2024

ISTI Annual Reports

AIMH Research Activities 2024

AIMH Lab., CNR-ISTI, Pisa, Italy

ISTI-AR-2024/001

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

AIMH Research Activities 2024
AIMH Lab.
ISTI-AR-2024/001

The AIMH (Artificial Intelligence for Media and Humanities) laboratory is committed to advancing the field of Artificial Intelligence, with a special emphasis on its applications in digital media and the humanities. The lab aims to improve AI technologies, particularly in areas such as deep learning, text analysis, computer vision, multimedia information retrieval, content analysis, recognition, and retrieval. This report summarizes the laboratory's achievements and activities over the course of 2024.

Keywords: Multimedia Information Retrieval, Artificial Intelligence, Computer Vision, Similarity Search, Machine Learning, Text Classification, Quantification, Deep Learning, Transfer learning, Representation Learning, Knowledge Representation, Digital Humanities.

Citation
AIMH Lab. *AIMH Research Activities 2024*, ISTI Annual Reports 2024/001. DOI: 10.32079/ISTI-AR-2024/001.

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" Area della Ricerca CNR di Pisa
Via G. Moruzzi 1 56124, Pisa Italy
http://www.isti.cnr.it

Annual Report, DOI: 10.32079/ISTI-AR-2024/001, December 2024
Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

AIMH Research Activities 2024

Nicola Aloia, Giuseppe Amato, Valentina Bartalesi, Lorenzo Bianchi, Paolo Bolettieri, Catherine Bosio, Michele Carraglia, Fabio Carrara, Vittore Casarosa, Maria Cassese, Luca Ciampi, Davide Alessandro Coccomini, Cesare Concordia, Richard Connor, Silvia Corbara, Claudio De Martino, Marco Di Benedetto, Andrea Esuli, Fabrizio Falchi, Edoardo Fazzari, Claudio Gennaro, Ludovico Iannello, Kajal Negi, Gabriele Lagani, Emanuele Lenzi, Martina Leocata, Marco Malvaldi, Carlo Meghini, Nicola Messina, Alejandro Moreo, Alessandro Nardi, Giacomo Pacini, Andrea Pedrotti, Nicolò Pratelli, Giovanni Puccetti, Fausto Rabitti, Pasquale Savino, Francesca Scotti, Fabrizio Sebastiani, Gianluca Sperduti, Costantino Thanos, Luca Trupiano, Lucia Vadicamo, Claudio Vairo, Loredana Versienti, Lorenzo Volpi.

Abstract
The AIMH (Artificial Intelligence for Media and Humanities) laboratory is committed to advancing the field of Artificial Intelligence, with a special emphasis on its applications in digital media and the humanities. The lab aims to improve AI technologies, particularly in areas such as deep learning, text analysis, computer vision, multimedia information retrieval, content analysis, recognition, and retrieval. This report summarizes the laboratory's achievements and activities over the course of 2024.

Keywords
Multimedia Information Retrieval – Artificial Intelligence — Computer Vision — Similarity Search – Machine Learning – Text Classification – Quantification – Deep Learning – Transfer learning – Representation Learning – Knowledge Representation – Digital Humanities

[1] *AIMH Lab, ISTI-CNR, via Giuseppe Moruzzi, 1 - 56124 Pisa, Italy*
*__Corresponding author__: giuseppe.amato@isti.cnr.it

Contents

AIMH
Artificial Intelligence for Media and Humanities

http://aimh.isti.cnr.it

Introduction

The Artificial Intelligence for Media and Humanities laboratory (AIMH) of the Information Science and Technologies Institute "Alessandro Faedo" (ISTI) of the Italian National Research Council (CNR) located in Pisa, has the mission to investigate and advance the state-of-the-art in the Artificial Intelligence field, specifically addressing applications to digital media and digital humanities, and also taking into account issues related to scalability.

The laboratory is composed of four research groups for a total, at the end of 2024, of 45 people:

- 17 Researchers
  - *2 Directors of Research*
  - *9 Senior Researchers*
  - *1 Researchers*
  - *5 Temp. Researchers*
- 2 Technologists
  - *1 Temp. Senior Technologist*
  - *1 Technologists*
- 2 Technicians
- 11 PhD Students
- 3 Graduate Fellows
- 6 Research Associates

AI4Text

The AI4Text group is active in the area at the crossroads of machine learning and text analysis; it investigates novel algorithms and methodologies, and novel applications of these to different realms of text analysis. Topics within the above-mentioned area that are actively researched within the group include representation learning applied to text, language modeling, learning to quantify, cross-lingual and cross-domain transfer learning, sequence learning for information extraction, transductive learning, cost-sensitive learning. During this year, the group consisted of Fabrizio Sebastiani (Director of Research), Andrea Esuli (Senior Researcher), Alejandro Moreo, Giovanni Puccetti, Andrea Pedrotti (Researchers), Silvia Corbara, Maria Cassese, Kajal Negi, Martina Leocata, Lorenzo Volpi, and Gianluca Sperduti (PhD Students), and is led by Fabrizio Sebastiani.

Digital Humanities

Investigating AI-based solutions to represent, access, archive, and manage tangible and intangible cultural heritage data. This includes solutions based on Semantic Web technologies and ontologies, with a special focus on narratives and geospatial data, and solutions based on data analysis, recognition, and retrieval. During this year, the group consisted of Valentina Bartalesi, Cesare Concordia (Researchers), Michele Carraglia (Senior Technologist) Luca Trupiano (Technologist), Claudio De Martino (Graduate Fellows), Emanuele Lenzi, Nicolò Pratelli (PhD Students), Nicola Aloia, Vittore Casarosa, Carlo Meghini, and Costantino Thanos (Research Associates), and is led by Valentina Bartalesi.

Large-scale IR

Investigating efficient, effective, and scalable AI-based solutions for searching multimedia content in large datasets of non-annotated data. This includes techniques for multimedia content extraction and representation, scalable access methods for similarity search, multimedia database management. During this year, the group consisted of Claudio Gennaro, Pasquale Savino (Senior Researchers), Nicola Messina, Lucia Vadicamo, Claudio Vairo (Researchers), Paolo Bolettieri (Technician), Fausto Rabitti (Research Associate), Giulio Federico (PhD Student), and is led by Claudio Gennaro.

