Vinitra Swamy

Curriculum Vitae

vinitra.swamy@epfl.ch

vinitra

Lausanne, Switzerland

Human-Centric ML (education), eXplainable AI, Multimodality, Natural Language Processing (LLMs)

Google Scholar, Twitter, LinkedIn, Website

EDUCATION

Current **Ph.D. Computer Science** Swiss Federal Institute of Technology (EPFL)
Advisor: Tanja Käser (ML4ED), Martin Jaggi (MLO)

4th year PhD on trustworthy, explainable AI for human-centric data. Recipient of the EDIC Fellowship, awarded the IC Distinguished Service Award twice ('21, '22).

2018 **M.S. Electrical Engineering and Computer Science** University of California, Berkeley
Advisor: David Culler (RISELab)

Research degree with emphasis in deep learning / AI, Graduate Opportunity fellow (full tuition award).

2017 **B.A. Computer Science** University of California, Berkeley

Graduated 2 years early, awarded EECS Award for Excellence in undergraduate teaching and leadership.

2015 **High School Diploma** Cupertino High School

Relevant Coursework: Fairness in ML+, Designing and Visualizing Deep Neural Networks+, Special Topics in Deep Learning+, Natural Language Processing+, ML for Education+, Databases+, Artificial Intelligence, Algorithms, Advanced Data Science, Operating Systems, Linear Algebra, Linguistics

PROFESSIONAL EXPERIENCE

2018 - 2020 **Microsoft AI** Bellevue, WA
AI Software Engineer

Worked on the Core Engineering team of the Open Neural Network eXchange (ONNX) Standard (founded by Microsoft, Facebook, and Amazon and now extended to 30+ companies) to enable deep learning framework interoperability. Released ONNX Runtime 1.0 for efficient model inference.

2020 **University of Washington, Paul G. Allen School of CSE** Seattle, WA
Part-Time Faculty, Machine Learning

Lecturer for CSE/STAT 416: Introduction to Machine Learning to 100+ UW Seattle upper-division undergraduate and graduate students, with a teaching team of 4.

2018 **UC Berkeley Division of Data Sciences** Berkeley, CA
Lecturer, Data Science

Lecturer for DATA 8: Foundations of Data Science to over 250 UC Berkeley summer students, managed a course staff of 30.

+ Graduate Coursework

2018 **Real-time Intelligent Secure Explainable (RISE) Lab** Berkeley, CA
Graduate Research Assistant

Worked on projects in AI + Systems with an application area of scaling tools for data science education (JupyterHub architecture, OkPy autograding, deep knowledge tracing).

2017 **IBM Research** Yorktown Heights, NY
Research Scientist Intern, Machine Learning

Developed ensemble-based ML algorithm for inactive VM identification, filed 2 patents.

2015 **LinkedIn** Mountain View, CA
Software Engineering Intern, Search Engine Optimization (SEO) Team

Data pipelining and automated testing for LinkedIn public profile pages, UX research experiments.

2011 **Google** Mountain View, CA
Computing and Programming Experience (CAPE) Intern

Summer shadowing software engineers at Google's Mountain View HQ as a high school intern.

RESEARCH

NeurIPS 2023 **MultiModN- Multimodal, Multi-Task, Interpretable Modular Networks.** Vinitra Swamy*, Malika Satayeva*, Jibril Frej, Thierry Bossy, Thijs Vogels, Martin Jaggi, Tanja Käser*, Mary-Anne Hartley*.

We present MultiModN, a multimodal, modular network that fuses latent representations in a sequence of any number, combination, or type of modality while providing granular real-time predictive feedback on any number or combination of predictive tasks**. MultiModN's composable pipeline is interpretable-by-design, as well as innately multi-task and robust to the fundamental issue of biased missingness.** We perform four experiments on several benchmark MM datasets across 10 real-world tasks (predicting medical diagnoses, academic performance, and weather), and show that MultiModN's sequential MM fusion does not compromise performance compared with a baseline of parallel fusion.

