

Josua Krause

email: josua.krause@gmail.com
web: <https://www.josuakrause.com/>
phone: +1 (917) 525 0807

Education

PhD in Computer Science, NYU Tandon School of Engineering

dissertation: “Using Visual Analytics to Explain Black-box Machine Learning”

advisor: Prof. Dr. Enrico Bertini

New York, USA – 2013 - 2018

MSc in Computer Science, University of Konstanz

thesis: “Graph Comics: Interactive Staging for Exploring Dynamic Graphs”

advisor: Prof. Dr. Oliver Deussen, Prof. Dr. Enrico Bertini

Konstanz, Germany – 2011 - 2014

BSc in Information Engineering, University of Konstanz

thesis: “Annotation of Changes in Evolving Graphs”

advisor: Prof. Dr. Oliver Deussen, Prof. Dr. Ulrik Brandes

Konstanz, Germany – 2008 - 2011

Employments

UNDP, Accelerator Labs

role: NLP Researcher

focus on accessibility of LLMs in low-resource environments and facilitating information access in low-bandwidth communities through LLMs

major projects: online semantic search on document corpus; deep dive semantic search with offline LLM result refinement; LLM RAG for report generation

scale: field notes on grassroots innovation and UNDP published blogs, speeches, press releases, reports: 10k documents (200k chunks)

select AI technology: PyTorch, Hugging Face, llama.cpp, Llama 3, Gemma, ChatGPT 3.5/4, Project Jupyter, spaCy

select Cloud technology: Azure, qdrant, Docker Compose, Redis, PostgreSQL

select (public) URLs:

<https://insights.sdg-innovation-commons.org/> (Semantic Search)

<https://github.com/UNDP-Accelerator-Labs/nlpapi> (see p.7)

<https://github.com/JosuaKrause/scattermind> (see p.7)

<https://github.com/JosuaKrause/redipy> (see p.7)

New York, USA – 2023 - current

Accern, NLP startup in the FinTech space (\$20M Series B early 2022)

role: Vice President of Data Science

lead of research, development, and deployment of AI models; focus on deep representation learning, NLP, and adaptive learning at scale

major projects: progressively refining real-time custom NER and event extraction pipeline; subject dependent sentiment analysis; fine-tuned BERT for financial documents and LoRA for task specialization

team: 6 Data Scientists; 2 Data Engineers (excluding myself)

tenure: started during Pre-Series A as Senior Data Scientist

scale: 500k documents per day; repository of 4B historical documents

select AI technology: PyTorch, Hugging Face BERT / DistilBERT, NVIDIA Triton, scikit-learn, pandas, polars, Project Jupyter

select Cloud technology: AWS, GCP, DigitalOcean, Kubernetes, Prometheus, Grafana, Redis, PostgreSQL, Apache Spark

New York, USA – 2018 - 2023

NYU Tandon School of Engineering

role: Adjunct Professor

teaching: Foundations of Data Science

New York, USA – 2021

Pacific Northwest National Laboratory

role: PhD Intern, National Security Internship Program

Washington, USA – 2016 and 2017

IBM T. J. Watson Research Center

role: Research Summer Intern

Interacting with Predictions: Visual Inspection of Black-box ML Models

New York, USA – 2015

IBM T. J. Watson Research Center

role: Research Summer Intern

Supporting Iterative Cohort Construction with Visual Temporal Queries

New York, USA – 2014

NYU Tandon School of Engineering

role: Research Assistant, Teaching Assistant

New York, USA – 2013 - 2018

University of Konstanz

role: Research Assistant, Teaching Assistant

Konstanz, Germany – 2009 - 2013

Patents

Josua Krause, Kenney Ng, Adam Perer: “Identifying and ranking risk factors using trained predictive models”,