Vision and Deep Learning

Investigating novel AI-based solutions to image and video content analysis, understanding, and classification. This includes techniques for detection, recognition (object, pedestrian, face, etc), classification, counting, feature extraction (low- and high-level, relational, cross-media, etc), anomaly detection also considering adversarial machine learning threats). We also have specific AI research fields such as Bio-Inspired Deep Learning. The group consists of Giuseppe Amato (Director of Research), Fabrizio Falchi (Senior Researcher), Marco Di Benedetto, Fabio Carrara, Luca Ciampi (Researchers), Alessandro Nardi (Technician), Lorenzo Bianchi, Davide Alessandro Coccomini, Edoardo Fazzari, Giacomo Pacini (PhD Students), Marco Malvaldi (Research Associate), and is led by Fabrizio Falchi.

The rest of the report is organized as follows. In Section **??**, we summarize the research conducted on our main research fields. In Section 1, we describe the projects in which we were involved during the year. We report the complete list of papers we published in 2024, together with their abstract, in Section 2. The list of theses on which we were involved can be found in Section 3. In Section 4.2 we highlight the datasets we created and made publicly available during 2024.

1. Projects & Activities

1.1 EU Projects

AI4media

Artificial Intelligence for the Society and the Media Industry (AI4Media) is a network of research excellence centres delivering advances in AI technology in the media sector. Funded under H2020-EU.2.1.1., AI4Media started in September 2020 and will end in August 2024.

Motivated by the challenges, risks and opportunities that the wide use of AI brings to media, society and politics, AI4Media aspires to become a centre of excellence and a

wide network of researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the key sector of Media, to make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and to reimagine AI as a crucial beneficial enabling technology in the service of Society and Media.

The leader of the AIMH team participating in AI4Media is Fabrizio Sebastiani.

The Craeft project aims to advance our understanding of the various aspects of crafts as a living and developing heritage, a sustainable source of income, and a means of expressing the mind through "imagery, technology, and sedimented knowledge". Drawing on disciplines such as Anthropology, Knowledge Representation, Cognitive Science, Art History, Advanced Digitisation, Audiovisual & Haptic Immersivity, and Computational Intelligence, the project will take a generative approach that can accommodate digital conservation, reenactable preservation, and scaling of approaches for different materials and techniques.

The leader of the AIMH team participating in Craeft is Valentina Bartalesi.

SoBigData++ is a project funded by the European Commission under the H2020 Programme INFRAIA-2019-1, started Jan 1 2020 and ending Dec 31, 2023. SoBigData++ proposes to create the Social Mining and Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by “big data”. SoBigData plans to open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. It plans to not only strengthen the existing clusters of excellence in social data mining research, but also create a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities.

The leader of the AIMH team participating in SoBigData++ is Alejandro Moreo.

Social and hUman ceNtered XR (SUN) is a project funded by the European Commission under the H2020 Programme HORIZON-CL4-2022-HUMAN-01-14, started Dec 1 2022 and ending Nov 30 2025. SUN aims at investigating and developing extended reality (XR) solutions that integrate the physical and the virtual world in a convincing way, from a human and social perspective. The virtual world will be a means to augment the physical world with new opportunities for social and human interaction.

Our institute is the leading partner of the project and the coordinator is Giuseppe Amato.

1.2 NextGenerationEU PNRR National Projects

Then extended partnership titled Future Artificial Intelligence Research (hereafter FAIR) is the response of the Italian AI scientific community to the National Strategic Program. FAIR takes on the challenge to set the agenda of frontier research for the AI methodologies and techniques of tomorrow.

Well beyond currently available technologies, we need AI systems capable of interacting and collaborating with humans, of perceiving and acting within evolving contexts, of being aware of their own limitations and able to adapt to new situations, and interact appropriately in complex social settings, of being aware of their perimeters of security and trust, and of being attentive to the environmental and social impact that their implementation and execution may entail. In short, we need an AI that does not yet exist.

ITSERR is a project designed according to the needs of the Religious Studies scientific community to support the existing national infrastructure and bring it to a higher level of maturity, in terms of involvement of technology and ability to increase the innovation, quality and variety of the knowledge produced by the community of Religious Studies. The research aims at the development of Digital Maktaba (in Arabic "maktaba", "library") whose aim is to establish procedures for the extraction, management of libraries and archives and to develop virtuous models in the field of cataloguing that can accommodate texts written in non-Latin alphabets, starting with the case of the Arabic alphabet and testing it also with other alphabets.

MOST - Centro Nazionale per la mobilità sostenibile through collaboration with 24 universities, CNR and 24 large companies, has the mission of implementing modern, sustainable and inclusive solutions for the entire national territory.

The areas and technological fields of greatest interest in the project are: air mobility, sustainable road vehicles, water transport, rail transport, light vehicles and active mobility. The National Center will take care of making the mobility system more "green" as a whole and more "digital" in its management. It will do so through lightweight solutions and electric and hydrogen propulsion systems; digital systems for the reduction of accidents; more effective solutions for public transport and logistics; a new model of mobility, as a service, accessible and inclusive.

MUCES

In this project ("A MUltimedia platform for Content Enrichment and Search in audiovisual archives"), we aim to develop advanced visual analysis and retrieval methodologies that can make unlabeled Italian audiovisual archives searchable through natural language and exemplar queries in a personalized manner. These methodologies will be multi-modal, adaptable to long-tail concepts, and efficient for large-scale archives. The project aims to bridge cutting-edge research in Computer Vision and Content-based Image Retrieval to enhance the accessibility of audiovisual cultural heritage in Italy. At the core of the project lies a new unifying synergy between cutting-edge research in Computer Vision, Machine Learning, and large-scale Content-Based Retrieval. The project brings together the research experiences and expertise of two internationally-recognized research teams: the AImageLab research group at UNIMORE and the Artificial Intelligence for Media and Humanities laboratory at ISTI CNR, encompassing years of expertise in Multimedia, Similarity Search, and Computer Vision.

