KENZA AMARA

PERSONAL INFORMATION

Born in France, 21 February 1997

webpage k-amara.github.io

email kenza.amara@ai.ethz.ch

phone (M) +41 77 265 54 38

INTEREST

Explainable AI for Graph Neural Networks, Large Language Models, Generative Models and its applications in health sciences, environmental sciences and linguistic.

EDUCATION

2021-2024 ETH AI Center, ETH University, Zurich

PhD in Computer Science

Description: This thesis explores the research field of explainable AI and its specific use for graph neural networks. Current work consists in re-defining explainability and building a comparative framework of the existing methods. Advisors: Prof. Andreas Krause, Prof. Menna El-Assady & Assoc. Prof. Ulrik Brandes, Prof. Sebastian Schemm

2019-2021 ETH University, Zurich

Master in Environmental Sciences

Specialization: Environmental Systems and Policy
Description: This degree brought an in-depth understanding of environmental systems science with the knowledge and skills required to identify solution strategies that are environmentally, technically, and socially robust.

2016-2019 Ecole Polytechnique, Paris

Master in Computer Science

Description: Core areas in Computer Science and Applied Mathematics. Electives in Physics and Economy.

2014-2016 Lycee Henri IV, Paris

Scientific Preparatory Program

Description: Intensive undergraduate-level curriculum in advanced mathematics and physics, leading to competitive entrance examinations to the French Grandes Ecoles

RESEARCH EXPERIENCE

July-Oct 2022 Research Internship, Microsoft, Cambridge

Microsoft Research (MSR) Center

Improved feature attribution methods in drug discovery by optimizing GNN training with a substructure-aware loss
Reference: Jose Salvador Jimenez Luna, Raquel Rodriguez Perez

May-Sept 2021 Research Internship, FAIR, Paris

Facebook AI Research Center

Worked on data compression retrieval to increase the decoding accuracy for efficient vector indexing
Reference: Hervé Jégou, Matthijs Douze

June-Jan 2020/21 — Master Thesis, Crowther Lab & DS3Lab, Zurich

ETH University

Developed OneForest, a tool that uses graph attention networks and optimal transport to merge drone imagery and citizen science, and build informative maps of our forests worldwide.
Reference: Tom Crowther

Mar-Aug 2019 — Research Internship, Earth Institute, NYC

Columbia University

Predicted the probability distribution of wildfires in Western United States with Mixture Density Networks.
Reference: Pierre Gentine

June-Sept 2018 — Engineer Internship, Daikin, Osaka

DAIKIN Japan

Predicted power consumption of air-conditioning systems with an optimized machine learning model regularized by thermodynamics laws.
Reference: Haruo Kochi

Sept-Mar 2017/18 — Scientific Project, LIX, Paris

Ecole Polytechnique

Developed a machine learning program embedded in a drone for endemic plants detection.
Reference: Sylvie Putot

PUBLICATIONS

October 2022 — A substructure-aware loss for feature attribution in drug discovery

Journal of Cheminformatics

Explainable machine learning is increasingly used in drug discovery to help ratio- nalize compound property predictions. In this work we present a modification to the regression objective of GNNs to specifically account for common core structures between pairs of molecules. The proposed approach shows higher accuracy on a recently-proposed explainability benchmark. We expect this methodology to be useful in drug discovery pipelines, and specifically in lead optimization efforts where specific chemical series of related compounds are investigated.
Authors: Kenza Amara, Jose Jimenez Luna, Raquel Rodriguez Perez
10.26434/chemrxiv-2022-qxq56-v2 - *preprint*

July 2022 — GraphFramEx: Towards Systematic Evaluation of Explainability Methods for Graph Neural Networks

Learning on Graphs (LoG)

We propose the first systematic evaluation framework for GNN explainability: it uses a unique metric that combines the fidelity measures and classify explanations based on their quality of being sufficient or necessary. This work shows that structure-based methods give the best explanations on the widely used synthetic benchmarks because of their simple structure and node dependency. This paper and the webpage about GraphFramEx also aim to facilitate the assessment of future explainability methods that can be easily integrated to the framework.
Authors: Kenza Amara, Rex Ying, Zitao Zhang, Zhihao Han, Yinan Shan, Ulrik Brandes, Sebastian Schemm, Ce Zhang
https://arxiv.org/abs/2206.09677 - *preprint*

March 2022 — Nearest neighbor search with compact codes: A decoder perspective

International Conference on Multimedia Retrieval (ICMR)

We re-interpret popular methods such as binary hashing or product quantizers as auto-encoders, and point out that they implicitly make suboptimal assumptions on the form of the decoder. We design backward-compatible decoders that improve the reconstruction of the vectors from the same codes, which translates to a better performance in nearest neighbor search. Our

method significantly improves over binary hashing methods or product quantization on popular benchmarks.
Authors: Kenza Amara, Matthijs Douze, Alexandre Sablayrolles, Hervé Jégou
https://dl.acm.org/doi/pdf/10.1145/3512527.3531408 - *accepted to ICMR 2022*

January 2022 ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery

Association for the Advancement of Artificial Intelligence (AAAI)

We present ReforesTree, a benchmark dataset of forest carbon stock in six agro-forestry carbon offsetting sites in Ecuador. Furthermore, we show that a deep learning based end-to-end model using individual tree detection from low cost RGB-only drone imagery is accurately estimating forest carbon stock within official carbon offsetting certification standards.
Authors: Gyri Reiersen, David Dao, Bjorn Lütjens, Konstantin Klemmer, Kenza Amara, Attila Steinegger, Ce Zhang, Xiaoxiang Zhu
https://arxiv.org/pdf/2201.11192.pdf

August 2020 OneForest: Towards a Global Species Dataset by Fusing Remote Sensing and Citizen Science Data with Graph Neural Networks

Knowledge Discovery and Data Mining (KDD) - Fragile Earth Workshop

We propose OneForest, a method to create a tree species dataset by fusing unlabeled data from remote sensing, i.e. drone images of forests, with labeled citizen science data, i.e. ground measurements. In this paper, we compare methods that map drone imagery and ground observations based on nearest neighbors, registration, and optimal transport techniques. Our proposed matching algorithm classifies species with 81.5% accuracy on a dataset of 5061 trees in Ecuador by leveraging optimal transport and an initial CNN-based matching step. We additionally test our methods on a variety of datasets composed of different species and locations from South America (Ecuador) and North America (NEON).
Authors: Kenza Amara, David Dao, Charlotte Bunne, Bjorn Lütjens, Dava Newman, Ce Zhang and Tom Crowther

SKILLS

Code — Python, R, C/C++, JAVA

Communication — LaTeX, HTML/CSS/PHP, OpenOffice, Linux, Microsoft Windows

Languages — Fluent · English, French, German

Intermediate · Spanish, Japanese

Presentation — 2022 · Panel discussion at the workshop on Practical ML for Developing Countries, PML4DC ICLR 2022

2022 · Lightning talk on Advances of ML Approaches for Financial Decision Making & Time Series Analysis, AMLD EPFL 2022

Leadership — Oct-May 2016/17 · Zhuhai, China
Teaching Assistant at the Sino-French Institute of Nuclear Engineering and Technology.

Interests — Swimming · Cycling · Hiking · Skying

August 31, 2023