Liam Dugan

ldugan@seas.upenn.edu | linkedin.com/in/liam-dugan | liamdugan.com

Research Focus

I specialize in AI safety research, focusing on human and automated detection of AI-generated content. My work explores the limitations of detection and investigates strategies for deploying detectors with minimal harm. I have previously worked on evaluating applying large language models to tasks such as educational question generation, machine translation, and multi-document question-answering.

Education

University of Pennsylvania — Philadelphia, PA

Ph.D, Computer Science (Advisor: Chris Callison-Burch) — *Aug. 2021 – est. Aug 2026*

M.S.E, Robotics — *Aug. 2017 – Dec. 2020*

B.S.E, Computer Engineering & East Asian Studies — *Aug. 2015 – Aug. 2020*

Work Experience

Summer 2022 **Roblox** (San Mateo, CA) - *PhD Research Intern*
Hosts: Morgan McGuire & Victor Zordan
Project: *Real-Time Speech-to-Speech Translation*

Summer 2021 **John's Hopkins University** (Baltimore, MD) - *Visiting Research Scholar*
Hosts: Kevin Duh, Paul McNamee, Matt Post
Project: *Fine-Tuning MT for Cross-Lingual IR*

Summer 2019 **NVIDIA** (Santa Clara, CA) - *Autonomous Driving Software Intern*
Host: Gajanan Bhat
Project: *Docker Image Server for Self-Driving Cars*

Summer 2018 **Forterra** (Clarksburg, MD) - *Software Engineering Intern*
Host: Anne Schneider
Project: *Velodyne VLP-16 LIDAR Point Cloud Classifiers*

Publications

2024 Josh Ludan, Qing Lyu, Yue Yang, **Liam Dugan**, Mark Yatskar, and Chris Callison-Burch. Interpretable-by-Design Text Understanding with Iteratively Generated Concept Bottleneck. In *Proceedings of the 2nd Workshop on Interpretable AI: Past, Present and Future (IAI) @ NeurIPS 2024*, Vancouver, Canada, 2024. (In Submission)

Andrew Zhu, **Liam Dugan**, and Chris Callison-Burch. ReDel: A Toolkit for Recursive Multi-Agent Systems. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024 System Demonstrations*, Miami, Florida, November 2024. (To Appear)

Runsheng Huang, **Liam Dugan**, Yue Yang, and Chris Callison-Burch. MiRAGeNews: Multimodal Realistic AI-Generated News Detection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024*, Miami, Florida, November 2024. (To Appear)

Andrew Zhu, Alyssa Hwang, **Liam Dugan**, and Chris Callison-Burch. FanOutQA: A Multi-Hop, Multi-Document Question Answering Benchmark for Large Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 18–37, Bangkok, Thailand, August 2024

Liam Dugan, Alyssa Hwang, Filip Trhlík, Andrew Zhu, Josh Magnus Ludan, Hainiu Xu, Daphne Ippolito, and Chris Callison-Burch. RAID: A Shared Benchmark for Robust Evaluation of Machine-Generated Text Detectors. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12463–12492, Bangkok, Thailand, August 2024. Association for Computational Linguistics

2023 Andrew Zhu*, **Liam Dugan***, Alyssa Hwang, and Chris Callison-Burch. Kani: A Lightweight and Highly Hackable Framework for Building Language Model Applications. In *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)*, pages 65–77, Singapore, Singapore, December 2023. Empirical Methods in Natural Language Processing

Liam Dugan, Anshul Wadhawan, Kyle Spence, Chris Callison-Burch, Morgan McGuire, and Victor Zordan. Learning When to Speak: Latency and Quality Trade-offs for Simultaneous Speech-to-Speech Translation with Offline Models. In *Proc. INTERSPEECH 2023*, pages 5265–5266, August 2023

