



Blended-IP

Online meeting on thesis proposals

23. 11. 2023

humancentered-ai.eu

hcaim human centred
artificial intelligence
masters

Agenda

| Timeslot | Institution | Speaker |
|----------|----------------------------|---------------------------------------|
| 17.05 | CeADAR | Alireza Dehghani |
| 17.10 | HU | Huib Aldewereld |
| 17.15 | CNR | Francesco Gargiulo |
| 17.20 | ENEA | Gabriele Piantadosi & Saverio de Vito |
| 17.25 | Nathean | John Pugh |
| 17.30 | Meditech, TeaTek, NetGroup | Stefano Marrone |
| 17.35 | RealAI | Tarry Singh |
| 17.40 | Fiven | Alessandro Barducci |

HCAIM Thesis

CeADAR, Ireland's Centre for Applied AI

Dr Alireza Dehghani - AI Technical Programme Manager

Dr Oisín Boydell - Director of the Applied Research Group

1



Ethical Implications of AI Digital Twins in Healthcare

Develop ethical guidelines for the use of AI Digital Twins in healthcare, focusing on ensuring data privacy, informed consent, inclusivity, and bias mitigation to promote equitable and secure healthcare outcomes.

2



Public Engagement and Trust Building in AI Digital Twins Deployment

Investigate and enhance public engagement and trust-building strategies for AI Digital Twins deployment, aiming to create transparent, inclusive, and ethical practices that align with societal values and promote broader acceptance of AIDT technologies.

3



Survey on Human-Centric AI Digital Twins Regulatory Frameworks

Survey AIDT regulatory frameworks in healthcare to assess and enhance their human-centricity, ultimately developing policy recommendations to guide ethical and patient-focused AIDT deployment.

4



Ethical Guidelines for AI Digital Twins in a Specified Healthcare Scenario

Formulate a detailed set of ethical guidelines for the use of AI Digital Twins in a specific healthcare scenario, integrating stakeholder perspectives and ensuring responsible AIDT application in line with patient rights and societal values.

5



Human-Centric Data Privacy and Consent in Remote Working Hubs AI Digital Twin Applications

Create a human-centric framework for data privacy and informed consent in Remote Working Hubs' AI Digital Twin applications, balancing the utilization of AI/ML insights with ethical data management and user autonomy.

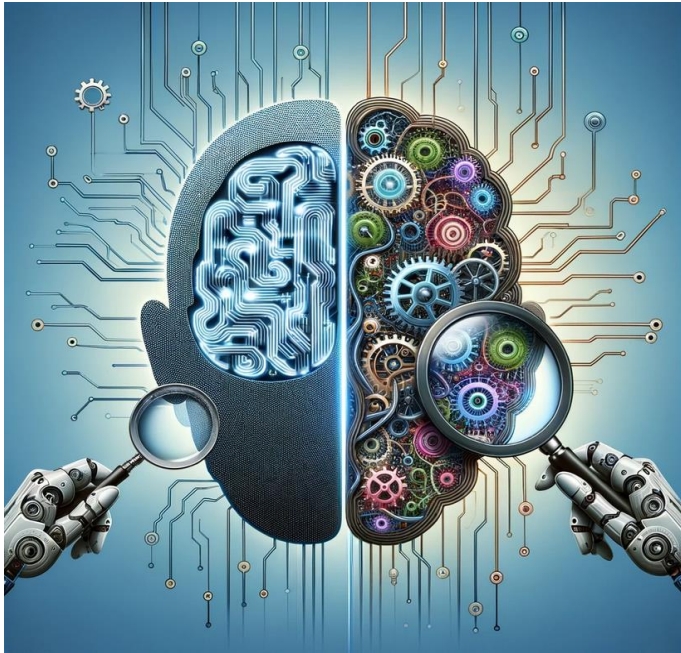
6



Human-in-the-loop AI for Personalised Experience in Remote Working Hubs

Create a Human-in-the-Loop AI framework for Remote Working Hubs that leverages real-time user feedback to provide a personalized and adaptive digital twin experience, focusing on user satisfaction, productivity, and ethical data handling.

7



Bias detection and Generative AI - can the latter enhance the former?

Harness the capabilities of generative AI to identify and mitigate biases in AI systems, focusing on race, gender, and nationality, and to develop ethical AI frameworks that enhance fairness and inclusivity in AI-driven applications.

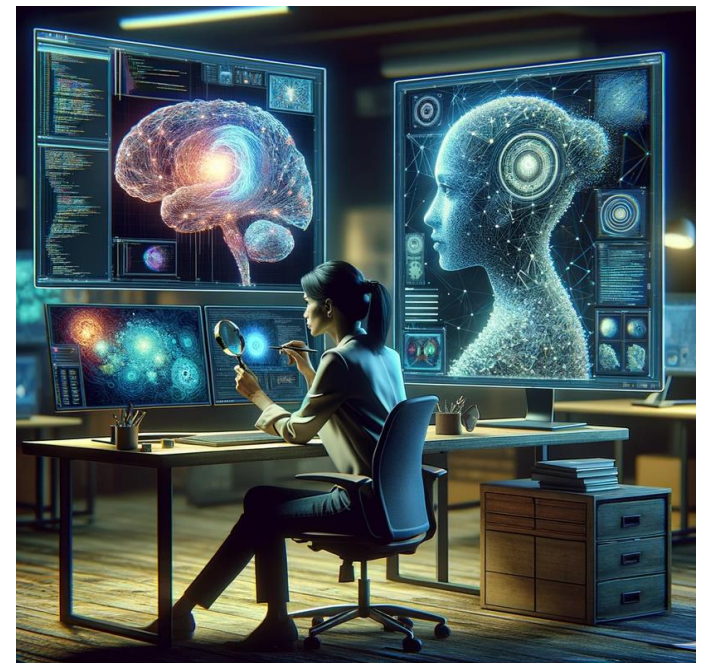
8



Bias detection and privacy enhancing technology - does the latter inhibit the former?

Critically evaluate the relationship between privacy-enhancing technologies and bias detection in AI systems, developing a framework that effectively balances the protection of individual privacy with the ethical need to detect and mitigate bias, particularly when handling sensitive attributes.

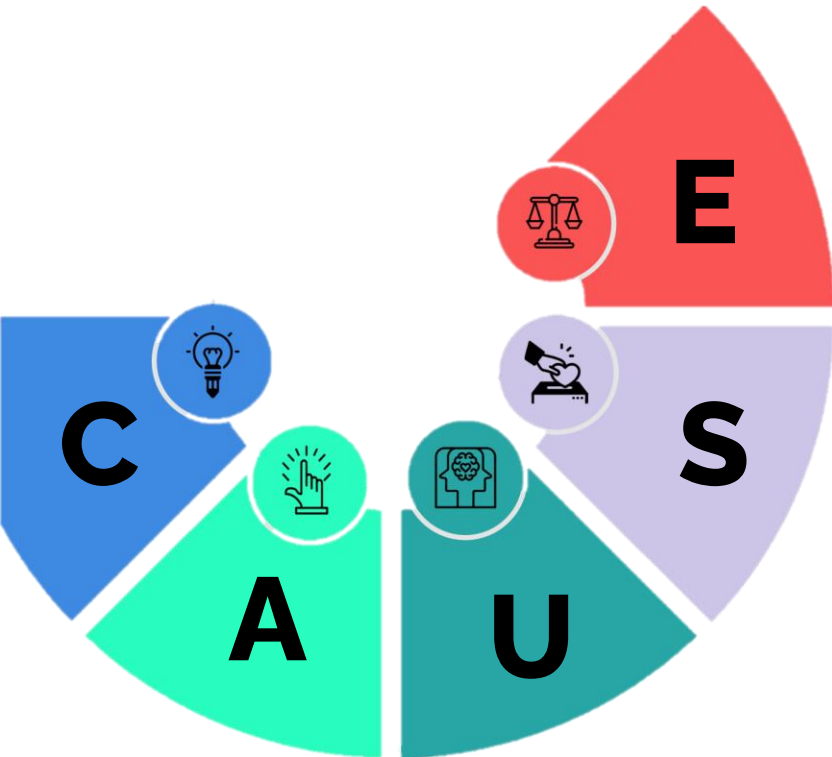
9



What is real? An exploration of tools and methods for identifying AI generated content

Investigate and evaluate the effectiveness and scalability of current and emerging technologies and methodologies for identifying AI-generated content, including watermarking, while considering the societal and ethical implications of supporting transparency and trust in digital media.

