Tailia Malloy, Ph.D.

Google Scholar | Email | Personal Site | ResearchGate | LinkedIn

Employment History

2025 – ..... **Research Associate** Natural Language Processing for Personality Diagnostics. University of Luxembourg, Trustworthy Software Engineering, Interdisciplinary Center for Security, Reliability, and Trust. Under Prof. Jaques Klein and Prof. Tegawendé F. Bissyande.

2022 – 2025 **Postdoctoral Researcher** Cognitive Modeling and Human-Computer Interaction. Carnegie Mellon University, Social and Decision Sciences, Dynamic Decision Making Lab. AI Institute for Societal Decision Making. Under Prof. Cleotilde Gonzalez.

2019 – 2022 **External Researcher** IBM Research AI, Reinforcement Learning Group, RPI-IBM Artificial Intelligence Research Collaboration, Under Dr. Tim Klinger and Dr. Gerald Tesauro.

2018 – 2022 **PhD Student/Research Assistant** Rensselaer Polytechnic Institute, Department of Cognitive Science, Adaptive Computational Cognition Lab, Under Assoc. Prof. Chris R. Sims.

2017 – 2018 **Research Assistant/Volunteer** University of British Columbia, Computer Science Department, Sensory Perception & Interaction Research Group, Under Prof. Karon MacLean. School of Population and Public Health Under Prof. Peter Danielson.

2016 – 2017 **Research Assistant/Lab Software Developer** University of Innsbruck, Institute for Experimental Physics Under Prof. Tracy Northup; Innsbruck Institute for Quantum Optics and Quantum Information Under Prof. Rainer Blatt.

Education

2022 **Ph.D., Rensselaer Polytechnic Institute** Cognitive Science. Advised by Chris R. Sims
Thesis title: *Resource Rational Cognitive Modelling: An Information Theoretic Approach.*

2020 **M.Sc., Rensselaer Polytechnic Institute** Cognitive Science. Advised by Chris R. Sims
Thesis title: *Modelling Learning and Decision Making Under Information Processing Constraints*.

2018 **B.Sc., University of British Columbia**
Cognitive Systems: Computational Intelligence and Design; Dual Degree in Philosophy

Research Publications

Journal Articles

1. Du, Yinuo, Baptiste Prebot, Tailia Malloy, Fei Fang, and Gonzalez. Cleotilde. "Experimental Evaluation of Cognitive Agents for Collaboration in Human-Autonomy Cyber Defense Teams". *Computers and Human Behavior: Artificial Agents*, 2025.
2. Malloy, Tailia, Laura Bernardy, Omar El Bachyr, Fred Phillipy, Jordan Samhi, Jacques Klein, and Tegawendé F Bissyande. "You Got Phished! Analyzing how to Provide Useful Feedback in Anti-Phishing Training with LLM Teacher Models". *under review for MDPI Electronics*, 2025.
3. Malloy, Tailia, Maria Ferriera, Fei Fang, and Gonzalez. Cleotilde. "Improving Online Anti-Phishing Training Using Cognitive Large Language Models". *Computers and Human Behavior*, 2025.

4 Du, Yinuo, Baptiste Prebot, Tailia Malloy, and Gonzalez. Cleotilde. "A Cyber-War Between Bots: Cognitive Attackers are More Challenging for Defenders than Strategic Attackers". *ACM Transactions on Social Computing*, 2024.

5 Malloy, Tailia, and Cleotilde Gonzalez. "Applying Generative Artificial Intelligence to Cognitive Models of Decision Making". *Frontiers in Psychology*, 2024.

6 Malloy, Tailia, and Chris R. Sims. "Efficient Visual Representations in Learning and Decision Making". *Psychological Review*, 2024.

7 Malloy, Tailia, and Chris R. Sims. "A Beta-Variational Auto-Encoder Model of Human Visual Representation Formation in Utility-Based Learning". *Journal of Vision*, 2022.

