Jason Lucas

Data Sci. & AI Research Scientist| PIKE Research Lab

Eric J. Barron Innovation Hub, Penn State| 954.864.4944 | jsl5710@psu.edu | www.jasonslucas.com

EDUCATION

THE PENNSYLVANIA STATE UNIVERSITY
PH.D. IN INFORMATICS
Data Science and AI
GPA: 3.91
Aug. 2021 - Dec 2025 | University Park, PA

ST. GEORGE'S UNIVERSITY
MASTER OF PUBLIC HEALTH
Epidemiology
GPA: 3.83
Jan. 2018 - May 2021 | Grenada, WI

BOSTON UNIVERSITY
M.SC. IN COMPUTER INFO. SYS.
Health Informatics
GPA: 3.64
Jan. 2012 - Sept. 2014 | Boston, MA

ST. GEORGE'S UNIVERSITY
B.SC. IN INFORMATION TECHNOLOGY
GPA: 3.64
Aug. 2007 - May 2010 | Grenada, WI

PENN STATE LINDIV
LINGUISTICS & LANGUAGE SCIENCE
Certificate
Aug. 2022 - present | University Park, PA

INTEREST

Natural Language processing
Low-Resource Multilingual NLP
Adversarial Machine Learning
Adaptative and Alignment Tuning
Disinformation generation & Detection
Large Language Model & Hallucination

SKILLS

ARTIFICIAL INTELLIGENCE
Core Expertise:
• NLP • ML • DL • LLM
DATA SCIENCES
• Wrangling • Feature Engineering
• Visualization • Analytics
• Scrapping • Statistical & Predictive Analysis
PROGRAMMING
Proficient:
Python • SQL • PyTorch • LaTeX • R
Intermediate:
• C++ • CSS • PHP • HTML • Java
• Shell Scripting • TensorFlow
TRANSFERABLE
• Scientific Presentation • Project
Management • Technical Writing
• Grant Writing • Training
• Science Communication
PLATFORM/FRAMEWORK
Domain Specific:
• Epi Info • GeoDa • SAS
Development
VS Codes• Google Collab • Hugging Face
IntelliJ • Github • Oracle • MSSQL

EXPERIENCE

THE PENNSYLVANIA STATE UNIVERSITY | GRADUATE ASSISTANT
Aug. 2021 - Present | PA, US
- **Research Assistant**
  - Leads ML/NLP research under **Dr. Dongwon Lee** stewardship for combat disinformation and deepfake text across languages and distribution gaps
  - Mentors 2 research interns, assists advisor with grant writing, and yielding 6 top-tier publications (95+ citations) in `ACL`, `EMNLP`, `NAACL`, `COLING`, and `IEEE` venues
  - Design ML pipelines using `PyTorch`, `Hugging Face`, `AdapterHub`,and `APIs` (e.g., OpenAI)
  - Apply transfer learning (fine-tuning), prompt engineering, & in-context learning (pre-trained LLMs), Retrieval-Augmented Generation (RAG) techniques for multilingual detection, generation, and adversarial attacks/robustness.
- **Teaching Assistant**
  - Managed course delivery, assessments, office hours, feedback & attendance for 3 computing classes (150+ students)
  - Enhanced student engagement 40% through innovative teaching approaches
- **NSF NRT Fellow**
  - Leads interdisciplinary & transdisciplinary AI research through **LINDIV** with $9K NSF grant funding
  - Collaborate with 10+ diverse researchers across 5 departments, advancing human-AI interaction

ST. GEORGE'S UNIVERSITY | FACULTY
June 2011 – July 2021 | Grenada, WI
- **Instructor, Head of Division & Technical Lead**
  - Led 6-person team supporting 4000+ medical students across 3 school terms
  - Achieved 99.0% uptime managing SimCapture/Examsoft evaluation platforms
  - Taughts 3+ courses across IT, Medicine, and Public Health programs

GOVERNMENT OF GRENADA | INFORMATION TECHNOLOGY INSTRUCTOR
Aug 2010 – Jul 2011 | Grenada, WI
- Supervise class & managed IT infrastructure for 500+ students.
- Achieved 95% pass rate in CSEC IT examinations through targeted instruction
- Implemented digital literacy program reaching 300+ students

INTERNSHIP

05/01/2024	**Interaction LLC** • NLP, Intelligent Virtual Assistant, LLMs, Prompt Engineering
05/01/2023	**Coalfire** • Cybersecurity, security compliance and standard, SOC, PCI
02/01/2010	**Flow Grenada** (LIME: Cable & Wireless) Networking, Mobile Tech, Database, IT Supp.

AWARD

2024	Pan-African Professional Alliance Leadership, PSU
2018	Chancellor's List for Academic Excellence, MPH
2012	Excellence and Dedication to OSCE, SGU
2009	Dean's List for Academic Excellence, B.Sc
2008	Chancellor's List for Academic Excellence, B.Sc

FELLOWSHIP

2022-24	Linguistic Diversity Across the Lifespan, National Science Foundation.
2021-22	Funds for Excellence in Graduate Recruitment, Penn State University.
2018-20	Graduate Faculty Fellowship, St. George's University
2007-10	Undergraduate Scholarship, Government of Grenada.

