Sander Schulhoff

sanderschulhoff@gmail.com | trigaten.github.io | 410-805-2290

SUMMARY

I am a natural language processing (NLP) and deep reinforcement learning (DRL) researcher. I wrote the first guide on prompt engineering on the Internet (Learn Prompting) and ran the first ever competition on prompt injection (HackAPrompt). I am the CEO of Learn Prompting and HackAPrompt as well as a MATS scholar.

RESEARCH

Lead Organizer of HackAPrompt
Spring 2023 - Summer 2023

Learn Prompting, University of Maryland

- Organized the first ever Prompt Injection competition
- Collected the largest ever dataset of prompt injections (600K+ adversarial prompts) and published the first taxonomical ontology of Prompt Hacking
- Raised 40K in sponsorship from OpenAI, Preamble, Scale AI, Stability AI, HuggingFace, and 7 other companies
- Won Best Theme Paper at EMNLP2023 out of 20K papers, presented live to 2K researchers[1]
- Cited by OpenAI, Google, Anthropic, MSFT, Meta, IBM, Tencent, Andreesen Horowitz, Salesforce, all Ivy League research institutions, and more.

Lead Author of The Prompt Report
Fall 2023 - Spring 2024

University of Maryland, Learn Prompting

- Led a team of 32 researchers from OpenAI, Google, Microsoft, Princeton, Stanford, Maryland, etc. to write the most comphrensive systematic literature review of prompting ever.
- Summarized 1000s of papers on prompting into a massive 80 page survey paper.
- Went viral on the Internet, with millions of views. This paper has inspired the creation of many Python libraries, been used in Enterprise Documentation, is used to interview job candidates, and was featured by UMD.
- #1 Paper of the Year (Latent Space), #1 Paper of the Day (HuggingFace), featured in multiple podcasts [1] [2] [3]

ML Alignment and Theory Scholar (MATS)
Summer 2025

Lighthaven, Professor Florian Tramèr and Daniel Paleka

- Performing agentic security research by adapting AgentDojo for a massive online human red-teaming study.

MineRL Organizing Team
Summer 2022 - Summer 2023

MineRL Labs

- Helped organize and promote a competition on training deep reinforcement learning agents in Minecraft
- Helped write relevant award-winning publication (Section 8) as well as project documentation

Stabilizing Hostilities through Arbitration and Diplomatic Engagement
Spring 2022 - Spring 2023

University of Maryland, Professor Jordan Boyd-Graber

- Managed team of graduate and undergraduate students on a DARPA funded multi-university project with the goal of building bots to play the boardgame Diplomacy and communicate with players
- Ideated and pitched 4 bot ideas: Janus Bot, Janus Bot [Shortened], Janus Bot [MVP], and SOA Bot
- Built CI pipelines, including Dockerized testing, precommit, and Vercel integration for the documentation website
- Wrote a report on DAIDE-English translation and built a DAIDE syntax parser package

Andreou Lab Internship
Summer 2021 - Fall 2021

Johns Hopkins University, Professor Andreas G. Andreou

- Built data pipelines to record data from Arduino chips and RealSense cameras and send it to Microsoft \Psi
- Implemented a CNN+GRU model from scratch with Pytorch for video classification on event camera data

Botany Researcher
Fall 2023

University of Maryland

- Identified and mapped 100s of wild plants on the UMD campus (report, website)
- Engaged 100s of campus stakeholders and successfully obtained a UMD sustainability grant
- Built a biogas reactor from Home Depot parts

CLIP Undergraduate Research
Fall 2020 – Summer 2021

University of Maryland, Professor Jordan Boyd-Graber
College Park, MD

---

[1] I was the second undergraduate ever to win this award.

