AI. My approach has also been adopted in other domains, e.g., scholars from Saarland University applied my framework to select abstraction predicates in neural policy safety verification [13].

Future work My work aims to build AI systems that make transparent decisions, thereby enhancing their trustworthiness. Beyond explainability, I am interested in exploring *interpretability*, which involves understanding how the components of the model, such as neurons and layers, contribute to the final output, thus making the decision-making process more transparent. I plan to investigate how interpretability can be enhanced from two perspectives: 1) neuron-level: *activation maximization* helps identify the input patterns that activate or deactivate certain neurons, clarifying the causal relationship between input features and neuron responses; and 2) layer-level: *relevance propagation* redistributes the prediction backward through the intermediate layers, highlighting the relative importance of each input feature in the final decision. Another potential future direction is to provide interpretability with provable guarantees, for instance, by demonstrating that a model's decision is definitively due to specific input features that activate or deactivate certain neurons in intermediate layers.

Conclusion and Future Work

The grand vision of my research is to forge a future where AI systems are inherently *safe and trustworthy*. My work aspires to set new standards in AI reliability and responsibility, building robust AI systems that people can trust to make transparent and fair decisions, thereby leading to widespread acceptance and integration across various aspects of daily life. My work can also pave the way for more informed AI regulations and policies, which are gaining more importance, as reflected by the UK AI Regulation, the EU AI Act, the US Executive Order on AI, and the Singapore AI Verify Foundation.

Future work In addition to the directions mentioned above, a key research direction is to generalize responsible AI to *generative models*. For instance, how can we make *large language models* (*LLMs)* robust, transparent, and fair? Because LLMs are substantially larger than conventional networks, applying current formal methods techniques faces significant scalability challenges. A possible solution is to apply these approaches at the encoder and/or decoder levels, ensuring they remain computationally manageable while still effective. My ongoing work on verifying *transformers*, the essence of encoders and decoders, provides a strong theoretical foundation for this direction.

Generative AI also faces unique challenges, such as *hallucination*, where the model generates plausible but incorrect information, and *jailbreaking*, where users manipulate the model to produce unintended or harmful outputs. Hallucination in LLMs can be mitigated by providing high-quality training data and continuously updating the model with up-to-date information to reduce the likelihood of generating incorrect or outdated content. From the *output* side, I plan to implement *automated post-processing checks* to verify the accuracy of generated content before it is presented to users. Take code generation as an example. Strategies from my domain knowledge—such as model checking, theorem proving, abstract interpretation, and symbolic execution—can ensure the correctness of programs produced by LLMs. As for jailbreaking, from the *input* side, filters can be implemented to detect and block adversarial prompts designed to manipulate the model. For code generation, beyond mere detection and blocking, I can also develop techniques to transform unstructured natural language into formal specifications that clearly and precisely outline the desired behavior of the program. By using these techniques, programs generated by LLMs can be verified before being presented to users and deployed in high-stakes applications, preventing severe economic or societal catastrophes.

References (* denotes alphabetical order)

[1] Pei Huang, Haoze Wu, Yuting Yang, Ieva Daukantas, **Min Wu**, Yedi Zhang, and Clark Barrett. Towards efficient verification of quantized neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pages 21152–21160, 2024.

[2] Pei Huang, Yuting Yang, Haoze Wu, Ieva Daukantas, **Min Wu**, Fuqi Jia, and Clark Barrett. Parallel verification for δ-equivalence of neural network quantization. In *International Symposium on AI Verification (SAIV)*, pages 78–99. Springer, 2024.

[3] Xiaowei Huang, Daniel Kroening, Wenjie Ruan, James Sharp, Youcheng Sun, Emese Thamo, **Min Wu***, and Xinping Yi. A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. *Computer Science Review*, 37:100270, 2020.

[4] Xiaowei Huang, Marta Kwiatkowska, Sen Wang, and **Min Wu***. Safety verification of deep neural networks. In *Proceedings of the 29th International Conference on Computer Aided Verification (CAV)*, pages 3–29. Springer, 2017.

[5] Emanuele La Malfa, **Min Wu**, Luca Laurenti, Benjie Wang, Anthony Hartshorn, and Marta Kwiatkowska. Assessing robustness of text classification through maximal safe radius computation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP): Findings*, pages 2949–2968, 2020.

[6] Wenjie Ruan, **Min Wu**, Youcheng Sun, Xiaowei Huang, Daniel Kroening, and Marta Kwiatkowska. Global robustness evaluation of deep neural networks with provable guarantees for the Hamming distance. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5944–5952, 2019.

[7] **Min Wu**. Robustness evaluation of deep neural networks with provable guarantees. *University of Oxford*, 2020.

[8] **Min Wu** and Marta Kwiatkowska. Robustness guarantees for deep neural networks on videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 311–320, 2020.

[9] **Min Wu**, Xiaofu Li, Haoze Wu, and Clark Barrett. Better verified explanations with applications to incorrectness and out-of-distribution detection. In *Submitted to the 42nd International Conference on Machine Learning (ICML)*, 2025.

[10] **Min Wu**, Tyron Louw, Morteza Lahijanian, Wenjie Ruan, Xiaowei Huang, Natasha Merat, and Marta Kwiatkowska. Gaze-based intention anticipation over driving manoeuvres in semi-autonomous vehicles. In *Proceedings of the 32nd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6210–6216, 2019.

[11] **Min Wu**, Matthew Wicker, Wenjie Ruan, Xiaowei Huang, and Marta Kwiatkowska. A game-based approximate verification of deep neural networks with provable guarantees. *Theoretical Computer Science*, 807:298 – 329, 2020.

[12] **Min Wu**, Haoze Wu, and Clark Barrett. VeriX: towards verified explainability of deep neural networks. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 22247–22268, 2023.

[13] Marcel Vinzent, **Min Wu**, and Jörg Hoffmann. Policy-specific abstraction predicate selection in neural policy safety verification. In *Proceedings of the 2nd Workshop on Reliable Data-Driven Planning and Scheduling, co-located with the International Conference on Automated Planning and Scheduling (RDDPS@ICAPS)*, 2023.

[14] Dennis Wei, Haoze Wu, **Min Wu**, Pin-Yu Chen, Clark Barrett, and Eitan Farchi. Convex bounds on the softmax function with applications to robustness verification. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 6853–6878. PMLR, 2023.

[15] Haoze Wu, Omri Isac, Aleksandar Zeljić, Teruhiro Tagomori, Matthew Daggitt, Wen Kokke, Idan Refaeli, Guy Amir, Kyle Julian, Shahaf Bassan, Pei Huang, Ori Lahav, **Min Wu**, Min Zhang, Ekaterina Komendantskaya, Guy Katz, and Clark Barrett. Marabou 2.0: A versatile formal analyzer of neural networks. In *Proceedings of the 36th International Conference on Computer Aided Verification (CAV)*, 2024.

[16] Haoze Wu, **Min Wu**, Dorsa Sadigh, and Clark Barrett. Soy: An efficient MILP solver for piecewise-affine systems. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6281–6288. IEEE, 2023.

"Pass the parcel. Take it, feel it, and pass it on.[1]"

— Alan Bennett, *The History Boys*

Philosophy

I believe that teachers have a profound responsibility to inspire the next generation of *thinkers*, *leaders*, and *innovators*. Every individual has the right to pursue the life they desire. For my students, I aspire to equip those seeking a comfortable and fulfilling life with the necessary skills to achieve it. I aim to raise awareness among those who wish to change the world about societal imperfections, empowering them to effect positive changes. For those intent on pushing the boundaries of human knowledge, my goal is to prepare them to "boldly go where no one has gone before.[2]"

Before every lecture, I ask myself the following questions:

1. *How do I present all the technical details in a deep and interesting way?*
   I strive to build **solid knowledge foundations** so that students can succeed academically and professionally. This foundation ensures they have the freedom to pursue their passions rather than being compelled to work solely for financial reasons.
2. *Have I addressed the **societal implications** of my discipline?*
   I aim to make students **aware of both the positive and negative impacts** of what they are learning. Knowledge can be a double-edged sword, and it is essential for students to understand the potential consequences of misused technology and to develop **empathy** for those affected by such misuse. This **awareness** will guide them in considering the broader benefits of their work on society.
3. *How can I encourage students to seek alternative solutions?*
   I nurture **critical thinking and curiosity**, hoping students will continually question existing solutions and seek better alternatives. This mindset prepares them to explore the unknown and innovate in their future endeavors.

Teaching and Mentoring Experience

Solidifying Knowledge Foundations

Conducting class tutorials at Oxford provided an invaluable chance to reinforce both my students' and my own foundational knowledge of the discipline. The tutorial format, with fewer than *10 students per class* and thus multiple classes if there were more than 10 students, allowed me to answer each student's questions carefully and always encourage alternative solutions.

From 2016 to 2019, I served as the class tutor and lab demonstrator for the "`Probabilistic Model Checking [Part C]`" course, instructed by Prof. Marta Kwiatkowska, Dr. Bruno Lacerda, and Dr. George Kenison. Attached on the right is feedback from students for this course, compiled by admin Jo and co-lecturer George.

From: Jo Ponting
Subject: Feedback for PMC class - MT2018
"I like how the tutor points out alternative methods to get to a solution. Tutor has great timing and is very well prepared. Thanks!"

From: George Kenison
Subject: PMC class feedback - MT2019
"I preferred Min Wu's use of the projector with examples from the class, it was easier to understand what she was referring to as she kept the notes open."

Tutor and Marker: Min Wu	1	2	3	4	5
The exercises in the problem sheets helped me to understand the topics covered in the lectures.				4	16
The tutor helped to give me a rounded understanding of the course material.			2	4	14
The tutor took time to understand any difficulties I had with the work.			2	4	14
The tutor worked hard to make the classes relevant and interesting.		1	2	6	11

Incorporating Societal Responsibility

During the Fall 2024 Quarter at Stanford, I delivered a guest lecture on "*Explainable AI and Verified Explainability of Neural Networks*" for the course "`Introduction to AI safety [CS120]`", instructed by Dr. Max Lamparth. I began by elucidating

[1] "*The History Boys*" is a play (also adapted into a film) about high school students preparing for Oxford and Cambridge entrance exams. In the final scene, Hector, an unconventional and eccentric teacher, tells his students: "Pass the parcel. Take it, feel it, and pass it on. Not for me, not for you, but someone, somewhere, one day. Pass it on, boys. That's the game I want you to learn. Pass it on."

[2] This phrase outlines the mission of the Starship Enterprise in the original *Star Trek* series (1966-1969). Captain James T. Kirk's complete speech is: "Space: the final frontier. These are the voyages of the Starship Enterprise. Its five-year mission: to explore strange new worlds; to seek out new life and new civilizations; to boldly go where no man has gone before."

the concept of explainable AI and its potential achievements, focusing on scenarios where the lack of transparency in AI decision-making processes could lead to catastrophic outcomes. For instance, in healthcare, researchers at the National Cancer Institute employ AI algorithms to highlight lesions that could signify cancer in MRI images of a patient's prostate. Similarly, when dermatologists utilize diagnostic models, the decisions must be based on relevant factors, such as skin lesions indicative of cancer, rather than skin tones.

In the upcoming Winter Quarter 2025, I will give a guest lecture on "*Neural Network Verification with Applications to Robustness Guarantees*" for the course "`Validation of Safety-Critical Systems [AA228/CS238]`", co-instructed by Prof. Mykel Kochenderfer and Dr. Sydney Katz. In safety-critical applications, such as autonomous driving, the integrity of learning components is paramount. Compromised components can lead to severe consequences—an early case is that a self-driving Uber test car killed a pedestrian in Arizona in March 2018, and a recent case is that a Tesla in autopilot mode crashed into a patrol car in California in June 2024. These incidents underscore the importance of robust verification in ensuring the safety and reliability of AI systems.

Encouraging Critical Thinking and Curiosity

I have had the privilege of mentoring quite a few talented students at both Stanford and Oxford. One notable mentee is Kaia Xiaofu Li, a Stanford undergraduate who was diagnosed with congenital hearing loss at the age of two and has since utilized a hearing aid and then a cochlear implant. Our mentorship began when Kaia joined my project on formal explainable AI through the Stanford CURIS summer research program. This initial summer internship evolved into a research collaboration spanning over a year, culminating in *an ongoing ICML submission*.

I mentored Kaia with the rigor and guidance typically reserved for future Ph.D. students, overseeing their progress through literature reviews, research proposals, and independent research. Our collaboration started when Kaia took `CS197C`, a course designed to provide an onramp to CURIS. I met with Kaia regularly, helping them with a literature review and a research proposal. Over the summer, Kaia had some preliminary results and presented them at the 2023 Stanford Center for AI Safety annual meeting. Subsequently, they enrolled in `CS195` and `CS199` to pursue independent research under my supervision, achieving further results presented at the Stanford Centaur seminar.

Apart from this, I have also mentored a few Oxford graduate and undergraduate students—Bill Daqian Shao (now a PhD Candidate at Oxford), Denitsa Markova (now at Jane Street), Jason Chak Yan Lam, and Elton Antonis.

Community Outreach

Delivering invited talks across various communities provides a valuable platform to teach those who may not be familiar with my entire body of work but have a specific interest in certain aspects. For example, when I presented on "*Safe Deep Learning: Verification, Robustness, and Explainability*" at Stanford's Department of Biomedical Data Science (invited by Prof. James Zou), the students focused on AI for healthcare were particularly interested in explainability, such as highlighting different lesions in MRI images. In contrast, when I delivered a similar talk at Stanford's Department of Aeronautics and Astronautics (invited by Prof. Mykel Kochenderfer), the students, who were more concerned with the safety of self-driving cars, were primarily interested in robustness, including the accurate recognition of traffic signs. A further iteration of this talk was given at the Institute of Science and Technology in Austria (invited by Prof. Tom Henzinger), where the audience, mainly consisting of students working on the theory of software systems, showed a keen interest in how formal methods could be applied to ensure the safety of deep learning models.

Teaching Plans

I am interested in developing new courses at the intersection of AI and formal methods, such as an undergraduate course on "**Introduction to AI Safety**" or a graduate course on "**Responsible AI: Robustness, Explainability, and Privacy**." These courses would explore the societal implications of AI and solutions to enhance AI safety and trustworthiness.

I am also happy to assist existing courses in AI and formal methods. For example, I could contribute to courses on ***Machine Learning***, ***Computer Vision***, and ***Natural Language Processing*** for AI, and ***Probabilistic Model Checking*** and ***Computer-Aided Verification*** for formal methods.

Application ID: 164019, Sajid Javed: Assistant and Associate Professor - Computer Science (100893-0325)

Personal Information

We take privacy seriously and will only use your personal information to administer your application. For more information please see our Data Protection Policies (https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection).

Title	Dr.
Preferred pronouns	He/him/his
Given Name(s)	Sajid
Preferred Name	Sajid Javed
Family Name	Javed
Email	sajid.javed@ku.ac.ae
Preferred Phone No	+971503308507

Additional Information

Are you currently employed by University of Warwick?

No

Will you now or in the future require a visa to obtain/continue to hold the right to work legally in the UK?

Yes

Reasonable Adjustments

We make intentional efforts to employ and retain people with disabilities. The following question will aid us to assist you should you need it during the application process.

Do you have any medical condition, special educational needs or disability that means that you may require reasonable adjustments made for you during either the online assessment, interview or assessment centre stages of our selection process?

No

Source

Where did you find the advert for this vacancy?

Jobs.ac.uk

What made you apply for this vacancy at the University of Warwick? Please select all that apply.

Career progression, University reputation

Is there anything further that prompted you to apply?

Yes

Please detail

My decision to apply to the University of Warwick is rooted not only in its global reputation for research excellence but also in my own strong academic connection to the institution. I had the privilege of completing my postdoctoral research at Warwick, which was a formative experience that significantly shaped my research direction and academic values. That experience left a lasting impression of the university's dynamic, interdisciplinary environment and commitment to impactful research. I'm also currently collaborating with Prof. Nasir Rajpoot on computational pathology projects, which has only deepened my appreciation for Warwick's leadership in AI-driven healthcare research. This ongoing partnership has reinforced my belief that Warwick is an ideal environment to pursue cutting-edge computer vision research with real-world impact. Joining the faculty would be both a return to a place that has been foundational to my career and an exciting step forward in contributing meaningfully to its future.

References

Reference 1

Title	Professor

First Name	Nasir
Last Name	Rajpoot
Email	n.m.rajpoot@warwick.ac.uk
Reference Type	Academic
Relationship	I have collaboration with Prof. Rajpoot. He is still my active mentor/influential peer in my academic journey.

Reference 2

Title	Professor
First Name	Michael
Last Name	Felsberg
Email	michael.felsberg@liu.se
Reference Type	Academic
Relationship	I have worked with Michael Felsberg on computer vision project. He is an eminent figure in the computer vision community.

Reference 3

Title	Professor
First Name	Mohammed
Last Name	Bennamoun
Email	mohammed.bennamoun@uwa.edu.au
Reference Type	Academic
Relationship	Worked with him as a collaborator. He is an eminent computer vision figure in Australia.

Research Statement

Dr. Sajid Javed

Good research advances knowledge; great research changes perspectives. My work in computer vision strives to do both—by building intelligent systems that not only process visual information but also learn to perceive, adapt, and reason.

