PERSONAL INFORMATION

Andrea Espis

Cagliari, Sardinia, Italy
[redacted]
[redacted]
[redacted]
[redacted]

Sex Male | Date of birth [redacted] | Nationality Italian

PREFERRED JOB POSITION

Research Scientist (AI for medicine)

I hold a peculiar perspective on what constitutes a fulfilling endeavour. While some individuals aspire to lead effortless lives, my satisfaction lies in unravelling enigmas, acquiring fresh knowledge, and conquering hurdles. I actively seek out challenging roles, not solely for the goals I aim to accomplish, but also because I derive immense pleasure from the entire process. The realm of AI and medicine aligns perfectly with my aspirations, as it embodies a harmonious blend of intricacy, ground-breaking advancements, and boundless possibilities. This never-ending source of stimulation sustains both my constant curiosity and unwavering commitment.

WORK EXPERIENCE

(Mar 2023 - present)

Doctoral Researcher (Ph.D. candidate)

University of Bologna (Alma Mater Studiorum) – Bologna, Italy.

Design and implementation of frameworks for the application of deep learning techniques beyond supervised learning (e.g., self-supervised, semi-supervised, multi-modal learning, and causal inference) to clinical data with the goal of helping the research community to get closer to the realization of precision medicine. Particularly focused on imaging, genomics, and neuroscience.
Title of the fellowship project: 'Beyond supervised deep learning framework for precision medicine'.

Business or sector Academic research, machine learning, deep learning, precision medicine.

(Dec 2022 - Mar 2023)

Research Fellow

University of Bologna (Alma Mater Studiorum) – Bologna, Italy.

Participation in the pilot project of a screening program for lung cancer integrated with smoking cessation: pathways, selection of subjects and diagnostic protocols, in view of an HTA evaluation funded in the CCM program (National Center for Disease Prevention and Control) 2019 by the Ministry of Health. Title of the fellowship project: 'Archiving of CT images and evaluation of diagnostic aid systems for lung cancer screening'.

Business or sector Academic research, data collection, data management.

(Apr 2022 - Oct 2022)

Scientific Researcher Intern

Amazon – Luxembourg, Luxembourg.

Design, development, and deployment of ML pipelines to optimize the spare parts stock levels and to detect anomalies in the Amazon Warehouses equipment. The project details are described in the project section.

Business or sector Logistics, computer vision, machine learning, optimization, data science.

(Dec 2021 - Jan 2022)

Academic Tutor (Machine Learning course)

University of Bologna (Bologna Business School) – Bologna, Italy.

Supporting the student during the hands-on activities and helping them for the development of the assessments or to clarify doubts on the topics explained during the course by Professor Claudio Sartori. More details in the teaching activities section.

Business or sector Machine learning, teaching.

(Oct 2021 - Jan 2022)

Machine Learning Engineer Intern

DataVision s.r.l. – Prague, Czech Republic.

Design and development of a pipeline for the semi-automatic creation of annotations on a dataset of images. Application of Object Detection and Semantic Segmentation algorithms for the automatic localization of markers on the surface of tires to identify certain types of wheels and determine actions such as where to place the valve to inflate the wheel.

Business or sector Computer vision, machine learning, semi-supervised learning, active learning.

(May 2016 –Oct 2021)

Private Tutor

Cagliari, Italy.

Private lessons for scientific subjects (physics, mathematics, computer science, chemistry) to students enrolled at the University, High School, or Middle School.

Business or sector Science, teaching.

(Apr 2019 - May 2019)

Curricular internship

EOLAB (University of Cagliari) – Cagliari, Italy.

Denoising, on MATLAB, leveraging the Wavelet transform and non-linear filters, of real neural signal, under the supervision of Professor Danilo Pani.

Business or sector Academic research, biosignal processing.

(Nov 2016 - Jan 2019)

Nightwatchman

Cagliari, Italy.