- Worked on Diplomacy boardgame NLP project developing a Discord bot used to collect player data and display machine learning predictions (Bot passed live Alpha and Beta tests with paid participants)
- Developed a simple full-stack website to collect data for a NLP analogy project
- Built an annotation workflow for the UMD QANTA project

Neurodata Lab Internship
Johns Hopkins University, Professor Joshua Vogelstein

Summer 2019

Baltimore, MD

- Wrote unit tests for a project converting clustering algorithm libraries written in R to Python
- Investigated different clustering metrics where ground truth is known based on the Zachary's Karate Club social network graph dataset

WORK EXPERIENCE

Learn Prompting | *Javascript, Markdown, NextJS*   Winter 2022 - Present
- Wrote the first ever guide on prompt engineering
- Reached 3 Million+ users, including people from every Fortune 500 and Generative AI company.
- Cited by NIST, Wikipedia, many research papers, OpenAI, and 100s of Youtube videos and articles

Semiotic Labs | *Python, Julia*   Summer 2022
- Developed software for performing deep reinforcement learning to price The Graph protocol subgraph queries
- Implemented continuous policy gradient bandits (VPG, PPO) in Julia for adaptive query pricing

Farama Foundation | *Python, Jekyll, Liquid, HTML, CSS*   Spring 2022 – Summer 2022
- Led development of new Gymnasium website: managed PRs and wrote scripts to generate pages, menus, and gifs
- Made a number of fixes to Gymnasium, built and refactored fantasia Jekyll theme from PettingZoo website, wrote and published gym-notices PyPI package, assisted in development of PettingZoo website

Teacher at Friends School of Baltimore | *HTML, CSS*   Fall 2021 – Spring 2022
- Co-taught an introductory web development/design course to highschoolers

Axidraw Control Software | *HTML, CSS, Javascript, Python*   Spring 2020
- Hired to build a Mac application for automating the process of writing physical notes

EDUCATION

University of Maryland, College Park   Fall 2020 - Spring 2024
Bachelor of Science in Computer Science, graduated on Dean's List, with Honors   *GPA: 3.8*

SELECTED PROJECTS

Wetland Mitigation Banking Data Analysis | *Python*   2022
- Co-founded a firm analyzing wetland performance using OMBIL Regulatory Module and Regulatory In-lieu fee and Bank Information Tracking System data
- Recommended investment locations based on country-wide analysis to clients

Teacher Recommender System | *Google Apps Scripts, HTML, MaterializeCSS*   Spring 2019 – Spring 2020
- Developed suite of scripts and menus to automate the process of matching students with teachers who will write their college recommendation letters
- Sold Alpha version and released Beta version as an official Google Sheet Add-On

Heart Heist App | *Swift, Objective-C*   Summer 2018
- Developed code, art, and music for a top down shooter app using XCode and deployed it on the Apple App Store

OPEN SOURCE CONTRIBUTIONS

MineRL (Minecraft Reinforcement Learning Library): Committed 1000+ lines of documentation, bug fixes, and feature additions, including a tutorial on custom environment building. Also ported (from Project Malmo) the ability to send chat messages in MineRL and wrote a tutorial page for it. This functionality allows significant speed ups for training agents.

OpenAI Gym/Gymnasium (Reinforcement Learning Library): Led the development of new website, organized various documentation additions, and made codebase changes.

Learn Prompting (Prompt Engineering Guide): Led the development of the first prompt engineering guide, managed 100+ contributors, gained 4K+ stars.

Technical Skills

Frequent Languages: Python, Google Apps Script, Javascript, Java, HTML, CSS, C, Markdown
Frequent Developer Tools: Visual Studio Code, Git, Docker, Adobe XD, Photoshop, Google Drive Scripting Environment
Natural Languages: English, Spanish

Infrequent Languages and Tools: PHP, MySQL, Bash, Objective-C, C++, C#, Eclipse, IntelliJ IDEA, Pycharm, MAMP, SQLPro, PHPMyAdmin, Platform IO, Visual Studio, Atom, Jupyter Notebook, Google Colab, XCode, Unity