My research career has been driven by a deep-seated passion for advancing the field of computer vision and applying its transformative potential to address critical societal challenges. As an Assistant Professor of Computer Vision at Khalifa University of Science and Technology (KUST), I have cultivated a robust research program characterized by the publication of high-impact work in top-tier computer vision venues considered as $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$, including ***TPAMI [16], [18], CVPR* [1]-[2], *TIP [4], [9], [19]-[20], [23], TNNLS [11]-[13]*, and *MEDIA [3], [7]- [8], [10]***, *etc*. My research to date has focused on developing innovative solutions within the domains of pathology image analysis, video surveillance, and marine vision. *I am now eager to bring my established expertise, a clear future research vision, and a strong collaborative spirit to the University of Warwick, where I aim to establish a leading research lab and contribute significantly to the creation of a new computer vision Centre of excellence. My research contributions have been significant across three key areas:*

Video Surveillance: Developing Intelligent Systems for Enhanced Security and Monitoring: My research in video surveillance has focused on developing robust and efficient algorithms for fundamental tasks such as *visual tracking, background subtraction, and video object segmentation*, which are essential for a wide range of security, safety, and monitoring applications. My publications in ***TPAMI*, *TIP*, *TCSVT, TMM*, *TCyb*,** and ***TNNLS*** showcase the development of innovative approaches designed to perform reliably in challenging real-world scenarios. A notable contribution, published in ***TIP*** *and* ***TCyb*,** involved the development of a novel transformer-based architecture for robust object tracking, leveraging the attention mechanism to effectively model long-range dependencies and handle occlusions and distractors. In the domain of background subtraction, my work published in ***TIP, TNNLS,*** and ***TCyb*** introduced a robust and adaptive background modeling technique based on tensor decomposition, effectively handling complex background variations, illumination changes, and moving objects. Moreover, my research in video object segmentation has explored the use of deep learning and conditional random fields to achieve pixel-accurate and temporally consistent segmentation of multiple moving objects in complex video sequences. These projects underscore my ability to develop cutting-edge algorithms that address the core challenges in video surveillance and contribute to the development of more intelligent and reliable monitoring systems.

Pathology Image Analysis: AI-Driven Solutions for Advance Cancer Diagnosis: My work in computational pathology has been dedicated to developing advanced computer vision and machine learning techniques for automated analysis of medical images, particularly histopathology slides. My research has focused on improving diagnostic accuracy, enhancing prognostic capabilities, and facilitating the discovery of novel disease biomarkers. For instance, my work published in ***CVPR 2024*** introduced a novel comprehensive vision-language model *(CPLIP)* for the classification of diverse cancer subtypes based on histopathological image features. This approach effectively captured the complex spatial relationships between different tissue components, leading to significantly improved classification accuracy compared to existing

methods. In another project, published in ***CVPR 2025***, I developed a multi-resolution foundational model *(MR-PLIP)* for the detection of clinically significant cancer in whole-slide images, addressing the challenge of limited pixel-level annotations by leveraging slide-level labels and achieving SOTA performance. Furthermore, I have explored the application of graph-based CNN techniques to improve the robustness and generalizability of computational pathology models across different imaging protocols and datasets. These contributions demonstrate my ability to tackle complex challenges in medical image analysis and develop practical AI solutions with the potential to significantly impact clinical practice.

Marine Vision: Unveiling the Depths with Advanced Visual Intelligence: Recognizing the growing importance of understanding and protecting our oceans, I have also dedicated research efforts to the emerging and challenging field of underwater computer vision. This domain presents unique difficulties due to the inherent optical properties of water, leading to significant image degradation. My work in this area has focused on developing effective underwater image enhancement and restoration techniques to improve the visual quality of underwater imagery. Furthermore, I have explored the application of deep learning for tasks such as automated marine species detection, classification, and behavior analysis by developing foundational model based on 2M image-text pairs. While this is a more recent focus of my research, it reflects my commitment to expanding my expertise to address critical environmental challenges and to explore the application of computer vision in novel and impactful domains, contributing to our ability to monitor and understand the underwater world.

Overall, my research journey has been significantly enriched through collaborations with leading international experts in computer vision, including *Prof.* ***Nasir Rajpoot****, Prof.* ***Jiri Matas****, Prof.* ***Thierry Bouwmans****, and Prof.* ***Michael Felsberg***. These collaborations have provided invaluable opportunities for knowledge exchange, access to diverse perspectives, and the development of impactful joint research projects. For example, my collaboration with *Prof.* ***Rajpoot*** has focused on pushing the boundaries of AI-powered diagnostics in computational pathology, while my interactions with *Prof.* ***Matas*** have advanced my work on robust and efficient visual tracking. My work with *Prof.* ***Bouwmans*** has explored novel and adaptive techniques for background subtraction in dynamic scenes, and my collaboration with *Prof.* ***Felsberg*** has focused on the theoretical underpinnings of robust computer vision algorithms. These ongoing collaborations are a testament to my ability to build and maintain strong and productive research partnerships within the global computer vision community.

Future Research Directions

Pioneering Conversational and Agent-Based Intelligent Vision Systems: Building upon my established research foundation and recognizing the transformative potential of recent breakthroughs in *Large Language Models* (LLMs) and *Vision-Language Models* (VLMs), my future research will be strategically directed towards addressing generic computer vision problems within my core areas of interest through these innovative paradigms. I aim to pioneer the development of *conversational computer vision models* and *intelligent AI agents* that can significantly enhance the way we interact with and extract meaningful insights from visual data, leading to transformative applications in healthcare, security, and environmental understanding.

Revolutionizing Computational Pathology with Conversational and Proactive AI Agents: In computational pathology, my future research will focus on developing intelligent systems that can engage in sophisticated dialogues with pathologists, providing a new level of interactive assistance in diagnosis, education, and research. I envision creating conversational AI agents that can understand complex natural

language queries related to histopathology images, such as "*Identify and describe the key cytological and architectural features indicative of this specific grade of tumor*," or "*Compare the immunohistochemical staining patterns in this region with those typically observed in benign lesions*." By leveraging the powerful language understanding and generation capabilities of LLMs and their grounding in visual information through VLMs, these agents will be able to provide detailed textual explanations, highlight relevant regions of interest within the image, and even retrieve and present pertinent medical literature or case studies. This will not only enhance the interpretability and transparency of AI-driven analyses but also serve as an invaluable educational resource for pathology trainees and facilitate more effective collaboration among pathologists. Furthermore, I plan to explore the development of proactive AI agents that can continuously analyze patient image data, identify subtle patterns indicative of disease progression or potential treatment resistance that might be missed by human observers, and suggest further investigations or consultations. These agents could also assist in automating routine tasks such as preliminary report generation and annotation of large image datasets, freeing up pathologists to focus on more complex and critical cases. Key research challenges include developing robust mechanisms for accurate visual grounding of natural language in the complex domain of medical imaging, ensuring the reliability and trustworthiness of the generated information, and addressing the ethical considerations associated with the deployment of conversational AI in healthcare settings.

Developing Intelligent Video Surveillance with Natural Language Interaction and Autonomous Response Agents: In video surveillance, my future research will focus on creating intelligent systems that offer more intuitive and natural human-machine interaction and enhanced autonomous capabilities for proactive monitoring and response. I aim to develop conversational interfaces that allow security personnel to interact with surveillance systems using natural language commands and questions, such as "*Show me all instances of individuals wearing a red jacket entering the building after 6 PM*," or "*Describe the sequence of events leading up to the unauthorized access alarm in Sector 4*." By integrating LLMs for sophisticated natural language understanding and VLMs for grounding these queries in the rich visual context of video streams, these systems will provide a significantly more efficient and user-friendly way to access, analyze, and manage vast amounts of surveillance data. Beyond conversational interfaces, I plan to investigate the development of autonomous AI agents for video surveillance that can proactively monitor scenes for anomalous activities, learn patterns of normal behavior, and generate timely and context-aware alerts for potential security threats. These agents could also be designed to perform more complex tasks such as intelligent multi-camera tracking of suspicious individuals, prediction of potential security breaches based on observed behaviors and historical data, and even autonomous initiation of pre-defined response protocols in certain critical situations. Significant research challenges include developing robust methods for understanding complex temporal and spatial relationships in dynamic video scenes, ensuring the privacy and ethical use of such advanced surveillance technologies, and creating AI agents that are both reliable and adaptable to changing environmental conditions and evolving security threats.

Transforming Marine Vision with Language-Enabled Understanding and Autonomous Exploration Agents: In the challenging and increasingly important field of underwater computer vision, my future research will focus on leveraging the unique capabilities of LLMs and VLMs to unlock new possibilities for understanding, monitoring, and protecting our oceans. I plan to explore the development of VLMs that can learn to associate visual features of diverse marine life and complex underwater habitats with rich textual descriptions, ecological knowledge, and scientific data. This could enable transformative applications such as automated identification of marine species based on visual input and natural language

queries about their characteristics or behavior, as well as the generation of detailed reports on the health, biodiversity, and ecological status of underwater environments. Furthermore, I envision developing intelligent AI agents for autonomous underwater vehicles (AUVs) that can understand high-level mission commands expressed in natural language, such as "*Survey the coral reef in this designated area and report on the presence of any signs of coral bleaching or* plastic *debris*," or "*Track the migration patterns of this specific species of fish and provide visual and textual summaries of their movements*." These agents would need to interpret noisy and degraded visual data in real-time to navigate, identify relevant features, and generate comprehensive textual summaries of their observations and findings. This will significantly enhance our ability to explore and monitor underwater environments, providing invaluable data for marine biology, conservation efforts, and sustainable resource management. Key research challenges include overcoming the inherent limitations of underwater image quality, developing robust methods for visual understanding in these challenging conditions, and creating AI agents that can operate autonomously and reliably in complex underwater environments while effectively integrating visual and textual information for comprehensive environmental understanding.

Towards a Leading Computer Vision Centre of Excellence at the University of Warwick: To realize this ambitious and impactful research vision, I am committed to actively seeking substantial external funding from prestigious UK research councils such as the EPSRC, MRC, and Innovate UK. The University of Warwick, with its strong interdisciplinary research culture, its commitment to research excellence, and its strategic location within a vibrant technological ecosystem, provides an ideal environment for me to establish a leading research lab in computer vision. My immediate priorities upon joining Warwick will be to recruit talented and motivated postgraduate students, secure funding for state-of-the-art computational resources and experimental equipment, and foster a collaborative, innovative, and impactful research culture within my lab. I envision this lab serving as the foundational pillar for the development of a new computer vision Centre of excellence at the University of Warwick. This Centre would aim to bring together researchers from across different disciplines within the university to tackle fundamental and applied challenges in computer vision, fostering interdisciplinary collaborations with departments such as medicine, engineering, life sciences, and environmental science, and attracting further significant research funding, top-tier talent, and impactful industrial partnerships. My long-term vision is to position the University of Warwick as a national and international leader in computer vision research, driving innovation, fostering the next generation of computer vision experts, and contributing significantly to addressing some of the most pressing societal challenges in healthcare, security, and environmental sustainability. My established track record of high-quality research, my clearly articulated and compelling future research vision, my experience in building successful international collaborations, and my strong commitment to securing external funding make me confident in my ability to achieve these goals and make a significant and lasting contribution to the University of Warwick.

Yours Sincerely

Sajid Javed

Dr. Sajid Javed
Assistant Professor
Khalifa University of Science & Technology

Relevant Publications in Pathology Image Analysis

1. S. Albastaki, A. Sohail, I. I. Ganapathi, B. Alawode, A. Khan, ***S. Javed***, N. Werghi, M. Bennamoun, and A. Mahmood, "*Multi-Resolution Pathology-Language Pre-training Model with Text-Guided Visual Representation*", ***Computer Vision and Pattern Recognition (CVPR), 2025.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
2. ***S. Javed***, A. Mahmood, M. Bennamoun, "*CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment*", ***Computer Vision and Pattern Recognition (CVPR), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
3. ***S. Javed***, A. Mahmood, T. Qaiser, and N. Rajpoot, "*Unsupervised Mutual Transformer Learning for Multi-Gigapixel Whole Slide Image Classification*", ***Medical Image Analysis (MEDIA), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
4. T. Hassan, Z. Li, ***S. Javed***, J. Dias, and N. Werghi, "*Neural Graph Refinement for Robust Nuclei Recognition in Histopathological Landscape*", ***Transactions on Image Processing (TIP), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
5. T. Mahbub, A. S. Obeid***, S. Javed***, J. M.M. Dias, T. Hassan, N. Werghi, "Center-Focused Affinity Loss for Class Imbalanced Histology Image Classification", ***IEEE Journal of Biomedical and Health Informatics (JBHI), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
6. ***S. Javed,*** A. Mahmood, T. Qaiser, and N. Werghi, "*Knowledge Distillation in Histology Landscape by Multi-Layer Features Supervision*", ***IEEE Journal of Biomedical and Health Informatics (JBHI), 2023***
7. Taimur Hassan, ***S. Javed***, A. Mahmood, N. Werghi, and Nasir Rajpoot, "Nucleus Classification in Histology Image Using Message Passing Networks", ***Medical Image Analysis (MEDIA), 2022.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
8. ***S. Javed***, A. Mahmood, N. Rajpoot, J. Dias, and N. Werghi, "*Spatially Constraint Context-Aware Hierarchical Deep Correlation Filters for Nucleus Detection in Histology Images*", ***Medical Image Analysis (MEDIA), 2021.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
9. ***S. Javed***, A. Mahmood, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "*Multiplex Cellular Communities for Tissue Phenotyping in Colon Cancer Histology Images*", ***Transactions on Image Processing (TIP), 2020.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
10. ***S. Javed***, A. Mahmood, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "*Cellular Community Detection for Tissue Phenotyping in Colon Cancer Histology Images*", ***Medical Image Analysis (MEDIA)***, **2020.** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$

Relevant Publications in Computer Vision

11. B. Alawode and ***S. Javed***, "*Learning Spaital-Temporal Robust Tensor Sparce RPCA for Background Subtraction*", ***Transactions on Neural Networks and Learning Systems (TNNLS), 2025.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
12. I. I. Ganapathi, F. A. Dharejo, ***S. Javed***, S. S. Ali, and N. Werghi, "*Unsupervised Dual Transformer Learning for 3D Textured Surface Segmentation*", ***Transactions on Neural Networks and Learning Systems (TNNLS), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
13. Y. Alkendi, R. Azzam, A. Ayyad, ***S. Javed***, L. Seneviratne, and Y. Zweiri, "*Neuromorphic Camera Denoising Using Graph Neural Network-Driven Transformers*", ***Transactions on Neural Networks and Learning Systems (TNNLS), 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$

14. F. Ali, N. Werghi, and ***S. Javed***, "*SwinWave-Transform: Texture and Detail-Preserving Lightweight Network for Underwater Image Super-Resolution*", ***Information Fusion, 2024.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
15. X. Zhang, A. Ghimire, J. Dias, ***S. Javed***, and N. Werghi, "*Robot-Person Tracking in a Uniform Appearance Scenarios: A New Dataset and Challenges*", ***Transactions on Human Machine Systems (THMS)***, ***2023.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
16. ***S. Javed***, M. Danelljan, F. S. Khan, M. H. Khan, M. Felsberg, and J. Matas, "*Visual Object Tracking with Discriminative Filters and Siamese Networks: A Survey and Outlook*", ***Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2022.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
17. ***S. Javed***, A. Mahmood, J. Dias, L. Seneviratne, and N. Werghi, "Hierarchical Spatiotemporal Graph Regularized Discriminative Correlation Filters for Visual Object Tracking", ***Transactions on Cybernetics (TCyb), 2022.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
18. ***S. Javed***, J. Ziuologa, T. Bouwmans, "*Graph Moving Object Segmentation*", ***Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
19. ***S. Javed***, A. Mahmood, J. Dias, and N. Werghi, "*Robust Structural Low-Tracking*", ***Transactions on Image Processing (TIP), 2020.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
20. ***S. Javed***, A. Mahmood, T. Bouwmans, S. A. Madeed, and S. K. Jung, "*Moving Object Detection in Complex Scenes using Spatiotemporal Structured Sparse RPCA*", ***Transactions on Image Processing (TIP), 2019.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
21. M. Fiaz, A. Mahmood, ***S. Javed***, and S. K. Jung, "*Handcrafted and Deep Feature Trackers: A Review on Recent Visual Object Tracking Algorithms*", ***ACM Computing Surveys, 2019.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
22. ***S. Javed***, T. Bouwmans, and S. K. Jung, "*Spatiotemporal Low-rank Modeling for Complex Scene Background Initialization*", ***Transactions on Circuits, Systems, and Video Technology (TCSVT), 2018.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$
23. ***S. Javed***, A. Mahmood, T. Bouwmans, and S. K. Jung, "*Background-Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering*", ***Transactions on Image Processing (TIP), 2017.*** $\mathbf{4}^{*}$ $\boldsymbol{REF\ publication}$

Teaching Statement

Dr. Sajid Javed

My best teacher's words still resonate in my ears— "A truly effective teacher doesn't simply present the material, but rather opens the door to new ideas, allowing students to step through and explore the depths on their own!"

Teaching, to me, is not just about being prepared and organized with the material, eloquently delivering lectures and making yourself available outside class, but it is also about demonstrating to students your genuine passion for the subject, thereby kindling their interest. I believe that explaining the context before concept, emphasize learning not scoring, evaluate and re-orient, and encourage open-ended projects improve the holistic learning experience.

