Coen D. Needell

☎ +1 (970) 456 2411
✉ coen@needell.org
🌐 coen.needell.org
⌨ SoyBison
in coen-needell-b4503216a

▬▬▬ Education

2019 – 2021   **Master of Arts in Computational Social Science**, *University of Chicago*, Chicago, IL
Thesis on Deep Learning and Human Memory.

2015 – 2019   **Bachelor of Arts in Economics and Physics**, *Washington University in St. Louis*, St. Louis, MO
Minor in the Philosophy of Science

▬▬▬ Experience

2023 –   **Audio Data Lead / Data Engineering**, *Reality Defender, Inc.*, New York, NY

Built scaleable tools for the AI model development cycle, designed for the specific needs of Deepfake Detection models. Designed and implemented dataset pathways for a rapidly growing, research and development orientated startup. Worked directly with AI scientists both implementing tools for them to execute research projects more quickly and provide more information for the research iteration process. Worked with scientists to build novel, research-driven and domain-focused analytics tooling for improving production models. Served as technical lead during zero-to-one development on a generalized Audio-AI research pipeline. Specifics include:

- Took on the lead role for the organization's dataset management and model develoment pipeline. Scaled the organization's research and development operation from training on datasets with hundreds of hours of data to hundreds of thousands of hours of data and beyond.
  - Built a scaleable, asynchronous pipeline for generating voice clones from arbitrary audio across all commercial platforms, utilizing functional and type-driven programming techniques to minimize the labor cost of implementing new AI features.
  - Designed and implemented a cloud-native framework for scaling and deploying content generation models and AI-powered analysis tools across different modalities and techniques. Designed with experimental models in mind, for rapid research on detection effectiveness, and if needed, rapid deployment of updated model checkpoints.
- Directed and established a pipeline and process for evaluating the impact of focused dataset expansion work as well as tooling for analyzing the impact of modeling changes.
  - Directed the use of internal data science and engineering tooling for product deployments, ensuring that all models are tested in-situ on unseen data before and after deployment. These tools and processes ensured client trust in deployments to high-cybersecurity environments.
  - Using the information from these evaluations, synthesized and executed plans for improving the training set using data procurement, augmentation, and the implementation of additional generative models.
  - This process resulted in the rapid improvement of the audio detection model from the low 80% accuracy range on public benchmarks to the 95%+ accuracy range. This led to the company winning client competitions handedly, and after a few iterations, the company won first place in the ASVSpoof 5 single-model competition.
- Overhauled the research team's audio preprocessing pipeline from a collection of single threaded scripts to a highly modular and cloud-level parallelized system of serverless microservices, enabling research scientists to perform augmentation experiments at terabyte-scale, just by providing simple python functions. Coupled with a friendly interface, this also improved the real-world speed of the research-to-production iteration loop from taking months to weeks.
- Developed a novel, computational-linguistics-driven approach to Answering Machine Detection.
- Developed a novel approach to analyzing the real-world representativeness of an arbitrary dataset, leveraging information theoretic tools and high dimensional embedding spaces.
- Designed and Implemented a tool for scaling and automating the simulation of phone calls at scale for research and production AI-model training and evaluation.
- Built and Maintained an internal knowledge-base for the organization's datasets, model iterations, production systems, and known deepfake generation tools. Designed socio-technical integrations with the knowledge-base to ensure project visibility and provide semi-automated "sanity checks" throughout the research pipeline.
- Oversaw the work of other Data Engineers, from designing and executing crowdsourcing studies, to implementing internal research tooling, to scaling novel data augmentations for research and development.

2023   **Freelance Researcher / Research Engineer**, *University of Chicago, Department of Psychology*, Chicago, IL

Developed a novel approach using deep learning for estimating auditory memorability with Wilma A. Bainbridge. Built large scale training tools for this purpose to reduce the Lab's iteration time, decrease compute costs, and execute modern AutoML practices.