HU University of Applied Sciences



Creativity
Autonomy
Understanding
Sentience
Ethics

Future Machine Learning

Generative AI Life sciences & chemistry

In vitro assays

Transparency & XAI

Governance Healthcare

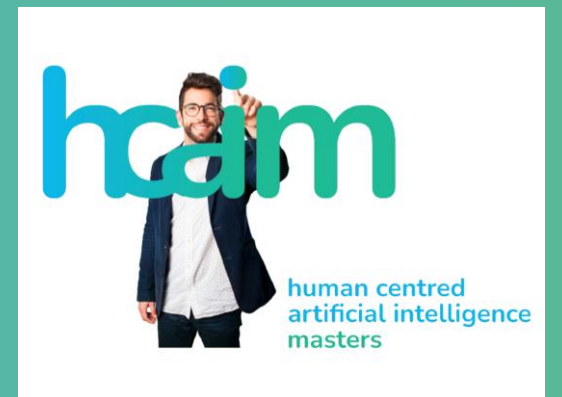
Cooperative AI & AI Governance

Media Public safety



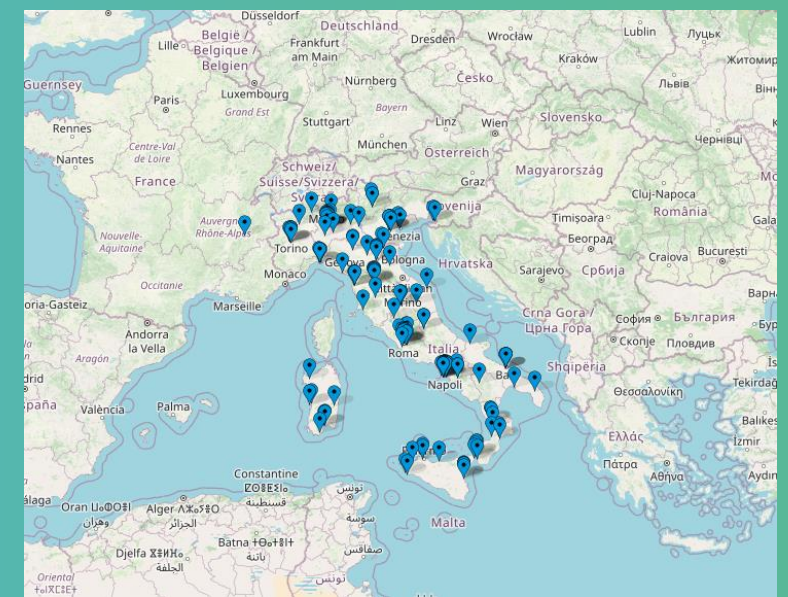
CNR - Research Topics for HCAIM

National Research Council (CNR)





Consiglio Nazionale delle Ricerche



The CNR is a public national research center with multidisciplinary skills, supervised by the Ministry of University and Research (MUR) founded in 1923.

- 7 Departments for macro-thematic areas.
- 88 research institutes
- with 8,500 employees operating throughout the national territory
- of which over 5,600 engaged in research and research support activities.



ICAR → LANGUAGE AND KNOWLEDGE ENGINEERING (LKE)



Dr. Massimo Esposito



Dr. Raffaele Guarasci



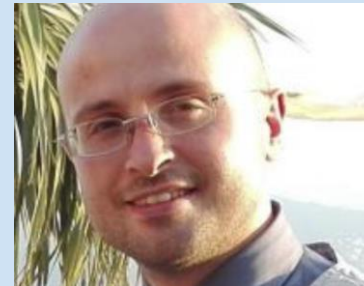
Dr. Aniello Minutolo



Dr. Maria Antonietta Panza



Dr. Marco Pota



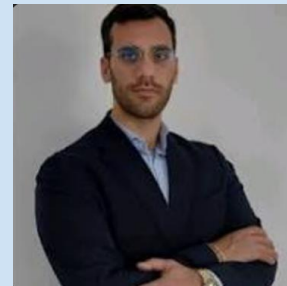
Dr. Francesco Gargiulo



Dr. Giuseppe Buonaiuto



Dr. Chiara Marullo



Dr. Ciro Mennella

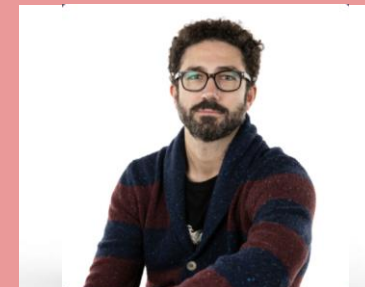


ICAR → HUMAN-ROBOT INTERACTION (HRI)



