Danilo Franco

SHORT BIO

Danilo Franco was born in in 1993.
He received his BSc in Computer Science and MSc in Data Science and Engineering from the University of Genoa.
Currently, He is pursuing his PhD in Security, Risk and Vulnerability with the curriculum "Cybersecurity and Reliable Artificial Intelligence" at the same university.
His main research interests revolve around AI ethics with a focus on Fairness, Explainability, Privacy, and Robustness.

EDUCATION

B.S. in Computer Science

University of Genoa
October 2012 - July 2017

Programming & Software Development (Java, C++, Python, C#, Matlab, JavaScript, PHP), Database Design & Administration, Calculus, Operating System, Basic Probability, Basic Statistical Inference, Basic Information Theory, Software Engineering, Web Development.

Final thesis:

WannaCry: Ransomware and Spread Analysis

M.S. in Data Science and Engineering

University of Genoa
October 2017 - July 2020

Advanced Data Management (Big Data, NoSql), Machine Learning, Signal Processing, Computer Vision, Computational Biology & Advanced Machine Learning, Data Mining & Analysis, Data Visualisation, High-Performance Computing, Graph Analysis, Probability Theory, Statistical Inference.

Final thesis:

Algorithmic Fairness: Learning Fair Representation

Thesis Abstract:

Artificial Intelligence (AI) methods have been employed to help almost every aspect of the human decision-making process. Some of them are proven to outperform human capabilities. This increasing ubiquity has led the scientific community to address the possibility of discrimination in autonomous systems that directly or indirectly harm minorities, fighting the misconception that AI is absolutely objective. In the past ten years, the research has investigated the causes of such discrimination, proposing methods for addressing and removing the bias produced by bad labelling, sampling or simply reflecting a controversial state of our society. This thesis points to applying three well-established methods in the fairness literature in the context of age prediction from RGB face images. In particular, we will exploit a recent IBM dataset (Diversity in Faces) to learn fair features mappings derived from convolutional neural networks. This new data representation is forced to discard the information related to the membership to a particular protected population, leading to an increase in the adopted fairness rate.

PhD in Security, Risk, and Vulnerability

Curriculum in Cybersecurity and Reliable Artificial Intelligence
University of Genoa
November 2020 - Current Date

Project Abstract:

In the past years, the scientific community has increasingly driven its attention to the development of machine learning methodologies, to the point that the topic of data science has become one of the clusters of the research world. Moreover, the widespread of cloud services along with the recent advances in technology allowed these new techniques to obtain extraordinary performances in different fields (e.g. computer vision, natural language processing, bioinformatics) and, consequently, to be employed by a wide range of end-users, whether they are individuals, universities, or industries. Nevertheless, researchers brought attention to the wide impact that the deployment of these models will have on the broad audience and, more generally, on people's lives. Governments along with the scientific community have formulated, among the others, four main aspects for dealing with trustworthy Artificial Intelligence: privacy, fairness, explainability and robustness. Arguably, every framework deployed to the public need to satisfy these mandatory regulations, both on an ethical and lawful level. This project targets the study of these topics, the implementation of state-of-the-art methods on usable scenarios and the development of new techniques related to the handling and processing of data.

PUBLICATIONS

Journals

Deep Fair Models for Complex Data: Graphs Labeling and Explainable Face Recognition

D., Franco and N., Navarin and M., Donini and D., Anguita and L., Oneto
Neurocomputing - Volume 470, Pages 318-334
January 2022 - Link to the paper

Paper Abstract:
The central goal of algorithmic fairness is to develop AI-based systems which do not discriminate subgroups in the population with respect to one or multiple notions of inequity, knowing that data is often humanly biased. Researchers are racing to develop AI-based systems able to reach superior performance in terms of accuracy, increasing the risk of inheriting the human biases hidden in the data. An obvious tension exists between these two lines of research that are currently colliding due to increasing concerns regarding the widespread adoption of these systems and their ethical impact. The problem is even more challenging when the input data is complex (e.g. graphs, trees, or images) and deep uninterpretable models need to be employed to achieve satisfactory performance. In fact, it is required to develop a deep architecture to learn a data representation able, from one side, to be expressive enough to describe the data and lead to highly accurate models and, from the other side, to discard all the information which may lead to unfair behaviour. In this work we measure fairness according to Demographic Parity, requiring the probability of the model decisions to be independent of the sensitive information. We investigate how to impose this constraint in the different layers of deep neural networks for complex data, with particular reference to deep networks for graph and face recognition. We present experiments on different real-world datasets, showing the effectiveness of our proposal both quantitatively by means of accuracy and fairness metrics and qualitatively by means of visual explanation.