EMNLP Findings 2023 **Unraveling Downstream Gender Bias from Large Language Models: A Study on AI Educational Writing Assistance.** Thiemo Wambsganss, Xiaotian Su, Vinitra Swamy, Seyed Parsa Neshaei, Roman Rietsche, Tanja Käser.

We investigate how bias transfers through an AI writing support pipeline through a large scale user study with 231 students writing business case peer reviews in German. Students are divided into five groups with different levels of writing support (traditional ML suggestions, control group with no assistance, finetuned versions of GPT2, GPT 3, and GPT3.5). Using GenBit, WEAT, and SEAT, we evaluate the gender bias at various stages of the pipeline: in model embeddings, in suggestions generated by the models, and in reviews written by students. Our results indicate no significant difference in gender bias between the peer reviews of groups with and without LLM suggestions. **Our research is therefore optimistic about the use of AI writing support in the classroom, showcasing a context where bias in LLMs does not transfer to students' responses**.

JAIR Preprint **The future of human-centric eXplainable Artificial Intelligence (XAI) is not post-hoc explanations.** Vinitra Swamy, Jibril Frej, Tanja Käser.

Current approaches in human-centric XAI (e.g. predictive tasks in healthcare, education, or personalized ads) tend to rely on a single explainer. This is a concerning trend given systematic disagreement in explainability methods applied to the same points and underlying black-box models. We propose to shift from post-hoc explainability to designing interpretable neural network architectures; moving away from approximation techniques in human-centric and high impact applications. **We identify five needs of human-centric XAI (real-time, accurate, actionable, human-interpretable, and consistent) and propose two schemes for interpretable-by-design neural network workflows** (adaptive routing for interpretable conditional computation and diagnostic benchmarks for iterative model learning). We postulate that the future of human-centric

XAI is neither in explaining black-boxes nor in reverting to traditional, interpretable models, but in neural networks that are intrinsically interpretable.

LAK 2023 Honorable Mention

Trusting the Explainers: Teacher Validation of Explainable Artificial Intelligence for Course Design. Vinitra Swamy, Sijia Du, Mirko Marras, Tanja Käser.

We use human experts to validate explainable AI approaches in the context of student success prediction. Our pairwise analyses cover five course pairs (nine datasets from Coursera, EdX, and Courseware) that differ in one educationally relevant aspect and popular instance-based explainers. We quantitatively compare the distances between the explanations across courses and methods, then validate the explanations of LIME, SHAP, and a counterfactual-based confounder with 26 semi-structured interviews of university-level educators regarding which features they believe contribute most to student success, which explanations they trust most, and how they could transform these insights into actionable course design decisions. **Our results show that quantitatively, explainers significantly disagree with each other about what is important, and qualitatively, experts themselves do not agree on which explanations are most trustworthy**.

AAAI 2023

RIPPLE: Concept-Based Interpretation for Raw Time Series Models in Education. Mohammad Asadi, Vinitra Swamy, Jibril Frej, Julien Vignoud, Mirko Marras, Tanja Käser. *Educational Advances in AI Symposium @ AAAI 2023.*

We present RIPPLE, utilizing irregular multivariate time series modeling with graph neural networks to achieve comparable or better accuracy with raw time series clickstreams in comparison to hand-crafted features. Furthermore, we extend concept activation vectors for interpretability in raw time series models. Our experimental analysis on 23 MOOCs with millions of combined interactions over six behavioral dimensions show that **models designed with our approach can (i) beat state-of-the-art time series baselines with no feature extraction and (ii) provide interpretable insights for personalized interventions.**

COLING 2022

Bias at a Second Glance: A Deep Dive into Bias for German Educational Peer-Review Data Modeling. Thiemo Wambsganss*, Vinitra Swamy*, Roman Rietsche, Tanja Käser.