Conference Proceedings

1 Malloy, Tailia, Hagmann David, and Gonzalez. Cleotilde. "Hierarchical Instance-Based Learning for Decision Making from Delayed Feedback". *2025 Conference in Cognitive Science*. 2025.

2 Malloy, Tailia, Maria Ferriera, Fei Fang, and Gonzalez. Cleotilde. "Using Cognitive Models to Improve Training Against Human and GPT-4 Generated Social Engineering Attacks". *International Conference on Human Computer Interaction*. 2025.

3 Malloy, Tailia, Roderick Seow, and Cleotilde Gonzalez. "Modeling Attention during Dimensional Shifts with Counterfactual and Delayed Feedback". *Reinforcement Learning and Decision Making*. 2025.

4 Malloy, Tailia, Maria Ferriera, Fei Fang, and Gonzalez. Cleotilde. "Leveraging a Cognitive Model to Measure Subjective Similarity of Human and GPT-4 Written Content". *Conference on Natural Language Learning*. 2024.

5 Malloy, Tailia, and Gonzalez. Cleotilde. "Learning to Defend by Attacking (and Vice-Versa): Transfer Learning in Cyber-Security Games". *IEEE European Symposium on Security and Privacy Workshop Series*. 2023.

6 Malloy, Tailia, Yinuo Du, Fei Fang, and Gonzalez Cleotilde. "Accounting for Transfer of Learning in Human Behavior Models". *Human Computation and Crowdsourcing*. 2023.

7 Malloy, Tailia, Yinuo Du, Fei Fang, and Gonzalez. Cleotilde. "Generative Environment-Representation Instance-Based Learning: A Cognitive Model". *AAAI Symposium on Integration of Cognitive Architectures and Generative Models*. 2023.

8 Malloy, Tailia, Tim Klinger, and Chris R Sims. "Modeling Human Reinforcement Learning with Disentangled Visual Representations". *Reinforcement Learning and Decision Making*. 2022.

9 Malloy, Tailia, Chris R Sims, Tim Klinger, Matthew D Riemer, Miao Liu, and Gerald Tesauro. "Learning in Factored Domains with Information-Constrained Visual Representations". *NeurIPS 2022 Workshop on Information-Theoretic Principles in Cognitive Systems*. 2022.

10 Malloy, Tailia, Tim Klinger, Miao Liu, Matthew Riemer, Gerald Tesauro, and Chris R Sims. "Consolidation via Policy Information Regularization in Deep RL for Multi-Agent Games". *AAAI Reinforcement Learning in Games Workshop*. 2021.

11 Malloy, Tailia, Tim Klinger, Miao Liu, Gerald Tesauro, Matthew Riemer, and Chris R Sims. "Modeling Capacity-Limited Decision Making Using a Variational Autoencoder". *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2021.

12 Malloy, Tailia, Tim Klinger, Miao Liu, Gerald Tesauro, Matthew Riemer, and Chris R Sims. "RL Generalization in a Theory of Mind Game Through a Sleep Metaphor (Student Abstract)". *AAAI Student Abstract Workshop*. 2021.

13 Malloy, Tailia, and Chris R Sims. "Modelling Visual Decision Making Using a Variational Autoencoder". *Proceedings of the 19th International Conference on Cognitive Modelling*. 2021, pp. 177–83.

14 Malloy, Tailia, Chris R Sims, Tim Klinger, Miao Liu, Matthew Riemer, and Gerald Tesauro. "Capacity-Limited Decentralized Actor-Critic for Multi-Agent Games". *IEEE Conference on Games*. 2021.

15 Malloy, Tailia, and Chris R Sims. "Modelling Human Information Processing Limitations in Learning Tasks with Reinforcement Learning". *Proceedings of the 18th International Conference on Cognitive Modelling*. Applied Cognitive Science Lab, PennState, 2020, pp. 159–60.