Dr. Umberto Maniscalco

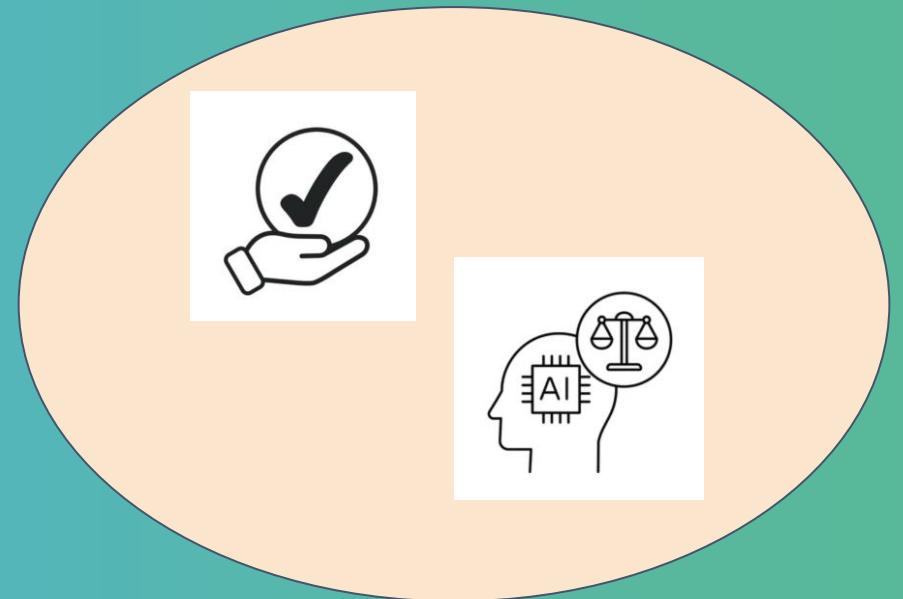
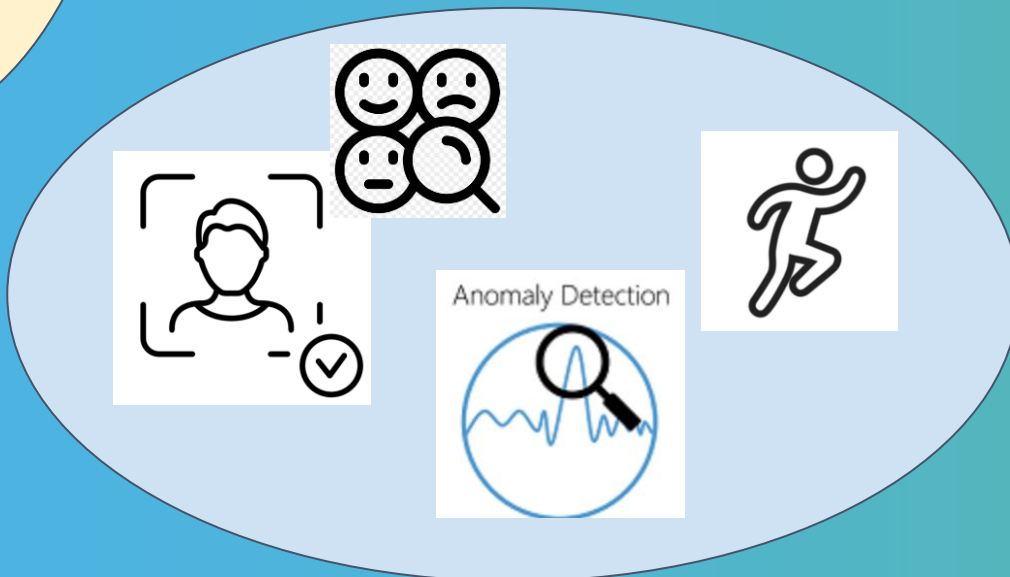
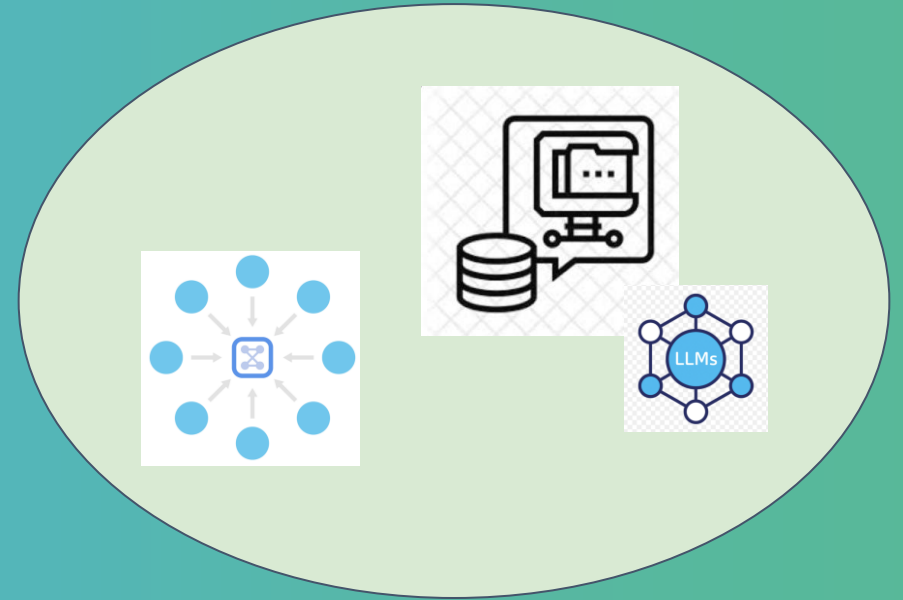
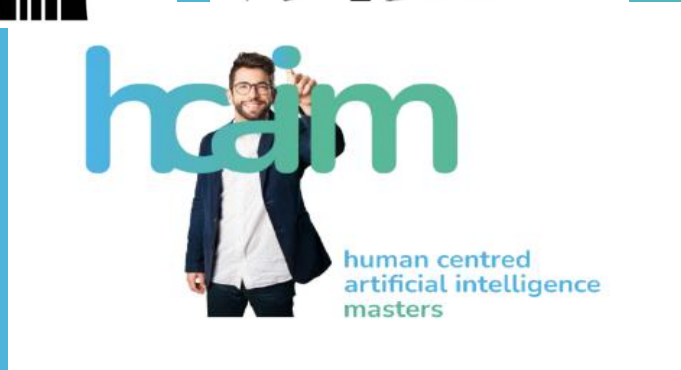
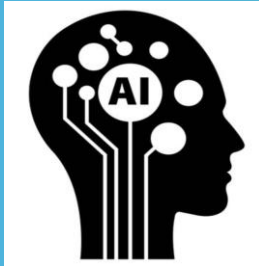
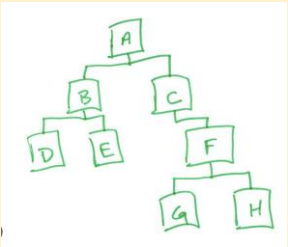
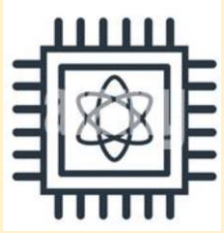
INSTITUTE OF INFORMATICS AND TELEMATICS (IIT)



Dr. Lorenzo Valerio



Research Topics for HCAIM



1. **Quantum Machine Learning - Generative Models: Practical and Ethical Issues.**
 - Quantum Computing; Generative Models; Ethical Considerations; Regulatory Compliance; Bias and Fairness
 - Dr. Giuseppe Buonaiuto (giuseppe.buonaiuto@icar.cnr.it), Dr. Francesco Gargiulo (francesco.gargiulo@icar.cnr.it)
2. **How deep neural networks learn hierarchical data?**
 - Deep networks, Disordered systems, structured data, Ethical Issue
 - Dr. Chiara Marullo (chiara.marullo@icar.cnr.it), Dr. Giuseppe Buonaiuto (giuseppe.buonaiuto@icar.cnr.it)
3. **Enhancing Fairness in Face Recognition and Emotion Detection**
 - Computer vision, Face recognition, Emotion detection, Fairness, Privacy, Data protection
 - Dr. Ciro Mennella (ciro.mennella@icar.cnr.it), Dr. Massimo Esposito (massimo.esposito@icar.cnr.it)
4. **AI for human activity monitoring and evaluation systems**
 - Pose estimation, Pattern Recognition, Motion Analysis, Privacy
 - Dr. Ciro Mennella (ciro.mennella@icar.cnr.it), Dr. Umberto Maniscalco (umberto.maniscalco@icar.cnr.it)
5. **Fair anomaly detection in industry**
 - anomaly detection, deep learning, multi-sensor data, fairness
 - Dr. Maria Antonietta Panza (mariaantonieta.panza@icar.cnr.it), Dr. Massimo Esposito (massimo.esposito@icar.cnr.it)
6. **Comparative Analysis of Generative AI Models' Fairness**
 - Generative Models, Justice, Trustworthy, Fairness, Measure
 - Dr. Francesco Gargiulo (francesco.gargiulo@icar.cnr.it)
7. **Trustworthiness in AI applications**
 - Trustworthiness; Accuracy, Decision Support Systems.
 - Dr. Marco Pota (marco.pota@icar.cnr.it)
8. **Federated Learning**
 - Deep Learning, Federated Learning, Human-centred AI, Fairness
 - Dr. Lorenzo Valerio (lorenzo.valerio@iit.cnr.it), Dr. Massimo Esposito (massimo.esposito@icar.cnr.it)
9. **Impact of model compression on Human-centered aspects**
 - NLP, Language Models, Transformers, Knowledge Distillation, Quantization, Low-rank Optimization
 - Dr. Aniello Minutolo (aniello.minutolo@icar.cnr.it)
10. **Model Compression in Large Language Models**
 - NLP, Language Models, Transformers, Knowledge Distillation, Quantization, Low-rank Optimization
 - Dr. Giuseppe Buonaiuto (giuseppe.buonaiuto@icar.cnr.it), Dr. Massimo Esposito (massimo.esposito@icar.cnr.it)



Italian National Agency for New Technologies, Energy and Sustainable Economic Development

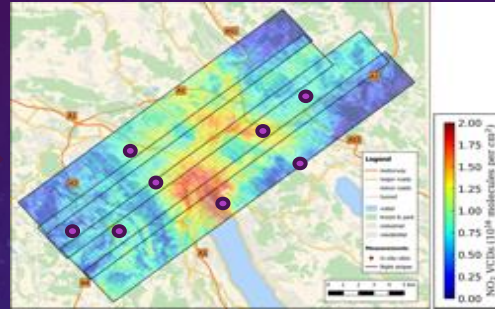
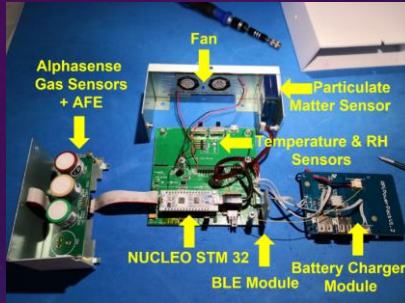


ENEA: National Agency for New Technologies, Energy and Sustainable Economic Development, a public agency aimed at research, technological innovation and the provision of advanced services to enterprises, public administration and citizens in the sectors of energy, the environment and sustainable economic development

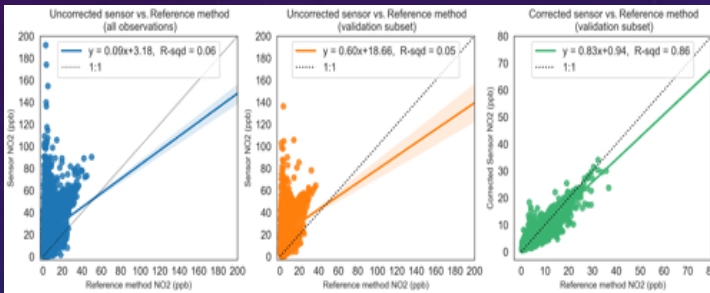
ENEA C.R. Portici: The center is located in Portici (NA), facing the sea, in front of the Bourbon port, less than 100m from the Portici-Ercolano FS national train station and less than 2km from the Portici-Via Libertà local train station. The center has free parking for thesis students. Food services are available in the area. Part of the thesis can be carried out remotely.