Hannah Gonzalez, **Liam Dugan**, Eleni Miltsakaki, Zhiqi Cui, Jiaxuan Ren, Bryan Li, Shriyash Upadhyay, Etan Ginsberg, and Chris Callison-Burch. Enhancing Human Summaries for Question-Answer Generation in Education. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 108–118, Toronto, Canada, July 2023. Association for Computational Linguistics

Li Zhang*, **Liam Dugan***, Hainiu Xu*, and Chris Callison-Burch. Exploring the Curious Case of Code Prompts. In *Proceedings of the 1st Workshop on Natural Language Reasoning and Structured Explanations (NLRSE)*, pages 9–17, Toronto, Canada, June 2023. Association for Computational Linguistics **Selected for Oral Presentation**

Liam Dugan*, Daphne Ippolito*, Arun Kirubarajan, Sherry Shi, Chris Callison-Burch. Real or Fake Text?: Investigating Human Ability to Detect Boundaries between Human-Written and Machine-Generated Text. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37(11), pages 12763–12771, Washington, D.C., Jun. 2023. **Selected for Oral Presentation**

Aarohi Srivastava, Abhinav Rastogi, and **(440 others)**. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*, May 2023

2022 S. Upadhyay, E. Ginsberg, **Liam Dugan**, E. Miltsakaki, H. Gonzalez, D. Choi, C. Yuan, and C. Callison-Burch. Question Generation for Textbook Flashcards. In *EDULEARN22 Proceedings*, 14th International Conference on Education and New Learning Technologies, page 3412. IATED, July 2022

Daphne Ippolito, **Liam Dugan**, Emily Reif, Ann Yuan, Andy Coenen, and Chris Callison-Burch. The Case for a Single Model that can Both Generate Continuations and Fill-in-the-Blank. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2421–2432, Seattle, United States, July 2022

Liam Dugan, Eleni Miltsakaki, Shriyash Upadhyay, Etan Ginsberg, Hannah Gonzalez, DaHyeon Choi, Chuning Yuan, and Chris Callison-Burch. A Feasibility Study of Answer-Agnostic Question Generation for Education. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1919–1926, Dublin, Ireland, May 2022

2020 **Liam Dugan**. Learning Formality from Japanese-English Parallel Corpora. Master's thesis, University of Pennsylvania, December 2020

Liam Dugan*, Daphne Ippolito*, Arun Kirubarajan*, and Chris Callison-Burch. RoFT: A Tool for Evaluating Human Detection of Machine-Generated Text. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 189–196, Online, October 2020. Association for Computational Linguistics

2019 Zhengyi Luo, Austin Small, **Liam Dugan** and Stephen Lane. Cloud Chaser: Real Time Deep Learning Computer Vision on Low Computing Power Devices. In *Eleventh International Conference on Machine Vision (ICMV 2018)*, volume 11041, page 110412Q. International Society for Optics and Photonics, SPIE, 2019

Invited Talks

2024 RAID: A Shared Benchmark for Robust Evaluation of Machine-Generated Text Detectors: University of Maryland, College Park MD, June 2024 (slides)

2023 Should we still use Text for Speech-to-Speech Translation? Promise meets Practice: John's Hopkins University, Baltimore MD, May 2023 (slides)

Real or Fake Text: Investigating Human Ability to Detect Boundaries between Human-Written and Machine Generated Text: Brown University, Providence RI, March 2023 (slides)

Detecting Generated Text from ChatGPT and other LLMs: University of Pennsylvania, Philadelphia PA, February 2023 (slides)

2022 Intro to Machine Learning and AI Research: St. Joseph's Preparatory High School, Philadelphia PA, Feb. 2022 (slides)

Are Humans Able to Detect Boundaries between Human-Written and Machine-Generated Text?: University of Pennsylvania Computational Linguistics Lunch (CLUNCH), Philadelphia PA, Jan. 2022