PAPYRI

The PRIN PNRR Reconstructing Fragmentary Papyri through Human-Machine Interaction project investigates the application of Artificial Intelligence to the reconstruction of specific lots of papyrus fragments from two Italian papyrological collections: the Papiri collection of the Società Italiana, stored at the Istituto Papirologico “G. Vitelli” (University of Florence), and the Papiri collection of the University of Genova. Following an innovative and interdisciplinary approach, the two papyrological teams will work closely together with ISTI in implementing an already prototyped interactive software aimed to assist papyrologists in the screening phase and, mainly, in the matching of fragments, allowing the user to evaluate and revise multiple hypotheses of reconstruction. The system will take advantage of visual information of both the front and the back of the fragments by exploiting the continuity of the fibre patterns and by taking into account positional information and additional constraints supplied by the expert.

AIMH collaborates to SEcurity and Rights in the CyberSpace (SERICS)

AIMH is involved in particular in Spoke 1 and Spoke 2:

Spoke 1. Human, Social, and Legal Aspects - Spoke 1 aims at protecting social rights and values in the Cyberspace. The research will involve public and private stakeholders in the implementation of innovative technological, legal, ethical and organizational solutions, in order to strengthen the resilience and digital sovereignty of the public and private sectors and, therefore, of the country system. The Spoke 1 has two main projects: Cyberights and DiSe; the former is devoted to study and promote legal and ethical aspects for cyberspace while the latter to digital sovereignty aspects that affect also the underlying digital technologies."

Spoke 2. Misinformation and Fakes The Spoke will establish an excellent multidisciplinary structure that, leveraging intelligence analysis, artificial intelligence, political analysis, data science, and web intelligence capabilities, employs suitable tools and methods to support information disorder awareness. The Spoke 2 has four main projects: DETERRENCE, FF4LL, HUMANE, IDA.

The objective of Spoke 8 is to leverage the existing critical mass of neuroscientists in Tuscany, which has a long and well-established tradition and strength and a vast breadth of highly specialized expertise and tools, integrating it with the recruitment of a diverse set of interdisciplinary knowledge, ranging from chemistry to computational sciences, data sciences, synthetic biology, bioinformatics, high-throughput analysis of gene expression (-omics), imaging and others. Due to its intrinsic multidisciplinary nature, the spoke will naturally have strong links with other spokes in the overall project.

ISTI is active in particular Sub-project 8 - Patient-derived stem cells and “brain-in-a-dish” cultures: a cellular platform for target validation and drug screening. The aim of this subproject is to establish cell cultures of neurons from patients of Retinite Pigmentosa (RP) and Autism Spectrum Diseases (ASD) aimed at their diagnosis by Artificial Intelligence (AI) and at their pharmacological and cell therapy. Moreover, solutions will also be developed to use Cultured Neuronal Networks (CultNN), that is biological cultures of real neuronal cells or brain-in-a-dish, to execute Artificial Intelligence (AI) tasks. CultNN will be directly used for AI, rather than bioinspired devices or software implementations of neural networks. To do so, CultNN of neurons derived from patient reprogrammed cells (hiPSCs) will be established, and com-

putational models of the neural activity-designed AI training methods will be developed to be applied to CultNN. Finally, imaging solutions will be developed, based on AI, to analyze the pupil area and the retina, and CultNN used to screen for drug panels and validate targets.

QuaDaSh

The "Quantification in the Context of Dataset Shift" (QuaDaSh) PRIN PNRR project (P2022TB5JF) is a joint project of the AIMH's AI4Text group at ISTI-CNR (coordinating unit, formed by Fabrizio Sebastiani, Andrea Esuli, and Alejandro Moreo), with the University of Pisa (lead by Nicola Salvati) and the University of Padova (lead by Gian Antonio Susto).

QuaDaSh aims to advance research in quantification. In QuaDaSh, we go beyond the conventional focus on prior probability shift (which deals with variations in class proportions) and consider broader challenges associated with dataset shift (scenarios where the training dataset may also exhibit variations in other aspects). For example, the training set might be in a different language or consist of labelled opinions about a different topic, introducing what is known as covariate shift. By addressing these broader challenges, QuaDaSh aims to devise improved solutions for real applications of quantification that may encounter diverse types of dataset shift. Examples of these include seabed cover mapping, small area estimation, or fairness auditing, to name a few. Alejandro Moreo is the PI of this project.

1.3 Other National Projects

In the industrial realm, it's well known that many accidents involving machine operators in production processes are somehow linked to the operator's behavior and various types of errors. No matter how sophisticated, traditional control and supervision systems have been unable to entirely eliminate or satisfactorily manage these risks. This issue becomes particularly acute when machines malfunction, such as product jams that need to be cleared or other breakdowns, requiring operators to access dangerous areas of the machine and actively intervene. Predicting these operational conditions linked to faults, malfunctions, and operator errors is challenging during the design phase. The required activities may not be easily definable and hence can lead to accidents or near-miss incidents. In this context, equipping machines with networks of sensors and Artificial Intelligence (AI) systems capable of interpreting even new operational situations, recognizing potential risk conditions for operators, and generating appropriate commands for the machines is seen as an effective path towards enhancing user safety. AISAFETY steps into this challenging scenario with a groundbreaking approach. By integrating AI, RFID technology, and a network of intelligent cameras, it offers a proactive solution to enhance workplace safety. The system's AI brain analyzes data from the cameras and RFID tags worn by operators, understanding the workspace dynamics in real-time. If it detects any danger, it can instantly instruct the machines to stop or adjust their operation, often before any human can react. Moreover, AISAFETY respects the indispensable role of human judgment. Supervisors monitor the system, ensuring that the balance between automated safety measures and human oversight is maintained. Compliant with strict safety and data protection regulations, AISAFETY is not just about employing advanced technology; it's about responsibly creating a safer industrial environment where technology and human expertise collaborate to prevent accidents. Claudio Gennaro is the scientific responsible for ISTI.