Teaching has always been an important part of my academic life. I find teaching, both as a process of learning and the passing on of knowledge, extremely challenging, and yet, equally rewarding. I am excited by the opportunity to be a professor because it holds out the promise and privilege of engaging students in these activities. My goal in teaching is to develop students as critical thinkers, providing them with a solid technical foundation and illuminating the societal implications of the computer science discipline. I want my students to approach and think about problems considering the technical aspect and the broader implications to society. As an assistant professor of computer science at Khalifa University of Science and Technology (KUST), I have had the privilege of teaching a wide array of core computer science courses, including HCI, programming, image processing, computer vision, operating systems, AI, data structures, software engineering, and deep learning. These courses have been offered both at the undergraduate and postgraduate levels, allowing me to engage with students from various academic backgrounds and research interests. In addition, I serve as the **course coordinator** for operating systems, HCI, and deep learning, overseeing course development, implementation, and assessment across different sections. In all courses, I made an intentional effort to provide the students perspectives on how the solid understanding of the basic computer science concepts are essential for analysing complex problems. I have developed my teaching skills and interests through teaching, mentoring students, studying the scholarship of teaching and learning in STEM through my current academic position at KUST, and more recently supporting TAs in developing their teaching skills and adapting to online teaching.

Teaching Philosophy: Core Pedagogy, Student-Centric Approach, and Course Coordination Expertise

My teaching philosophy is based on the following principles: **(i)** *At the heart of my teaching philosophy is the belief that students should not only learn the what and how of computer science but also the why behind the concepts they are studying*. My goal as an educator is to help students develop a deep understanding of the subject matter and provide them with the tools to think critically, solve problems creatively, and approach new challenges with confidence. I aim to cultivate an environment that supports active learning, encourages collaboration, and fosters a sense of ownership over the students' learning process. This approach is not just about transmitting knowledge; it's about inspiring students to explore, question, and apply what they've learned in meaningful ways. **(ii)** *My teaching philosophy emphasizes the integration of theoretical foundations with practical applications, fostering critical thinking, and*

encouraging lifelong learning. I am committed to creating an inclusive and supportive learning environment where all students, regardless of their backgrounds or prior experience, can thrive. I strive to provide students with not only the technical knowledge and skills necessary for success in computer science but also the intellectual curiosity and creativity that will enable them to become leaders and innovators in their respective fields. My student evaluations consistently reflect the effectiveness of my teaching approach, with an average score of 4.5/5. These evaluations have been instrumental in shaping my approach to teaching, as they allow me to assess the effectiveness of my methods and continually refine my pedagogical strategies to meet the needs of my students. **(iii)** *My teaching philosophy is also deeply rooted in the belief that effective computer science education transcends the mere transmission of technical knowledge.* It necessitates the cultivation of critical thinking, problem-solving abilities, and a profound appreciation for the discipline's transformative impact. As an Assistant Professor of Computer Science at KUST, I have consistently prioritized *student-centered learning*, resulting in an **average student evaluation of 4.5/5** across a diverse spectrum of courses. I am a strong advocate for experiential learning, where students actively engage with concepts through hands-on projects, real-world case studies, and collaborative problem-solving. For example, in my data structures and algorithms courses, I emphasize the practical application of theoretical concepts by challenging students to implement efficient solutions for real-world problems. In deep learning, students design and train neural networks to solve complex tasks, fostering a deeper understanding of the field's practical applications.

Comprehensive Teaching Experience, Curriculum Development, and Course Coordination Impact

My extensive experience teaching a broad spectrum of core computer science courses has provided me with a deep understanding of curriculum development and pedagogical best practices.

- In ***Programming,*** I focus on building a strong foundation in fundamental concepts, emphasizing *problem-solving* and *algorithmic thinking*. I introduce students to various programming paradigms and languages, equipping them with the versatility to tackle diverse challenges.
- In ***Data Structures and Algorithms***, I emphasize the importance of efficiency and scalability. Students learn to analyze the performance of different data structures and algorithms, developing the ability to design and implement efficient solutions for complex problems.
- In ***Operating systems***, I delve into the intricacies of system architecture, process management, and memory allocation, providing students with a comprehensive understanding of how software interacts with hardware.
- ***AI and deep learning*** have allowed me to explore the exciting frontier of machine intelligence. I introduce students to the fundamental principles of machine learning, including supervised, unsupervised, and reinforcement learning. I also cover advanced topics such as DNNs, CNNs, and RNNs, providing students with the tools to build and deploy intelligent systems.
- ***Image processing and Computer Vision*** allow students to explore the intersection of computer science and visual perception, applying computational techniques to analyze and interpret images and videos.
- ***HC*I** provides me with a good understanding of the importance of human-computer interactions. I stress the importance of user-centered design and iterative development.

As **course coordinator** for operating systems, HCI, and deep learning, I have significantly contributed to the improvement of these courses. I have implemented innovative teaching strategies, such as flipped classrooms and project-based learning, which have resulted in increased student engagement and

improved learning outcomes. I have also developed comprehensive course materials, including lecture notes, assignments, and projects, ensuring consistency and quality across all sections of the courses. I have also been responsible for managing teaching assistants, providing them with guidance and support to ensure effective instruction. To ensure continuous improvement in my teaching, I actively solicit feedback from students and peers. I analyze student evaluations to identify areas for improvement and adapt my teaching strategies accordingly. I also participate in teaching workshops and conferences to stay abreast of the latest pedagogical innovations. I believe in the importance of reflective practice and am committed to continuously refining my teaching methods to enhance student learning outcomes.

Vision for Computer Science Education at the *University of Warwick* and Future Contributions

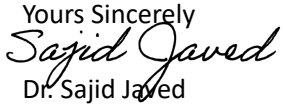
Joining the faculty at the Warwick University presents an exciting opportunity to contribute to the advancement of computer science education and research. My vision for teaching computer science course is to create a dynamic and innovative learning environment that empowers students to become leaders in the field. I propose to develop new courses and modules that reflect the latest advancements in computer science. This could include courses on topics such as:

- **Advance Large Language Models for Computer Scientists:** Focusing on modern LLM architectures, their design and implementation.
- **Explainable AI and Responsible Machine Learning:** Addressing the critical issues of transparency, fairness, and accountability in AI systems.
- **Advanced Topics in Computer Vision and Robotics:** Exploring the intersection of computer vision with autonomous systems, including robotics and autonomous vehicles.

I also plan to foster strong industry collaborations, providing students with opportunities to work on real-world projects and gain practical experience. This could involve inviting industry experts to give guest lectures, organizing industry-sponsored workshops, and facilitating internships. I would also seek to establish a computer science research group where students can participate in cutting-edge research projects. Furthermore, I am committed to promoting diversity and inclusion in computer science education. I will actively seek to create a welcoming and supportive environment for students from all backgrounds. I will also work to increase the representation of women and underrepresented minorities in the field by organizing outreach activities and mentoring programs. I am eager to collaborate with faculty members across different departments to develop interdisciplinary projects that leverage the power of computer science. This could include projects in areas such as bioinformatics, digital humanities, and environmental science. I believe that computer science has the potential to transform many fields, and I am excited to contribute to its advancement at Warwick.

In summary, my teaching philosophy, extensive experience, and vision for the future, coupled with my course coordination expertise, align perfectly with the goals of the *University of Warwick*. I am confident that I can make a significant contribution to the university's computer science program and inspire the next generation of computer scientists.

Yours Sincerely



Dr. Sajid Javed
Assistant Professor, Khalifa University of Science & Technology

Dr. Sajid Javed

Assistant Professor of Computer Vision
Khalifa University of Science and Technology, UAE (QS Ranking 181).
Phone: +971-50-3308507
Email: sajid.javed@ku.ac.ae
Google Scholar Citations: 4,300+ as of 25-Apr-2025 (h-index 28, i10-index 59)
Web: https://scholar.google.com/citations?user=6qvbEhUAAAAJ&hl=en

Professional Summary

I am an active researcher, educator, and innovator in **computer vision**, **machine learning**, and **generative AI**. I have a strong track record of cutting-edge research, interdisciplinary collaborations, and high-impact factor publications. My expertise spans deep learning, image and video analysis, pathology image analysis, generative models (e.g., GANs and diffusion models), and AI-driven visual perception, with applications in underwater and marine vision, medical imaging, and video surveillance.

I have a proven track record of securing competitive research funding and leading large-scale projects, including those integrating AI, big data analytics, and computational modeling for complex visual tasks. I have extensive experience **mentoring PhD students and early career researchers**, fostering innovation, and developing curriculum in computer vision, AI, and data science. I am a strong advocator for AI-driven scientific discovery and real-world applications, committed to advancing state-of-the-art techniques in generative AI for visual content synthesis, anomaly detection, and domain adaptation. I actively collaborate with industry partners, government agencies, and international research institutions to translate research into practical solutions. I am dedicated to enhancing academic excellence, contributing to institutional growth, and shaping the future of computer vision and AI through impactful research, teaching, and leadership.

My research to date has focused on developing innovative solutions within the domains of **pathology image analysis**, **resilient video surveillance**, and **marine vision**. My publications are published in the top-notch computer vision venues, including ***TPAMI, CVPR, TNNLS, MEDIA, TCSVT, TIP,*** **and** ***TCyb*****,** etc.

Work History

August 2021-Current: Assistant Professor of Computer Vision

Khalifa University of Science and Technology, UAE (QS Ranking 201)

I conduct research, teaching, and other administrative activities in the computer science department. I am a member of the KUCARS Research Centre. I teach core computer science courses both at the undergraduate and postgraduate levels. This includes operating systems, HCI, machine learning, computer vision, and deep learning systems design**.** I am supervising **6 PhD students and 10 MSc students.** My computer vision group develops novel deep learning techniques for image and video analysis and medical image analysis applications using next-generation machine learning methods such as vision transformers and GenAI. I regularly publish high quality and impact publications in computer vision venues including **TPAMI, TIP, TNNLS, TCyb, and CVPR**. In my professional capacity, my main role involves:

- Assisting with various departmental duties and providing academic support to Professors and other staff.
- Recruiting, training, and mentoring new TAs and other junior staff.
- Conducting research and publishing papers in academic journals.
- Representing the university at conferences and delivering presentations when necessary.
- Teaching and supervising undergraduate and graduate students.
- Providing demonstrations and supervising experiments and investigations.
- Answering questions in class or via email or telephone.
- Providing Professors and Department Heads with feedback on student progress.
- Writing proposals to secure funding for research.
- Attending faculty and departmental meetings and voicing concerns or providing suggestions for improvement**.**

Feb 2019 - August 2021: Research Fellow

Khalifa University of Science and Technology, UAE (QS Ranking 201)

I developed a novel deep learning algorithms for video object tracking in the wild. This included deep correlation filters learning in an end-to-end manner, using pre-trained models, and detection strategies for long term. The second part of the project included multiple object tracking and person re-identification problems. Within this resilient video surveillance project, my main role included:

- Designed research projects and alternative approaches and discussed results with research director of the center.

- Maintained accurate records of research findings and provided statistical analysis of data results.
- Contributed to and actively participated in research conception, design and execution to address defined problems.
- Leveraged interpersonal and communication skills to mentor PhD, graduate and undergraduate students.
- Developed unique research methodologies into MBZIRC challenge-1 solution.
- Created and executed scenarios for analysis projects and interpreted results to address relevant questions.
- Collaborated with all KUCARS multidisciplinary team members to accomplish research goals.
- Published research results in peer-reviewed prestigious journals and presented at seminars and internal meetings.
- Conducted independent research and development to attain short and long-term objectives.
- Liaised with research director to review grants and manuscripts to develop research projects.
- Pursued independent and complementary research interests to achieve project objectives.
- Supervised and mentored 18 students in conducting targeted experiments and gathering data and analyzing results to obtain valuable and actionable conclusions.
- Provided a novel solution for challenge-1 to the KU team for participating in Muhammad Bin Zayed International Robotics Challenge, 2020.

Oct 2017 – Dec 2018: Research Fellow

***University of Warwick, United Kingdom** (QS Ranking 69)*

I worked on novel micro-community analytics for histology landscapes project supervised with **Prof. Nasir Rajpoot**. The main aim of the project was to analyze the social-behavior of the cancer cells using multi-giga pixel histology images funded by the Medical Research Council (MRC), UK. The project main projective was to estimate cellular level tissue communities for better cancer grading and diagnoses. We have developed a novel tissue phenotyping methods and nucleus detection methods using deep neural networks. During this position, I have contributed in the following research-related activities:

- Designed research projects and alternative approaches and discussed results with supervisor.
- Maintained accurate records of research findings and provided statistical analysis of data results.
- Contributed to and actively participated in research conception, design and execution to address defined problems.

- Collaborated with computer science department multidisciplinary team members to accomplish research goals.
- Conducted independent cancer image analysis research and development to attain short and long-term objectives.
- Upheld all university and regulatory requirements for documentation of findings.
- Authored professional scientific papers for publishing in peer-reviewed journals.
- Collaborated with a team of expert pathologists at University Hospital Coventry and Warwickshire for domain knowledge discussion, tissue image annotations, and results verifications.
- Travelled outside to present project outcomes to external project consortiums.

Education Details

Sept 2014-Aug 2017: Ph.D.: Computer Science and Engineering

Kyungpook National University, South Korea

- Supervised by Soon Ki Jung and Thierry Bouwmans.
- Thesis: ***Structured Low-Rank Robust PCA for Background-Foreground Modelling***.
- Graduated with 4.2 GPA.
- Coursework includes Advanced Computer Vision, Advanced Machine Learning, Artificial Intelligence, and Augmented Reality.

Aug 2012 - Aug 2014: Master in Computer Science and Engineering

Kyungpook National University, South Korea

- Supervised by Soon Ki Jung and Thierry Bouwmans.
- Thesis: ***Foreground Object Detection in Video Surveillance for Activity Analysis***.
- Graduated with 4.1 GPA.
- Coursework includes Advanced Computer Vision, Advanced Machine Learning, Artificial Intelligence, and Augmented Reality.

Aug 2007 – July 2010: B.Sc. (Hons): Computer Science

University of Hertfordshire, United Kingdom

- Graduated with 2.1 UK degree class.
- Dissertation: ***Hospital patients information system***.

- Coursework includes Advanced Programming, Advanced Databases, Computer Networks, and Image Processing Human Computer Interaction.

Secured Project Funding

- **Project Title:** Towards the Development of AI System for Detecting Early-stage Lung Cancer Using Histopathology Images of UAE Patients
    - **Principle Investigator:** Sajid Javed
    - **Source:** ASPIRE
    - **Total Award:** 2.5M AED (equivalent to 600K GPB)
    - **Postdocs:** 01
    - **PhDs:** 02
    - **Duration:** 2 Years (August 2023-2025)

- **Project Title:** Artificial Intelligence for Ocean Surveillance
    - **Principle Investigator:** Sajid Javed
    - **Source:** Faculty Startup Funds from Khalifa University
    - **Total Award:** 1.1M AED (equivalent to 250K GBP)
    - **Postdocs:** 01
    - **PhDs:** 01
    - **Duration:** 2 Years (August 2021-2023)

- **Project Title:** Development of Traffic Surveillance System for Abu Dhabi Police
    - **Principle Investigator:** Sajid Javed
    - **Source:** ADNEC
    - **Total Award:** 2.5M AED (equivalent to 600K GBP)
    - **Duration:** 2 Years (August 2023-2025)

Key Publications

1. S. Albastaki, A. Sohail, I. I. Ganapathi, B. Alawode, A. Khan, **S. Javed**, N. Werghi, M. Bennamoun, and A. Mahmood, Multi-Resolution Pathology-Language Pre-training Model with Text-Guided Visual Representation, ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*** 2025. $\mathbf{4}^{*}$ ***REF publication.***

2. B. Alawode and **S. Javed**, *Learning Spaital-Temporal Robust Tensor Sparce RPCA for Background Subtraction*, ***IEEE Transactions on Neural Networks and Learning Systems***, 2025. **(Top 1%)**. $\mathbf{4}^*$ ***REF publication.***
3. I. I. Ganapathi, F. A. Dharejo, **S. Javed**, S. S. Ali, and N. Werghi, *Unsupervised Dual Transformer Learning for 3D Textured Surface Segmentation*, ***IEEE Transactions on Neural Networks and Learning Systems***, 2024. **(Top 1%)**. $\mathbf{4}^*$ ***REF publication.***
4. **S. Javed**, A. Mahmood, M. Bennamoun, CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment, ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*** 2024. $\mathbf{4}^*$ ***REF publication.***
5. **S. Javed**, M. Danelljan, F. S. Khan, M. H. Khan, M. Felsberg, and J. Matas, **Visual Object Tracking with Discriminative Filters and Siamese Networks: A Survey and Outlook**, ***IEEE Transactions on Patterns Analysis and Machine Intelligence***, 2022. $\mathbf{4}^*$ ***REF publication.***
6. **S. Javed**, J. Ziuologa, T. Bouwmans, **Graph Moving Object Segmentation**, ***IEEE Transactions on Patterns Analysis and Machine Intelligence (IEEE T-PAMI*)**, 2020. $\mathbf{4}^*$ ***REF publication.***
7. **S. Javed**, A. Mahmood, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "**Multiplex Cellular Communities for Tissue Phenotyping in Colon Cancer Histology Images**", ***IEEE Transactions on Image Processing (IEEE T-IP),* 2020.** $\mathbf{4}^*$ ***REF publication.***
8. **S. Javed**, A. Mahmood, T. Bouwmans, and S. K. Jung, "**Background-Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering**", ***IEEE Transactions on Image Processing (IEEE T-IP),*** *2017*. $\mathbf{4}^*$ ***REF publication.***
9. **S. Javed**, A. Mahmood, T. Bouwmans, S. A. Madeed, and S. K. Jung, "**Moving Object Detection in Complex Scenes using Spatiotemporal Structured Sparse RPCA**", ***IEEE Transactions on Image Processing (IEEE T-IP),*** *2019*. $\mathbf{4}^*$ ***REF publication.***