US Patent 11,355,245 and 11,355,246 – submitted 2017, accepted 2022

Invited Talks / Teaching

mentor: Capstone Project

Bringing Structure to Emergent Taxonomies from Open-Ended CMS Tags

NYU Center for Data Science – 2023

hackathon: UN Datathon 2023

Best European Team; Best Solution using AIS Data

United Nations – 2023

blog: A Backend Agnostic Redis Interface

<https://medium.com/stackademic/a-backend-agnostic-redis-interface-9fdeb8641bc5>

blog: Dot Product is a Bad Distance Function

<https://medium.josuakrause.com/dot-product-is-a-bad-distance-function-aff7667da6cc>

blog: Asymmetric Topic Models

<https://medium.josuakrause.com/asymmetric-topic-models-199a0e1d03fc>

blog: The Problem with Teaching Language Models about the World

<https://medium.josuakrause.com/the-problem-with-teaching-language-models-about-the-world-5e024b408711>

Medium – 2023

invited talk: AI4CI: Artificial Intelligence for Collective Intelligence

United Nations Development Programme – 2022

webinar: How to Spot and Understand AI Bias within Financial Services

Accern – 2022

blog: Adaptive Modeling: Building Models that Last

Accern – 2021

teaching: Foundations of Data Science

NYU Tandon School of Engineering – 2021

mentor: Capstone Project
Beyond Bert-based Financial Sentimental Classification: Label Noise and Company Information
NYU Center for Data Science – 2020

mentor: Capstone Project
Visualizing and Analyzing Accern Signals as Knowledge-Graph
Columbia University – 2020

mentor: Capstone Project
Predicting Stock Market Movements using Public Sentiment Data & Sequential Deep Learning Models
NYU Center for Data Science – 2019

teaching: VIS for Practice Workshop – Vis. for the Web: Front- & Backends
Universität Konstanz – 2018

teaching assistant: Foundations of Data Science *NYU Tandon* – 2017

teaching assistant: Information Visualization *NYU Tandon* – 2015 - 2016

teaching assistant: Concepts of Computer Science *Univ. Konstanz* – 2012

teaching assistant: Concepts of Programming *Univ. Konstanz* – 2011

Committee Memberships

session chair: “Visualizing Machine Learning”
IEEE VIS Short Papers Track – 2020

associate chair: Special Applications Subcommittee
ACM CHI Conference on Human Factors in Computing Systems – 2019

reviewer: OPUS Grant Proposal
National Science Center, Poland – 2018

chair: *Workshop on Visual Analytics in Healthcare (VAHC)* – 2017

journal reviewer
Journal of Computational and Graphical Statistics – 2021 - 2022
Transactions on Visualization and Computer Graphics – 2015, 2018 - 2021
IEEE Transactions on Big Data – 2015, 2021
ACM Transactions on Interactive Intelligent Systems – 2021
IEEE Computer Graphics and Applications – 2015

program committee / reviewer

IEEE VIS – 2020 - 2022, 2024

IEEE VIS Short Papers Track – 2020 - 2022, 2024

EuroVis – 2016, 2019 - 2020, 2022

ACM Symposium on User Interface Software and Technology (UIST) – 2020

IEEE Information Visualization (InfoVis) – 2015, 2017 - 2019

IEEE Visual Analytics Science and Technology (VAST) – 2016 - 2019

ICLR Wrkshp. on Debugging Machine Learning Models (DEBUGML) – 2019

KDD Wrkshp. on Interactive Data Exploration and Analytics (IDEA) – 2018

IEEE Pacific Visualization Symposium (PacificVis) – 2018

ACM CHI Conference on Human Factors in Computing Systems – 2016

Publications

Josua Krause, Adam Perer, Enrico Bertini: “A User Study on the Effect of Aggregating Explanations for Interpreting Machine Learning Models”,

KDD Workshop on Interactive Data Exploration and Analytics (IDEA) 2018

Josua Krause, Aritra Dasgupta, Jordan Swartz, Yindalon Aphinyanaphongs, Enrico Bertini: “A Workflow for Visual Diagnostics of Binary Classifiers using Instance-Level Explanations”,

IEEE Transactions on Visualization and Computer Graphics (TVCG – VAST) 2017

Paolo Tamagnini, Josua Krause, Aritra Dasgupta, Enrico Bertini: “Interpreting Black-Box Classifiers Using Instance-Level Visual Explanations”,