Equipment: Personal workstation in open space. Access to Intelligent Sensing and GIS lab equipped with several WS with NVIDIA A5000 GPUs. Availability of use of CRESCO supercomputing facilities. Access to datasets relevant to the subject application from both public (free to use, institutional - WFS, ECMWF, Copernicus) and commercial sources as well as from international collaborations on the subject.

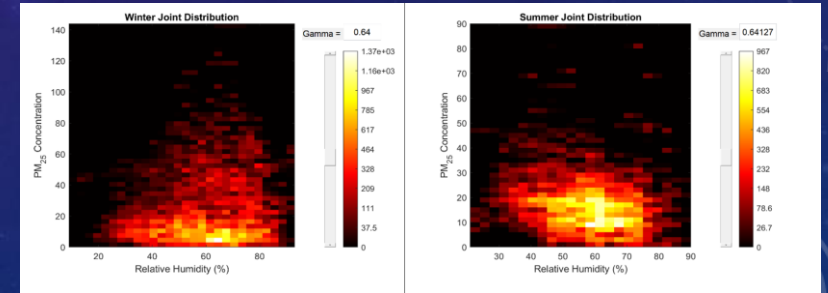
IA FOR IOT: AIR QUALITY SENSORS IN-NETWORK CALIBRATION



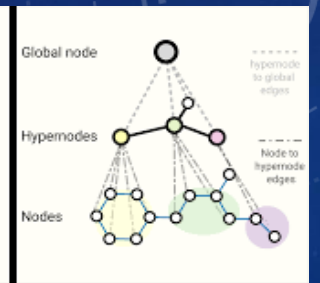
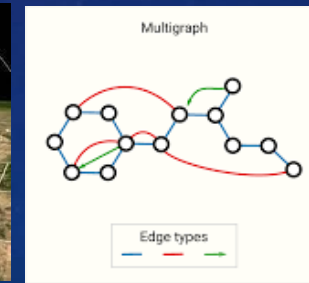
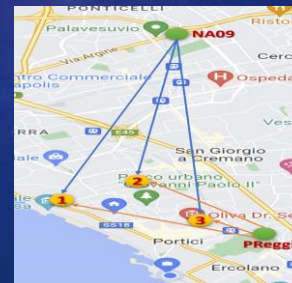
- IoT Air Quality multisensor devices can bring high spatial resolution to the sparse network of costly regulatory grade analysers currently unable to cope with AQ spatial variance.
- Unfortunately, these IoT device suffer from lack of accuracy due to inherent sensors non-specificity, non-linearities, instabilities, and are affected by environmental conditions.



Machine learning techniques are paramount in deriving accurate concentrations estimation from raw data. However, systems are needed to learn in a dynamic environment under constantly changing conditions from co-located ground truth stations which is unfeasible.



OSINT data (Conventional AQ Network, Satellite data) begins to be exploited to guarantee continuous recalibration in a dynamically changing environment with sensors drift. However, we currently lack structured, globally applicable approaches. We believe that solutions lies in extracting useful content from both OSINT data and raw sensor responses.

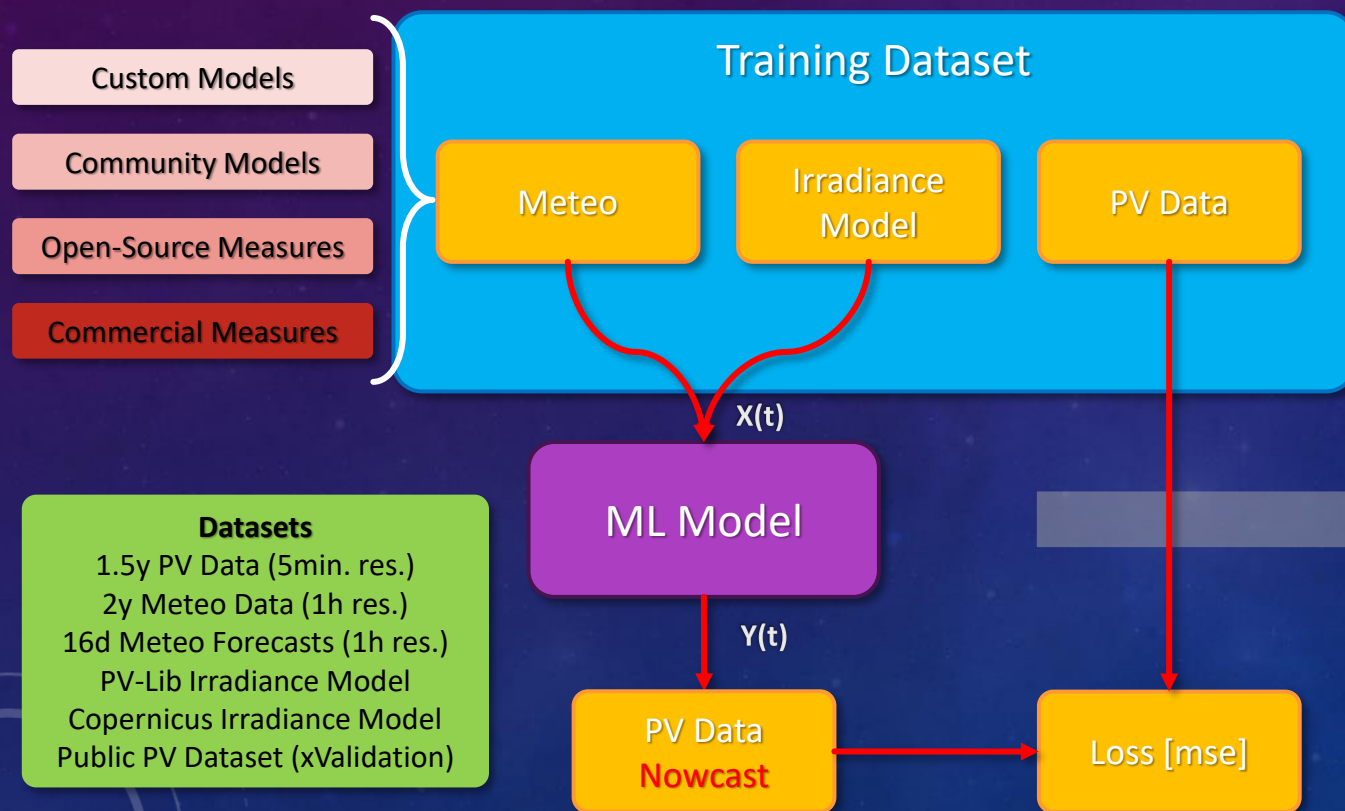


...as such we want to explore GNN and Transformer architecture to provide a general solution.