Journal Articles

- B. Alawode and **S. Javed**, *Learning Spaital-Temporal Robust Tensor Sparce RPCA for Background Subtraction*, ***IEEE Transactions on Neural Networks and Learning Systems***, 2025. **(Top 1%)**. $\mathbf{4}^*$ ***REF publication.***
- M. Alansari and **S. Javed,** EfficientFaceV2S: A Lightweight Model and Benchmarking Approach for Drone Captured Face Recognition, ***Expert Systems and Applications***, 2025.
- I. Ganapathi, F. A. Dharejo, **S. Javed**, S. S. Ali, and N. Werghi, *Unsupervised Dual Transformer Learning for 3D Textured Surface Segmentation*, ***IEEE Transactions on Neural Networks and Learning Systems***, 2024. **(Top 1%)**. $\mathbf{4}^*$ ***REF publication.***
- Y. Alkendi, R. Azzam, A. Ayyad, **S. Javed**, L. Seneviratne, and Y. Zweiri, *Neuromorphic Camera Denoising Using Graph Neural Network-Driven Transformers*, ***IEEE Transactions on Neural Networks and Learning Systems***, 2024. **(Top 1%)**. $\mathbf{4}^*$ ***REF publication***
- **S. Javed**, A. Mahmood, T. Qaiser, and N. Rajpoot, Unsupervised Mutual Transformer Learning for Multi-Gigapixel Whole Slide Image Classification**, Medical Image Analysis**, 2024. [**Q1 Journal Top 1%**]. $\mathbf{4}^*$ ***REF publication***

- F. Ali, N. Werghi, and **S. Javed**, SwinWave-Transform: Texture and Detail-Preserving Lightweight Network for Underwater Image Super-Resolution***, Information Fusion***, 2024. [**Q1 Journal Top 1%**]. 4^* ***REF publication***
- M. Alansari, O. A. Hay, S. Alansari, S. Javed, A. Shoufan, Y. Zweiri, N. Werghi, Drone-Person Tracking in Uniform Appearance Crowd: A New Dataset, ***Nature Scientific Data***, 2024. [**Q1 Journal Top 1%**]. 4^* ***REF publication***
- T. Hassan, Z. Li, **S. Javed**, J. Dias, and N. Werghi, Neural Graph Refinement for Robust Nuclei Recognition in Histopathological Landscape, ***IEEE Transactions on Image Processing***, 2024. [**Q1 Journal Top 2%**]. 4^* ***REF publication***
- N. Nida, M. H. Yousaf, A. Istaza, **S. Javed**, S. A. Velastin, Spatial Deep Feature Augmentation Technique for FER using Genetic Algorithm, ***Neural Computing and Applications***, 2024. [**Q1 Journal Top 12%**].
- L. Wang, Y. Shi, G. Mao, F. A. Dharejo**, S. Javed**, and M. Alathbah, Consumer-Centric Insights into Resillient Small Object Detection: SCIoU Loss and Recursive Transformer Network, ***IEEE Transactions on Consumer Electronics***, 2024. [**Q1 Journal Top 8%**].
- Y. Alkendi, R. Azzam, **S. Javed**, L. Seneviratne, and Y. Zweiri, *Neuromorphic Vision-based Motion Segmentation with Graph Transformer Neural Network*, ***IEEE Transactions on Multimedia***, 2024. **(Top 4%)**. 4^* ***REF publication***
- I. Ehtesham, K. S. Ullah, **S. Javed**, B. Moyo, Y. Zweiri, and A. Yusra, *Multi-Scale Feature Reconstruction Network for Industrial Anomaly Detection*, ***Knowledge-Based Systems***, 2024. (**Top 6%**)
- T. Mahbub, A. S. Obeid, **S. Javed**, J. M.M. Dias, T. Hassan, N. Werghi, Center-Focused Affinity Loss for Class Imbalanced Histology Image Classification, ***IEEE Journal of Biomedical and Health Informatics***, 2024. [**Q1 Journal Top 6%**].
- F. A. Dharejo, B. Alawode, I. I. Ganapathi, L. Wang, A. Mian, R. Timofte, and **S. Javed**, *Multi-Distillation Underwater Image Super-Resolution via Wavelet Transform*, ***IEEE ACCESS***, 2024 (**Top 8%**).
- M. Alansari, A. Ahmed, K. Alnuaimi, D. Velayudhan, T. Hassan, **S. Javed**, M. Bennamoun, and N. Werghi, *Multi-Scale Hierarchical Contour Framework for Detecting Cluttered Threats in Baggage Security*, ***IEEE ACCESS***, 2024. **(Top 8%)**.
- M. Radi, P. Li, S. Boumaraf, J. Dias, N. Werghi, H. Karki, and **S. Javed**, *AI-Enhanced Gas Flares Remote Sensing and Visual Inspection: Trends and Challenges*, ***IEEE ACCESS***, 2024. **(Top 8%)**.
- A. B. Bakht, **S. Javed**, S. Q. Gillani, H. Karki, M. Muneeb, and N. Werghi, DeepBLS: Deep Feature-Based Broad Learning System for Tissue Phenotyping in Colorectal Cancer WSIs, ***Journal of Digital Imaging***, 2023. [**Q1 Journal Top 10%**].
- M.Y. Alansari, O. A. Hay, **S. Javed**, A. Shoufan, Y. Zweiri, and N. Werghi, GhostFaceNets: Lightweight Face Recognition Model from Cheap Operations, ***IEEE ACCESS***, 2023. [**Q1 Journal Top 8%**].

- M. Ummar, F. Ali, T. Mahbub, and **S. Javed**, UwTGAN: Underwater Image Enhancement Using a Window-based Transformer Generative Adversarial Network, ***Engineering Applications of Artificial Intelligence (EAAI)***, 2023. [**Q1 Journal Top 7%**]. 4^* ***REF publication***
- K. Ganesan, I. I. Ganapathi**, S. Javed**, N. Werghi, Multimodal Hybrid Fusion in 3D Ear Recognition, ***Applied Intelligence***, 2023. [**Q1 Journal Top 20%**].
- X. Zhang, A. Ghimire, J. Dias, **S. Javed**, and N. Werghi, "**Robot-Person Tracking in a Uniform Appearance Scenarios: A New Dataset and Challenges**", ***IEEE Transactions on Human Machine Systems***, **[Impact Factor: 4.13],** 2023. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, T. Qaiser, and N. Werghi, **Knowledge Distillation in Histology Landscape by Multi-Layer Features Supervision**, ***IEEE Journal of Biomedical and Health Informatics,* [Impact Factor: 7.021]**, 2023.
- **S. Javed**, M. Danelljan, F. S. Khan, M. H. Khan, M. Felsberg, and J. Matas, **Visual Object Tracking with Discriminative Filters and Siamese Networks: A Survey and Outlook**, ***IEEE Transactions on Patterns Analysis and Machine Intelligence***, **[Impact Factor: 24.134]**, 2022 4^* ***REF publication***.
- Taimur Hassan, **S. Javed**, A. Mahmood, N. Werghi, and Nasir Rajpoot, **Nucleus Classification in Histology Image Using Message Passing Networks**, ***Medical Image Analysis,* [Impact Factor: 13.82]**, 2022. 4^* ***REF publication***
- X. Huang, R. Azzam, **S. Javed**, D. Gan, L. Seneviratne, A. Abusafieh, and Y. Zweiri, **CM-UNET: ConvMixer UNET for Segmentation of Unknown Objects in Cluttered Scenes**, ***IEEE ACCESS***, 2022.
- K. Ganesan, I. Ganapathi, **S. Javed**, and N. Werghi, **Multimodal Hybrid Features in 3D Ear Recognition**, **Applied Intelligence**, 2022.
- Iyyakutti Iyappan Ganapathi, Syed Sadaf Ali, Arif Mahmood, **Sajid Javed**, Ngoc-Son Vu, and Naoufel Werghi, **Learning to Localize Forgery Using End-to-End Attention Network**, **Neuro-computing [Impact Factor: 5.77]**, 2022
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, **Multi-level Feature Fusion for Nucleus Detection in Histopathology Images Using Correlation Filters**, ***Computers in Biology and Medicine* [Impact Factor: 6.69]**, 2022.
- Sufian Badawi, Muhammad Mozam Fraz, Muhammad Shehzad, Imran Mahmood, **Sajid Javed**, Emad Mosalam, Ajay Kamath Nileshwar, **Detection and Grading of Hypertensive Retinopathy Using Vessels Tortousity and Arteriovenous Ratio**, ***Journal of Digital Imaging* [Impact Factor: 4.90]**, 2022.
- **S. Javed**, A. Mahmood, I. Ullah, T. Bouwmans, M. Khonji, J. Dias, and N. Werghi, **A Novel Algorithm Based on a Common Subspace Fusion for Visual Object Tracking**, ***IEEE Access* [Impact Factor: 3.476]**, 2022.
- **S. Javed**, A. Mahmood, J. Dias, L. Seneviratne, and N. Werghi, **Hierarchical Spatiotemporal Graph Regularized Discriminative Correlation Filters for Visual Object Tracking**, ***IEEE Transactions on Cybernetics*** (***IEEE T-Cyb***) **[Impact Factor: 19.11]**, 2022. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, N. Rajpoot, J. Dias, and N. Werghi, **Spatially Constraint Context-Aware Hierarchical Deep Correlation Filters for Nucleus Detection in Histology Images**, ***Medical Image Analysis*** (MEDIA) **[Impact Factor: 13.82]**, 2021. 4^* ***REF publication***

- Xiaxiong Zhang, **Sajid Javed**, Jorge Dias, and Naoufel Werghi, **Person Gender Classification on RGB-D Data with Self-Joint Attention**, ***IEEE Access*** **[Impact Factor: 3.476]**, 2021.
- **S. Javed**, J. Ziuologa, T. Bouwmans, **Graph Moving Object Segmentation**, ***IEEE Transactions on Patterns Analysis and Machine Intelligence (IEEE T-PAMI*****) [Impact Factor: 19.42]**, 2020. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "**Multiplex Cellular Communities for Tissue Phenotyping in Colon Cancer Histology Images**", ***IEEE Transactions on Image Processing (IEEE T-IP)*** **[Impact Factor: 11.04], 2020.** 4^* ***REF publication***
- **S. Javed**, A. Mahmood, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "**Cellular Community Detection for Tissue Phenotyping in Colon Cancer Histology Images**", ***Medical Image Analysis (MEDIA)*** **[Impact Factor: 13.82]**, *2020*. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, "**Robust Structural Low-Tracking**", ***IEEE Transactions on Image Processing (IEEE T-IP)*** **[Impact Factor: 11.04]**, *2020*. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, T. Bouwmans, S. A. Madeed, and S. K. Jung, "**Moving Object Detection in Complex Scenes using Spatiotemporal Structured Sparse RPCA**", ***IEEE Transactions on Image Processing (IEEE T-IP)*** **[Impact Factor: 11.04],** *2019*. 4^* ***REF publication***
- **S. Javed**, T. Bouwmans, and S. K. Jung, "**Spatiotemporal Low-rank Modeling for Complex Scene Background Initialization**", ***IEEE Transactions on Circuits, Systems, and Video Technology (IEEE TCSVT)*** **[Impact Factor: 5.08]**, *2018*. 4^* ***REF publication***
- **S. Javed**, A. Mahmood, T. Bouwmans, and S. K. Jung, "**Background-Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering**", ***IEEE Transactions on Image Processing (IEEE T-IP)*** **[Impact Factor: 11.04],** *2017*. 4^* ***REF publication***
- **S. Javed**, T. Bouwmans, and S. K. Jung, "**Robust Background Subtraction to Global Illumination Changes via Multiple Features Based OR-PCA with MRF**", ***Journal of Electronic Imaging (JEI)*** **[Impact Factor: 0.94],** *2015*.
- T. Bouwmans, **S. Javed**, M. Sultana, and S. K. Jung, "**Deep Neural Network Concepts for Background Subtraction: A Systematic Review and Comparative Evaluation**", ***Neural Networks*** **[Impact Factor: 9.67]**, *2019*. 4^* ***REF publication***
- M. Sultana, A. Mahmood, **S. Javed**, S. K. Jung, "**Unsupervised Deep Context Prediction for Background Estimation and Foreground Segmentation**", ***Machine Vision and Application*** **[Impact Factor: 2.98]**, *2019*.
- M. Fiaz, A. Mahmood, **S. Javed**, and S. K. Jung, "**Handcrafted and Deep Feature Trackers: A Review on Recent Visual Object Tracking Algorithms**", ***ACM Computing Surveys*** **[Impact Factor: 14.32]**, *2019*. 4^* ***REF publication***
- N. Vaswami, T. Bouwmans, **S. Javed**, and P. Narayanamurthy, "**Robust PCA and Robust Subspace Tracking**", ***IEEE Signal Processing Magazine*** **[Impact Factor: 15.20]**, *2018*. 4^* ***REF publication***
- T. Bouwmans, **S. Javed**, H. Zhang, Z. Lin, and R. Otazo, "**On the Applications of Robust PCA in Image and Video Processing and 3D Computer Vision**", ***IEEE Proceedings*** **[Impact Factor: 14.91]**, *2018*. 4^* ***REF publication***

- M. Sultana, **S. Javed**, N. Bhatti, S. K. Jung, "**Local Binary Pattern Variants based Adaptive Texture Features Analysis for Posed and Non-Posed Facial Expression Recognition**", **Journal of Electronic Imaging [Impact Factor: 0.94]**, *2017.*
- T. Bouwmans, A. Sobral, **S. Javed**, S. K. Jung, E. H. Z. Zah, "**Background/Foreground Separation via Decomposition into Low Rank Plus Additive Matrices: A review for a Comparative Evaluation with a Large-Scale Dataset**", ***Computer Science Review*** **[Impact Factor: 8.75]**, *2016.* $\mathbf{4}^*$ ***REF publication***

Conference Papers

- S. Albastaki, A. Sohail, I. I. Ganapathi, B. Alawode, A. Khan, **S. Javed**, N. Werghi, M. Bennamoun, and A. Mahmood, Multi-Resolution Pathology-Language Pre-training Model with Text-Guided Visual Representation, ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*** 2025. $\mathbf{4}^*$ ***REF publication***
- **S. Javed**, A. Mahmood, M. Bennamoun, CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment, ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*** 2024. $\mathbf{4}^*$ ***REF publication***
- M. Yonathan, M. Alansari, and **S. Javed**, *Text-Guided Multi-Modal Fusion for Underwater Visual Tracking*, ***Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS),*** 2024.
- J. Liu, D. Zhu, and **S. Javed**, *Visual-Language Alignment for Background Subtraction*, ***IEEE International Conference on Multimedia and Expo (ICME),*** 2024. $\mathbf{4}^*$ ***REF publication***
- M. Alansari, H. Alremeithi, S. Alansari, N. Werghi, and **S. Javed**, *Performance Analysis of Synthetic Events Via Visual Object Trackers*, ***Science and Information Conference (LNCS),*** 2024.
- I. I. Ganapathi, F. A. Dharejo, **S. Javed**, A. S. Sadaf, and N. Werghi, *3D-TexSeg: Unsupervised Segmentation of 3D Texture Using Mutual Transformer Learning*, ***International Conference on 3D Vision (3DV),*** 2024.
- M. Alansari, H. Alremeithi, H. Bilal, S. Alansari, J. Dias, M. Khonji, N. Werghi, and **S. Javed**, Vision-Perceptual Transformer Network for Semantic Scene Understanding, ***Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)***, 2024.
- J.H. Giraldo, **S. Javed**, A. Mahmood, F.D. Malliaros, and T. Bouwmans, Higher-order Sparse Convolutions in Graph Neural Networks, ***IEEE International Conference on Acoustics, Speech, and Speech Processing*** (ICASSP), 2023. [**Q1 Conference**].
- M. Nagy, M. Khonji, J. Dias, and **S. Javed**, DFR-FastMOT: Detection Failure Resistant Tracker for Fast Multi-Object Tracking Based on Sensor Fusion, ***2023 IEEE International Conference on Robotics and Automation (ICRA)***, 2023. [**Q1 Conference**]. $\mathbf{4}^*$ ***REF publication***

- I.I. Ganapathi, S. S. Ali, M. Owais, N. Gour, **S. Javed**, and N. Werghi, Facet-Level Segmentation of 3D Textures on Cultural Heritage Objects, ***2023 IEEE International Conference on Image Processing*** (ICIP), 2023. [**Q1 Conference**].
- M.A.Radi, P. Li, H. Karki, N. Werghi, **S. Javed**, N. Werghi, Multi-view Inspection of Flare Stacks Operations using a Vision-Controlled Autonomous UAV, ***IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society***, 2023. [**Q1 Conference**].
- M.A. Hafeez, S. Javed, M. Madden, and I. Ullah, Unsupervised End-to-End Transformer based Approach for Video Anomaly Detection, ***38th International Conference on Image and Vision Computing*** (IVCZN), 2023. [**Q1 Conference**].
- A. Khan, A. Haq, and S. Javed, Accurate and Efficient Urban Street Tree Inventory with Deep Learning on Mobile Phone Imagery, ***Digital Image Computing: Techniques and Applications*** (DICTA), 2023.[**Q1 Conference**].
- B. Alawode, Y. Guo, M. Ummar, N. Werghi, J. Dias, A. Mian, **S. Javed**, **UTB180: A High-quality Benchmark for Underwater Tracking**, **Proceedings of the Asian Conference on Computer Vision**, 2022. 4^* ***REF publication***
- Iyyakutti Iyappan Ganapathi, **Sajid Javed**, and Naoufel Werghi, **Graph Based Texture Classification**, **International Conference on Virtual Reality**, 2022.
- Iyyakutti Iyappan Ganapathi, **Sajid Javed**, Taimur Hassan, and Naoufel Werghi, **Detecting 3D Texture on Cultural Heritage Artifacts**, **International Conference on Pattern Recognition**, 2022. 4^* ***REF publication***
- V. Sudevan, N. Mankovski, S. Javed, H. Karki, G.D. Massi, and J. Dias, **Multi-sensor Fusion for Marine Infrastructure Inspection and Safety**, **OCEANS**, 2022.
- M. Radi, H. Karki, N. Werghi, S. Javed, and J. Dias, **Video Analysis of Flare Stacks with an Autonomous Low-cost Aerial System**, **ADIPEC**, 2022.
- M. Radi, H. Karki, N. Werghi, S. Javed, and J. Dias, **Vision-based Inspection of Flare Stacks Operating Using a Visual Servoing Controlled Autonomous Unmanned Aerial Vehicle**, **IECON2022-48th Annual Conference of the IEEE Industrial Electronics Society**, 2022.
- Muaz Al Radi, Hamad Karki, Naoufel Werghi, **Sajid Javed**, and Jorge Dias, **Autonomous Inspection of Flare Stacks using an Unmanned Aerial System**, Book Chapter Springer, 2022
- Taslim Mahbub, Ahmad Obeid, **Sajid Javed**, Jorge Dias, and Naoufel Werghi, **Class-Balanced Affinity Loss for Highly Imbalanaced Tissue Classification in Computational Pathology**, **International Conference on Pattern Recognition**, 2022.
- Ahmad Obeid, Taslim Mahbub, **Sajid Javed**, and Naoufel Werghi, **NucDETR: End-to-End Transformer for Nucleus Detection in Histopathology Images**, **International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI)**, 2022.
- Adarsh Ghimire, Xiaoxiong Zhang, **Sajid Javed**, Jorge Dias, and Naoufel Werghi, **Robot Person Following in Uniform Crowd Environment**, **International Conference on Robotics Automation (ICRA**), 2022.
- Divya Velayudhan, Naoufel Werghi, and **Sajid Javed**, **Underwater Fish Tracking-by-Detection**, **International Conference on Pattern Recognition**, 2022.