CY4Gate/SWOAD

The project involves the analysis, study, and implementation of a system for searching and recognizing images depicting works of art. The project includes the development of advanced solutions for the analysis and extraction of information from images, image search and recognition, and the indexing of information extracted from images. Specifically, the image analysis component extracts information (visual features) that can be used for image search and recognition based on visual content, using state-of-the-art artificial intelligence and computer vision techniques, including cutting-edge deep neural networks. The image search and recognition component receives images used as queries, sends them to the image analysis component, compares them with those in the dataset, and returns images whose visual content is most similar to the query. Finally, the image indexing component operates incrementally, coordinating with the image analysis component to extract information from images to be included in the system. It creates a database of visual features for comparison and recognition, allowing for fast and efficient search. The project is carried out within a collaboration between CNR-ISTI and the Comando Carabinieri per la Tutela del Patrimonio Culturale. It is funded through a sub-contract given to CNR-ISTI by the company CY4Gate, as part of the SWOAD (Stolen Work Of Art Detection) project.

HDN

Hypermedia Dante Network (HDN) is a three year (2020-2024) Italian National Research Project (PRIN) which aims to extend the ontology and tools developed by AIMH team to represent the sources of Dante Alighieri's minor works to the more complex world of the Divine Comedy. In particular, HDN aims to enrich the functionalities of the DanteSources Web application (https://dantesources.dantenetwork.it/) in order to efficiently recover knowledge about the Divine Comedy. Relying on some of the most important scientific institutions for Dante studies, such as the Italian Dante Society of Florence, HDN makes use of specialized skills, essential for the population of ontology and the consequent creation of a complete and reliable knowledge base. Knowledge will be published on the Web as Linked Open Data and will be access

through a user-friendly Web application.

IMAGO

The IMAGO (Index Medii Aevi Geographiae Operum) is a three year (2020-2024) Italian National Research Project (PRIN) that aims at creating a knowledge base of the critical editions of Medieval and Humanistic Latin geographical works (VI-XV centuries). Up to now, this knowledge has been collected in many paper books or several databases, making it difficult for scholars to retrieve it easily and to produce a complete overview of these data. The goal of the project is to develop new tools that satisfy the needs of the academic research community, especially for scholars interested in Medieval and Renaissance Humanism geography. Using Semantic Web technologies, AIMH team will develop an ontology providing the terms to represent this knowledge in a machine-readable form. A semi-automatic tool will help the scholars to populate the ontology with the data included in authoritative critical editions. Afterwards, the tool will automatically save the resulting graph into a triple store. On top of this graph, a Web application will be developed, which will allow users to extract and display the information stored in the knowledge base in the form of maps, charts, and tables.

The leader of the AIMH team participating in IMAGO is Valentina Bartalesi.

INAROS

INtelligenza ARtificiale per il mOnitoraggio e Supporto agli anziani (INAROS) is a 2-year project funded by Regione Toscana, Istituto di Scienza e Tecnologie dell'Informazione "A.Faedo" (ISTI) del CNR, Visual Engines srl. The main goal of the INAROS project is to build solutions for monitoring and surveillance of the elderly based on the use of autonomous smart cameras. Computer vision algorithms will be developed by leveraging artificial intelligence, in particular deep learning to automatically and in real time analyze video streams from smart cameras positioned in the home environment. To achieve these results, techniques will be developed for the tracking and detection of the elderly person's activity in the home environment and for the discovery of new activities and abnormalities of the elderly through off-line analysis of temporal patterns of learned events. Claudio Gennaro is the scientific coordinator of the project.

MIGHT

Gut Microbiota as a bioremediator for gut-health in infants (MIGHT) is a 2 years project funded by the 'Progetti@CNR' (Area: Tecnologie a supporto delle fasce più fragili: giovani e anziani) program. The understanding of the relationship among diet, metabolites, and host/microbiota is a key challenge to investigate personalized nutrition for the most fragile segments of the population, and the modulation of the gut microbiota through dietary interventions is one of the most promising approaches. MIGHT project has the ambition to disentangle key research questions behind food proteins modifications and the effects on the host microbiota. MIGHT is interlinked with a recently granted project from the EU (MAMMAL, EIT Food, Innovation). While the MAMMAL proposal focuses on biochemical aspects, MIGHT enlightens IT solutions for the organization, management, and access to the data produced, and for their exploration to translate the experimental evidence from newborns to the general population, in particular to elderly. Partners of the project are: Istituto di Biologia e Biotecnologia Agraria (IBBA) - CNR, Istituto Sistema di Produzione Animale in Ambiente Mediterraneo (ISPAAM) - CNR, Istituto di Scienza e Tecnologie dell'Informazione (ISTI), CNR. Cesare Concordia is the scientific responsible for ISTI.

WEMB

The "Word EMBeddings: From Cognitive Linguistics to Language Engineering, and Back" (WEMB) PRIN project (2022EPTPJ9) is a joint collaboration of AIMH's AI4Text group (Fabrizio Sebastiani, Andrea Esuli, and Alejandro Moreo) with the University of Bologna (lead by Marianna Bolognesi).

The goal of WEMB is twofold: (1) reaching a better understanding of how word embeddings relate to language processing in the human mind, and (2) using this understanding in order to contribute to the development of a new generation of word embeddings that can be applied to the NLP / text mining tasks having to do with the semantic analysis of text. Examples of these tasks may include text classification, word sense disambiguation, machine translation, text summarization, question answering, and sentiment analysis. Fabrizio Sebastiani is the PI of this project.

The Cultural Heritage Cloud (ECCCH) is a shared platform designed to provide heritage professionals and researchers with access to data, scientific resources, training, and advanced digital tools tailored to suit their needs. This platform is developed by ECHOES (European Cloud for Heritage OpEn Science), a project funded by the European Commission and UK Research and Innovation (UKRI) that brings together fragmented communities of the Cultural Heritage field into a new community around the Digital Commons.