LINKS

SERVICE

Committee Member • MASC-SLL '25
Committee Member • COLING '24
Committee Member • IST GBoR'24
Peer Reviewer • ACL' 24
IT Officer •Pan-African Prof. Alliance' 23
President • Graduate Student Assc' 22.
Outreach Coordinator •LinDiv' 22
Volunteer •EMNLP Volunteer' 23
IT Officer • Red Cross Grenada' 20

MENTORSHIP

Lia Carin Djaouga •Millennium Scholar
Kendall Reed II •Millennium Scholar

TRAINING

Interdesciplinary Research -IST PSU
• Data Mining • Social Informatics
• Human Computer Interactions
Transdesciplinery Research -LinDiv
• Generative Syntax • Linguistics
• Transdesciplinery Team Research
Multilingual NLP -CLS & IST PSU
• Transfer Learning • Prompt Engineering
• Adversarial ML & DL • Generative AI
• Conversational AI Epidemiology -SGU
• Data Management & Analysis
• Chronic & Infectious Disease Epi
Health Informatics -BU
• Information Structures in Python
• Advance Health Informatics

INVITED TALK

WXXI News NPR •Guest Speaker'24
Cohere For AI •Guest Speaker'24
St. George's University •Invited Talk'23
CRA-WP IDEALS •Lightening Talk'23
NRT Annual Meeting •Lightening Talk'23

TAUGHT COURSES

Application Development • IST 140
Organization of Data •IST 210
Into. Public Health •PUBH 302
Database System •Comp 420

PUBLICATIONS

[1] Ekaterina Artemova, **Lucas, Jason**, Saranya Venkatraman, Jooyoung Lee, Sergei Tilga, Adaku Uchendu, and Vladislav Mikhailov. Beemo: Benchmark of expert-edited machine-generated outputs. *arXiv preprint*, 2025. Accepted at NAACL 2025.

[2] Ali Al Lawati, **Jason Lucas**, and Prasenjit Mitra. **Semantic Captioning: Benchmark Dataset and Graph-Aware Few-Shot In-Context Learning for SQL2Text**. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8026–8042, 2025. COLING 2025.

[3] Dominik Macko, Robert Moro, Adaku Uchendu, Ivan Srba, **Lucas, Jason Samuel**, Michiharu Yamashita, Nafis Irtiza Tripto, Dongwon Lee, Jakub Simko, and Maria Bielikova. **Authorship Obfuscation in Multilingual Machine-Generated Text Detection.** In *2024 Findings of the Association for Computational Linguistics: EMNLP*, pages 13706–13724. **Association for Computational Linguistics**, 2024.

[4] **Lucas, Jason Samuel**, Maung Maung Barani, Tabar Maryam, McBride Keegan, and Lee Dongwon. **The Longtail Impact of Generative AI on Disinformation: Harmonizing Dichotomous Perspectives.** *IEEE Intelligent Systems*, 2024.

[5] Maung Maung Barani, McBride Keegan, **Lucas, Jason Samuel**, Tabar Maryam, and Lee Dongwon. **Generative AI Disproportionately Harms Long Tail Users**. *IEEE Computer*, *Volume 57, Issue 11*, 2024.

[6] **Lucas, Jason**, Adaku Uchendu, Michiharu Yamashita, Jooyoung Lee, Shaurya Rohatgi, and Dongwon Lee. **Fighting Fire with Fire: The Dual Role of LLMs in Crafting and Detecting Elusive Disinformation**. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: EMNLP*, pages 14279–14305, Singapore, December 2023. **Association for Computational Linguistics**.

[7] Dominik Macko, Robert Moro, Adaku Uchendu, **Lucas, Jason**, Michiharu Yamashita, Matúš Pikuliak, Ivan Srba, Thai Le, Dongwon Lee, Jakub Simko, and Maria Bielikova. **MULTITuDE: Large-Scale Multilingual Machine-Generated Text Detection Benchmark**. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: EMNLP*, pages 9960–9987, Singapore, December 2023. **Association for Computational Linguistics**.

[8] **Lucas, Jason**, Limeng Cui, Thai Le, and Dongwon Lee. **Detecting False Claims in Low-Resource Regions: A Case Study of Caribbean Islands**. In Tanmoy Chakraborty, Md. Shad Akhtar, Kai Shu, H. Russell Bernard, Maria Liakata, Preslav Nakov, and Aseem Srivastava, editors, *Proceedings of the Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situations: ACL*, pages 95–102, Dublin, Ireland, May 2022. **Association for Computational Linguistics**.

PROJECTS

RESEARCH INNOVATION PROJECTS | PENN STATE UNIVERSITY LINDIV NSF NRT
Lead Researcher & App Developer | 2023 - Present
- BAbSANT: AI-Powered Speech Pathology Platform
    – Engineered multilingual conversational-AI platform integrating GPT-4 & Whisper for speech pathology across 5+ languages
    – Built Streamlit interface reducing assessment time by 60% while expanding accessibility
    – Achieved 98% accuracy in simulation & controllability through in-context learning optimization

HEALTHCARE TECHNOLOGY INITIATIVES | ST. GEORGE'S UNIVERSITY
Technical Lead & Project Manager | 2019 - 2021
- Online Clinical Evaluation Exercise (OCEX)
    – Led university-wide virtual clinical examination system supporting 3000+ students
    – Accelerated examination processing by 40% through 5-platform integration
    – Achieved 98% satisfaction rating from students & standardized patients
- Simulation Lab Modernization
    – Managed B-Line Medical lab expansion serving 4K+ students annually
    – Led 15-member team across 4 departments, increasing capacity by 250%
    – Optimized space utilization enabling multi-disciplinary simulation training