2022 – 2023   **Predoctoral Researcher / Research Engineer**, *Computational Social Science, University of Pennsylvania*, Philadelphia, PA

Researched topics relating to the News, especially how the apparent "Speed of News" has been increasing. Developed the *Living Journal Toolkit*, a collection of tools for publishing live updating dashboards associated with papers, blog posts using a novel statistical and visualization package, and interactive papers. Developed the *News Observatory*, a scaled, cloud-native system for monitoring news websites live and collecting analyzable data about their publication behavior.

2021 – 2022   **Predoctoral Researcher / Research Engineer**, *Microsoft Research Lab – New York City*, New York, NY

Researched topics related to the News, such as how events are framed by different publishers, the length of the news cycle, and people's perceptions of factualness. Researched the process of conversion due to advertising. Teaching assistant for the MSR NYC Data Science Summer School. Built custom analysis and modeling tools to fit researchers' specific mathematical theories about marketing, the news cycle, and perceptions of bias.

2020 – 2021   **Research Assistant**, *University of Chicago: Brain Bridge Lab*, Chicago, IL

Researched the efficacy of deep learning techniques to estimate the probability that a subject will remember an image. Used these models to create a better understanding of the features of an image that are common among highly memorable images. Developed *ResMem*, a novel deep-learning based model to estimate the memorability of images. With Prof. Wilma A. Bainbridge.

2020 – 2021   **Research Assistant**, *University of Chicago: Memory Lab*, Chicago, IL

Developed an online experiment to generate pilot data for the development of a computer assisted testing program for cognitive decline. Built in JavaScript, the tooling was used for studies on prolific and a selection of older adults through a partnership with Rush University. The apparatus was designed to evaluate the participants' cognitive state and possibly diagnose Mild Cognitive Impairment. With Prof. David Gallo.

Publications

2025   **Search conversion journeys and the missed opportunity of associated keywords**, David Rothschild, Coen D. Needell, Joe Veverka, Elad Yom-Tov, *Frontiers in Communications*
Published

2024   **Framing in the Presence of Supporting Data: A Case Study in U.S. Economic News**, Alexandria Leto, Elliot Pickens, Coen D. Needell, David Rothschild, Maria Leonor Pacheco, *Association for Computational Linguistics*
Published

2023   **Mainstream Media Producing and Distributing Misinformation Through Stenography and "Opinions"**,
David Rothschild, Elliot E. Pickens, Tin Orešković, Katelyn Morrison, Markus Mobius, Coen D. Needell, Jared Katzman, Duncan J. Watts
Forthcoming

2021   **Embracing New Techniques in Deep Learning for Estimating Image Memorability**, Coen D. Needell, Wilma A. Bainbridge, *Computational Brain & Behavior*
Published

Presentations

2023   **The News Observatory: A collection and Storage System for Past and Future News Media Content**,
Coen D. Needell, Elliot E. Pickens, David Rothschild, Duncan J. Watts, *International Conference on Computational Social Science*
Poster

2021   **Memorability: A Stimulus-Centric Framework for Analyzing Memory Performance**, Wilma A. Bainbridge, Paige Hanson, Max Kramer, Coen D. Needell, Xinyue Li, *Interdisciplinary Graduate Conference, UChicago*
Panel

2021   **Embracing New Techniques in Deep Learning for Predicting Image Memorability**, Coen D. Needell, Wilma A. Bainbridge, *Annual Meeting of the Vision Sciences Society*
Poster

Skills

Machine Learning	Natural Language Processing	Data Mining
Network Analysis	Statistics and Statistical Learning	Deep Learning
Cloud Computing	Web Development	Unix
Data Visualization	Data Scraping	Philosophy of Science
Econometric Models	Systems Analysis	Advanced Mathematics
Functional Programming	Systems Design	Data Engineering
Technical Leadership	Project Management	Human-Computer Interaction

Languages, Packages, and Frameworks

Python	JavaScript	Julia
NLTK	PyTorch	Sci-kit Learn
SciPy	Numpy	Pandas
D3.js	MatPlotLib	ggplot2
Linux	Rust	OpenCL
SQL	AWS	Kubernetes
Polars	Postgres	Databricks
Terraform	Docker	Spark