- Abderrahmene Boudiaf, **Sajid Javed**, Jorge Dias, Giulia De Masi, and Naoufel Werghi, **Underwater Image Enhancement Using Pre-Trained Transformer**, ***International Conference on Image Analysis and Processing***, 2022.
- A. Bakht, **S. Javed**, H. AlMazrouqi, A. Khandokar and N. Werghi**, "**Colorectal Cancer Tissue Classification Using Semi-Supervised Hypergraph Convolutional Network**", IEEE ISBI,** 2021**.**
- **S. Javed**, X. Zhang, J. Dias, L. Seneviratne, and N. Werghi, "Spatial Graph Regularized Correlation Filters for Visual Object Tracking", **SoCPaR**, 2020.
- A. Bakht, **S. Javed**, R. Dina, H. AlMazrouqi, and A. Khandokar**, "**Thyroid Nodule Cell Classification In Cytology Images Using Transfer Learning Approach**", SoCPaR,** 2020**.**
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, "CS-RPCA: Clustered Sparse RPCA for Moving Object Detection", ***IEEE ICIP***, 2020.
- **S. Javed**, X. Zhang, L. Seneviratne, J. Dias, and N. Werghi, "Deep Bidirectional Correlation Filters for Visual Object Tracking", ***IEEE Fusion***, 2020.
- X. Zhang**, S. Javed**, O. Ahmed, J. Dias, and N. Werghi, "Gender Recognition Using RGB-D Images", ***IEEE ICIP***, 2020.
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, "Deep Multiresolution Cellular Communities for Semantic Segmentation of Multi-Gigapixel Histology Images ", ***IEEE ICCV Workshop***, *2019*.
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, "Tensor Low-Rank Tracking", ***IEEE ICCV Workshop***, *2019*.
- **S. Javed**, A. Mahmood, J. Dias, and N. Werghi, "Structural Low-Rank Tracking", ***IEEE Advanced Video and Signal-based Surveillance (IEEE-AVSS***), *2019*.
- **S. Javed**, M. Fraz, D. Snead, D. Epstein, and N. Rajpoot, "Cellular Community Detection for Tissue Classification", ***MICCAI Computational Pathology Workshop***, *2018*.
- M. Sultana, A. Mahmood, **S. Javed**, and Soon Ki Jung, "Unsupervised RGB-D Video Object Segmentation Using GANs", ***Asian Conference on Computer Vision RGB-D Workshop***, *2018*.
- W. Ansar, M.M. Fraz, M. Shahzad, I. Gohar, **S. Javed**, and S. K. Jung, "Two Stream Deep CNN-RNN Attentive Pooling Architecture for Video-based Person Identification", ***Ibero-American Congress on Pattern Recognition***, *2018*.
- **S. Javed**, P. Narayanamurthy, T. Bouwmans, and N. Vaswami, "Robust PCA and Robust Subspace Tracking: A Comparative Evaluation", ***IEEE Special Session on Signal Processing***, June, *2018*.
- M. Fiaz, **S. Javed**, A. Mahmood, and S. K. Jung, "Comparative Study of ECO and CFNET Trackers in Noisy Environment", ***The 3rd International Conference on Next Generation Computing***, *2018*.
- **S. Javed**, T. Bouwmans, and S. K. Jung, "Superpixels-based Manifold Structured Sparse RPCA for Moving Object Detection", ***British Machine Vision Conference, BMVC***, September, *2017*.
- **S. Javed**, T. Bouwmans, and S. K. Jung, "Moving Object Detection of RGB-D Videos using Spatiotemporal RPCA Algorithm", ***International Conference on Image Analysis and Processing, ICIAP***, *2017*.

- **S. Javed**, T. Bouwmans, and S. K. Jung, "Stationary Background Model Initialization Based on Low-rank Tensor Decomposition", ***The 33th ACM/SIGAPP Symposium on Applied Computing (SAC'17),*** *2017*.
- **S. Javed**, T. Bouwmans, and S. K. Jung, "Spatially-consistent Smoothed RPCA for Highly Dynamic Scene Background Subtraction", ***The 33th ACM/SIGAPP Symposium on Applied Computing (SAC'17),*** *2017*.
- **S. Javed**, A. Mahmood, T. Bouwmans, and S. K. Jung, "Motion-Aware Graph Regularized RPCA for Background Modeling of Complex Scene", ***International Conference on Pattern Recognition, ICPR***, *2016*.
- **S. Javed**, O. Seonho, A. Sobral, T. Bouwmans, and S. K. Jung, "Background Subtraction via Superpixel-based Online Matrix Decomposition with Structured Foreground Constraints", ***ICCV Workshop: Robust Subspace Learning and Computer Vision***, *2015*.
- A. Sobral, **S. Javed**, S. K. Jung, E. ZahZah, and T. Bouwmans, "Online Tensor Decomposition for Background Subtraction in Multispectral Video Sequences", ***ICCV Workshop: Robust Subspace Learning and Computer Vision***, *2015*.
- **S. Javed**, T. Bouwmans, and S. K. Jung, "Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction", ***IEEE 6th International Conference on Imaging for Crime Prevention and Detection, (ICDP-15),*** *2015*.
- **S. Javed**, T. Bouwmans and S. K. Jung, "Combining ARF and OR-PCA for Background Subtraction Robust to Noisy Videos", ***18th International Conference on Image Analysis and Processing, (ICIAP'15),*** *2015*.
- **S. Javed**, A. Sobral, T. Bouwmans, and S. K. Jung, "OR-PCA with Dynamic Feature Selection for Robust Background Subtraction", ***The 30th ACM/SIGAPP Symposium on Applied Computing (SAC'15)***, *2015*.
- **S. Javed**, T. Bouwmans, and S. K. Jung, "Depths Extended OR-PCA with Spatiotemporal Constraints for Robust Background Subtraction", ***IEEE 21st Japan-Korea Joint Workshop on Frontiers of Computer Vision (FCV'15),*** *2015*.
- **S. Javed**, S. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "OR-PCA with MRF for Robust Foreground detection in Highly Dynamic Backgrounds", ***12th Asian Conference on Computer Vision (ACCV-2014),*** *2014*.
- **S. Javed**, S. Oh, J. Heo, and S. K. Jung, "Robust Background Subtraction using Online Robust PCA via Image Decomposition", ***International Conference on Research in Adaptive and Convergent System (RACS'14),*** 2014.
- **S. Javed**, S. Oh, S. K. Jung, "Foreground Object Detection and Tracking for Visual Surveillance System: A Hybrid Approach", ***IEEE 11th International Conference on Frontiers of Information Technology (FIT),*** *2013.*

Book Chapters

- **S. Javed** and S. K. Jung, "Stochastic RPCA for Background/Foreground Detection", ***Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video***

Processing, Part V, Chapter 3, CRC Press, Taylor and Francis Group, *2015*, https://sites.google.com/site/lowranksparsedecomposition/

Invited Talks

- **S. Javed**, "CPLIP: Zero-shot Learning in Histology Images", ***University of Warwick***, U.K, *2024*.
- **S. Javed**, "Cancer Image Analytics and Its Footprints in Histopathological Landscapes", ***Kyungpook National University***, South Korea, *2019*.
- **S. Javed**, "Robust Principal Component Analysis for Moving Object Detection", ***University of Warwick***, United Kingdom, *2017*.
- **S. Javed**, "Machine Learning its Applications in Computer Vision and Medical Imaging", ***Khalifa University Centre for Autonomous Robotics Systems (KUCARS),*** United Arab Emirates, *2019*.
- **S. Javed**, "Statistical Machine Learning and Its Applications in Computer Vision", ***National University of Computer and Emerging Sciences (NUCES),*** Pakistan, *2019*.
- **S. Javed**, "Recent Advancements in Intelligent Video Surveillance", ***City University of Science & Information Technology (CUSIT),*** Pakistan, *2014*.

Scholarships & Awards

- ***Best Researcher Award*** *for BK (21+) project at* ***Kyungpook National University****, 2016*
- ***Merit-Based Scholarship*** *for combined* ***Master's & Doctoral Program*** (BK+), Brain Korea 21, KNU, *2013*
- ***Participate Researcher*** (OZ002), Brain Korea 21, ***Kyungpook National University***, Korea, March *2012-2015*.
- ***University of Hertfordshire*****,** UK*,* ***Chancellor Scholarships*** *for International Students 2010*, Ministry of education science and technology UK.
- ***Teaching Assistant*** (IM001), ***Kyungpook National University***, Korea, *2012*.
- ***Research Grants*** for i-LIDS UK ***Visual Surveillance System,*** *2015****.***

Patent Applications

- Soon Ki Jung, Oh Seon Ho, **Sajid Javed**, "*Foreground Detection Using Image Decomposition & Online RPCA"*, **Registration No**. 10-1583787, **Registration Date** *2016.01.04*, Republic of Korea.
- Soon Ki Jung, Ja Hu, Oh Seon Ho, **Sajid Javed**, "*Apparatus and Method for Processing Image to Adaptively Enhance Low Contrast and Detecting Object from Night Scenes"*, **Registration No**. PCT/KR2016/009394, **Registration Date** *2016.08.24*, Republic of Korea.

Professional Services

- Reviewer of CVPR 2022, CVPR2024, CVPR2025, and ICCV2025
- Area Chair of ACCV-2022
- Organizer of "1st Deep learning Workshop on Underwater Video Tracking", ACC-2022
- Organizer of "Computer Vision for Coastal and Marine Environment Monitoring", ICIAP-2022.
- Organizer of "Towards Practical Deep Learning Algorithms for Moving Object Detection for Challenging Environments", ICIP-2022.
- Coordinator for the computer vision and artificial intelligence courses in KU.
- Member of the MBZIRC-2020 committee
- Guided students to win MBZIRC-2020 challenge
- Organizer of Robust Subspace Learning and Computer Vision (RSL-CV) in **ICCV**-2019.
- Organizer of Robust Subspace Learning and Computer Vision (RSL-CV) in **ICCV**-2017.
- Organizer of Background Modeling/Reconstruction in Complex Environment special session in **ICIP**-2020.
- Organizer of **MBZIRC**-2020.
- Reviewer for the following journals: (**1**) IEEE Transactions on Pattern Analysis and Machine Intelligence. (**2**) IEEE Transactions on Circuits and Systems for Video Technology. (**3**) IEEE Transactions on Image Processing. (**4**) IEEE Transactions on Industrial Electronics. (**5**) IEEE Transactions on Multimedia. (**6**) IEEE Transactions on Signal Processing. (**7**) Pattern Recognition. Pattern Recognition letters. (8) Neurocomputing. (9) Journal of Visual Communications and Image Representation. (**10**) Computer Vision and Image Understanding. (**11**) Journal of Electronic Imaging. (**12**) Image and Vision Computing. (**13**) Machine Vision and Applications. (**14**) Robust Subspace Learning and Computer Vision in Conjunction with ICCV. (**15**) International Conference on Image Analysis and Processing. (**16**) International Conference on Advanced Video and Signal Based Surveillance. (**17**) International conference on Image Processing. (**18**) International Conference on Pattern Recognition. (**19**) Journal of Intelligent Systems. (**20**) IEEE Transactions on Evolutionary Computations.
- Reviewer for the following conferences: (**1**) Computer Vision and Pattern Recognition. (**2**) European Conference on Computer Vision. (**3**) Asian Conference on Computer Vision. (**4**) International Conference on Computer Vision. (**5**) British Machine Vision Conference.

Teaching Experience

I have taught the following core computer courses at the BS, MS, and PhD. levels during my current position:

- Introduction to Computer Vision (overall student evaluation of 4/5)
- Advanced Image Processing (overall student evaluation of 4.5/5)
- Artificial Intelligence (overall student evaluation of 5/5)
- Operating Systems (overall student evaluation of 4.3/5)
- Deep Learning Systems Design (overall student evaluation of 4.5/5)
- C++ Programming (overall student evaluate of 4.4/5)
- Graph Theory (overall student evaluate of 4.4/5)
- Human-Computer Interfaces (overall student evaluate of 4/5)
- Data structure and Algorithms (overall student evaluate of 3.9/5)
- Database Management System (overall student evaluate of 4.6/5)

Projects & Thesis Supervision

- **Date:** June**,** 2019- March**,** 2020**, 06 National Services Associates** from United Arab Emirates, **MBZIRC Project**: Challenge-1 solution for ball capturing and piercing balloons using UAVs.
- **Date:** Feb, 2019- Current**, Ph.D. Student**: Xiaoxiong Zhang from Khalifa University of Science and Technology, **Title**: Resilient Video Surveillance.
- **Date:** Nov, 2019- Current**, Masters Student**: Ahsan Akbar from Khalifa University of Science and Technology, **Title**: Tissue Classification and Nucleus Detection using histology images.
- **Date:** Aug, 2019- Dec, 2019**, Masters Student**: Muhammad from Khalifa University of Science and Technology, **Title**: Visual Object Tracking in the wild.
- **Date:** Sept, 2019- Jan**,** 2020, **Masters Student**: Mazin Debe from Khalifa University of Science and Technology, **Title**: Visual Object Tracking.
- **Date:** June, 2019- Sept**,** 2019**, Ph.D. Student**: Nadia Abdullah from Khalifa University of Science and Technology, **Title**: Tissue Deep Learning for Image Segmentation.
- **Date:** Feb, 2019- Current**, Master Student**: Ahmed Abdullah from Khalifa University of Science and Technology, **Title**: Face recognition using 3D mesh and deep learning.
- **Date:** Feb, 2019- Feb 2020**, Master Student**: Ahmed Umais from Khalifa University of Science and Technology, **Title**: Video Object Detection and Tracking.
- **Date:** Feb**,** 2020- Current**, Group of 12 Master Students** from Khalifa University of Science and Technology, **SDP Project Title**: Visual object Tracking using deep neural networks.
- **Date: March, 2**019- Sept, 2019**, Master Student:** Abdullah Albreiki from Khalifa University of Science and Technology, **Title**: Deep neural networks for moving object detection.

References

1. Prof. Nasir Rajpoot, University of Warwick, United Kingdom.
2. Prof. Jiri Matas, Czech Republic University, Prague.
3. Prof. Thierry BOUWMANS, Université De La Rochelle.
4. Prof. Michael Felsberg, Linköping University, Sweden.
5. Prof. Muhammad Bennamoun, University of Western Australia, Australia

Dr. Sajid Javed
Assistant Professor
Department of Computer Science
Khalifa University of Science and Technology, UAE
sajid.javed@ku.ac.ae
+971-503308507
April 4th, 2025

Hiring Committee
University of Warwick
Department of Computer Science
CV4 7AL, Coventry, United Kingdom

Subject: Application for Academic Position in Computer Science

Dear Hiring Committee,

I am writing to express my strong interest in the academic position in computer science at the University of Warwick. As an experienced researcher and educator in computer vision, machine learning, and deep learning, I am excited about the opportunity to contribute to your department's research excellence and academic leadership. *Having closely followed your institution's groundbreaking work in the computer vision field*, I am confident that my extensive track record in high-impact research, international collaborations, and successful research funding aligns well with the vision and ambitions of your institution.