Toward Learning Trustworthily from Data Combining Privacy, Fairness, and Explainability: An Application to Face Recognition

D., Franco and L., Oneto and N., Navarin and D., Anguita
Entropy 2021 - Volume 23 (Number 8), Pages 1047
August 2021 - Link to the paper

Paper Abstract:

In many decision-making scenarios, ranging from recreational activities to healthcare and policing, the use of artificial intelligence coupled with the ability to learn from historical data is becoming ubiquitous. This widespread adoption of automated systems is accompanied by increasing concerns regarding their ethical implications. Fundamental rights, such as the ones that require the preservation of privacy, do not discriminate based on sensible attributes (e.g., gender, ethnicity, political/sexual orientation), or require one to provide an explanation for a decision, are daily undermined by the use of increasingly complex and less understandable yet more accurate learning algorithms. For this purpose, in this work, we work toward the development of systems able to ensure trustworthiness by delivering privacy, fairness, and explainability by design. In particular, we show that it is possible to simultaneously learn from data while preserving the privacy of the individuals thanks to the use of Homomorphic Encryption, ensuring fairness by learning a fair representation from the data, and ensuring explainable decisions with local and global explanations without compromising the accuracy of the final models. We test our approach on a widespread but still controversial application, namely face recognition, using the recent FairFace dataset to prove the validity of our approach.

Conferences

Fair Empirical Risk Minimization Revised

17th International Work-conference on Artificial Neural Networks (IWANN)
June 2023

Paper Abstract:
Artificial Intelligence is nowadays ubiquitous, thanks to a continuous process of commodification, revolutionizing but also impacting society at large. In this paper, we address the problem of algorithmic fairness in Machine Learning: ensuring that sensitive information does not unfairly influence the outcome of a classifier. We extend the Fair Empirical Risk Minimization framework where the fair risk minimizer is estimated via constrained empirical risk minimization. In particular, we first propose a new, more general, notion of fairness which translates into a fairness constraint. Then, we propose a new convex relaxation with stronger consistency properties deriving both risk and fairness bounds. By extending our approach to kernel methods, we will also show that the proposal empirically over-performs the state-of-the-art Fair Empirical Risk Minimization approach on several real-world datasets.

Learn and Visually Explain Deep Fair Models: An Application to Face Recognition

International Joint Conference on Neural Networks (IJCNN)
July 2021 - Link to the Paper

Paper Abstract:
Trustworthiness, and in particular Algorithmic Fairness, is emerging as one of the most trending topics in Machine Learning (ML). In fact, ML is now ubiquitous in decision-making scenarios, highlighting the necessity of discovering and correcting unfair treatments of (historically discriminated) subgroups in the population (e.g., based on gender, ethnicity, political and sexual orientation). This necessity is even more compelling and challenging when unexplainable black-box Deep Neural Networks (DNN) are exploited. An emblematic example of this necessity is provided by the detected unfair behaviour of the ML-based face recognition systems exploited by law enforcement agencies in the United States. To tackle these issues, we first propose different (un)fairness mitigation regularizers in the training process of DNNs. We then study where these regularizers should be applied to make them as effective as possible. We finally measure, by means of different accuracy and fairness metrics and different visual explanation strategies, the ability of the resulting DNNs in learning the desired task while, simultaneously, behaving fairly. Results on the recent FairFace dataset prove the validity of our approach.

FURTHER EDUCATION

Regularization Methods for Machine Learning (RegML) 2018

School Attendant
June 2018

RegML is a 22 hours advanced machine learning course including theory classes and practical laboratory sessions. The course covers foundations as well as recent advances in Machine Learning with emphasis on high dimensional data and a core set of techniques, namely regularisation methods. Focus on Tikhonov Regularisation and Kernels, Early Stopping and Spectral Regularisation, Regularisation for Multi-task Learning, Sparsity Based Regularisation, Structured Sparsity, Data Representation: Dictionary Learning, Data Representation: Deep Learning.

DeepLearn Summer School 2018

School Attendant
July 2018

DeepLearn is a 24 hours research training event with a global scope aiming at updating participants on the most recent advances in the critical and fast-developing area of deep learning.

Research Scholarship

University of Genoa
April 2020 - October 2020

This scholarship aims at developing automated systems that are at the same time accurate and not biased against protected populations. The research scope starts as a review of my master thesis and extends it with new fairness metrics from the literature and broader use cases.