16 Malloy, Tailia, Chris R Sims, Tim Klinger, Miao Liu, Matthew Riemer, and Gerald Tesauro. "Deep RL with information constrained policies: Generalization in continuous control". *arXiv preprint arXiv:2010.04646*. 2020.

Experience

Grants and Fellowships

2024-2025 — $180,000 Co-PI Invited Presentation
Carnegie Bosch Institute Postdoctoral Research Grant. Title: Improving Cyber Defense Against Foundation Models with Cognitive Models. This proposal introduces a method of improving cyber defense capabilities by integrating foundation models with cognitive modeling approaches. The motivation for this is in mitigating the potential dangers introduced by the misuse of foundation models by cyber-attackers.

2023-2024 — $20,000 PI **Awarded**
Microsoft Accelerating Foundation Models Research Grant, Lead researcher and grant author on a study of human educational tools utilizing ChatGPT and cognitive modeling. Purpose: to study socio-technical impacts of large language models. Title: Personalized Education with Foundation Models via Cognitive Modeling. Grant Amount: $20,000 azure credit grant.

2024-2025 — $180,000 Co-PI Applied
Oak Ridge Institute for Science and Education Postdoc Fellowship. Title: Applying Cognitive Models and Generative Artificial Intelligence onto Countering Automated Online Influence Campaigns. This proposal introduced a method of improving defense against automated online influence campaigns that utilize generative models to create fake accounts, videos, news stories, and other types of disinformation.

$180,000 Co-PI Applied
Oak Ridge Institute for Science and Education Postdoc Fellowship Title: Improving Human-AI Collaboration in Intelligence Analysis by Integrating Cognitive Models and Generative Artificial Intelligence. This proposal introduces a method of improving human-AI collaboration in intelligence analysis by integrating foundation models with cognitive modeling approaches.

2019-2022 — ∼$300,000. Co-PI **Awarded**
RPI-IBM Artificial Intelligence Research Collaboration Fellowship. Title: Human Inspired Reinforcement Learning. RPI-IBM Artificial Intelligence Research Collaboration Scholarship providing tuition and salary for the duration of my PhD studies totaling approximately

Teaching

2025 — **Code, Create, & Innovate with AI**: Over the course, students will progress from learning Python programming and prompt engineering to more advanced topics like using large language models (LLMs) and API calls in their Python code and learning about retrieval-augmented generation (RAG) and its use.

2024 — **Human and Machine Decisions from Experience**: Carnegie Mellon University, Department of Social and Decision Sciences. Head lecturer and course designer for a new upper year undergraduate course. This course investigates the similarities and differences between human and machine cognition in the context of decisions from experience.

Experience (continued)

Awards and Keynote Presentations

2025 — **Keynote Presentation** on the topic of Human-AI Interdependence in Security and Privacy given at the TU Delft Workshop on Generative AI and Privacy/Security.

— **Best Paper Award** for the paper entitled 'Using Cognitive Models to Improve Training Against Human and GPT-4 Generated Social Engineering Attacks' presented at the Human Computer Interaction International Conference area on Cybersecurity, Privacy, and Trust.

Mentoring

2023-2025 — **Mentor for Graduate Students** Served as a mentor for graduate students in the Adaptive Computational Cognition Lab at RPI and the Dynamic Decision Making lab at CMU. This included several publications where I aided in writing grant proposals, designing studies, preparing manuscripts, or analyzing data.

2022-2025 — **Mentor for Undergraduate Students** Advisor for undergraduate theses for students at Carnegie Mellon and New York University. Four years serving as a senior mentor for the University of British Columbia Cognitive Systems mentoring program. Advised undergraduate students on research, internships, and industry work.

Service

Editorial Board — Frontiers in Computer Science Research Topic Editor (2026) Human Computer Interaction International - Cybersecurity, Privacy, and Trust (2026).

Committee — Association for the Advancement of Artificial Intelligence Program Committee (2025). Association for the Advancement of Artificial Intelligence Demonstration Program

Journal Reviewer — Nature Communications Psychology, Institute of Electrical and Electronics Engineers, Journal of Cognitive Science, Nature Scientific Reports, Discover Sustainability Journal, PLOS Computational Biology.