News Articles

(9/9/24) AI Detectors are Easily Fooled, Researchers Find - EdScoop
(9/3/24) Teachers still can't trust AI text checkers - Axios
(8/14/24) Putting AI Text Detectors to the Test: From Hype to Hard Data - SafeAI@Penn Newsletter
(8/12/24) Detecting Machine-Generated Text: An Arms Race With the Advancements of Large Language Models - Penn Engineering Today
(5/21/24) Originality.ai is the Most Accurate AI Detector According to an Extensive Study "RAID" - Originality.ai Blog
(1/2/24) Can Humans Learn To Spot Fake Text? - Penn Engineering Magazine
(9/20/23) Unlocking AI Potential: Unveiling Kani, the Groundbreaking Open-Source Framework Revolutionizing Large Language Model Applications - CJ&CO
(9/19/23) Researchers from the University of Pennsylvania Introduce Kani: A Lightweight, Flexible, and Model-Agnostic Open-Source AI Framework for Building Language Model Applications - MarkTechPost
(9/19/23) Kani: A Lightweight and Customizable Framework for Language Model Applications - TS2
(8/7/23) AI 'Watermarking' Tools Emerging to Tag Machine-Made Content - Bloomberg Law
(7/27/23) CNN Features Penn Engineering AI Research - Penn Engineering Today
(7/19/23) Academic Integrity and AI: Is Detection the Answer? - Temple Center for Teaching
(7/11/23) Bot or not? How to tell when you're reading something written by AI - CNN
(5/18/23) NewsChannel12 Investigates: Artificial Intelligence Part III - ABC News North Carolina
(4/26/23) Alien Minds, Immaculate Bullshit, Outstanding Questions - The Pennsylvania Gazette
(4/18/23) How can people navigate AI-generated misinformation? - Canvas 8
(4/11/23) Reddit Moderators Brace for a ChatGPT Spam Apocalypse - Vice
(3/10/23) Real or fake text? We can learn to spot the difference - Penn Today
(3/8/23) A Bot Isn't Going to Take Your Place, But AI Will Make Your Job Harder - Corporate Compliance Insights
(3/8/23) New Study Shows People Can Learn to Spot Machine-Generated Text - UniteAI
(3/6/23) How can humans detect AI writing? These Penn researchers have some tips - Technically Philly
(3/3/23) Can Humans Detect Text by AI Chatbot GPT? - Psychology Today
(3/2/23) People can learn to detect AI writing - Cosmos Magazine
(2/27/23) Real or Fake Text? We Can Learn to Spot the Difference - Penn Engineering Today
(12/19/22) How to spot AI-generated Text - MIT Technology Review
(1/23/18) Object-Seeking Robot Wins PennApps XVII - Penn Engineering Today
(9/10/17) At PennApps XVI, students made inter-dimensional robots and hung out with the founder of Quora - The Daily Pennsylvanian

Teaching

Summer 2023 **Teaching Assistant for CIS530, *Computational Linguistics***
Taught by Chris Callison-Burch. Wrote homework "Fine-Tuning Pre-Trained Language Models"

Fall 2022 **Teaching Assistant for CIS700, *Research Practicum***
Taught by Chris Callison-Burch

Spring 2022 **Teaching Assistant for CIS700, *Interactive Fiction & Text Generation***
Co-Taught by Chris Callison-Burch and Lara Martin

Fall 2021 **Teaching Assistant for CIS565, *GPU Programming & Architecture***
Taught by Shehzan Mohammed. Gave two guest lectures, "Optimizing Machine Learning with CUDA" and "Introduction to Machine Learning". Mentored students with ML final projects

Fall 2020 **Teaching Assistant for CIS530, *Computational Linguistics***
Taught by Clayton Greenberg. Wrote homework "Transformers and State-of-the-Art Models"

Spring 2020 **Teaching Assistant for CIS530, *Computational Linguistics***
Taught by Chris Callison-Burch. Wrote homework "Neural Machine Translation"

Fall 2019 **Head Teaching Assistant for CIS380, *Operating Systems***
Taught by Boon Thau Loo. Re-wrote homework write-ups and developed autograders
Gave guest lecture "Linux Page Replacement Algorithms and Belady's Anomaly"
Achieved highest ever course rating in TA Quality (3.37/4), and Overall Quality (3.29/4)

Spring 2019 **Teaching Assistant for CIS548, *Operating Systems***
Taught by Boon Thau Loo.