Night guard activity for a house, where an elderly (autonomous) woman was/is living and keeping her company.

Business or sector Social activity.

(Jun 2018 - Sep 2018)

Receptionist

B&B Agnese – Cagliari, Italy.

Welcoming the guests.

Business or sector Customer service, tourism.

(Jun 2014 - Aug 2014)

Waiter

Palm Beach Bar – Cagliari, Italy.

Serving customers, cleaning the beach.

Business or sector Customer service, tourism.

TEACHING ACTIVITIES

(Mar 2023 - Jun 2023)

Academic Tutor (Biomedical Signal Processing and Machine Learning)

University of Bologna (Bologna Business School) – Bologna, Italy.

Description: Assisting Prof. Chiara Marzi to support the students during the hands-on activities of the course Biomedical Signal Processing and Machine Learning held by Prof. Stefano Diciotti. Total amount of hours: 15.

Main topics: Python, signal processing, machine learning.

(Dec 2021 - Jan 2022)

Academic Tutor (Machine Learning)

University of Bologna (Bologna Business School) – Bologna, Italy.

Description: Supporting the students during the hands-on activities and helping them for the development of the assessments or to clarify doubts on the topics explained during the course held by Professor Claudio Sartori.Total amount of hours: 35.

Topics: Theory of learning, supervised vs unsupervised learning, classification and regression, model selection, validation, and cross-validation; linear discrimination, decision trees, Bayesian inference, Support Vector Machines, k-nearest neighbours, logistic regression, random forests, adaboost, ensemble learning, boosting, bagging, association rules and the apriori algorithm, clustering with k-means, dbscan, expectation maximization, hierarchical methods, kernel methods, analysis of case studies, CRISP-DM methodology.

EDUCATION

(Sep 2019 – Mar 2022)

Master's degree (2 years) – Artificial Intelligence

Level QEQ 7

University of Bologna (Alma Mater Studiorum) – Bologna, Italy.

Main topics: Traditional Artificial Intelligence, Programming, Neuroscience, Cognition, Optimization, Constrained Optimization, Machine Learning, Deep Learning, Logic, Computer Vision, Natural Language Processing, Ethics, Big Data, AI in Industry.

Thesis title: "Object detection and semantic segmentation for assisted data labeling".

Thesis abstract: "The automation of data labelling tasks is a solution to errors and time costs related to manual labeling. In this thesis work CenterNet, DeepLabV3 and K-Means applied in the RGB color space, are implemented to build a pipeline for assisted data labeling: a semi-automatic process to iteratively improve the quality of annotations. The proposed pipeline showed a total of 1547 incorrect and missing annotations when applied to a dataset originally containing 8,300 annotations. In addition, the quality of each annotation has been drastically improved and at the same time more than 600 hours of work have been saved. The same models were also used to tackle real-time tire inspection with regard to detecting markers on the surface of tires. According to the experiments, the combination of the output of DeepLabV3 and a post-processing based on the properties of the detected blobs, reaches a maximum of average Precision 0.992, with average Recall 0.982, and a maximum of average Recall 0.998, with Average Precision 0.960".

Final mark: 110/110 cum laude (GPA: 4.0).

(Sep 2016 – Mar 2019)

Bachelor's degree (3 years) – Biomedical Engineering

Level QEQ 6

University of Cagliari – Cagliari, Italy.

Main topics: Mathematics, physics, and chemistry. Architecture of computers. Study of materials and biomechanics. Biology, anatomy, and physiology. Knowledge of the principles of electronics. Development of algorithms in C, Matlab and Python. Development of algorithms aimed at biosignal processing and clustering.

Thesis title: "Development of automatic algorithms for spike sorting based on Self-Organizing Map".