We contribute to the fourth UN sustainability goal (quality education) with a novel dataset of 9,165 German peer-reviews, an understanding of bias in natural language education data, and the potential **harms of not counteracting biases in language models** (BERT, T5, GPT) for educational tasks.

EDM 2022

Evaluating the Explainers: Black-Box Explainable Machine Learning for Student Success Prediction in MOOCs. Vinitra Swamy, Bahar Radhmehr, Natasa Krco, Mirko Marras, Tanja Käser.

We compare five explainers for black-box neural nets (LIME, PermutationSHAP, KernelSHAP, DiCE, CEM) on the downstream task of student performance prediction for five massive open online courses. Our experiments demonstrate that the families of explainers **do not agree** with each other on feature importance for the same Bidirectional LSTM models with the same representative set of students. We use Principal Component Analysis, Jensen-Shannon distance, and Spearman's rank-order correlation to quantitatively cross-examine explanations across methods and courses. **Our results come to the concerning conclusion that the choice of explainer contains systematic bias and is in fact paramount to the interpretation of the predictive results, even more so than the data the model is trained on**.

ACM L@S 2022

Meta Transfer Learning for Early Success Prediction in MOOCs. Vinitra Swamy, Mirko Marras, Tanja Käser.

We tackle the problem of transferability across MOOCs from different domains and topics, focusing on models for early success prediction. In this paper, we present and analyze three novel strategies to creating generalizable models: 1) pre-training a model on a large set of diverse courses, 2) leveraging the pre-trained model by including meta features about courses to orient downstream tasks, and 3) fine-tuning the meta transfer learning model on previous course iterations. Our experiments on 26 MOOCs with over 145,000 combined enrollments and millions of interactions show

that models combining interaction clickstreams and meta information have comparable or better performance than models which have access to previous iterations of the course. We enable educators to **warm-start their predictions** for new and ongoing courses.

NeurIPS 2021 **Interpreting Language Models Through Knowledge Graph Extraction.** Vinitra Swamy, Angelika Romanou, Martin Jaggi. *XAI4Debugging Workshop @ NeurIPS 2021*.

While transformer-based language models are undeniably useful, it is a challenge to quantify their performance beyond traditional accuracy metrics. In this paper, we compare BERT-based language models (DistilBERT, BERT, RoBERTa) through snapshots of acquired knowledge at sequential stages of the training process. We contribute a quantitative framework to compare language models through knowledge graph extraction and showcase a part-of-speech analysis to identify the linguistic strengths of each model variant. Using these metrics, machine learning practitioners can **compare models, diagnose their models' behavioral strengths and weaknesses, and identify new targeted datasets to improve model performance**.

KDD 2019 **Machine Learning for Humanitarian Data: Tag Prediction using the HXL Standard.** Vinitra Swamy, Elisa Chen, Anish Vankayalapati, Abhay Aggarwal, Chloe Liu, Vani Mandava, Simon Johnson. *Social Impact Track @ KDD 2019.*

We present a deep learning model to predict standardized Humanitarian eXchange Language (HXL) tags on datasets from the United Nations' open data platform. Our workflow provides a 14% accuracy increase over prior work and is a **novel case study of using ML to enhance humanitarian data**.

AIED 2018 **Deep Knowledge Tracing for Free-Form Student Code Progression.** Vinitra Swamy, Allen Guo, Samuel Lau, Wilton Wu, Madeline Wu, Zachary Pardos, David Culler.

Knowledge Tracing is a body of learning science literature that seeks to model student knowledge acquisition through their interaction with coursework. This paper uses LSTMs and free-form code attempts to **model student knowledge in large scale computer science classes**.

MSc Thesis **Pedagogy, Infrastructure, and Analytics for Data Science Education at Scale.** Vinitra Swamy, David Culler. *M.S. Thesis, UC Berkeley 2018*.

