Connor Shorten

connorshorten300@gmail.com 📨
cshorten.github.io 🏡

Employment History

2022-present **Research Scientist at Weaviate**

Working across three roles: Research Scientist, Product Engineer and Podcast Host, contributing to research, product development, and technical communication. Progressed from the Product Research team to Applied Research and subsequently, the Agents team.

Led research on filtered vector search algorithms including ACORN, IVF^2, and Beta Window Search. Profiled CPU performance using Go's `pprof` to investigate computational bottlenecks and benchmarked query throughput (QPS) versus exact nearest neighbor recall. Led research and implementation of text-to-SQL optimization, MCP, multi-stage reranking, predicted filtering, suggested queries, diversity ranking, data sampling for query understanding, multimodal retrieval, and model evaluation, alongside production engineering and operational improvements.

My work on ultrahigh effort in Weaviate's Query Agent improved nDCG@10 on the BRIGHT benchmark from 17.8 with hybrid BM25 and vector search alone to 47.6.

2019-2023 **Research Assistant at Florida Atlantic University**

Worked with Memorial Healthcare System to develop Machine Learning Systems for COVID-19 Patient Prognosis prediction. Worked in FAU's Civil, Environmental, and Geomatics Engineering lab on Database Systems for Spatial Data. Worked on algorithms for Data Augmentation and Class Imbalance.

Education

2023 **Florida Atlantic University**, Boca Raton, FL

Ph.D. in Computer Science, Advised by Taghi M. Khoshgoftaar
Thesis: Data Augmentation in Deep Learning

2019 **Florida Atlantic University**, Boca Raton, FL

M.Sc. in Computer Science

2018 **Florida Atlantic University,** Boca Raton, FL

B.S. in Computer Science

Papers & Manuscripts

Preprint 2026 Scaling Reranked Retrieval.
C Shorten, DJ Williams, DM Jones, A Skaburskas, J Trengrove, E Dilocker, BV Luijt

ArXiv 2026 IRPAPERS: A Visual Document Benchmark for Scientific Retrieval and Question Answering.
C Shorten, A Skaburskas, DM Jones, C Pierse, R Esposito, J Trengrove, E Dilocker, BV Luijt

ArXiv 2025 Querying Databases with Function Calling.
C Shorten, C Pierse, TB Smith, K D'Oosterlinck, T Celik, E Shorten, L Monigatti, MS Hasan, E Schmuhl, DJ Williams, A Kesiraju, BV Luijt

ArXiv 2025 StructuredRAG: JSON response formatting with Large Language Models.
C Shorten, C Pierse, TB Smith, E Shorten, A Sharma, J Trengrove, BV Luijt

Frontiers in Digital Health 2023 Using machine learning to identify patient characteristics to predict mortality of in-patients with COVID-19 in South Florida.
D Datta, SG Dalmida, L Martinez, D Newman, J Hashemi, TM Khoshgoftaar, **C Shorten**, C Sareli, P Eckardt

Springer Nature 2022 Language Models for Deep Learning Programming: A Case Study with Keras.
C Shorten, TM Khoshgoftaar

IEEE ICMLA 2022 Exploring Language-Interfaced Fine-Tuning for COVID-19 Patient Survival Classification.
C Shorten, E Shorten, TM Khoshgoftaar, J Hashemi, SG Dalmida, D Newman, D Datta, L Martinez, C Sareli, P Eckward

FLAIRS 2022 An Exploration of Consistency Learning with Data Augmentation.
C Shorten, TM Khoshgoftaar

FLAIRS 2022 A Comparison of House Price Classification with Structured and Unstructured Text Data.
E Shorten, **C Shorten,** TM Khoshgoftaar, B Furht

FLAIRS 2022 Predicting the Severity of COVID-19 Respiratory Illness with Deep Learning.
C Shorten, TM Khoshgoftaar, J Hashemi, SG Dalmida, D Newman, D Datta, L Martinez, C Sareli, P Eckward

IEEE ICMLA 2021 KerasBERT: Modeling the Keras Language.
C Shorten, TM Khoshgoftaar

Springer Nature 2021	Text Data Augmentation for Deep Learning. **C Shorten,** TM Khoshgoftaar, B Furht (Cited 1,064 times)
Springer Nature 2021	Deep Learning Applications for COVID-19. **C Shorten,** TM Khoshgoftaar, B Furht (Cited 484 times)
IEEE ICTAI 2021	Investigating the Generalization of Image Classifiers with Augmented Test Sets. **C Shorten,** TM Khoshgoftaar (Best Student Paper Award)
Springer Nature 2019	A Survey on Image Data Augmentation for Deep Learning. **C Shorten,** TM Khoshgoftaar (Cited 19,933 times)

Talks and Technical Blogs

Weaviate Podcast.
I have hosted and published 147 episodes of the Weaviate Podcast featuring over 300 unique guests. The podcast has achieved over 500,000 views across YouTube and Spotify.

Generative Feedback Loops.
Presented at the Databricks Data + AI Summit's Compound AI Systems workshop, 2024.

DSPy Explained!
Published on Connor Shorten YouTube, 2024. (Viewed 83K times)

GEPA Explained!
Published on Weaviate YouTube, 2025. (Viewed 20K times)

Scaling Test-Time Compute in Search Mode.
Published on the Weaviate Blog 2026.

Scaling Test-Time Compute in Search Mode.
Published on Weaviate YouTube, 2026.

Search Mode Benchmarking.
Published on the Weaviate Blog 2025.

DSPy End-to-End San Francisco.
Organized in-person event with collaborators from Cohere, Arize AI, and Stanford University. Hosted 115 guests.

DSPy Boston.
Organized in-person event with collaborators from AWS, MIT, and PyData. Hosted 80 guests.

MemGPT Explained!
Published on Weaviate YouTube, 2024. (Viewed 22K times)

REFRAG Explained!
Published on Weaviate YouTube, 2025.

Generative Feedback Loops with LLMs for Vector Databases.
By Connor Shorten, published on the Weaviate Blog 2023.

Large Language Models and Search.
By Connor Shorten and Erika Shorten, published on the Weaviate Blog 2023.

Weaviate Gorilla Part 1: GraphQL APIs.
By Connor Shorten and Sam Stoelinga, published on the Weaviate Blog 2023. Invited and Attended U.C. Berkeley Sky Camp 2025.

Retrieval-Augmented Generation (RAG).
Published on Connor Shorten YouTube, 2021. (Viewed 34K times)

RAPIDS Feature Engineering for the NFL Data Bowl
Published on Connor Shorten YouTube, 2019.
Collaboration with NVIDIA and Digital Storm.

ResNet Explained!
Published on Connor Shorten YouTube, 2019. (Viewed 128K times)

Technical Skills

Programming Languages: Python, Go, TypeScript

Database Systems: Vector Indexing (HNSW, IVF, DiskANN, HFresh), LSM Storage Engines, Inverted Indexing, Replication, Concurrency Control, Multi-Tenancy, Storage Tiers.

AI Agents: Frameworks (Pydantic AI, LangChain, DSPy), Model Evaluation, MCP, Retrieval-Augmented Generation, Search.

Infrastructure and Observability: Prometheus, Grafana, Dash0, Descope, Kubernetes, Terraform, AWS, Temporal, Redis, Postgres.