PV PRODUCTION NOWCASTING & FORECASTING

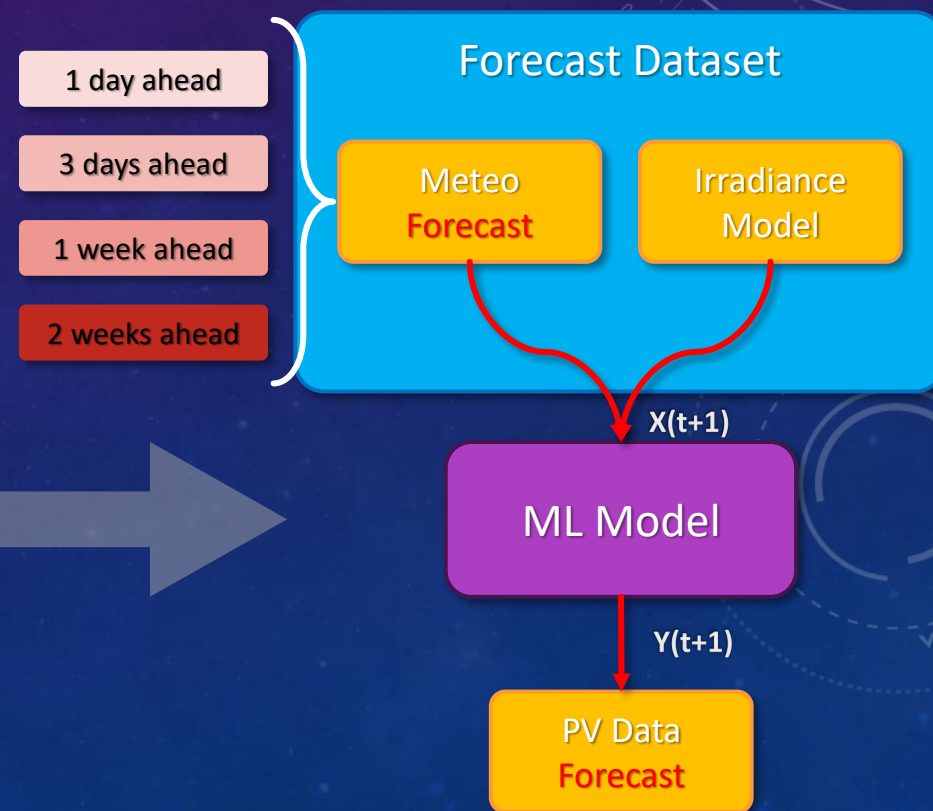
nowcasting

Training Environment (1day horizon)



forecasting

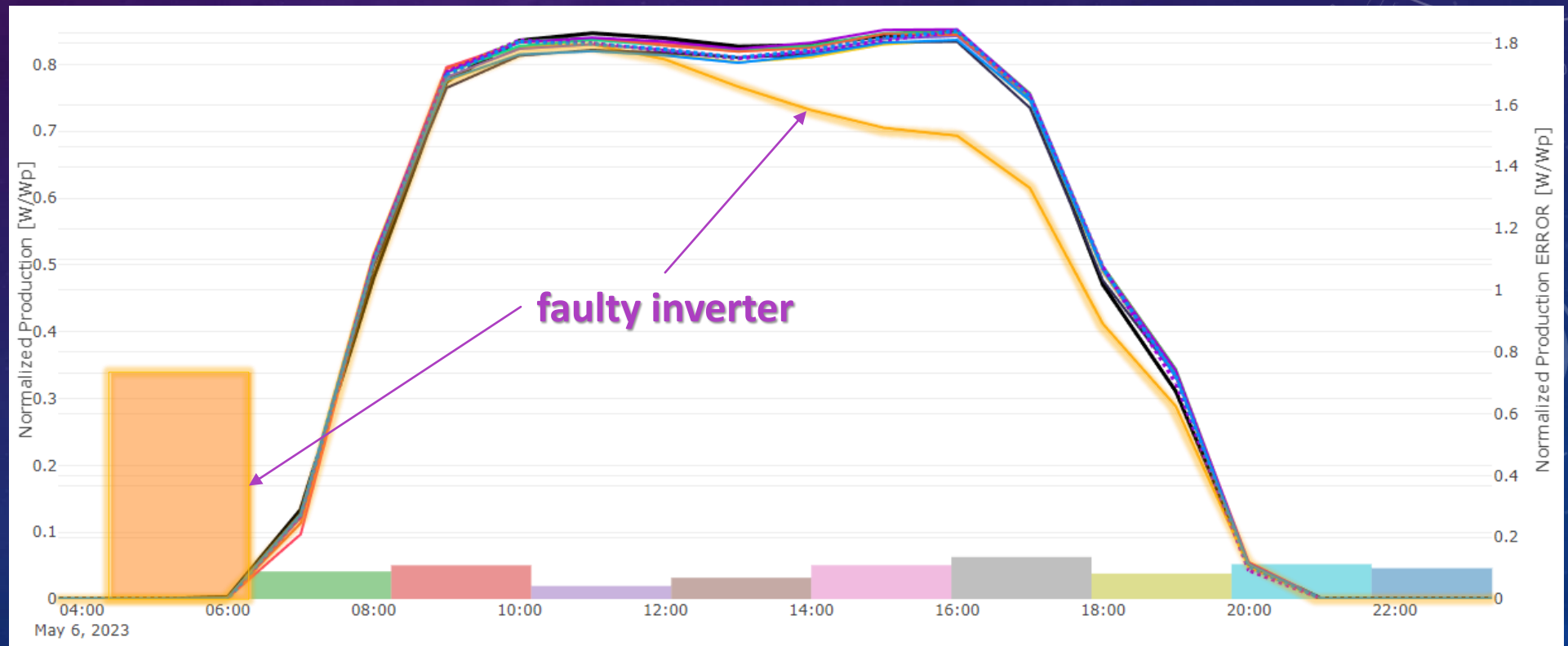
Production Environment



PV ANOMALY DETECTION

Key idea:

- Well defined **nowcasting model** is able to provide approximately less than 1% error
- By **comparing the forecasted and the measured** values on a whole day inverter by inverter is possible to **discriminate faulty inverters**



Preliminary results obtained by total daily Normalized Production Error with respect to the predicted production is able to detect an anomaly at inverter-level.



Nathean Thesis #2 for HCAIM 2023

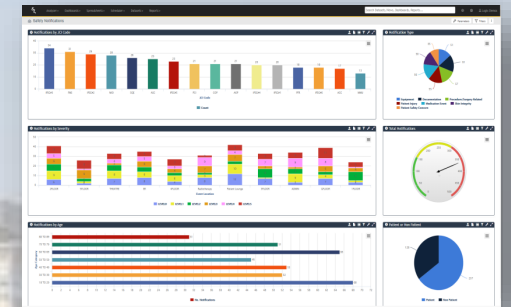
Topic: Protecting Privacy: A Critical Evaluation of Advanced Algorithms and LLMs for Data Re-identification and Effective Countermeasures



MAURICE LYNCH & JOHN PUGH | 13.01.2023



- Founded in Dublin in 2001
- Data Analytics Software & Services company
- Cross-sector global customers
- Industry founding member of CeADAR



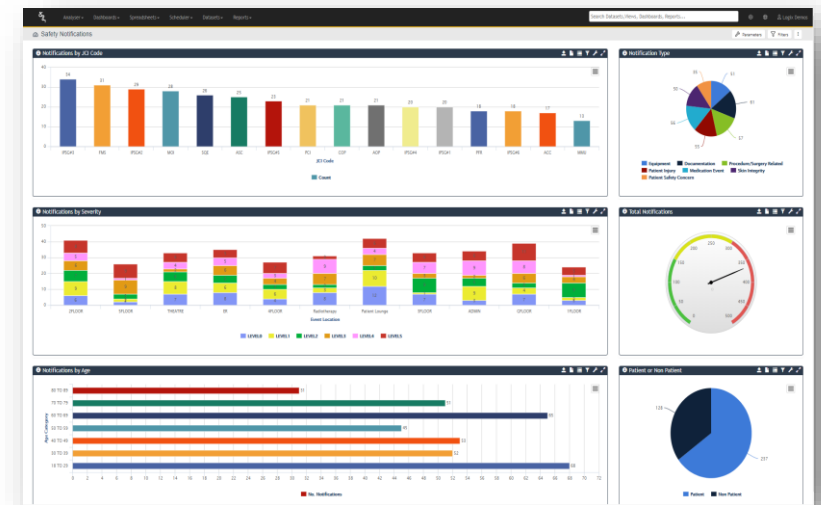
Topic: Protecting Privacy: A Critical Evaluation of Advanced Algorithms and LLMs for **Data Re-identification** and Effective Countermeasures



Aim: Evaluation of technologies and techniques for re-identification of private data including the use of LLMs

Definition: A method of training ML models without exposing data to developers or users

Investigation: Comparison of homomorphic encryption, differential privacy, multi-party computation and federated learning



Evaluation of technologies and techniques for countering the threat of data being re-identified :

- **Understanding re-identification** This involves looking into how anonymised data can be cross-referenced with other data sources to identify individuals.
- Investigate how the various PPML techniques can **protect against re-identification**

- Accessing and using significant datasets from **real-world patient data**
- Ensuring data is acquired with consent and **obfuscated** for model development.