Thesis abstract: "Within the framework of this thesis work, the problem to which we propose to give a solution is to deduce from the neural signal the neuron from which each spike derives. This problem is known as 'spike sorting'. The aim of the thesis is to evaluate the effectiveness of alternative solutions compared to those discussed in the literature. The proposed and implemented solutions exploit Artificial Intelligence algorithms, in particular the neural networks known as Self-Organizing Map (SOM) and K-Means, as well as signal denoising using the Wavelet transform. Results based on synthetic data indicate that the performance of the algorithms developed is comparable with state-of-the-art algorithms, but that improvements are needed".

Final mark: 110/110 cum laude (GPA: 4.0).

(Sep 2011 – Jul 2016)

High school diploma (5 years) – Scientific school

Level QEQ 4

Liceo Scientifico Michelangelo – Cagliari, Italy.

Main topics: Mathematics, physics, chemistry, literature, Latin, history, geography, English.

Thesis title: "The destabilization of the poet in the twentieth century".

Final mark: 100/100.

TRAINING

(Jun 2023)

Ellis Summer School on Machine Learning for Healthcare and Biology

University of Manchester – Manchester, UK

Description: three days at the University of Manchester to discuss the latest research in these areas with leading machine learning scientists and to attend lectures by renowned researchers at the intersection of ML and Healthcare and Biology, and with an excellent track record of delivering educational content.

Main topics: machine learning for healthcare and biology, gaussian processes, topological deep learning, deep generative models, causal machine learning, ML for genomics, ML for drug discovery.

(Feb 2023 – May 2023)

Deep learning

University of Bologna (Alma Mater Studiorum) – Rome, Italy

Description: 40 hours participation to the course “deep learning” held by Prof. Matteo Ferrara for the master’s degree in computer science.

Main topics: Linear algebra, calculus and automatic differentiation, neural networks, backpropagation, optimization algorithms CNN, RNN, transformers, autoencoders, generative models, reinforcement learning, NLP.

(Feb 2023 – Mar 2023)

Deep learning for biomedical applications

University of Rome (Tor Vergata) – Rome, Italy

Description: Online participation to the course “deep learning for biomedical applications” of 15 hours, held by Prof. Nicola Toschi and Dr. Matteo Ferrante.

Main topics: Linear regression, logistic regression, SVM, XGBoost, random forest, neural networks, CNN, transfer learning, knowledge distillation, generative models, DALL-E, ChatGPT, reinforcement learning, graph learning, contrastive learning, multiple-instance learning, physically constrained neural networks, diffusion models, spiking neural networks.

(Jul 2017 – Aug 2017)

English intensive course II

Interactive English Language School Ltd – Brighton, UK

Description: Living for 30 days in a British native family. During the stay, I followed an intensive course of English language, level B2, at the Interactive English Language School Ltd, morning and afternoon from Monday to Friday.

(Jul 2016 – Aug 2016)

English intensive course I

Interactive English Language School Ltd – Brighton, UK

Description: Living for 30 days in a British native family. During the stay, I followed an intensive course of English language, level B1, at the Interactive English Language School Ltd, morning and afternoon from Monday to Friday.

FUNDED PROJECTS

(Dec 2022 - Mar 2023)

CT images storage and evaluation of auxiliary systems for lung cancer diagnosis

Project developed as a research fellow, at the University of Bologna. Verification of the correct configuration and installation of a PACS system, consisting of Conquest and K-PACS, on the computers of various participating hospitals. Evaluation of diagnostic aid systems for lung cancer screening.

(Apr 2022 - Oct 2022)

Optimization of the level of spare parts in warehouse stocks

Project developed during the internship at Amazon. Invention and implementation of two independent solutions for optimizing the level of spare parts in warehouse stocks to reduce stock-out (spare parts missing when needed) and hold-out (too many spare parts in stock, therefore spaces and funds spent inefficiently). The solutions are based on regression and statistical models. Both solutions do not require historical data of spare parts whose levels are to be optimized, but only properties such as weight, price, or other characteristics, so that spare parts levels for new machinery can also be optimized. A hybrid of the two solutions was used for a pilot test on 8 warehouses for a new machine; this saved over $0.5M compared to the expense that would have been faced compared to the recommendations proposed by the machine sellers. The pipeline, with some modifications, is integrated within the system controlling and helping the decision-making process of the spare parts stocks levels across all European Amazon warehouses.

