# Andrea Bosia

🏠 Italy    📞 +39 339-722-6659    ✉ [andrea.bosia@hotmail.it](mailto:andrea.bosia@hotmail.it)    in [LinkedIn](#)    🌐 [PersonalPage](#)    ⌂ [Git](#)

## Summary

Detail-oriented Data Scientist with a strong foundation in AI-powered tool development and interactive data visualizations. Expertise in machine learning, data analysis, and creating actionable insights from complex datasets. Experienced in collaborating with multidisciplinary teams to drive innovation and enhance decision-making processes. Skilled in presenting complex results to both technical and non-technical stakeholders, with a passion for leveraging data to solve business challenges and optimize performance.

## Working Experience

**OreFox Ai**                                                                                          Feb 2024 – June 2024
*Data Scientist*                                                                                          *Sydney, Australia*
- Led the development of an AI-powered news aggregation tool for OreFox, an AI-focused startup, as part of my master's thesis, aimed at providing strategic insights for stakeholders.
- Engineered a GenAI tool that effectively aggregates and summarizes industry-relevant news articles, providing OreFox with timely and actionable insights via a user-friendly news website.
- Implemented a sophisticated model pipeline using a Retrieval Augmented Generation (RAG) system, enhancing the tool's ability to query a vector base and deliver precise and relevant information to support informed decision-making.

**Brain and Mind Center (University of Sydney)**                                                          Dec 2023 – Feb 2024
*Research Intern*                                                                                          *Sydney, Australia*
- Collaborated with a multidisciplinary team to advance results in reconstructing natural images from brain activities, driving innovation in neuroimaging and AI integration for medical applications.
- Implemented and fine-tuned a text-conditioned latent diffusion model for image generation using fMRI brain signals, enhancing the accuracy of recreating visual stimuli perceived by subjects.
- Presented findings at a high-profile poster session, effectively communicating complex results to both academic and industry stakeholders, showcasing potential commercial and research applications in the medical field.

## Education

**University of Sydney**                                                                                   Feb 2023 - June 2024
*Master of Science in Data Science (Major in Machine Learning)*                                            *Sydney, Australia*
- **Relevant Coursework:** Advanced Machine Learning, Deep Learning, Database Management Systems, Game Theory, Online Learning, Computer Security.

**Politecnico di Milano**                                                                                  Sept 2017 - Mar 2021
*Bachelor Degree in Computer Science and Engineering*                                                      *Milan, Italy*
- **Relevant Coursework:** Data Base, Software Engineering,Technologies for the Web, Data Structures and Algorithms, Fundamentals of Automatic Control.

## Projects

**Class-Conditional Label Noise Robustness**
- Designed and implemented an image classifier robust to class-conditional label noise.
- Tested two distinguished CNN architectures, ResNet18 and GoogLeNet, in combination with various transition matrix based algorithms to train unbiased estimators of the data distribution.
- Evaluated the performances considering Accuracy, Precision, Recall and F1 score (achieving values $\sim 0.9$)

**Ensemble Model for Multi Label Classification**
- Developed a multi-modal ensemble model for multi label image classification using PyTorch.
- Leveraged the bi-modal input (images and annotations) I implemented 2 classification models, one for the images (a CNN based on ResNet-34) and one for the text (a LSTM).
- Following series of ablation studies and hyper-parameters tuning the model achieved an F1-score of 86 % finishing in the top 10 of the Kaggle competition.

**CO2 Emissions Visualization Dashboard**
- Developed an interactive business intelligence dashboard using Tableau and Plotly Dash to visualize CO2 emissions and socioeconomic data, enabling stakeholders to derive actionable insights.
- Provided actionable insights that informed the development of strategic CO2 reduction plans, focusing on high-impact areas such as energy efficiency, renewable energy adoption, and transportation policies.
- Presented findings to academic audience, demonstrating the potential of data-driven approaches to enhance environmental policy-making and resource management.

## Skills

**Programming Languages**: C, Java, JavaScript, Python, R, SQL, HTML, CSS, Bash.
**Data Science Tools**: Numpy, Pandas, Scikit-Learn, Keras, PyTorch, TensorFlow, Seaborn, SciPy, Matplotlib, Tableau, Selenium, BeautifulSoup, HuggingFace
**DevOps & Frameworks** : Git Version Control, React, Tailwind CSS, Django
**Cloud Technologies**: AWS EC2, S3, Amplify
**Certifications**: Coursera Unsupervised Learning, Recommenders, Reinforcement Learning (Mar. 2023)
**Other Tools**: PowerPoint, Excel, LaTeX, Slack, Notion
**Languages**: English (110/120 TOEFL iBT), Italian (Native)