I am currently working as an *assistant professor of computer vision* at *Khalifa University of Science and Technology*, where I lead cutting-edge research at the intersection of computer vision, deep learning, and real-world applications across healthcare, computer vision, and marine environments.

My work has been published in top-tier venues including ***CVPR, TPAMI, TIP, TNNLS, Information Fusion,*** *and* ***MedIA*** considered as a $\mathbf{4}^{*}$ ***REF publication***. I have a strong successful track record of securing competitive funding, including AI for video surveillance, developing AI system for cancer diagnosis, and have led interdisciplinary projects that bridge AI with healthcare and environmental monitoring. Notably, my collaborative work with **Prof. Jiri Matas** and **Prof. Michael Felsberg** has driven advances in robust tracking and domain adaptation, while my

contributions with **Prof. Nasir Rajpoot** have helped integrate AI into computational pathology—a direction that aligns closely with Warwick's strategic vision.

What excites me most about *Warwick* is its reputation for excellence in both foundational AI and translational research. The opportunity to engage with the ***Department of Computer Science***, contribute to ***Warwick's AI initiative***, and foster impactful collaborations with ***WMG, Warwick Medical School***, and industry partners is deeply appealing. I am equally committed to nurturing the next generation of AI researchers through inclusive, research-led teaching and mentoring.

Research Excellence and Contributions

My research focuses on developing ***cutting-edge GenAI techniques for image and video analysis, medical imaging, and AI-driven vision systems***. I have explored fundamental aspects of representation learning, vision-language models, and generative models with applications in autonomous systems, healthcare, and industrial automation. My work has been regularly published in top-tier venues ($\mathbf{4}^{*}$ ***REF publication***), including ***CVPR, TPAMI, TIP, TNNLS, MEDIA,*** and ***TCyb***, reflecting the significance and impact of my contributions to the field.

A key aspect of my research agenda is bridging the gap between fundamental AI advancements and real-world applications. In recent years, my team has focused on *next-generation AI models such as Vision Transformers and Generative AI for medical image analysis*. Our work has resulted in innovative deep-learning architectures that improve interpretability, robustness, and efficiency in complex vision tasks.

Furthermore, I have successfully *secured competitive research funding* during my academic career, supporting state-of-the-art developments in computer vision. These funding initiatives have enabled collaborative projects with academic and industrial partners, fostering an ecosystem for cutting-edge research translation.

International Collaborations and Academic Network

Collaboration has been a cornerstone of my research philosophy. I have had the privilege of working with some of the world's leading, most distinguished researchers in the field, including ***Prof. Jiri Matas*** and ***Prof. Michael Felsberg***, on fundamental advancements in deep learning-based visual recognition. These collaborations have strengthened my expertise in object detection, tracking, and marine vision, with applications ranging from visual tracking to underwater perception.

Additionally, I have collaborated with ***Prof. Nasir Rajpoot***, a globally recognized leader in computational pathology, who is currently affiliated with your esteemed institution. Our joint research has focused on leveraging AI for medical imaging applications, particularly in histopathological analysis. This work has demonstrated the potential of deep learning in transforming digital pathology workflows and has laid the foundation for further interdisciplinary research at the intersection of AI and healthcare. My work, co-authored with ***Prof. Rajpoot***, has

been continuously reported four times in the ***Medical Image Analysis*** journal and one time in the ***Transactions on Image Processing*** journal. These venues are considered as a $\mathbf{4}^{*}$ **REF publications.**

I believe that these collaborations, coupled with my extensive experience in developing AI-driven computer vision solutions, will be instrumental in strengthening the *University of Warwick* research network and fostering new partnerships.

Teaching and Mentorship

Beyond research, I am deeply committed to teaching and mentoring the next generation of AI and computer vision researchers. I have extensive experience teaching core computer science and computer vision courses at undergraduate and postgraduate levels, covering subjects such as *Machine Learning, Human-Computer Interaction, Operating Systems, and Deep Learning System Design*.

I am currently supervising four Ph.D. students and five M.Sc. students, guiding them on cutting-edge topics such as generative models, self-supervised learning, and vision transformers. Many of my former students have successfully transitioned into leading research roles in academia and industry, which I consider a testament to my dedication to student mentorship. I take great pride in fostering a research culture that emphasizes critical thinking, innovation, and ethical AI development.

Future Vision and Contributions to the University of Warwick

If given the opportunity to join the *University of Warwick,* I am eager to contribute to its world-class reputation in computer vision and AI. I aim to:

- **Establish a research lab focused on GenAI for computer vision**, advancing research in areas such as generative models, robust deep learning, and interdisciplinary AI applications. I aim to design vision-based AI agents on the industrial scale to perform automatic operations.
- **Strengthen collaborations between academia and industry**, securing external funding to support high-impact research in autonomous computer vision systems, healthcare AI, and marine vision.
- **Contribute to curriculum development**, introducing advanced topics in AI, deep learning, and ethical AI considerations to prepare students for the evolving landscape of computer vision research and applications.
- **Engage in interdisciplinary collaborations**, particularly in computational pathology and biomedical AI, aligning with the research strengths of your institution.

I am excited about the prospect of joining the *University of Warwick* and contributing to its vision of research excellence, impactful teaching, and interdisciplinary innovation. My experience in high-impact research, strong academic collaborations, and successful research funding acquisition makes me well-suited for this role.

I would welcome the opportunity to further discuss my candidacy and explore how my expertise aligns with the goals of your department. Thank you for your time and consideration. I have attached my CV, teaching statement, and research statement for your review and look forward to the possibility of an interview.

Sincerely,

Sajid Javed

Dr. Sajid Javed
Assistant Professor of Computer Vision
Khalifa University of Science and Technology
UAE

Application ID: 157327, Gecia Bravo-Hermsdorff: Assistant and Associate Professor - Computer Science (100893-0325)

Personal Information

We take privacy seriously and will only use your personal information to administer your application. For more information please see our Data Protection Policies (https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection).

Title	Dr.
Preferred pronouns	She/her/hers
Given Name(s)	Gecia
Family Name	Bravo-Hermsdorff
Email	gecia.bravo@gmail.com
Preferred Phone No	07780881912

Additional Information

Are you currently employed by University of Warwick?

No

Will you now or in the future require a visa to obtain/continue to hold the right to work legally in the UK?

Yes

Reasonable Adjustments

We make intentional efforts to employ and retain people with disabilities. The following question will aid us to assist you should you need it during the application process.

Do you have any medical condition, special educational needs or disability that means that you may require reasonable adjustments made for you during either the online assessment, interview or assessment centre stages of our selection process?

No

Source

Where did you find the advert for this vacancy?

Other

What made you apply for this vacancy at the University of Warwick? Please select all that apply.

Career progression, Personal recommendation, University reputation

Is there anything further that prompted you to apply?

Yes

Please detail

Yes, I visited the department last year to give a talk, and had a truly fantastic time. I thoroughly enjoyed my discussions with faculty members and postdocs, and left with exciting new ideas to think about. I am confident that working in such an environment would be both scientifically and personally rewarding, and I look forward to the collaborations it would foster.

References

Reference 1

Title	Professor
First Name	Ricardo
Last Name	Silva
Email	ricardo.silva@ucl.ac.uk
Mobile	07780881912
Reference Type	Academic

Relationship	Previous mentor and current collaborator at the Department of Statistical Science at UCL

Reference 2

Title	Professor
First Name	Kayvan
Last Name	Sadeghi
Email	k.sadeghi@ucl.ac.uk
Mobile	02076795977
Reference Type	Academic
Relationship	Previous mentor and current collaborator at the Department of Statistical Science at UCL

Reference 3

Title	Professor
First Name	Carey
Last Name	E. Priebe
Email	cep@jhu.edu
Reference Type	Academic
Relationship	Collaborator at the Department of Applied Mathematics and Statics at the University of Johns Hopkins

Gecia Bravo-Hermsdorff

Research Fellow · School of Informatics @ University of Edinburgh
Google Scholar | gecia.bravo@gmail.com | gecia.github.io

Education

Princeton University
Princeton, NJ, USA

PhD in Quantitative & Computational Neuroscience *(link)* *2020*

- *Dissertation*: "Quantifying human priors over abstract relational structures"
- *Selected courses (hyperlinked):* Random graphs and networks, Mathematical physics, Natural algorithms, Theory of deep learning, Complex analysis, Statistical learning and nonparametric estimation, Machine learning & pattern recognition, Optimal learning, Abstract algebra, Computational complexity, Statistical optimization and reinforcement learning, Stochastic processes on graphs

École Normale Supérieure (ENS Ulm)
Paris, France

Research Master in Cognitive Sciences and Neuroscience *(link)* *2011*

- *Dissertation*: "Neural basis of self-contingency detection in 5-month-old babies"

Diplôme de l'ENS *(link)* *2011*

- Admitted via the "International Selection in Science" *(link)*
- Three-year multidisciplinary program, coursework included: Computational neuroscience, Cognitive science, Decision theory, Biophysics, Logic, Mathematics, Statistics, Modeling, Ecology and evolutionary biology, Philosophy of science, Theoretical chemistry

Bachelor of Science *(link)* *2009*

Employment

School of Informatics – University of Edinburgh
Edinburgh, Scotland, UK

Research Fellow *Oct 2024–now*

Dept of Statistical Science – University College London
London, UK

Research Fellow *2022–2024*

- Causal Graphical Models for Relational Data ($\geq$ Oct 2023, with Kayvan Sadeghi):
  Systematically categorizing the conditional independence structures induced by natural models of growing random networks, analyzing their asymptotic distributions, and characterizing the relevant notions of intervention *(link)*.
- Extrapolating Interventions ($\leq$ Oct 2023, with Ricardo Silva):
  Using our framework of *interventional factor models*, we show when and how data collected under different interventions can be used to characterize the outcome of (unseen) combinations of interventions *(link)*.
  We also worked on principled methods for causal-effect estimation using *imperfect instrumental variables* *(link1,link2)*.

Google Research – Algorithms and Theory
US (remote) and London, UK

AI Resident *2020–2022*

- Privacy: Led a project that provided a class of data collection protocols that ensure anonymity to users without the need for a fully-trusted central entity performing the anonymization *(link)*.
- Geometric Deep Learning: Developed and implemented (in TensorFlow) a permutation and rotation equivariant neural network architecture for analyzing images of cells from drug discovery experiments, with the aim of identifying compounds that are likely to be effective against the dormant phase of the parasite responsible for Malaria.
- Graphs: Implemented (production code in C++) a method for scalable computation of graph cumulants with up to 3 edges [G&B-H 2020 *(link)*], with the aim of detecting atypical patterns of information propagation in social networks.
- Entropy: Developed private and communication efficient algorithms for estimating entropies (work published at NeurIPS, 2022). Implemented the algorithm from [Jian et al 2015 *(link)*] for computing entropy in the sparse data regime (production code in C++).

Publications

Causal Models for Growing Networks *(link)*
G Bravo-Hermsdorff, LM Gunderson, & K Sadeghi. *arXiv, 2025*

BudgetIV: Optimal Partial Identification of Causal Effects with Mostly Invalid Instruments *(link)*
J Penn, G Bravo-Hermsdorff, LM Gunderson, R Silva & DS Watson. *AISTATS, 2025* *(code)*

Bounding Causal Effects with Leaky Instruments *(link)*
DS Watson, J Penn, LM Gunderson, G Bravo-Hermsdorff, A Mastouri & R Silva. *UAI, 2024* *(code)*

Intervention Generalization: A View from Factor Graph Models *(link)*
G Bravo-Hermsdorff, DS Watson, J Yu, J Zeitler & R Silva. *NeurIPS, 2023* *(code, poster)*

The Graph Pencil Method: Mapping Subgraph Densities to Stochastic Block Models *(link)*
LM Gunderson, G Bravo-Hermsdorff & P Orbanz. *NeurIPS, 2023* *(poster)*

Quantifying Network Similarity using Graph Cumulants *(link)*
G Bravo-Hermsdorff*, LM Gunderson*, PA Maugis & CE Priebe.
Journal of Machine Learning Research (JMLR), 24(187):1-27, 2023 *(poster)*

Quantifying Human Priors over Social and Navigation Networks *(link)*
G Bravo-Hermsdorff. *ICML, 2023* *(demo, poster)*

Private and Communication-Efficient Algorithms for Entropy Estimation *(link)*
G Bravo-Hermsdorff, R Busa-Fekete, M Ghavamzadeh, A Muños Medina & U Syed. *NeurIPS, 2022* *(video)*

Statistical Anonymity: Quantifying Reidentification Risks Without Reidentifying Users *(link)*
G Bravo-Hermsdorff, R Busa-Fekete, LM Gunderson, A Muños Medina & U Syed. *arXiv, 2022*

Introducing Graph Cumulants: What is the Variance of your Social Network? *(link)*
LM Gunderson* & G Bravo-Hermsdorff*. *arXiv, 2020* *(video, code)*

Quantifying Human Priors over Abstract Relational Structures *(link)*
G Bravo-Hermsdorff. *Ph.D. dissertation, Princeton University, 2020* *(slides, demos)*

A Unifying Framework for Spectrum-Preserving Graph Sparsification and Coarsening *(link)*
G Bravo-Hermsdorff* & LM Gunderson*. *NeurIPS, 2019* *(video, demos, code, poster)*

Gender and Collaboration Patterns in a Temporal Scientific Authorship Network *(link)*
G Bravo-Hermsdorff, V Felso, E Ray, LM Gunderson, ME Helander, J Maria & Y Niv.
Applied Network Science, 4(1), 2019 *(dataset)*

Modeling the Hemodynamic Response Function for Prediction Errors in the Ventral Striatum *(link)*
G Bravo-Hermsdorff & Y Niv. *bioRxiv, 2019*

Quantifying Humans' Priors over Graphical Representations of Tasks *(link)*
G Bravo-Hermsdorff, TD Pereira & Y Niv.
Unifying Themes in Complex Systems IX. ICCS, Springer Proceedings in Complexity, 281–290, 2018

**denotes equal contribution*

Teaching

Probability and Statistics II STAT0005 (University College London) *(link)* *Fall 2023*

- Last mandatory course for bachelors and masters students in the Statistical Science Department, lectures given by Kayvan Sadeghi. Topics included: transformation of random variables, relations between standard distributions, statistical estimation, consistency, method of moments, Bayesian inference, conjugate priors, asymptotic guarantees.
- I held two one-hour tutorial sessions per week covering the students' homework and questions.

BioMath Bootcamp (Princeton University) *(link)* *Summer 2016*

- Month-long training in mathematical and computational tools for incoming PhD students in the computational neuroscience and computational biology programs, organized by Carlos Brody.
- I lectured for the probability module, and held afternoon sessions for exercises in: programming (Python), linear algebra, ordinary differential equations (ODEs), nonlinear dynamical systems, probability, Fourier transforms, convolutions, and signal processing.

Introduction to cognitive neuroscience (Princeton University) *(link)* *Spring 2015*

- I held three one-hour sessions per week discussing relevant journal publications, and helped construct and grade the exams.

Lab for introduction to experimental psychology (Princeton University) *Fall 2014*

- I held a weekly three-hour lab session with introductory lectures and exercises in: statistical analysis, MRI, EEG, psychophysics, experimental design, programming, computational modeling, and game theory.

Awards and fellowships

- **Top reviewer award for Learning on Graphs (LoG) Conference (prize of $1,500)** *2023*
- **Google AI residency NYC (Algorithms and Theory branch)** *2020–2022*
  Competitive position for exploring research at Google
- **Independent research grant ($5,000)** *2019*
  Funding awarded by the Princeton Cognitive Science Department to selected research proposals
- **Scholarship for Lake Como School of Advanced Studies in Complex Networks** *May, 2016*
- **Cognitive Science Graduate Fellowship** *2016–2017*
- **Scholarship for Brains, Minds and Machines summer school** *August, 2015*
- **Scholarship for SAMSI Bayesian Nonparametrics workshop** *July, 2015*
- **Princeton PhD Fellowship** *2013–2019*
- **École Normale Supérieure (ENS Ulm) "International Selection in Science"** *2008*
- **Scholarship by the Lions Club to study French literature in France** *Summer, 2006*
- **Brazilian CNPq "scientific initiation" scholarship** *2006–2008*
- **Entrance exam for Biomedical Sciences degree at the Universidade Federal do Rio de Janeiro** *2006*
  Top Brazilian undergraduate program in biomedical sciences, completed 2 of 4 years before moving to France
- **99th percentile at the Exame Nacional de Ensino Médio (ENEM)** *2005*
  Nationwide exam for Brazilian students after high school
- **Travel awards for conferences and workshops:**
  Eurandom Workshop on Graph Laplacians, Multivariate Extremes and Algebraic Statistics *(link)* *TU Eindhoven, Netherlands, 2024*
  Foundations of Quantum Computing (FQC2024) Workshop *(link)* *Royal Holloway, University of London, 2024*
  YES Causal Inference Workshop *(link)* *Eurandom, Eindhoven University of Technology, 2023*
  Neural Information Processing Systems (NeurIPS) Scholar Award, *2022* and *2019*
  International Conference on Complex Systems (ICCS), *2018*
  NeurIPS Women in Machine Learning, *2018* and *2017*
  Society for Industrial and Applied Mathematics (SIAM) Annual Meeting, *2018*
  Multidisciplinary Conference in Reinforcement Learning and Decision Making (RLDM), *2017* and *2015*
  International Conference on Mathematical Neuroscience (ICMNS), *2017*
  Austin Memory & Learning Conference, *2015*

Academic service and activities

Science outreach

Volunteer teacher for the In2STEM-In2ScienceUK outreach initiative *(link)* — *London, August 2024*

- Designed and taught three 2-hour classes to nine high-school students:
  The mathematics of card magic tricks — based on the wonderful book "Magical Mathematics" by Diaconis and Graham *(link)*;
  Optimal betting — derive the Kelly criterion *(link)* using hands-on simulations and a story about exploiting a broken arcade game;
  The mathematics of cooperation — a friendly introduction to evolutionary game theory using as entry point the classical work by Axelrod on tournaments of the repeated Prisoner's Dilemma game *(link)*.
- Our placement (which was hosted by Alex Watson) was picked as "host of the week" thanks to nominations from our students *(link)*.