(Apr 2022 - Oct 2022)

Detection of anomalies of used machinery in the warehouse

Project developed during the internship at Amazon. Invention, implementation, deployment in production of a pipeline that allows the detection of anomalies in the machinery before their breakdown. The input is a video consisting of an RGB image and a thermal image. The system automatically detects equipment anomalies and reports them with a description to the technicians, by automatic emails. Through technician feedbacks, the pipeline continuously improves. The algorithm has been estimated to lead to a minimum cost saving of $0.16M per year just considering the time saved by technicians. To this estimate should be added the costs saved by detecting anomalies that technicians without the pipeline might have not detected, but the internship ended before having the possibility to do a reliable estimation.

(Apr 2022 - Oct 2022)

Assisted data labelling

Project developed during the internship at DataVision. Invention and implementation of a pipeline that allows you to annotate a dataset in a semi-automatic way, saving time and producing higher quality annotations than manual work. The implemented pipeline showed a total of 1547 incorrect and missing annotations when applied to a dataset originally containing 8,300 annotations. In addition, the quality of each annotation has been drastically improved and at the same time more than 600 hours of work have been saved.

(SOME) NOT FUNDED PROJECTS

(2022)

Down syndrome feature selection

Application and modification of algorithms for the selection of features to find correlations, or causes, to the anomalous cognitive development of a subject affected by Down syndrome. Focus on the Boruta algorithm, modified to take into account the correlation between features when calculating their importance.
Python, tabular data, data science.

(2022)

Object detection

Implementation and training of a CNN (CenterNet) for the task of object detection of small objects (30x30 in a 1024x1024 image), and analysis of the performance depending on the resolution, batch size, and different training strategies (e.g., multi-resolution training).
Python, Pytorch, deep learning for computer vision.

(2021)

Conditional face generation

Implementation from scratch, by merging several state-of-the-art methods from different studies, of a GAN, a CGAN and an ACGAN for the generation of images representing faces. The generation is conditioned by 40 attributes such as 'sex', 'hair color', 'smiling', 'sunglasses', etc. Despite the use of a low resolution (64x64), due to the scarce availability of computational resources and time, the results are visually satisfactory. To the left, an example of a batch of generated images with the same attributes such as 'female', 'blonde', 'smiling', 'make-up'.
Python, Tensorflow, deep learning for computer vision.

(2021) **Question answering**

Implementation and training of different variants of an RNN with bidirectional attention, for the task of identifying the answer to a question, within a text. The most performant model implemented, was able to identify several errors in the SQuAD dataset.
Python, Pytorch, NLP.

(2021) **Sentiment analysis**

IMDb dataset sentiment analysis. Preliminary analysis of the dataset using Latent Semantic Analysis and SVD. Followed by the classification of the text using various techniques, including the Bernoulli model and Multinomial Naïve Bayes, and RNN.
Python, Pytorch, NLP.

(2021) **POS tagging**

Classification of each word of a text as one of the parts of speech (noun, adjective, verb etc.), through the implementation and training of an RNN with a custom weighted loss: referring to the idea of using a weighted loss to counteract the bias in the data, I implemented a weighted loss that increases the focus on instances belonging to classes more difficult to classify correctly.
Python, Pytorch, NLP.

(2021) **Fact-checking**

Classification of a sentence as correct or not, with respect to a text of reference. The dataset used is FEVER. Models were implemented from scratch and trained, consisting of two parts, an encoder (an NN) and a decoder (an NN or RNN) in numerous variants. The most performing model, with f1-score of 0.72, is able to identify errors in the dataset.
Python, Pytorch, NLP.