ECHOES will create a digital environment that enables the digitisation of existing knowledge and the collaborative analysis of cultural heritage assets, facts, and phenomena. In this context, actors – whether humans or Artificial Intelligence – can develop their interpretations, thereby enriching the knowledge of cultural heritage and their surroundings. The digital environment proposed by ECHOES will empower users to interact with, manipulate and enrich Digital Twins, fostering the creation of new, collaboratively developed scientific knowledge.

The leader of the AIMH team participating in ECHOES is Valentina Bartalesi.

2. Publications

In this section, we report the complete list of papers we published in 2024 organized in four categories: journals, proceedings, magazines, others, and pre-prints.

2.1 Journals

In this section, we report the paper we published (or accepted for publication) in journals during 2024, in alphabetic order of the first author. Our works were published in the following journals (ordered by Impact Factor):

- **IEEE Transactions on Information Forensics and Security (TIFS)**
  IEEE, IF 6.3: [12]
- **Scientific Data**
  Nature Publishing Group, IF 5.8: [3]
- **Neurocomputing**
  Elsevier, IF 5.5: [19]
- **ACM Transactions on Information Systems**
  Elsevier, IF 5.4: [14]
- **Neural Computing and Applications**
  Springer, IF 4.5: [28]
- **PeerJ Computer Science**
  PeerJ Inc., IF 3.8: [4, 11]
- **IEEE Access**
  IEEE, IF 3.4: [13]
- **Int. Journal of Machine Learning and Cybernetics**
  Springer, IF 3.1: [16]
- **Multimedia Tools and Applications (MTAP)**
  Springer, IF 3.0: [21]
- **Data Mining and Knowledge Discovery**
  Springer, IF 2.8: [9, 17, 22]
- **Applied Sciences**
  Multidisciplinary Digital Publishing Institute, IF 2.5: [37]
- **Multimodal Technologies and Interaction**
  Multidisciplinary Digital Publishing Institute, IF 2.4: [36]
- **Journal on Computing and Cultural Heritage (JOCCH)**
  Association for Computing Machinery (ACM), IF 2.4: [5, 34]
- **SN Computer Science**
  Springer [10]

2.1.1 A Noise-Oriented and Redundancy-Aware Instance Selection Framework

W. Cunha, A. Moreo, A. Esuli, F. Sebastiani, L. Rocha, M. André Gonçalves.
ACM Transactions on Information Systems [14]

Fine-tuning transformer-based deep-learning models are currently at the forefront of natural language processing (NLP) and information retrieval (IR) tasks. However, fine-tuning these transformers for specific tasks, especially when dealing with ever-expanding volumes of data, constant retraining requirements, and budget constraints, can be computationally and financially costly, requiring substantial energy consumption and contributing to carbon dioxide emissions. This article focuses on advancing the state-of-the-art (SOTA) on instance selection (IS)—a range of document filtering techniques designed to select the most representative documents for the sake of training. The objective is to either maintain or enhance classification effectiveness while reducing the overall training (fine-tuning) total processing time. In our prior research, we introduced the E2SC framework, a redundancy-oriented IS method focused on transformers and large datasets—currently the state-of-the-art in IS. Nonetheless, important research questions remained unanswered in our previous work, mostly due to E2SC's sole emphasis on redundancy. In this article, we take our research a step further by proposing biO-IS–an extended bi-objective instance selection solution, a novel IS framework aimed at simultaneously removing redundant and noisy instances from the training. biO-IS estimates redundancy based on scalable, fast, and calibrated weak classifiers and captures noise with the support of a new entropy-based step. We also propose a novel iterative process to estimate near-optimum reduction rates for both steps. Our extended solution is able to reduce the training sets by 41% on average (up to 60%) while maintaining the effectiveness in all tested datasets, with speedup gains of 1.67 on average (up to 2.46x). No other baseline, not even our previous SOTA solution, was capable of achieving results with this level of quality, considering the tradeoff among training reduction, effectiveness, and speedup. To ensure reproducibility, our documentation, code, and datasets can be accessed on GitHub–`https://github.com/waashk/bio-is`.

2.1.2 A semantic knowledge graph of European mountain value chains

V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, P. Pagano, M. Moretti, G. Brunori. Scientific Data, Nature Publishing Group [3]

The United Nations forecast a significant shift in global population distribution by 2050, with rural populations projected to decline. This decline will particularly challenge mountain areas' cultural heritage, well-being, and economic sustainability. Understanding the economic, environmental, and societal effects of rural population decline is particularly important in Europe, where mountainous regions are vital for supplying goods. The present paper describes a geospatially explicit semantic knowledge graph containing information on 454 European mountain value chains. It is the first large-size, structured collection of information on mountain value chains. Our graph, structured through ontology-based semantic modelling, offers representations of the value chains in the form of narratives. The graph was constructed semi-automatically from unstructured data provided by mountain-area expert scholars. It is accessible through a public repository and explorable through interactive Story Maps and a semantic Web service. Through semantic queries, we demonstrate that the graph allows for exploring territorial complexities and discovering new knowledge on mountain areas' environmental, societal, territory, and economic aspects that could help stem depopulation.

2.1.3
Binary quantification and dataset shift: an experimental investigation

P. González, A. Moreo, F. Sebastiani. Data Mining and Knowledge Discovery [17]

Quantification is the supervised learning task that consists of training predictors of the class prevalence values of sets of unlabelled data, and is of special interest when the labelled data on which the predictor has been trained and the unlabelled data are not IID, i.e., suffer from dataset shift. To date, quantification methods have mostly been tested only on a special case of dataset shift, i.e., prior probability shift; the relationship between quantification and other types of dataset shift remains, by and large, unexplored. In this work we carry out an experimental analysis of how current quantification algorithms behave under different types of dataset shift, in order to identify limitations of current approaches and hopefully pave the way for the development of more broadly applicable methods. We do this by proposing a fine-grained taxonomy of types of dataset shift, by establishing protocols for the generation of datasets affected by these types of shift, and by testing existing quantification methods on the datasets thus generated. One finding that results from this investigation is that many existing quantification methods that had been found robust to prior probability shift are not necessarily robust to other types of dataset shift. A second finding is that no existing quantification method seems to be robust enough to dealing with all the types of dataset shift we simulate in our experiments. The code needed to reproduce all our experiments is publicly available at `https://github.com/pglez82/quant_datasetshift`.