- Contribute to the knowledge of **PPML** and its potential for responsible and ethical use of ML in healthcare.
- Protect **privacy** of patients.



THESIS PROPOSAL



Name _____

Signature _____

Date _____

Prof. Flora Amato

Digitàlizzazione

Al fianco delle PMI

Accompagniamo la tua impresa nel processo di trasformazione
tecnologica

01
Al fianco delle
PMI

02
Crea valore con
Noi

03
Al passo con i
Tempi

Engaging Ethics and AI in the Exploration of Data Spaces: Balancing Technological Advancements with Moral Imperatives

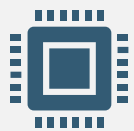
+

•

○



The thesis aims to analyze the multifaceted role of the Data Space in the AI ecosystem, from meeting the demands of data-hungry AI applications to addressing the challenges of data sharing, security, and governance.



By exploring these key issues, the research seeks to contribute to a deeper understanding of the Data Space as a critical infrastructure supporting AI advancements and providing guidance for its effective implementation and governance.

This thesis will regard
a comprehensive
investigation on the
methodologies for
comparing AI models
in terms of fairness



Data Sharing Challenges



Technology Framework for Data Space



Data Space Governance



Legal and Ethical Implications

tea tek

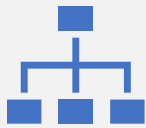
TECNOLOGIA • ENERGIA • AUTOMAZIONE

[CONTATTACI](#)

[SEGUICI SU LINKEDIN](#)

TEATEK
IMPIANTI ELETTRICI, FOTOVOLTAICI, MACCHINARI E SOFTWARE

A Trustworthy AI Approach to Navigating the Regulatory Framework in the Defence Domain



The Defence system is based on a well-structured and detailed regulatory framework. This network of standards, directives and protocols aims to ensure efficiency, security and consistency of operations.



The regulatory framework evolves rapidly, so navigating through all the documents and adaptations becomes problematic.



In addition, one can often face a multitude of documents, which can create confusion or overlap. It is crucial to find a solution that aims to give support in this context, in which it is essential to be informed and aware of the operations they perform and the choices they make.



This thesis studies trustworthy, robust, safe and fair AI methodology. The methodology will use Generative AI and NLP techniques to analyse and create content; it must respect the canons of Trustworthy, robust, safe and fair AI

Aim

- This thesis will regard a comprehensive investigation on the methodologies utilizing Trustworthy, Robust, Safe, and Fair Artificial Intelligence (AI), specifically incorporating Generative AI and Natural Language Processing (NLP) techniques, to streamline and optimize the navigation, understanding, and application of the rapidly evolving regulatory framework in Defence operations, thereby enhancing operational efficiency, security, and consistency.
- Legal and Ethical Implications:
 - What legal and ethical challenges arise in the deployment of generative AI models, particularly concerning fairness?
 - How can regulatory frameworks be adapted to address fairness issues in AI?

Innovazione per passione

Siamo completamente concentrati e pronti a rispondere alle sfide che il PNRR lancerà all'Italia nei prossimi anni

Scopri le nostre Expertise



Intelligent Chatbots in Support of Specialized Technical Operators: Performance and Ethical Implications

- A Specialized Technical Operator's role involves many responsibilities requiring precision, expertise, and quick decision-making.
- This research aims to study the potential of intelligent chatbots in aiding these operators to enhance their efficiency, precision, and responsiveness and explore the ethical aspects concerning the massive adoption of such technologies.

Aim

- This thesis aims to comprehensively investigate the capabilities and impact of intelligent chatbots in enhancing the performance attributes, such as efficiency, precision, and responsiveness, of specialised technical operators in their multifaceted roles.
- This thesis seeks to delve into the ethical considerations and implications arising from the widespread adoption of such digital aids in the operational realm, ensuring a holistic understanding of both the potential advantages and the challenges posed by this technological integration.
- Legal and Ethical Implications:
 - What legal and ethical challenges arise in the deployment of generative AI models, particularly giving support to Support of Specialized Technical Operators?

FIVEN

We re-define experiences
through design
and technology.

FIVEN
FUTURE DRIVEN



Embracing the Change

FIVEN
FUTURE DRIVEN

Future-Driven

Ready to embrace every change as an opportunity to learn and to deliver innovation and ideas

Ready to listen, to understand and to answer

Ready to explore, to discover and to share

Driven by and to a tomorrow where technology and design shape new experiences of choice and freedom for us all

MY ARTIFICIAL INTELLIGENCE PLATFORM

Embracing the Change

Future-Driven.

Ready to embrace every change as an opportunity to learn and to deliver innovation and ideas.

Ready to listen, to understand and to answer. Ready to explore, to discover and to share. Driven by and to a tomorrow where technology and design shape new experiences of choice and freedom for us all.

MyAiP: Artificial Intelligence as a Tool

MyAiP is the suite of AI solutions, available in the cloud or on-premise, designed to manage business processes and activities, from CV screening to omnichannel customer support.

FIVEN
FUTURE DRIVEN



Our Values

Innovation

We build and promote a true culture of innovation based on dynamism, fluid processes and ideas. We strive for excellence and newness: the ever-changing technological and methodological landscape is the perfect playground for our creativity, skills and accountability. We work to capture and create value in new, thrilling ways.

Positive Impact

We are here to make a difference. We are committed to the environmental, social and economic development of the territory we live in. We work to create and sustain a tangible, positive value proposition for each and every stakeholder, direct or indirect, from our employees to those who come in contact with our business and its purpose. We undertake to promote a culture of meritocracy, openness, equal opportunities, kindness and respect in our own organization and in the society at large.

FIVEN
FUTURE DRIVEN

Tailoring

Our Italian heritage is made of inventiveness, excellence, passion and eye for detail. We are partners to our clients and their customers. We build long-lasting relations based on trust, transparency and dialogue, operating with complete integrity to turn every project into a meaningful experience of quality, closeness and success.

Our updated 2024 Theses

Thesis 01 - Improving algorithms, explainability, information retrieval and accuracy measurement for real-world PJF CV extraction and ranking

Thesis 02 - Real-world chatbot performance measurement and self-training

Thesis 03 - Real-time sentiment analysis in real-world chatbot conversations

Thesis 04 - NLP and classification algorithms for mail dispatching

Improving algorithms, explainability, information retrieval and accuracy measurement for real-world PJF CV extraction and ranking

Fiven will accept multiple theses proposals in this research area, with a maximum of **six** different candidates.

PJF – Person-Job Fitting: paramount in current recruiting processes.

Companies are overwhelmed with a flood of CVs coming from everywhere in the world.

Most companies are forced to use PJF systems as first-level filters, providing a first, coarse ranking of the incoming CVs.

Current accuracy of these systems is generally quite low, and even measuring this accuracy is a challenging task.

Recruiting policies, and therefore ranking criteria, greatly varies between companies and HR experts.

CVs are presented in different formats.

Identifying whether some information is related to one or more soft or hard skills is often difficult.