A detailed research report on **autograding, analytics, and scaling JupyterHub infrastructure highlighted in use for thousands of students studying data science at UC Berkeley**. Thesis presented as a graduate student affiliated with RISELab, after helping develop UC Berkeley data science's software infrastructure stack including JupyterHub, autograding with OkPy, Gradescope, and authentication for 1000s of students.

Poster **CSi2: Machine Learning for Cloud Server Idleness Identification.** Vinitra Swamy, Neeraj Asthana, Sai Zeng, Alexei Karve, Aman Chanana, Ivan Dell'Era. *IBM Research, TJ Watson Research Center 2017*.

The CSi2 algorithm is an **ensemble machine learning algorithm to detect inactivity of VMs and suggest a course of action** like termination or snapshot. This project is projected to save $3.2M for IBM Research with 95.12% recall, 88% F1-score (>> industry standard).

Poster **Neural Style Transfer for Non-Parallel Text.** Vinitra Swamy, Vasilis Oikonomou, David Bamman. *NLP Research Showcase, UC Berkeley School of Information*, 2017.

We expand on an MIT CSAIL paper by Shen et. al. to **improve the accuracy of neural style transfer for unaligned text** using author disambiguation algorithms with VAEs.

SKILLS

Lang.	Python (Expert), Java (Advanced), C (Advanced), C++ (Advanced), SQL, C#, R
Data Science	Pandas, NLTK, NumPy, PlotLy, MatPlotLib, Jupyter, SKLearn, Seaborn

Deep Learning	PyTorch, TensorFlow, Keras, ONNX, ONNX Runtime
Infra.	Docker, Kubernetes, Prometheus, Hadoop, Spark, Azure / AWS / Google Cloud

TALKS

2023 **AWS Cloud Research Day @ EPFL**
Personalized, Trustworthy Human-Centric Computing

2023 **AAAI 2023 (EAAI Symposium)**
RIPPLE: Concept-Based interpretation for Raw Time Series

2023 **AAAI 2023 (AI4EDU Workshop)**
Encore Track talks (EDM 2022, L@S 2022)

2023 **LAK 2023: Learning Analytics and Knowledge (Honorable Mention)**
Trusting the Explainers

2022 **Oxford Machine Learning Summer Series**
Evaluating Explainable AI

2022 **EDM 2022: Educational Data Mining**
Evaluating the Explainers

2022 **EDM 2022: Opening Remarks at FATED Workshop**
Fairness, Accountability and Transparency in Education

2022 **L@S 2022: ACM Learning at Scale**
Meta Transfer Learning

2022 **ICML 2022: Opening Remarks at WiML "Un-Workshop"**

2022 **WIDS 2022: Women in Data Science, Silicon Valley (SAP)**
Fair and Explainable AI

2021 **NeurIPS 2021: Spotlight at eXplainable AI Workshop**
Interpreting LLMs through Knowledge Graph Extraction

2021 **Tamil Internet Conference (INFITT)**
TamilBERT: Natural Language Modeling for Tamil

2021 **EDIC Orientation for PhDs**
How to navigate a PhD at EPFL

2021 **UC Berkeley Data Science Alumni Panel**

2020 **Tech Gals Podcast (Episode 3)**
Microsoft AI and Creating Berkeley's Division of Data Sciences

2020 **ONNX Workshop**
ONNX Model Zoo, Tutorials, and Cloud Integrations

2020 **WIDS 2020: Women in Data Science: Silicon Valley (SAP)**

Interoperable AI

2020 **Linux Foundation AI Day**
ONNX Converter Updates

2019 **SF Artificial Intelligence Meetup (Microsoft Bay Area)**
Accelerate and Optimize Machine Learning Models with ONNX

2019 **eScience Institute (University of Washington)**
ONNX: An Interoperable Standard for AI Models

2019 **Machine Learning, AI, and Data Science Conference (Microsoft MLADS)**
Model operationalization and acceleration with ONNX