(2020) **Optimal rectangle packing**

Implementing a solution for the rectangle packing problem to identify in which position to put a set of rectangles, so as not to leave gaps between them, within a certain area. The solution was implemented in both SAT and CP language. Various experiments were carried out to reduce the search time of the solver.
MiniZinc, constraint optimization.

(2020) **Tablut AI player**

Development of an AI algorithm for the game Tablut, an ancient game with pawns, 9x9 board, and two players with opposite goals. The algorithm was implemented by exploiting the well-known Min-Max algorithm and the definition of ad-hoc heuristic functions for the game.
Python, traditional AI.

(2020) **Numpy neural network**

Implementation and training of a neural network and backpropagation using only Numpy. The network has been tested on various datasets.
Python, deep learning.

(2019) **Tools visual inspection**

Application of traditional computer vision techniques and implementation of new ideas for the inspection of objects in grayscale images. The inspection consists in identifying objects of interest and then calculating their properties, such as the position of the center of gravity, orientation, the presence of holes etc.
Python, OpenCV, computer vision.

(2019) **Bayesian network, COVID-19**

Implementation of a Bayesian network to answer questions such as 'what is the probability that a subject will get sick with covid (given a set of characteristics such as lifestyle, the average number of people frequented per day etc.) if he uses the mask? What if he doesn't wear a mask?' The algorithm is based on a set of simplifications and assumptions that do not make the algorithm's responses reliable.
Python, statistics.

(2019) **Spike sorting**

Implementation of various unsupervised learning techniques, for the clustering of action potentials (spikes) of the neural signal, with the aim of associating each spike with its own neuron of origin, to be able to exploit this information, for example, to control an artificial limb.
Matlab, unsupervised biosignal processing.

(2018) **ECG noise robust peak detector**

Implementation of various algorithms for the identification of the 'R' peak in the noisy ECG signal. Three algorithms have been implemented, based on the study of the derivative, the maxima and the minimums, and the threshold values.
Matlab, signal processing.

SEMINARS HELD BY ME

(31 Mar 2023) **Self-supervised learning**

I held a seminar on self-supervised learning (SSL) to the team I am member of (AI for medicine Research Group, supervised by Professor Stefano Diciotti, at the University of Bologna). The first part of the seminar was a high-level description of the trends of SSL from its origin until the state of the art. The second part of the seminar was focused on the description of the BEiT model and possible biomedical applications. The last part of the seminar was focused on the description of the VICReg model and a proposal and discussion of a modification it, for better representation learning. Number of people attending: 5.

(8 Sep 2022) **Machine Learning for Computer Vision**

I held a seminar on machine learning for computer vision with focus on semantic segmentation, a discussion of the DeepLabV3 model, and a use-case application. The seminar was held during the internship at Amazon, internally to the team I was member of, to share the knowledge on the topic. Number of people attending: 20.

SEMINARS / CONFERENCES I ATTENDED RECENTLY

(30 May 2023) **Managing and storing research data: privacy aspects**

Seminar held by Dr. Francesco Di Tano (ARIC, University of Bologna), 2 hours long.

(26 May 2023) **Continual Learning**

Seminar held by Dr. Lorenzo Pellegrini (University of Bologna), 2 hours long.

(23 May 2023) **Open science in Horizon Europe: research data management, and data management plan**

Seminar held by Dr. Sara Cuppini (University of Bologna), 2 hours long.

(12 May 2023) **AI Etichs**

Seminar held by Dr. Gabriele Graffieti (University of Bologna), 2 hours long.

(04 Apr 2023) **Open science in Horizon Europe: instructions for use**
Seminar held by Bianca Gualandi (ARIC, University of Bologna)and Irene Frascari (AlmaDL, University of Bologna), 2 hours long.

(27 Mar 2023) **Explainable Artificial Intelligence. Counterfactual Explanations of (some) Machine Learning Methods**
Seminar held by Prof. Fabrizio Silvestri (University of Sapienza), 2 hours long.