Achievements

Won UMD Hackathon 2022 (Best "Bitcamp" Hack) with Marshie's Adventure, a Bitcamp themed platformer
Won UMBC Hackathon 2020 (Cipher-Tech-Solutions challenge) with ForeTrackR, a novel application of Blockchain to Chain of Custody in digital forensics
Won True Bit design competition (UMBC Hackathon)
Won University of Maryland Writing Competition (Alternative Media Writing Award) for Learn Prompting

Selected Publications (Google Scholar)

The Prompt Report: A Systematic Survey of Prompting Techniques (arXiv)

Sander Schulhoff, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal, Chau Pham, Gerson Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Hevander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncearenco, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle, Philip Resnik

Ignore This Title and HackAPrompt: Exposing Systemic Vulnerabilities of LLMs through a Global Scale Prompt Hacking Competition (Best Theme Paper EMNLP2023)

Sander Schulhoff, Jeremy Pinto, Anaum Khan, Louis-François Bouchard, Chenglei Si, Svetlina Anati, Valen Tagliabue, Anson Liu Kost, Christopher Carnahan, Jordan Boyd-Graber

The MineRL BASALT Evaluation and Demonstrations Dataset (NeurIPS2023, Outstanding Paper ICML2024)
Stephanie Milani, Anssi Kanervisto, Karolis Ramanauskas, **Sander Schulhoff**, Brandon Houghton, Rohin Shah

Towards Solving Fuzzy Tasks with Human Feedback: A Retrospective of the MineRL BASALT 2022 Competition (NeurIPS2022)

Stephanie Milani, Anssi Kanervisto, Karolis Ramanauskas, **Sander Schulhoff**, Brandon Houghton, Sharada Mohanty, Byron Galbraith, Ke Chen, Yan Song, Tianze Zhou, Bingquan Yu, He Liu, Kai Guan, Yujing Hu, Tangjie Lv, Federico Malato, Florian Leopold, Amogh Raut, Ville Hautamäki, Andrew Melnik, Shu Ishida, João F. Henriques, Robert Klassert, Walter Laurito, Ellen Novoseller, Vinicius G. Goecks, Nicholas Waytowich, David Watkins, Josh Miller, Rohin Shah

GPT Deciphering Fedspeak: Quantifying Dissent Among Hawks and Doves (EMNLP Findings)

Denis Peskoff, Adam Visokay, **Sander Schulhoff**, Benjamin Wachspress, Alan Blinder, Brandon M. Stewart

Gymnasium (arXiv)

Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U. Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, Rodrigo Perez-Vicente, Andrea Pierré, **Sander Schulhoff**, Jun Jet Tai, Hannah Tan, Omar G. Younis

Invited Talks

The State of Prompt Hacking, given at OpenAI, ML Commons (AI Safety Group), and The Access Group, 100 attendees each.
Best Theme Paper Award Acceptance Speech, given at EMNLP2023, 2,000 attendees
The State of Prompt Engineering, given at AI4 2024 and Generative AI Week 2024, 200 attendees
The State of Prompt Engineering and Prompt Security, given at a Big 5 consulting company, 500 attendees
The State of Prompt Engineering and LLM Security, given at Microsoft, 1,700 attendees.
The State of Prompt Hacking and LLM Security, given at Stanford.
Lightening talk at UMD government language day on prompt engineering

Podcasts

The Ultimate Guide to Prompting: **Latent Space** (watch here)

Delving into The Prompt Report: **The Cognitive Revolution** (watch here)

Ignore Previous Instructions & Listen to This Interview, LearnPrompting.org: **The Cognitive Revolution** (watch here)

Become a Pro at Prompt Engineering: Exclusive Tips from LearnPrompting Creator: **What's AI** (watch here)

The Impact of Prompt Injection and HackAPrompt: **The Security Table** (watch here)

Generative AI Prompt Hacking and Its Impact on AI Security & Safety: **MLSecOps** (watch here)

Sander Schulhoff's Learn Prompting is teaching 3 million people how to use AI: **Prompt Security** (watch here)