Fall 2018 **Teaching Assistant for CIS380, *Operating Systems***
Taught by Boon Thau Loo.

Spring 2018 **Teaching Assistant for CIS240, *Intro to Computer Systems***
Taught by Thomas Farmer.

Fall 2017 **Teaching Assistant for CIS240, *Intro to Computer Systems***
Taught by Camillo Jose Taylor.

Fall 2017 **Teaching Assistant for SD4x, *Programming for the Web with JavaScript***
Co-Taught by Chris Murphy and Swapneel Sheth.

Spring 2017 **Teaching Assistant for CIS240, *Intro to Computer Systems***
Taught by Thomas Farmer.

Fellowships, Awards, And Honors

Aug 2022 **Roblox Research Grant**
Funding granted to continue work into speech-to-speech translation for the 2022-2023 academic year

Oct 2021 **Google Cloud Platform Research Grant**
For the server and compute costs of the Real or Fake Text website (http://roft.io)

May 2020 **Penn Engineering Exceptional Service Award**
For my work as Head Teaching Assistant for CIS380 (Operating Systems)

May 2019 **Penn Engineering Senior Design Award**
For my Senior Design Project (Scene++) [see video]

Oct 2018 **Foreign Language and Area Studies Undergraduate Fellowship**
Funding granted to continue my Master's Thesis research into east asian language NLP applications

Jan 2018 **Grand Prize & Best use of Cloud Hosting: PennApps XVII**
For my project Cloud Chaser (1st place out of 156 teams) [see video]

Sept 2017 **Third Prize: PennApps XVI**
For my project Todd: The Inter-Dimensional Robot (3rd place out of 158 teams) [see video]

Mentorship

2024 Tony An, Andrew Jiang, Ishaan Lal, Joseph Lee, Nathaniel Lao (Senior Design)
—Project: En Poisson **[Won Best Overall Project in CIS]**

Josh Magnus Ludan (Independent Study) — Current Position: PhD at University of Pennsylvania

Emma Jin (Independent Study)

Runsheng (Anson) Huang (Independent Study)

Filip Trhlik (Independent Study)

2023 Maya Guru, Yiran Chen, Sahit Penmatcha, Kaitlynn Soo, V. Veeramachaneni (Senior Design)

—Project: Dubble [**Won M&T Integration Lab Finalist & Judge Harold Berger Award**]

River Yijiang Dong (Independent Study) — Current Position: PhD at Cambridge University

Hainiu Xu (Independent Study) — Current Position: PhD at King's College London

Hannah Gonzalez (Independent Study) — Current Position: PhD at John's Hopkins University

Charlie Chen (Independent Study)

Anshul Wadhawan (Master's Thesis)

2022 Shriyash Upadhyay & Etan Ginsberg (Independent Study)

—Co-Founders at Martian [**Raised $9M Seed Funding**]

Service

Reviewing: EMNLP 2024, ACL 2024, ACL 2023 (Main Track + Demo Track), ACL 2021
Organization: CLunch Fall 2024, PennStatNLP Reading Group Fall 2024, GenAI Workshop @ COLING

Technical Skills

Natural Languages: English (native), Japanese (advanced, business fluent - 7+ years [JLPT N2])
Programming Languages: Python, C, C++, Java, bash, CUDA, MATLAB, JavaScript, HTML/CSS, Verilog, Go
Frameworks: HuggingFace, PyTorch, Pandas, Numpy, Tensorflow, OpenCV, DXR, Vulkan
Developer Tools: Git, Slurm, VS Code, emacs, tmux, Google Cloud Platform, Docker, QSub