2019 **Channel 9: Microsoft AI Show**
ONNX Runtime speeds up Bing Semantic Precise Image Search

2018 **UC Berkeley Data Science Pedagogy and Practice Workshop**
Autograding at scale in a Data Science classroom (OkPy)

2017 **Salesforce Dreamforce Conference 2017**
Opening Panelist: Creating your Legacy

2017 **JupyterCon 2017**
Data Science at UC Berkeley: 2,000 undergraduates, 50 majors, no command line

2015 **SF Business Times Conference**
Presenter, STEM Education Leadership Summit

Organizing Committee
WiML Program Chair @ ICML 2022
FATED Workshop Co-Chair @ EDM 2022

Reviewer / Program Committee
AIED Program Committee 2023
AIED 2021*, 2022* (Subreviewer for Tanja Käser)
EMNLP BlackBoxNLP 2021, 2022, 2023
EACL 2022
EDM 2023, Journal of Educational Data Mining (JEDM) 2022
LAK 2022*, 2023* (Subreviewer for Tanja Käser)
Editor for Springer Series on Big Data Management (Educational Data Science)

Working Groups
Fairness Working Group @ EDM 2022
WiML Workshop Team @ NeurIPS 2021
Lead of the 2020 ONNX SIG for Models and Tutorials

HONORS AND AWARDS

2022 **EPFL IC Distinguished Service Award** EPFL, Switzerland

2021 **President of EPFL PhDs of IC** EPFL, Switzerland

2021 **EPFL IC Distinguished Service Award** EPFL, Switzerland

2020 **EDIC Fellowship** EPFL, Switzerland

2018 **Graduate Opportunity Fellow, EECS** University of California, Berkeley

2018 **President of Computer Science Honor Society** University of California, Berkeley
Upsilon Pi Epsilon (UPE), Nu Chapter

2018 **Society of Learning Analytics Research (SoLAR)**
Member

2018 **International Artificial Intelligence in Education Society (IAIED)**
Member

2017 **Kairos Entrepreneurship Fellow** Kairos Society, Berkeley Chapter

2017 **EECS Department Award for Excellence in Teaching and Leadership** University of California, Berkeley

2015 **Cal Alumni Leadership Scholar** University of California, Berkeley

2014 **Google International Trailblazer in Computer Science** Google

TEACHING

2020 - Current **Machine Learning, Discrete Math, Databases** EPFL, Switzerland
TA for Profs. Tanja Käser, Bob West, Anastasia Alaimaki

Taught lab and exercise sessions for Advanced Information, Computing and Complexity (AICC), Machine Learning for Behavioral Data, Applied Data Analysis, Database Systems.

2020 **CSE/STAT 416: Introduction to ML** University of Washington, Seattle
Part-Time Faculty

Taught 100+ upper-division undergraduate and graduate students a practical introduction to machine learning. Modules include regression, classification, clustering, retrieval, recommender systems, and deep learning, with a focus on an intuitive understanding grounded in real-world applications.

2016 - 2018 **DATA 8: Foundations of Data Science** University of California, Berkeley
Lecturer, Head TA for Profs. John DeNero, David Wagner, Ani Adhikari

Head Graduate Student Instructor for over 1000 students per semester, managed course infrastructure (live deployment of JupyterHub at scale), student teaching staff of over 100. Lecturer to 250+ undergraduate students on fundamentals of statistical inference, computer programming, and inferential thinking.

2014 - 2015 **Space Cookies Robotics Outreach** NASA Ames Research Center, Mountain View
Outreach Team Curriculum Developer and Volunteer for FRC Robotics Team

Conducted community workshops for students from elementary school through high school to introduce concepts of robotics and Lego Mindstorms robotics.

2012 - 2014 **Bridging the Digital Divide** Cupertino, CA
Program Founder

Developed computer literacy fundamentals and basic computer programming course, taught in 3 senior retirement communities and 2 family housing shelters in the Bay Area.