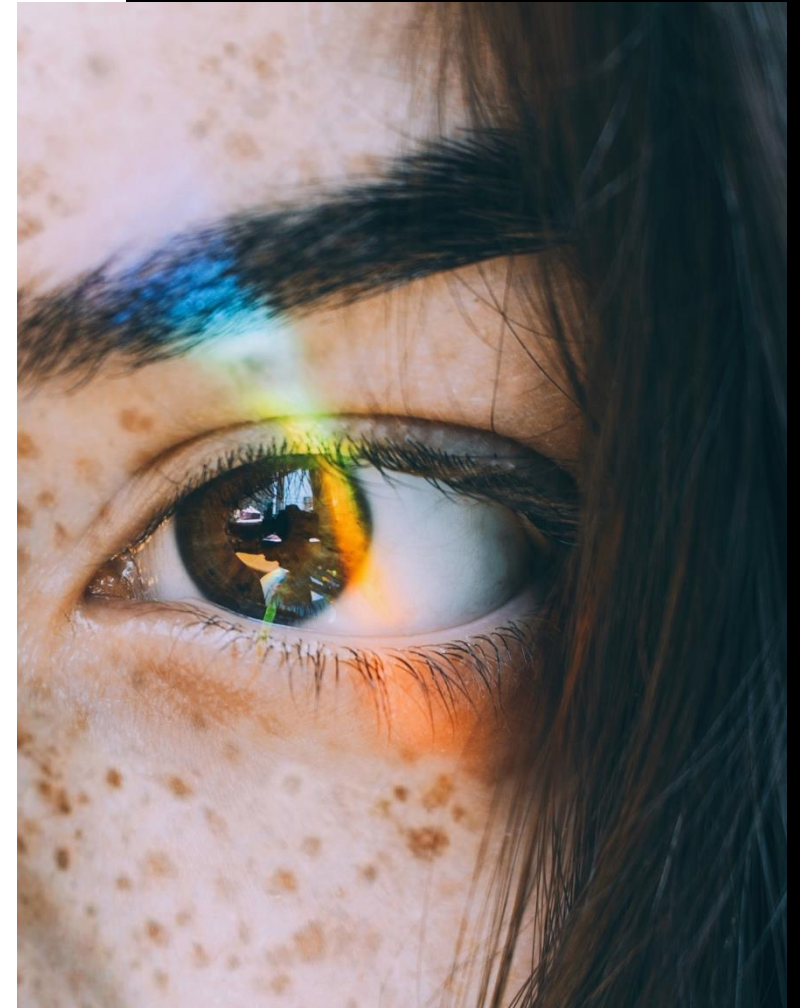
Real-world CVs strongly differs from benchmark texts used in academical research.

Candidates may use some “tricks” to achieve a better ranking.

Research Goals

Improving our current algorithms for CV retrieval and ranking. We have already achieved top-level results, but our goal is to improve even more.

- The student will focus on proposing and testing
 - one or more new algorithms for **improving CV extraction** accuracy.
 - one or more methods for **measuring information retrieval** accuracy.
 - one or more methods for **measuring candidates ranking** accuracy.
 - one or more methods for measuring accuracy of the explainability
- Besides classical LLM algorithms, **innovative hybrid algorithms** are welcome.
- Measuring methods may include **bias-detection** and **cheat-detection** algorithms.
- Expected outcome: one or more measuring methods for CVs ranking and/or extraction accuracy, for a set of given job positions, with their respective global estimated accuracy and as many partial accuracy estimations as possible.
- The solutions should fit in **a real-world software**, using **real-world CVs**.
- We strongly encourage collaboration between students working in projects in this Area.



Ethics and Human-Centred AI goals

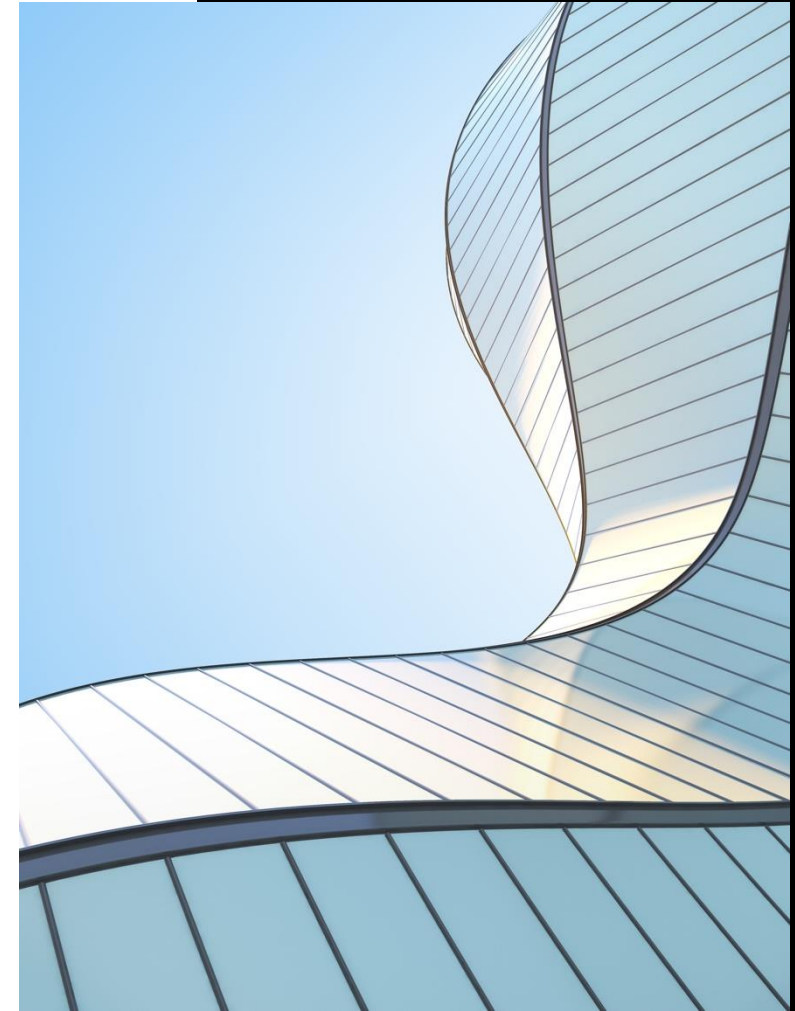
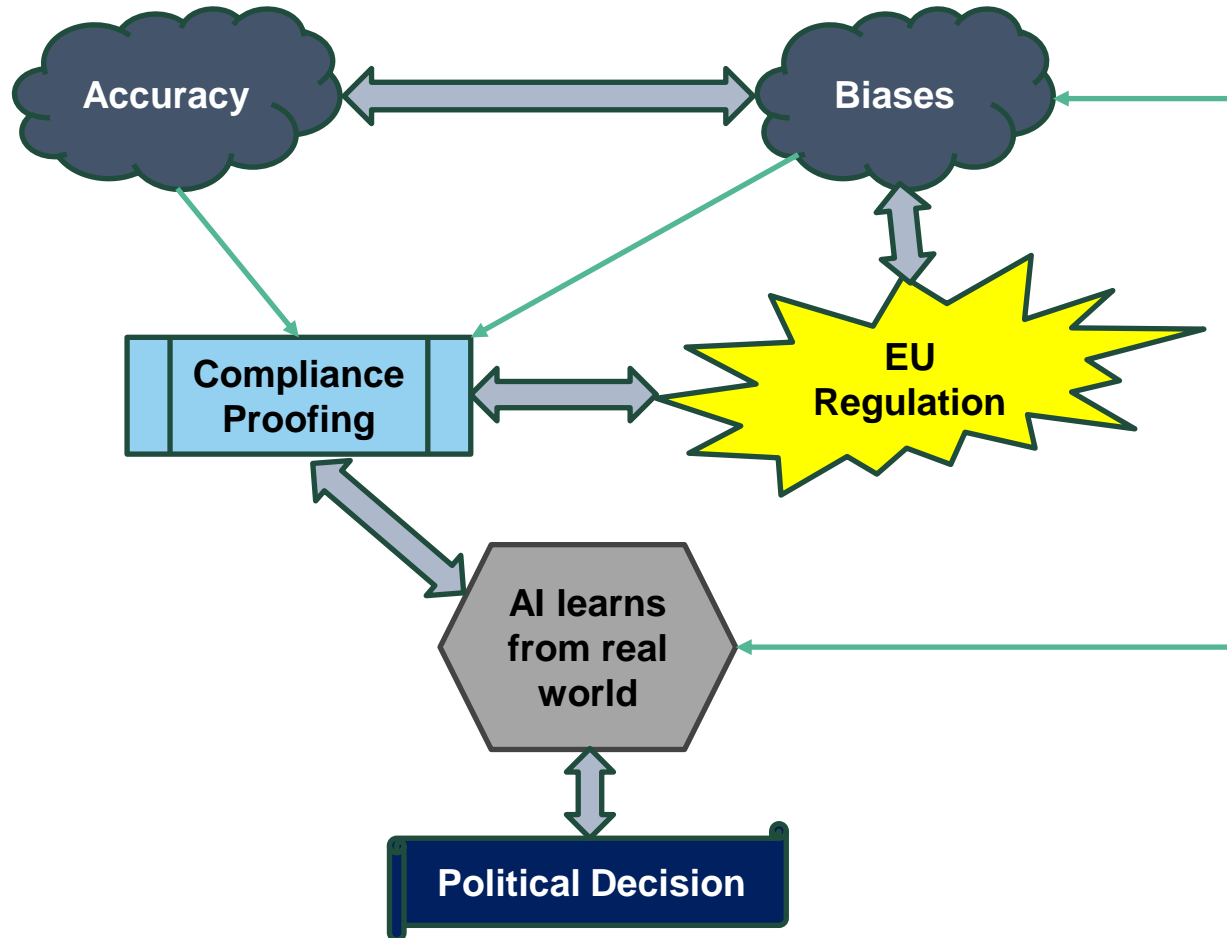
Accuracy in candidate information extraction is paramount in order to achieve a fair and complete candidates' ranking.