(24 Mar 2023) **Is attention all we need?**
Seminar held by Dott. Tommaso Boccato (University of Rome Tor Vergata), 2 hours long.

(10 Feb 2023) **How to give a scientific presentation**
Seminar held by Prof. Marco Viceconti (University of Bologna), 2 hours long.

(01 Feb 2023) **How to write and publish a scientific article**
Seminar held by Prof. Marco Viceconti (University of Bologna), 2 hours long.

(25 Jan 2023) **Artificial Intelligence – Webinar**
Seminar held by Prof. Stefano Diciotti and Prof. Michela Milano (University of Bologna), 2 hours long.

(25 Jan 2023) **Research integrity and data irreproducibility in a translational medicine perspective**
Seminar held by Prof. Laura Calzà (University of Bologna), 2 hours long.

(18 Jan 2023 – 24 Feb 2022) **Introduction to AI for Health and Well-being**
Series of seminars, held by Prof. Stefano Diciotti (University of Bologna), for a total of 7.5 hours.

(17 Jan 2023) **The mind's golden cage and the cognition in the wild**
Seminar held by Agustin Ibanez, 1.5 hours long.

(13 Jan 2023) **AI day**
National conference, organized by Prof. Michela Milano (University of Bologna), 4 hours long.

(12 Jan 2023) **Interpretation of BOLD fMRI signals**
Seminar held by Prof. Antonio Ferretti (University of Chieti), 1.5 hour long.

PERSONAL SKILLS

Mother tongue(s): Italian

Other language(s)

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	B2	B2	B2	B2	B2
	Replace with name of language certificate. Enter level if known.				

Levels: A1/2: Basic user - B1/2: Independent user - C1/2 Proficient user
Common European Framework of Reference for Languages

Soft skills

- Character inclined to collaboration, leadership, listening and empathy thanks to belonging to a large family, and working in team projects.
- Excellent communication skills thanks to the work activities carried out since adolescence, presentations held during the bachelor's and master's degree and during the working experience.
- Excellent ability to adapt to multicultural environments due to the interest in novelties and the passion for traveling abroad, in addition to the various work activities carried out abroad and with multinational teams.

Organisational / managerial skills

- High emotional intelligence, thanks to a lot of years of my childhood spent observing people rather than interacting with them due to extreme shyness, which is now overcome.
- Tendency towards the figure of leader, good ability to delegate tasks aimed at developing a complex project, due to the projects carried out with many different teams.
- Good capability to estimate the time needed to get the job done.
- Excellent organization of work and carrying out multiple projects in parallel within the pre-established time limits.

Computer science skills

- Programming languages: Python (Advanced), MATLAB (Intermediate), C (Intermediate), Scala (Base), Prolog (Base), SQL (Base), MiniZinc (Base).
- Good skills in using AWS: SageMaker, S3, CloudWatch, EC2, GroundTruth, ECR, Lambda, EventBridge, SES.
- Good skills in using MATLAB software. Good skills in using version control systems (specifically, Git).
- Limited skills in using Docker.
- Limited skills in using PAC systems (specifically, Conquest as database combined with K-PACS for visualization)

Other skills

- During my leisure time, I engage in various physical activities such as soccer, martial arts, swimming, dancing, cycling, and running. As a child, I used to play basketball. Although I don't excel in any of these activities, I pursue them purely for enjoyment.
- Both my work experiences and being part of a large family have instilled in me a remarkable level of patience.
- I possess an unwavering eagerness to acquire knowledge. As a result, over the years, I have developed a strong aptitude for quickly grasping new subjects and concepts.
- I approach interactions with others in a respectfully honest manner.
- I am highly motivated and determined.

Driving licence B

ADDITIONAL INFORMATION

Honours and awards Merit-based scholarship: student with the highest grades for the Biomedical Engineering bachelor's degree, University of Cagliari.

Memberships IEEE student member

28/06/2023