2.1.4
Cascaded transformer-based networks for wikipedia large-scale image-caption matching

N. Messina, D.A. Coccomini, A. Esuli, F. Falchi Multimedia Tools and Applications, Publisher] [21]

With the increasing importance of multimedia and multilingual data in online encyclopedias, novel methods are needed to fill domain gaps and automatically connect different modalities for increased accessibility. For example, Wikipedia is composed of millions of pages written in multiple languages. Images, when present, often lack textual context, thus remaining conceptually floating and harder to find and manage. In this work, we tackle the novel task of associating images from Wikipedia pages with the correct caption among a large pool of available ones written in multiple languages, as required by the image-caption matching Kaggle challenge organized by the Wikimedia Foundation. A system able to perform this task would improve the accessibility and completeness of the underlying multi-modal knowledge graph in online encyclopedias. We propose a cascade of two models powered by the recent Transformer networks able to efficiently and effectively infer a relevance score between the query image data and the captions. We verify through extensive experiments that the proposed cascaded approach effectively handles a large pool of images and captions while maintaining bounded the overall computational complexity at inference time. With respect to other approaches in the challenge leaderboard, we can achieve remarkable improvements over the previous proposals (+8% in nDCG@5 with respect to the sixth position) with constrained resources. The code is publicly available at `https://tinyurl.com/wiki-imcap`.

2.1.5
Detecting images generated by diffusers

D.A. Coccomini, A. Esuli1, F. Falchi1, C. Gennaro, G. Amato PeerJ Computer Science, PeerJ Inc. [11]

In recent years, the field of artificial intelligence has witnessed a remarkable surge in the generation of synthetic images, driven by advancements in deep learning techniques. These synthetic images, often created through complex algorithms, closely mimic real photographs, blurring the lines between reality and artificiality. This proliferation of synthetic visuals presents a pressing challenge: how to accurately and reliably distinguish between genuine and generated images. This article, in particular, explores the task of detecting images generated by text-to-image diffusion models, highlighting the challenges and peculiarities of this field. To evaluate this, we consider images generated from captions in the MSCOCO and Wikimedia datasets using two state-of-the-art models: Stable Diffusion and GLIDE. Our experiments show that it is possible to detect the generated images using simple multi-layer perceptrons (MLPs), starting from features extracted by CLIP or RoBERTa, or using traditional convolutional neural networks (CNNs). These latter models achieve remarkable performances in particular when pretrained on large datasets. We also observe that models trained on images generated by Stable Diffusion can occasionally detect images generated by GLIDE, but only on the MSCOCO dataset. However, the reverse is not true. Lastly, we find that incorporating the associated textual information with the images in some cases can lead to a better generalization capability, especially if textual features are closely related to visual ones. We also discovered that the type of subject depicted in the image can significantly impact performance. This work provides insights into the feasibility of detecting generated images and has implications for security and privacy concerns in real-world applications. The code to reproduce our results is available at: `https://github.com/davide-coccomini/Detecting-Images-Generated-by-Diffusers`.

2.1.6
Explainable Authorship Identification in Cultural Heritage Applications

M. Setzu, S. Corbara, A. Monreale, A. Moreo, F. Sebastiani. ACM Journal on Computing and Cultural Heritage [34]

While a substantial amount of work has recently been devoted to improving the accuracy of computational Authorship Identification (AId) systems for textual data, little to no attention has been paid to endowing AId systems with the ability to explain the reasons behind their predictions. This substantially hinders the practical application of AId methods, since the predictions returned by such systems are hardly useful unless they are supported by suitable explanations. In this article, we explore the applicability of existing general-purpose eXplainable Artificial Intelligence (XAI) techniques to AId, with a focus on explanations addressed to scholars working in cultural heritage. In particular, we assess the relative merits of three different types of XAI techniques (feature ranking, probing, factual and counterfactual selection) on three different AId tasks (authorship attribution, authorship verification and same-authorship verification)

by running experiments on real AId textual data. Our analysis shows that, while these techniques make important first steps towards XAI, more work remains to be done to provide tools that can be profitably integrated into the workflows of scholars.

2.1.7
Forging the Forger: An Attempt to Improve Authorship Verification via Data Augmentation

S. Corbara, A. Moreo. IEEE Access [13]

Authorship Verification (AV) is a text classification task concerned with inferring whether a candidate text has been written by one specific author (A) or by someone else ($\overline{A}$). It has been shown that many AV systems are vulnerable to adversarial attacks, where a malicious author actively tries to fool the classifier by either concealing their writing style, or by imitating the style of another author. In this paper, we investigate the potential benefits of augmenting the classifier training set with (negative) synthetic examples. These synthetic examples are generated to imitate the style of A. We analyze the improvements in the classifier predictions that this augmentation brings to bear in the task of AV in an adversarial setting. In particular, we experiment with three different generator architectures (one based on Recurrent Neural Networks, another based on small-scale transformers, and another based on the popular GPT model) and with two training strategies (one inspired by standard Language Models, and another inspired by Wasserstein Generative Adversarial Networks). We evaluate our hypothesis on five datasets (three of which have been specifically collected to represent an adversarial setting) and using two learning algorithms for the AV classifier (Support Vector Machines and Convolutional Neural Networks). This experimentation yields negative results, revealing that, although our methodology proves effective in many adversarial settings, its benefits are too sporadic for a pragmatical application.