Improving extraction and/or ranking accuracy, and accuracy measurement, provides a relevant contribution to AI ethical goals.

Theses dealing with accuracy improvements and/or new algorithms proposals should also verify – and possibly formally proof – that no Ai-ethics related bias of any kind is present.

Explainability is another key element of an ethical AI PJF system, since it strongly enhances our ability to verify its correctness and fairness.

What about Ethics?



Possible research questions

Besides the easier, most basic technical issues, we may want to face come more complex questions, as for example:

How can we decide that a CV is “well” or “poorly” written, and therefore “deserves” not being read correctly?

How can we detect a bias in a biased world?

Where is the frontier between legitimately stress candidate’s skills and achievements and cheating?

Until a certain level of ranking accuracy, we can measure it by simply making a comparison with a “perfect” human ranking. But at some point our software may become “better than human”. How can we prove this?

Do explainability definition has some extra nuances in PJF area? Can we improve its definition?

How can we measure explainability effectiveness? I.e., how can we objectively measure how understandable is our explanation?

Can we use a list of criteria and a table of word-matching count to build a more comprehensive and human-friendly explanation?

How can we justify our decisions without disclosing sensible information about the candidate?



Real-world chatbot performance measurement and self-training

Fiven will accept multiple theses proposals in this research area, with a maximum of **three** different candidates

Human language is astoundingly complex and diverse

Real-world conversations are very different from examples commonly used in research works

Within each language there is a unique set of grammar and syntax rules, terms and slang

NLP is important because it helps resolve ambiguity in language and adds useful numeric structure to the data for many downstream applications

Research Goals

Analyse whether is possible to implement some self-training functionality in a real-world chatbot with a reasonable implementation effort.

Find better automated methods to measure chatbot performances; compare accuracy with existing manual processes.

Data used: real-world chatbot conversation transcripts. Response rating both via complete manual analysis of every conversation and using statistical sampling methods.

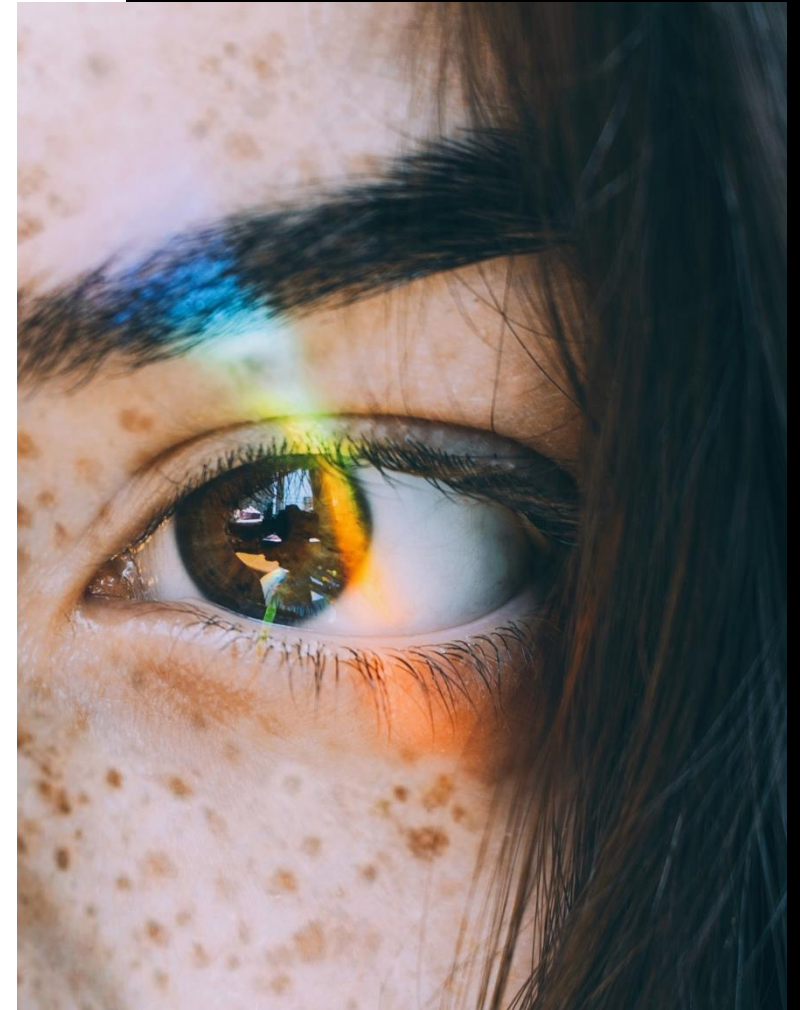
Expected outcomes:

- Realistic evaluation of the needed effort; prototype algorithm in the case of a positive answer.
- Automated AI rating methods for measuring chatbot performances with acceptable accuracy.

Activities:

- Analyse existing solutions both in the commercial and in the academic environments.
- Propose rating algorithms, possibly partially based on existing ones.
- Analyse estimated effort and performances, in order to obtain a cost/benefit ratio.
- Give a contribution to their implementation.
- Measure global and partial accuracy of the algorithms
- Provide further improvements based on performance and accuracy results

We strongly encourage tight collaboration between students working in projects in this Area.



Ethics and Human-Centred AI goals

Improving human language processing performances will provide better correctness to any AI NLP system, and will also provide a more natural and pleasant experience to the users. Furthermore, a correct, unbiased classification of the topics and of the user's feelings will greatly improve both man-machine interaction and the possible, subsequent escalation to a human operator.

Theses dealing with new algorithms proposals should obviously carefully verify and possibly demonstrate that no ethics-related bias has been introduced in the new algorithms.

Furthermore, in some cases users may be willing to "cheat", for example, by "convincing" the chatbot that they have paid the bill and so their internet service should be reactivated immediately. Such cases become even more relevant in modern chatbots which are often able to perform actions, besides just providing information.

Unavoidable trade-offs between performances and correctness may also rise ethical issues.

Possible research questions

How can we (re)define chatbot performance?

- o User satisfaction
- o Conversations correctly understood
- o Conversational flows correctly completed
- o Business goals achieved

Please note that some of these methods can be somewhat in conflict: for ex., a business goal may be to bar users with some pending debt with the company from accessing some service. If this task is performed correctly, user satisfaction will be likely low.

Another business goal may be of completing a conversation within a given time. This may be in contrast with achieving an high correctness of conversational flows.

Ethical issues may also arise: for the sake of efficiency, is acceptable to provide an higher percentage of wrong answers?

How can we measure errors severity? There should be a difference from a chatbot simply answering "I don't know", to another one providing a completely false, subtly misleading or even potentially dangerous information.



Host and supervision

Supporting Universities

The theses are available for students from:

- Università degli Studi di Napoli Federico II, under the supervision of Prof. Stefano Marrone
- Technological University Dublin
- Hogeschool Utrecht
- Budapesti Műszaki És Gazdaságtudományi Egyetem