SIGMOD Workshop on Human-In-the-Loop Data Analytics (HILDA) 2017

Josua Krause, Aritra Dasgupta, Enrico Bertini: “Explanatory Visual Analytics for Enhancing Human Interpretability of Machine Learning Models”,

Visualization in Data Science (VDS at IEEE VIS) 2016

Josua Krause, Adam Perer, Kenney Ng: “Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models”,

KDD Workshop on Interactive Data Exploration and Analytics (IDEA) 2016

Josua Krause, Adam Perer, Enrico Bertini: “Using Visual Analytics to Interpret Predictive Machine Learning Models”,

ICML Workshop on Human Interpretability in Machine Learning (WHI) 2016

Josua Krause, Aritra Dasgupta, Jean-Daniel Fekete, Enrico Bertini:
“SeekAView: An Intelligent Dimensionality Reduction Strategy for Navigating HD Data Spaces”,

IEEE Symposium on Large Data Analysis and Visualization (LDAV) 2016

Anshul Vikram Pandey, **Josua Krause**, Cristian Felix, Jeremy Boy, and Enrico Bertini: **“Towards Understanding Human Similarity Perception in the Analysis of Large Sets of Scatter Plots”** (Honorable Mention),

ACM Conference on Human Factors in Computing Systems (CHI) 2016

Josua Krause, Adam Perer, and Kenney Ng: **“Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models”**,

ACM Conference on Human Factors in Computing Systems (CHI) 2016

Josua Krause, Adam Perer, and Harry Stavropoulos: **“Supporting Iterative Cohort Construction with Visual Temporal Queries”**,

IEEE Transactions on Visualization and Computer Graphics (TVCG - VAST) 2015

Josua Krause, Narges Razavian, Enrico Bertini, and David Sontag: **“Visual Exploration of Temporal Data in Electronic Medical Records”**,

Poster Session I of the AMIA Annual Symposium; Poster 2015

Josua Krause and Adam Perer: **“Data-Driven Cohort Construction with Interactive Visual Queries”**,

2015 Workshop on Visual Analytics in Health Care (VAHC); Demo 2015

Josua Krause, Narges Razavian, Enrico Bertini, and David Sontag: **“Visual Inspection of Longitudinal Electronic Medical Records”**,

2015 Workshop on Visual Analytics in Health Care (VAHC); Demo 2015

Josua Krause, Adam Perer, and Enrico Bertini: **“INFUSE: Interactive Feature Selection for Predictive Modeling of High Dimensional Data”**,

IEEE Transactions on Visualization and Computer Graphics (TVCG – VAST) 2014

Josua Krause, Marc Spicker, Leonard Wörteler, Matthias Schäfer, Leishi Zhang, and Hendrik Strobel: **“Interactive Visualization for Real-time Public Transport Journey Planning”**,

Proceedings of SIGRAD 2012

Open Source Projects

UNDP AccLabs NLP API

API for various NLP tasks, including semantic search.

<https://github.com/UNDP-Accelerator-Labs/nlpapi>

Scattermind

Decentralized and distributed horizontally scalable ML model execution framework.

<https://github.com/JosuaKrause/scattermind>

RediPy

An in-memory Redis implementation in Python.

<https://github.com/JosuaKrause/redispy>

QuickServer

A quick to use and easy to set up Python server implementation.

https://github.com/JosuaKrause/quick_server

Searchspace

Visualization for various similarity functions (euclidean, cosine, dot, etc.).

<https://github.com/JosuaKrause/searchspace>

BubbleSets for JavaScript

An isosurface visualization for sets in JavaScript.

<https://github.com/JosuaKrause/bubblesets-js>

Projections for JavaScript

JavaScript implementation of MDS and PCA.

<https://github.com/JosuaKrause/mdsjs>

BubbleSets

An isosurface visualization for sets in Java.

<https://github.com/JosuaKrause/Bubble-Sets>

Explanation Explorer

A visual interface to explore ML explanations.

https://github.com/nyuvis/explanation_explorer

patient-viz

Visualization for electronic medical record patient histories.

<https://github.com/nyuvis/patient-viz>