2.1.8
In the Wild Video Violence Detection: An Unsupervised Domain Adaptation Approach

L. Ciampi, C. Santiago, F. Falchi, C. Gennaro, G. Amato SN Computer Science, Springer [10]

This work addresses the challenge of video violence detection in data-scarce scenarios, focusing on bridging the domain gap that often hinders the performance of deep learning models when applied to unseen domains. We present a novel unsupervised domain adaptation (UDA) scheme designed to effectively mitigate this gap by combining supervised learning in the train (source) domain with unlabeled test (target) data. We employ single-image classification and multiple instance learning (MIL) to select frames with the highest classification scores, and, upon this, we exploit UDA techniques to adapt the model to unlabeled target domains. We perform an extensive experimental evaluation, using general-context data as the source domain and target domain datasets collected in specific environments, such as violent/non-violent actions in hockey matches and public transport. The results demonstrate that our UDA pipeline substantially enhances model performances, improving their generalization capabilities in novel scenarios without requiring additional labeled data.

2.1.9
MINTIME: Multi-Identity Size-Invariant Video Deepfake Detection

D.A. Coccomini, G.K. Zilos; G. Amato; R. Caldelli, F. Falchi, S. Papadopoulos IEEE Transactions on Information Forensics and Security (TIFS), IEEE [12]

In this paper, we present MINTIME, a video deepfake detection method that effectively captures spatial and temporal inconsistencies in videos that depict multiple individuals and varying face sizes. Unlike previous approaches that either employ simplistic a-posteriori aggregation schemes, i.e., averaging or max operations, or only focus on the largest face in the video, our proposed method learns to accurately detect spatio-temporal inconsistencies across multiple identities in a video through a Spatio-Temporal Transformer combined with a Convolutional Neural Network backbone. This is achieved through an Identity-aware Attention mechanism that applies a masking operation on the face sequence to process each identity independently, which enables effective video-level aggregation. Furthermore, our system incorporates two novel embedding schemes: (i) the Temporal Coherent Positional Embedding, which encodes the temporal information of the face sequences of each identity, and (ii) the Size Embedding, which captures the relative sizes of the faces to the video frames. MINTIME achieves state-of-the-art performance on the ForgeryNet dataset, with a remarkable improvement of up to 14% AUC in videos containing multiple people. Moreover, it demonstrates very robust generalization capabilities in cross-forgery and cross-dataset settings. The code is publicly available at: `https://github.com/davide-coccomini/MINTIME-Multi-Identity-size-iNvariant-TIMEsformer-for-Video-Deepfake-Detection`.

2.1.10
Modelling and simulation of traditional craft actions

X. Zabulis, N. Partarakis, I. Demeridou, V. Bartalesi, N. Pratelli, C. Meghini, N. Nikolaou, P. Fallahian. Applied Sciences, Multidisciplinary Digital Publishing Institute [37]

The problem of modelling and simulating traditional crafting actions is addressed, motivated by the goals of craft understanding, documentation, and training. First, the physical entities involved in crafting actions are identified, physically, and semantically characterised, including causing entities, conditions, properties, and objects, as well as the space and time in which they occur. Actions are semantically classified into a taxonomy of four classes according to their goals, which are shown to exhibit similarities in their operation principles and utilised tools. This classification is employed to simplify the create archetypal simulators, based on the Finite Element Method, by developing archetypal simulators for each class and specialising them in craft-specific actions. The approach is validated by specialising the proposed archetypes into indicative craft actions and predicting their results in simulation. The simulated actions are rendered in 3D to create visual demonstrations and can be integrated into game engines for training applications.

2.1.11
Multimodal dictionaries for traditional craft education

X. Zabulis, N. Partarakis, V. Bartalesi, N. Pratelli, C. Meghini, A. Dubois, I. Moreno, S. Manitsaris. Multimodal Tech-

nologies and Interaction, Multidisciplinary Digital Publishing Institute [36]

We address the problem of systematizing the authoring of digital dictionaries for craft education from ethnographic studies and recordings. First, we present guidelines for the collection of ethnographic data using digital audio and video and identify terms that are central in the description of crafting actions, products, tools, and materials. Second, we present a classification scheme for craft terms and a way to semantically annotate them, using a multilingual and hierarchical thesaurus, which provides term definitions and a semantic hierarchy of these terms. Third, we link ethnographic resources and open-access data to the identified terms using an online platform for the representation of traditional crafts, associating their definition with illustrations, examples of use, and 3D models. We validate the efficacy of the approach by creating multimedia vocabularies for an online eLearning platform for introductory courses to nine traditional crafts.

2.1.12
Quantification using permutation-invariant networks based on histograms

O. Pérez-Mon, A. Moreo, J.J. del Coz, P. González. Neural Computing and Applications [28]

Quantification, also known as class prevalence estimation, is the supervised learning task in which a model is trained to predict the prevalence of each class in a given bag of examples. This paper investigates the application of deep neural networks for tasks of quantification in scenarios where it is possible to apply a symmetric supervised approach that eliminates the need for classification as an intermediate step, thus directly addressing the quantification problem. Additionally, it discusses existing permutation-invariant layers designed for set processing and assesses their suitability for quantification. Based on our analysis, we propose HistNetQ, a novel neural architecture that relies on a permutation-invariant representation based on histograms that is especially suited for quantification problems. Our experiments carried out in two standard competitions, which have become a reference in the quantification field, show that HistNetQ outperforms other deep neural network architectures designed for set processing, as well as the current state-of-the-art quantification methods. Furthermore, HistNetQ offers two significant advantages over traditional quantification methods: i) it does not require the labels of the training examples but only the prevalence values of a collection of training bags, making it applicable to new scenarios; and ii) it is able to optimize any custom quantification-oriented loss function.

2.1.13
Regularization-based methods for ordinal quantification

M. Bunse, A. Moreo, F. Sebastiani, M. Senz. Data Mining and Knowledge Discovery [9]

Quantification, i.e., the task of predicting the class prevalence values in bags of unlabeled data items, has received increased attention in recent years. However, most quantification research has concentrated on developing algorithms for binary and multi-class problems in which the classes are not ordered. Here, we study the ordinal case, i.e., the case in which a total order is defined on the set of $n > 2$ classes. We give three main contributions to this field. First, we create and make available two datasets for ordinal quantification (OQ) research that overcome the inadequacies of the previously available ones. Second, we experimentally compare the most important OQ algorithms proposed in the literature so far. To this end, we bring together algorithms proposed by authors from very different research fields, such as data mining and astrophysics, who were unaware of each others' developments. Third, we propose a novel class of regularized OQ algorithms, which outperforms existing algorithms in our experiments. The key to this gain in performance is that our regularization prevents ordinally implausible estimates, assuming that ordinal distributions tend to be smooth in practice. We informally verify this assumption for several real-world applications.