Pedagogy training

- One day workshop by the Alda Center for Communicating Science *(link)*. — *Princeton, October 2018*

Volunteer at the Princeton Neuroscience Fair — *Princeton, March 2018*

- Event with fun neuroscience demonstrations for 4th grade of low-income households from the Christina Seix Academy.

Science outreach volunteer at STEMCivics Charter School — *Ewing, NJ, April 2017*

- I performed a demonstration of a cerebellar illusion using glasses that distort ones aim as they throw a ball.

Invited workshops

Graph Laplacians, Multivariate Extremes and Algebraic Statistics *(link)* — *Eurandom, Nov 2024*

Foundations of Quantum Computing (FQC2024) Workshop *(link)* — *Royal Holloway, University of London, August 2024*

Networks and Time Meeting II *(link)* — *Northeastern University London, April 2024*

YES Causal Inference workshop *(link)* — *Eurandom, Eindhoven University of Technology, March 2023*

Mentoring

Graduate

- Katie Robertson — *Master student at University of Edinburgh* — *Feb 2025–now*
  I proposed and am supervising her Msc thesis project on efficiently counting medium-sized subgraphs.
- Jordan Penn — *PhD student at King's University College* — *Winter 2023–now*
  I am helping supervise his project on analyzing phase transitions in dynamical network models and his projects on methods for causal inference in the presence of unobserved confounders.
- Emma Graham (now PhD student at Dartmouth) — *Master's student at UCL Center for Artificial Intelligence* — *Summer 2023*
  I helped supervise her MSc thesis project on using Voronoi diagrams to study catastrophic forgetting in reinforcement learning.

Undergraduate

- Cristian Andronic — *Computer science major at Princeton University* — *2016*
  I proposed and supervised a project to build a navigation App to collect data for modeling naturalistic human mobility.
- Daniel J. Wilson (now PhD student at University of Toronto) — *Volunteer intern at Princeton Neuroscience Institute (PNI)* — *2015*
  I proposed and supervised a project involving coding a contextual bandit task in MTurk *(link)* to model human representation learning.
- Caitlyn Cap and Olamilekan Sule — *Summer interns at the Botvinick lab on PNI* — *2014*

Hackathons

Weekend-long MIT Quantum Hackathon (iQuHack) *(link)* — *Remote, Feb 2024*

- I worked on the IonQ challenge to reverse-engineer quantum circuits *(link)*.

Day-long biohackathon by PrecisionLife and ME Research UK *(link)* — *UCL, Feb 2024*

- Exploratory analysis of a large database of single nucleotide polymorphisms (SNPs) correlated with Myalgic Encephalomyelitis (ME).

Week-long PNI-Intel Hackathon on multivariate analysis of fMRI data — *Princeton, Jan 2017*

Reviewer

Journals

- Network Science *(2024)*; IEEE Transactions on Signal Processing *(2024)*; Scandinavian Journal of Statistics *(2024, 2023)*; Socio-Economic Planning Sciences *(2019)*; Trends in Cognitive Sciences *(2017)*

Conferences

- Learning on Graphs Conference (LoG) *(2024, 2023)*; International Conference on Machine Learning (ICML) *(2023, 2022)*; TheWebConf2023; Conference on Neural Information Processing Systems (NeurIPS) *(2022)*; WHMD *2021* NeurIPS workshop; NeurIPS Women in Machine Learning *(2017, 2018)*

Summer schools

IX Paraty Quantum Information School *(link)* — *upcoming: August 2025*
Quantum Paraty — *Paraty, Brazil*

Summer School on Mathematical Aspects of Quantum Information *(link)* — *upcoming: June, 2025*
Université Paris-Saclay and Institut Polytechnique de Paris — *Paris, France*

Quantum Machine Learning and Hamiltonian Simulation *(link)* — *upcoming: June 2025*
London Mathematical Society (LMS) — *Isle of Skye, Scotland*

Machine Learning Summer School (MLSS) *(link, my video ∼9min)* — *Summer 2020*
Max Planck Institute for Intelligent Systems — *Tübingen, Germany (virtual)*

Complex networks: theory, methods, and applications *(link)* — *May 2016*
Lake Como School of Advanced Studies — *Lake Como, Italy*

Brains, Minds and Machines (BMM) Summer Course *(link)* — *August 2015*
MIT Center for Brains, Minds and Machines — *Woods Hole, MA, USA*

Bayesian Nonparametrics: Synergies between Statistics, Prob and Math *(link)* — *June 2015*
Statistical and Applied Mathematical Sciences Institute (SAMSI) — *Durham, NC, USA*

Computational and Cognitive Neurobiology Summer School *(link)* — *July 2010*
Cold Spring Harbor Laboratory Asia — *Suzhou, China*

Languages

- **Human:** (Brazilian) Portuguese *(native),* English & French *(fluent),* Spanish *(basic)*
- **Computer:** Python & MATLAB *(fluent),* Mathematica, Qiskit, C++ & R *(functional),* Javascript & HTML *(basic)*

Other research experiences

PhD candidate at the Niv lab *(link)* — *Princeton University, 2014–2019*

- I developed methods to efficiently quantify human priors over relational data by exploiting the relevant underlying symmetry *(link)*.

PhD research rotation at the Botvinick lab — *Princeton University, 2013–2014*

Research scholar in neuroeconomics at the Montague lab *(link)* — *Virginia Tech, 2011–2013*

- I developed and validated computational models to explain human behavioral data from various neuroeconomic experiments, such as, multi-armed bandit tasks and the repeated ultimatum game. *Advisors*: Read Montague and Terry Lohrenz

Master's student at the Cognitive Science and Psycholinguistic lab *(link)* — *ENS Ulm, Paris, 2011*

- Researched the neural substrates of self-contingency detection in babies using functional near-infrared spectroscopy (fNIRS). I designed, coded, and built the experimental apparatus, and recorded and analyzed data from 61 babies. *Advisor*: Emmanuel Dupoux

Research internship at Caltech Emotion and Social Cognition lab *(link)* *Caltech, Pasadena, Spring 2010*

- I designed and carried out behavioral experiments to analyze whether humans express values learned via Pavlovian conditioning in an unrelated task without their conscious awareness. *Advisors*: Naotsugu Tsuchiya and Ralph Adolphs

Research internship at the Development and Neuropharmacology lab *(link)* *Collège de France, Paris, 2009*

- Researched the molecular mechanisms involved in the emergence of cellular territories during the morphogenesis of the neural tube. *Advisors*: Elizabeth Di Lullo and Alain Prochiantz

Undergraduate student at the Physiology of Cognition lab *(link)* *UFRJ, Brazil, 2007–2008*

- Studied the physiology of the visual system in monkeys (using intracranial recordings) and humans (using EEG). *Advisor*: Mário Fiorani

Research internship at the Institute of Neurobiology Alfred Fessard *(link)* *CNRS, Gif-sur-Yvette, Summer 2007*

- Researched the development of the neural crest by grafting quail and chick embryos *in ovo*. *Advisors*: Sophie Creuzet

Selected talks

- **What is the variance (and skew, kurtosis, etc) of a network? Graph cumulants for network analysis**
  Networks Seminar Series, Mathematical Institute, University of Oxford, Dec 2024 *(link)*
- **Stochastic graph reduction via combinatorial invariants** *(link)*
  Eurandom Workshop on Graph Laplacians, Multivariate Extremes and Algebraic Statistics, Eindhoven, Nov 2024
- **Principled Processing of Relational Data**
  University of Southampton, School of Mathematics, July 2024
- **What is the variance (and skew, kurtosis, etc) of a network? Introducing graph cumulants for network analysis**
  University of Warwick, Department of Computer Science, May 2024
- **A taxonomy of causal models for growing networks**
  Northeastern University London Meeting on Networks and Time II, April 2024 *(link)*
- **Cumulants for networks**
  Algebraic and Combinatorial Perspectives in the Mathematical Sciences (ACPMS) Seminar, Online, 2022 *(link)*
- **Graph cumulants: What is the variance of your social network?**
  Learning with Graphs Summit (Google), 2022
- **Statistical anonymity: Quantifying reidentification risk without reidentifying users.**
  Chrome Privacy Budget Meeting, 2021
- **Using graph cumulants to detect atypical patterns of information spread in social networks**
  MML Eng Meeting (Google), 2021
- **Entropy estimation of high-dimensional sparse datasets.**
  Chrome Privacy Budget Meeting, 2021
- **Graph reduction by edge deletion and edge contraction**
  International Conference on Complex Systems (ICCS), Cambridge, MA, 2018
- **Quantifying people's priors over graphical representations of tasks**
  International Conference on Complex Systems (ICCS), Cambridge, MA, 2018
- **Graph reduction by edge deletion and edge contraction.**
  SIAM Workshop on Network Science (SIAMNS18), Portland, Oregon, 2018
- **Assessing decision-making in patients with insula lesion using various neuroeconomic tasks**
  Regional Conference in Neuroeconomics at the Duke Center for Interdisciplinary Decision Sciences, 2016

Selected publications

Quantifying Network Similarity using Graph Cumulants *(link)*
G Bravo-Hermsdorff*, LM Gunderson*, PA Maugis & CE Priebe.
Journal of Machine Learning Research (JMLR), 24(187):1-27, 2023 *(video, code, poster)*

A Unifying Framework for Spectrum-Preserving Graph Sparsification and Coarsening *(link)*
G Bravo-Hermsdorff* & LM Gunderson*. *NeurIPS, 2019* *(video, demos, code, poster)*

Causal Models for Growing Networks *(link)*
G Bravo-Hermsdorff, LM Gunderson, & K Sadeghi. *arXiv, 2025*

Quantifying Human Priors over Social and Navigation Networks *(link)*
G Bravo-Hermsdorff. *ICML, 2023* *(demo, poster)*

Basic Principles of Information Processing by Complex Systems

The rate of adoption of machine learning and artificial intelligence models has increased rapidly in the last decade. At the same time, the models being employed are more complex, with "state-of-the-art" large language models using trillions of parameters. Unfortunately, our understanding of such models is largely based on heuristics and intuition. Not only is this lack of knowledge intellectually unsatisfying, it can also be harmful; these black-box AI models trained on big data are also notorious for their numerous issues relating to misinformation, privacy violations, and fairness.

Just like the industrial revolution was made possible by understanding of thermodynamics, the principles of transforming energy into useful work; the current AI revolution will only be made possible by understanding the principles of transforming data into useful information. Building a physically-grounded mathematical theory of information processing by complex systems is a formidable goal, one that I am passionate about contributing to.

Such a framework must necessarily span many domains of research and applications. My intellectual curiosity has led me to find such connections in the fields of: machine learning and AI, statistics, graph and network theory, information theory, causal inference, computational neuroscience, and quantum information. In particular, my work has contributed principled models, algorithms, and statistical tools for: quantifying *network structure* [1–4], extracting useful statistical *information* in various settings [5–7], and addressing problems in *causal inference* [8–11].

Networks, because most meaningful data are fundamentally relational; from social networks, transport networks, the internet, brains, DNAs, proteins, all the way down to quantum entanglement. ***Information***, because $|\mathrm{b}\rangle$it[1] offers a universal currency for the dynamics of learning and evolution of such interacting systems. And ***Causality***, because a satisfactory model is explanatory, predicting what would happen under *new* contexts and interventions. These themes have provided conceptual and technical guidance throughout my work. In what follows, I describe how they have shaped my past research, how they inform my current projects, and how I envision developing them further as a faculty member at Warwick.

The Three Threads (Previous Work)

*1 — **Networks***

It is widely acknowledged that the structure of a system is related to its function (consider for example the large-scale effort poured into various connectome initiatives [13]). However, characterizing precisely how and to what extent is a current topic of research and much debate, and is almost certainly system-dependent. Real networks often exhibit **structure** over a wide range of scales; in order to make progress on this important topic, it is necessary to have principled methods for characterizing this structure. A major focus of my research has been on developing such methods. In fact, two of my earliest (and favorite) contributions addressed this problem from two complementary viewpoints.

In my first[2] publication [1], we introduced an elegant framework for reducing the size of a graph while preserving its **global structure** with respect to the long-time dynamics of diffusion.[3] In the process, we discovered a surprising connection between preserving the Laplacian (pseudo)inverse [14] and the dual operations [15] of edge deletion (the limit of zero edge weight) and edge contraction (the limit of infinite edge weight). Through the lens of the graph Laplacian pseudoinverse, we analytically incorporated these two operations in a *single* natural objective function for reducing a (undirected positively weighted) graph, unifying work on graph sparsification [16–18] (reducing a graph by deleting edges) and graph coarsening [19–21] (reducing a graph by merging nodes), which had thus far been treated as separated algorithmic primitives.

In our second paper, we introduced graph cumulants [3], a principled and interpretable set of subgraph-based statistics that provides a hierarchical description of networks based on **local correlations** between an increasing number of connections.[4] Graph cumulants are a generalization of the classical cumulants (mean, (co)variance, skew, kurtosis, etc) for networks. Intuitively, they quantify the propensity (if positive) or aversiveness (if negative) for the appearance of any particular subgraph in a larger network. They share the defining properties of cumulants, with attractive statistical properties that provide clever shortcuts for certain computations. And in [2], we convincingly show that they lead to more powerful statistical tests (compared to the bare subgraph-densities on which they are based) for detecting differences between distributions of networks, even when given only a single network per sample.

In [4], we provide a method that analytically relates the local "texture" of a network to its global "shape". Given some observed subgraph densities, our algorithm **analytically infers** the parameters of a stochastic block model (SBM) that matches those subgraph densities, using a clever "graphical method of moments". SBMs are highly-popular simple random graph models, in which each node is assigned to a latent block/group, and connectivity between nodes is determined by their latent assignments. Our method solves two

[1] *"It from Bit"* [12].
[2] Independent exploratory project with Lee Gunderson (my now husband and also main collaborator) during our PhDs at Princeton.
[3] Click here for a playlist with videos of our algorithm and here for our summary video.
[4] Click here for another fun summary video.

important problems: *1)* efficient parameter estimation: in contrast to other algorithms for obtaining SBMs and graphons, it requires no numerical fitting; and *2)* efficient sampling: it allows one to efficiently sample from a network model that matches low level statistics of the data (without suffering from the degeneracy problem [22, 23] that plagues many intuitively attractive network models).

$|2\rangle$ *Information*

Another major focus of my research has been on designing innovative protocols to extract useful statistical information indirectly in various settings. For instance, taking inspiration from the insights afforded to neuroscience by the characterization of visual priors, in [5], I **quantified human priors** over small social and navigation networks.[5] To this end, I designed a large-scale online experimental platform[6] and developed a modeling framework to meaningfully compare and summarize such high-dimensional structured priors. In order to sample the (super-exponentially large [25]) space of graphs, the platform implemented an MCMC sampling algorithm based on participants' responses, employing a gamified interface that still engaged their attention (despite them technically reasoning about such a large number of possibilities). To meaningfully summarize and compare the priors, I used our graph cumulants (which, in fact, we created as a solution to this problem).

During my time at Google, motivated by the problem of browser fingerprinting, I led a project [7] that provided a class of data collection protocols for ensuring a statistical notion of k-anonymity [26], while reducing the amount of trust placed in the central entity performing the **anonymization**, and in certain cases completely removing the need for such an entity. We also [6] developed novel, more sample-efficient private and communication-efficient algorithms for estimating the Shannon [27], Gini [28], and collision [29] **entropies** of a distribution in a distributed setting, where each sample belongs to a single user who shares their data with a central server.

3 $\rightarrow$ *Causality*

The other major focus of my work has been on developing tools for causal inference under more realistic assumptions than typically assumed. Causal models often take the form of a directed acyclic graph (DAG) graphical model [30, 31]. These models are well-suited for modeling situations where variables have a clear (but potentially unknown) causal ordering. However, many situations, such as biological experiments, fundamentally rely on feedback and equilibration processes. Such situations are better modeled by graphical models called chain graphs, which incorporate both directed and undirected edges [32]. A main application of causal modeling is to predict the outcome of interventions to the system. In [9], we derived the conditions under which datasets from different (combinations of) interventions can be used to infer the outcome of **new combinations of interventions** for general chain graphs, and provided efficient algorithms for doing so.

Another important application of causal modeling is to estimate the causal effect of a treatment X on an outcome Y. Randomized controlled trials (RCTs) are ideal for these aplication, as they eliminate correlations caused by unobserved confounders, but are often infeasible due to ethical or practical constraints. In such cases, instrumental variables (IVs) offer a popular alternative, enabling causal estimation despite unobserved confounding. However, IV models rely on strong, often unrealistic, assumptions (e.g., an IV affects Y only through the treatment X and shares no common cause with Y). In [10, 11], we developed methods for estimating the set of feasible causal effects from observational data in the presence of unobserved confounding, even when only **imperfect IVs** [33] are available.