Book Reviewer — CRC Press: Taylor & Francis Group

Memberships — Cognitive Science Society, American Psychological Association, Institute of Electrical and Electronics Engineers, Association for the Advancement of Artificial Intelligence, Association for Computing Machinery, Association for Computational Linguistics

Presentations and Invited Lectures

2025 — "Human-AI Interdependence in Security and Privacy" TU Delft Workshop on Generative AI and Privacy/Security.

— "Hierarchical Instance-Based Learning for Decision Making from Delayed Feedback". Cognitive Science Society Conference

— "Using Cognitive Models to Improve Training Against Human and GPT-4 Generated Social Engineering Attacks" Human Computer Interaction International thematic area in Cybersecurity, Privacy, and Trust.

— "Modeling Attention during Dimensional Shifts with Counterfactual and Delayed Feedback" Reinforcement Learning and Decision Making Conference.

2024 — "Grounding Human-AI Collaboration with Cognitive Models" Leiden Institute of Advanced Computer Science University of Leiden

Experience (continued)

- "Leveraging a Cognitive Model to Measure Subjective Similarity of Human & GPT-4 Written Content" Conference on Natural Language Learning co-located with Empirical Methods in Natural Language Processing
- "Efficient Visual Representations for Learning and Decision Making" Human and Machine Cognition Lab University of Tubingen
- "Cognitive Large Language Models" Dynamic Decision Making Lab Carnegie Mellon University
- "Improving Human-LLM Collaboration in Cybersecurity using Cognitive Models" Software Engineering Institute Carnegie Mellon University
- "Human AI Teaming in Cyberdefense" Cybersecurity Assurance for Teams of Computers and Humans Multi-University Research Initiative
- "Efficient Visual Representations for Learning and Decision Making" Dynamic Decision Making Lab Carnegie Mellon University
- "Improving Cyber Defense Against Foundation Models with Cognitive Models" Carnegie Bosch Institute
- "Cognitive Modeling of Decision Making in Cybersecurity" Cognitive Security Institute
- "Computational Behavior Modeling" University of Virginia Course on Computational Behavior Modeling

2023

- "Accounting for Transfer of Learning Using Human Behavior Models" AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2024)
- "Phishing Education and Detection: Integrating Large Language & Cognitive Models" Conference of the Army Research Office
- "Learning to Defend by Attacking (and Vice-Versa): Transfer of Learning in CyberSecurity Games" IEEE European Conference on Security and Privacy
- "Designing Effective Masking Strategies for Cyberdefense through Human Experimentation and Cognitive Models in CyberVan" Conference of the Army Research Laboratory.
- "Comparing Models of Dynamic Adaptation from Crowd-Sourced Human Decision-Making" University of Alicante
- "Overcoming Biases and Constraints in Multi-Agent Games" University of Alicante
- "Generative Environment-Representation Instance-Based Learning: A Cognitive Model" AAAI Conference on Artificial Intelligence Fall Symposium Series

2022

- "Resource Rational Cognitive Modeling: an Information Theoretic Approach" PhD Thesis Defense Rensselaer Polytechnic Institute
- "A Beta-Variational Auto-Encoder Model of Human Visual Representation Formation in Utility-Based Learning" Vision Sciences Society.
- "Learning in Factored Domains with Information-Constrained Visual Representations" NeurIPS Conference Workshop on Information Theoretic Principals in Cognitive Systems

2021

- "Theory of Mind Planning" Neuro-Symbolic AI Workshop: IBM Research AI.
- "Human Inspired RL: Theory of Mind Games" RPI-IBM Artificial Intelligence Horizons Network Scholar Seminar Series.

2020

- "Deep RL with information constrained policies: Generalization in continuous control." Pompeu Fabra University: Theoretical and Cognitive Neuroscience