Tying Threads Thoughtfully (Ongoing Research)

In an exciting recent work [8] (Fig. 1), we developed a statistical framework for modeling relational data that naturally unifies these three threads by proposing a natural class of **causal models for growing networks**. Real-world networks do not typically pop into existence fully-developed. Much like the assumption of a "Last Universal Common Ancestor" in evolutionary biology, or the "Past Hypothesis" in cosmology, to better understand the state of a system at any given point in time, it is often insightful to model the history leading up to that point. This temporal evolution introduces a notion of causality. However, standard statistical models for networks are frequently based on node exchangeability, and are therefore not well-suited for modeling networks that grow over time.

Our framework [8] instead builds statistical models for growing networks based on invariance of the causal mechanisms *generating* the edges (with respect to node deletion and marginalization) instead of invariance of the *distribution* of edges (with respect to node permutation and subsampling). This perspective leads to distributed and asynchronous generative mechanisms for networks that give rise to a surprising diversity of tractable asymptotic behavior. With fewer parameters and increased conditional independence, our framework offers flexible and natural baseline models for causal inference in relational data.

In promising ongoing work, I have been developing methods for **certifying causal structure** through statistical inequalities, bridging ideas from causal inference and extremal combinatorics [34]. To answer causal questions, one often posits a causal structure that generates the variables of interest, and uses observational and experimental data to estimate the strength of those causal influences. However, in reality, the causal structure *itself* is uncertain. This raises a fundamental question: to what extent can we certify or falsify

[5] This paper is the culmination of the neuroscience part of my PhD thesis [24] and I am quite proud of the final product.
[6] Click here for a video demonstration of the experimental platform.

the claim that a given causal structure generated the observed data?

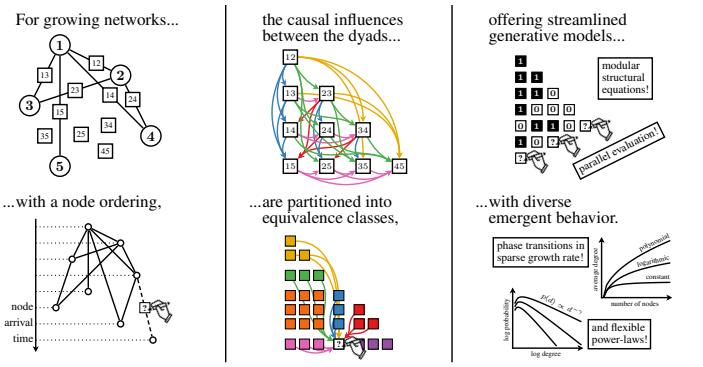


Fig. 1: **Summary of our work in [8] on causal models for growing networks.** *Left column:* The nodes of the growing network, represented as circles, have a total ordering. The variables in the model are indexed by the dyads (pairs of nodes), represented as squares. *Middle:* Causal relationships between these dyad variables are represented as arrows in a causal DAG describing the generative process of the growing network. We classify the relevant types of such causal arrows, represented by colors. *Right:* Some combinations of these causal arrows are remarkably amenable to parallel, distributed, and asynchronous generation, while exhibiting a range of asymptotic behaviors.

The **causal compatibility** problem asks whether a given probability distribution over the *observable* variables is compatible with a given causal arrangement of observable and latent variables. Unobserved latent variables introduce a further source of uncertainty; distributions resulting from such causal scenarios can be shown to obey nontrivial inequalities not captured by conditional independence statements [35–37]. One such result is the famous Bell inequality [38], which bounds the correlations that can arise from classical latent variables in a basic two-party configuration.

We have developed an adaptation of the celebrated method of flag algebras [39–41] to streamline the derivation of such causal inequalities, with promising initial results [34] appearing in an upcoming poster at QPL 2025. Looking ahead, I want to develop this work into a flexible toolkit for causal compatibility problems, in both classical and quantum settings, where the latent variables represent entanglement between quantum states. This has the potential to streamline development of certification protocols for entanglement resources, a vital component of quantum key distribution and other related cryptographic protocols [42–44].

Theoretical Trefoil Triumph (Future Directions)

Together, these projects reflect my commitment to developing flexible, principled frameworks for statistical reasoning about systems of **relations**. I am excited to expand this agenda in several directions, a few of which I describe below.

The applicability of our graph cumulants depends on the ability to **quickly count subgraphs** of interest g in the large network G being analyzed, which in the worst case can take $\mathcal{O}(|V(G)|^{|v(g)|})$. However, the minimum exponents are not known in general, and reducing them for large classes of medium-sized subgraphs would be a very useful tool for mining patterns in real-world networks. Subgraphs that have a series-parallel decomposition [45] can be counted in time $\mathcal{O}(|V(G)|^{\omega})$ (where ω is the matrix multiplication exponent) [46], and I am currently supervising a master thesis project whose goal is to implement such an algorithm. While this would be quite a useful contribution, for some large-scale applications, matrix multiplication can be too costly. Besides, one often does not need the exact counts; approximations using smart sampling schemes to control the error suffice for most of these applications. In future work, I would like to leverage the fact that deep neural networks are universal function approximators, to build a probabilistic model based on a **graph neural network** that *directly* outputs estimators of the graph **cumulants** of a network, while keeping track of the uncertainty of those estimators.[7]

The transformation we described of (injective homomorphism) subgraph densities to obtain our graph cumulants is part of the much more general framework of incidence algebras, which describe ways to transform and combine functions defined over (locally finite) partially ordered set (i.e., **posets**).[8] This suggests a general framework for "cumulantification" of other types of combinatorial structures, offering an avenue for obtaining similarly useful statistics for higher-order relational data, such as hypergraphs. Moreover, since incidence algebras also have a natural notion of convolution, an intriguing direction would be to construct CNNs for signals over combinatorial spaces with convolutional filters associated to (possibly multiple notions of) relevant posets. Posets can be used to describe the structure of nearly any data, and if this direction pays off, such **flexible representations** could significantly improve the ability for machine learning models to identify and generalize interesting structural patterns.

Another avenue that I am currently working on is applying our sparsification and coarsening method to directed graphs representing causal models (both cyclic and acyclic). Most work on **causal abstraction** currently uses slightly ad-hoc cost functions to ensure that the high-level model is useful for describing the low-level model. Adapting our method to these more general settings would offer a principled algorithm for doing this.

A guiding principle of my research is that clarity about structure, information processing, and causality is essential not only for understanding intelligent systems, but also for effectively designing and fruitfully interacting with them. I have already made promising progress in this direction, and am excited by the possibility of continuing to do so with the faculty, students, and postdocs at Warwick.

[7] This idea was suggested to me by Professor Theo Damoulas during my visit at Warwick last year, and is a collaboration I look forward to.

[8] In our case, the relevant poset is over partitions of the edges, and converting subgraph densities to graph cumulants is an example of a Möbius inversion, which is an operation that behaves much like a derivative with respect to a poset. As another example for graphs, the conversion between induced subgraph densities and injective homomorphism subgraph densities is given by a Möbius inversion with respect to the poset of edge deletion.

References

[1] **Gecia Bravo-Hermsdorff*** and Lee M Gunderson*. A unifying framework for spectrum-preserving graph sparsification and coarsening. *Neural Information Processing Systems (NeurIPS)*, 33, 2019.

[2] **Gecia Bravo-Hermsdorff***, Lee M Gunderson*, Pierre-André Maugis, and Carey E Priebe. Quantifying network similarity using graph cumulants. *Journal of Machine Learning Research (JMLR)*, 24(187):1–27, 2023.

[3] Lee M Gunderson* and **Gecia Bravo-Hermsdorff***. Introducing graph cumulants: what is the variance of your social network? *arXiv preprint arXiv:2002.03959*, 2020.

[4] Lee M. Gunderson, **Gecia Bravo-Hermsdorff**, and Peter Orbanz. The graph pencil method: mapping subgraph densities to stochastic block models. *Neural Information Processing Systems (NeurIPS)*, 37, 2023.

[5] **Gecia Bravo-Hermsdorff**. Quantifying human priors over social and navigation networks. *International Conference on Machine Learning (ICML)*, 40, 2023.

[6] **Gecia Bravo-Hermsdorff**, Robert Busa-Fekete, Mohammad Ghavamzadeh, Andrés Munõs Medina, and Umar Syed. Private and communication-efficient algorithms for entropy estimation. *Neural Information Processing Systems (NeurIPS)*, 36, 2022.

[7] **Gecia Bravo-Hermsdorff**, Robert Busa-Fekete, Lee M. Gunderson, Andrés Munõs Medina, and Umar Syed. Statistical anonymity: quantifying reidentification risks without reidentifying users. *arXiv preprint arXiv:2201.12306*, 2022.

[8] **Gecia Bravo-Hermsdorff**, Lee M Gunderson, and Kayvan Sadeghi. Causal models for growing networks. *arXiv preprint arXiv:2504.01012 (under review at UAI)*, 2025.

[9] **Gecia Bravo-Hermsdorff**, David S. Watson, Jialin Yu, Jakob Zeitler, and Ricardo Silva. Intervention generalization: a view from factor graph models. *Neural Information Processing Systems (NeurIPS)*, 37, 2023.

[10] David S. Watson, Jordan Penn, Lee M. Gunderson, **Gecia Bravo-Hermsdorff**, Afsaneh Mastouri, and Ricardo Silva. Bounding causal effects with leaky instruments. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 40, 2024.

[11] Jordan Penn, **Gecia Bravo-Hermsdorff**, Lee M. Gunderson, Ricardo Silva, and David S. Watson. BudgetIV: optimal partial identification of causal effects with mostly invalid instruments. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 28, 2025.

[12] John Archibald Wheeler. Information, physics, quantum: the search for links. *Proc. 3rd Int. Symp. Foundations of Quantum Mechanics*, pages 354–368, 1989.

[13] Michael Eisenstein. A milestone map of mouse-brain connectivity reveals challenging new terrain for scientists. *Nature*, 628(8008):677–679, 2024.

[14] Nisheeth K Vishnoi et al. Lx= b. *Foundations and Trends® in Theoretical Computer Science*, 8(1–2):1–141, 2013.

[15] James G Oxley. *Matroid theory*, volume 3. Oxford University Press, USA, 2006.

[16] Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 563–568, 2008.

[17] Joshua Batson, Daniel A Spielman, Nikhil Srivastava, and Shang-Hua Teng. Spectral sparsification of graphs: theory and algorithms. *Communications of the ACM*, 56(8):87–94, 2013.

[18] Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[19] Ilya Safro, Peter Sanders, and Christian Schulz. Advanced coarsening schemes for graph partitioning. In *International Symposium on Experimental Algorithms*, pages 369–380. Springer, 2012.

[20] Jie Chen, Yousef Saad, and Zechen Zhang. Graph coarsening: from scientific computing to machine learning. *SeMA Journal*, 79(1):187–223, 2022.

[21] Mohammad Hashemi, Shengbo Gong, Juntong Ni, Wenqi Fan, B Aditya Prakash, and Wei Jin. A comprehensive survey on graph reduction: sparsification, coarsening, and condensation. *arXiv preprint arXiv:2402.03358*, 2024.

[22] Mei Yin, Alessandro Rinaldo, and Sukhada Fadnavis. Asymptotic quantization of exponential random graphs. *The Annals of Applied Probability*, 26(6):3251–3285, 2016.

[23] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *Annals of Statistics*, 41(5):2428–2461, 2013.

[24] **Gecia Bravo-Hermsdorff**. *Quantifying human priors over abstract relational structures*. PhD dissertation, Princeton University, 2020.

[25] *The online encyclopedia of integer sequences (OEIS), entry A000088*, founded by Neil Sloane in 1964.

[26] Latanya Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002.

[27] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[28] Constantino Tsallis. *Introduction to Nonextensive Statistical Mechanics*. Springer New York, NY, Greece, 2009.

[29] Alfréd Rényi. On measures of information and entropy. In *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561, 1960.

[30] Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

[31] Judea Pearl. *Causality*. Cambridge university press, 2009.

[32] Steffen L Lauritzen and Thomas S Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):321–348, 2002.

[33] Nicolas Bastardoz, Michael J Matthews, Gwendolin B Sajons, Tyler Ransom, Thomas K Kelemen, and Samuel H Matthews. Instrumental variables estimation: assumptions, pitfalls, and guidelines. *The Leadership Quarterly*, 34(1):101673, 2023.

[34] Lee M. Gunderson, Davide Poderini, and **Gecia Bravo-Hermsdorff**. Causal bounds via subgraph inequalities. *Abstract for the poster we will present at the 22nd International Conference on Quantum Physics and Logic (QPL)*, July 2025.

[35] Elie Wolfe, Robert W Spekkens, and Tobias Fritz. The inflation technique for causal inference with latent variables. *Journal of Causal Inference*, 7(2):20170020, 2019.

[36] Armin Tavakoli, Alejandro Pozas-Kerstjens, Marc-Olivier Renou, et al. Bell nonlocality in networks. *Reports on Progress in Physics*, 2021.

[37] Elie Wolfe, Alejandro Pozas-Kerstjens, Matan Grinberg, Denis Rosset, Antonio Acín, and Miguel Navascués. Quantum inflation: a general approach to quantum causal compatibility. *Physical Review X*, 11(2):021043, 2021.

[38] John S Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[39] Alexander A Razborov. Flag algebras. *The Journal of Symbolic Logic*, 72(4):1239–1282, 2007.

[40] Alexander A Razborov. Flag algebras: an interim report. In *The Mathematics of Paul Erdős II*, pages 207–232. Springer, 2013.

[41] Alexander A Razborov. What is a flag algebra. *Notices of the AMS*, 60(10):1324–1327, 2013.

[42] Scott Aaronson. *Quantum computing since Democritus*. Cambridge University Press, 2013.

[43] Miralem Mehic, Marcin Niemiec, Stefan Rass, Jiajun Ma, Momtchil Peev, Alejandro Aguado, Vicente Martin, Stefan Schauer, Andreas Poppe, Christoph Pacher, et al. Quantum key distribution: a networking perspective. *ACM Computing Surveys (CSUR)*, 53(5):1–41, 2020.

[44] Peter W Shor and John Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Physical review letters*, 85(2):441, 2000.

[45] Xin He. Efficient parallel algorithms for series parallel graphs. *Journal of Algorithms*, 12(3):409–430, 1991.

[46] PA Maugis, Sofia C Olhede, and Patrick J Wolfe. Fast counting of medium-sized rooted subgraphs. *arXiv:1701.00177*, 2016.

Gecia Bravo-Hermsdorff
Ph.D. Princeton University
School of Informatics, University of Edinburgh
+44 07780881912 — gecia.bravo@gmail.com
gecia.github.io

Department of Computer Science
University of Warwick

April 28, 2025

Dear Members of the Search Committee,

I am writing to apply for the position of Assistant/Associate Professor in the Department of Computer Science at the University of Warwick (100893). After receiving a BS and MS from the École Normale Supérieure (ENS, Paris) and a PhD from Princeton University, I worked with the Algorithms and Theory team at Google Research (NYC), before returning to academia as a research fellow in the Department of Statistical Science at University College London (UCL), and now at the School of Informatics, University of Edinburgh. With a solid background in AI and ML, statistics, applied mathematics, and interdisciplinary research, as well as a true passion for research and teaching, I would be eager to contribute to the Department.

My research aims at distilling physically grounded mathematical principles for understanding and designing intelligent systems. I have contributed with principled models, algorithms, and statistical tools, to problems related to three interconnected themes: the structure of relational data (*Networks*), the extraction and quantification of useful statistics (*Information*), and the modeling and prediction of interventions (*Causality*). For more information, see my Research Statement. I am also deeply committed to teaching, mentoring, and outreach. I have led tutorials, delivered lectures, and designed hands-on sessions across a range of topics, from probability and statistics to computational modeling. I have also mentored several graduate students and volunteered in science outreach programs. For more information, see my Teaching Statement.

I had the opportunity to visit the Department last year, invited by Professor Long Tran-Thanh, and came away impressed by the atmosphere, both intellectual and social, and felt that the goals of the Department resonated strongly with my research interests. I am very excited by the possibility to participate in the Digital, Data Science & AI Spotlight initiative. Through my research, teaching, and mentoring, and other academic activities, I would help to continue and enhance Warwick's excellence in theoretical computer science, machine learning, data science, and interdisciplinary collaboration. I also look forward to being a part of the Warwick Institute for Data Science (WIDS), the Center for Discrete Mathematics and its Applications (DIMAP), and the University's partnership with the Alan Turing Institute, as well as working with member of the department such as Professors Tran-Thanh, Theo Damoulas, and many others.

I have presented and published my work in leading journals and conferences in the field, such as JMLR, NeurIPS, ICML, and AISTATS, and I have secured funding for all my education, as well as additional funding from various prestigious sources, such as an independent research grant from Princeton Cognitive Science Department and scholarships for workshops at institutions such as the Lake Como School of Advanced Studies. I am committed to sustaining a dynamic, externally funded research program that combines clarifying theory with meaningful applications, and to substantially contributing to Warwick's next REF submissions. To support this vision, within my first year I intend to apply for an EPSRC New Investigator Award, whose focus would be on further developing flexible, principled statistical frameworks for reasoning under uncertainty in relational systems.

In summary, I am a highly motivated and passionate academic eager to join Warwick's Department of Computer Science. I have a proven record of high-quality research with a distinctive research program that is fundamentally aligned with those of the department, as well as an intrinsic passion and commitment to teaching and mentoring. I have enclosed my Research Statement, CV, and Teaching Statement, and I would be very happy to provide any further information. Thank you very much for your time and consideration. I look forward to hearing from you.

Sincerely,