2.1.14
Representing geospatial knowledge in narratives

V. Bartalesi, N. Pratelli. Journal on Computing and Cultural Heritage (JOCCH), Association for Computing Machinery (ACM) [5]

This paper explores the representation of geospatial knowledge within narratives through a Semantic Web approach. We introduce the NOnt+Space (NOnt+S) model, an extension of the CIDOC CRM-based Narrative Ontology, which allows the representation of narratives and their geospatial aspects. By leveraging standards such as CRMgeo and GeoSPARQL, NOnt+S ensures systematic and interoperable geospatial representation in narratives, enabling geospatial queries on knowledge graphs. We present an assessment of NOnt+S utilising data from the H2020 MOVING European project (2021-2024), which collected knowledge about European mountain value chains intended as Cultural Heritage. We have represented this knowledge as geospatial narratives using NOnt+S. GeoSPARQL queries and semantic reasoning applied to the created KG reveal the ontology ability to infer new geospatial knowledge. Our work contributes to the ongoing efforts in the Semantic Web community to integrate and represent geospatial information within narratives, promoting collaboration and interoperability across various scientific domains.

2.1.15
SAL_τ: efficiently stopping TAR by improving priors estimates

A. Molinari, A. Esuli. Data Mining and Knowledge Discovery [22]

In high recall retrieval tasks, human experts review a large pool of documents with the goal of satisfying an information need. Documents are prioritized for review through an active learning policy, and the process is usually referred to as Technology-Assisted Review (TAR). TAR tasks also aim to stop the review process once the target recall is achieved to minimize the annotation cost. In this paper, we introduce a new stopping rule called SAL^R_τ (SLD for Active Learning), a modified version of the Saerens–Latinne–Decaestecker algorithm (SLD) that has been adapted for use in active learning. Experiments show that our algorithm stops the review well ahead of the current state-of-the-art methods, while providing the same guarantees of achieving the target recall. Code is available at `https://github.com/levnikmyskin/salt`

2.1.16

Scalable bio-inspired training of Deep Neural Networks with FastHebb

G. Lagani, F. Falchi, C. Gennaro, H. Fassold, G. Amato Neurocomputing, Elsevier [19]

Recent work on sample efficient training of Deep Neural Networks (DNNs) proposed a semi-supervised methodology based on biologically inspired Hebbian learning, combined with traditional backprop-based training. Promising results were achieved on various computer vision benchmarks, in scenarios of scarce labeled data availability. However, current Hebbian learning solutions can hardly address large-scale scenarios due to their demanding computational cost. In order to tackle this limitation, in this contribution, we investigate a novel solution, named FastHebb (FH), based on the reformulation of Hebbian learning rules in terms of matrix multiplications, which can be executed more efficiently on GPU. Starting from Soft-Winner-Takes-All (SWTA) and Hebbian Principal Component Analysis (HPCA) learning rules, we formulate their improved FH versions: SWTA-FH and HPCA-FH. We experimentally show that the proposed approach accelerates training speed up to 70 times, allowing us to gracefully scale Hebbian learning experiments on large datasets and network architectures such as ImageNet and VGG.

2.1.17

Using AI to decode the behavioral responses of an insect to chemical stimuli: towards machine-animal computational technologies

E. Fazzari, F. Falchi, C. Stefanini, D. Romano Int. Journal of Machine Learning and Cybernetics, Springer [16]

Orthoptera are insects with excellent olfactory sense abilities due to their antennae richly equipped with receptors. This makes them interesting model organisms to be used as biosensors for environmental and agricultural monitoring. Herein, we investigated if the house cricket Acheta domesticus can be used to detect different chemical cues by examining the movements of their antennae and attempting to identify specific antennal displays associated to different chemical cues exposed (e.g., sucrose or ammonia powder). A neural network based on state-of-the-art techniques (i.e., SLEAP) for pose estimation was built to identify the proximal and distal ends of the antennae. The network was optimised via grid search, resulting in a mean Average Precision (mAP) of 83.74%. To classify the stimulus type, another network was employed to take in a series of keypoint sequences, and output the stimulus classification. To find the best one-dimensional convolutional and recurrent neural networks, a genetic algorithm-based optimisation method was used. These networks were validated with iterated K-fold validation, obtaining an average accuracy of 45.33% for the former and 44% for the latter. Notably, we published and introduced the first dataset on cricket recordings that relate this animal's behaviour to chemical stimuli. Overall, this study proposes a novel and simple automated method that can be extended to other animals for the creation of Biohybrid Intelligent Sensing Systems (e.g., automated video-analysis of an organism's behaviour) to be exploited in various ecological scenarios.

2.1.18

Using large language models to create narrative events

V. Bartalesi, E. Lenzi, C. De Martino. PeerJ Computer Science, PeerJ Inc. [4]

Narratives play a crucial role in human communication, serving as a means to convey experiences, perspectives, and meanings across various domains. They are particularly significant in scientific communities, where narratives are often utilized to explain complex phenomena and share knowledge. This article explores the possibility of integrating large language models (LLMs) into a workflow that, exploiting the Semantic Web technologies, transforms raw textual data gathered by scientific communities into narratives. In particular, we focus on using LLMs to automatically create narrative events, maintaining the reliability of the generated texts. The study provides a conceptual definition of narrative events and evaluates the performance of different smaller LLMs compared to the requirements we identified. A key aspect of the experiment is the emphasis on maintaining the integrity of the original narratives in the LLM outputs, as experts often review texts produced by scientific communities to ensure their accuracy and reliability. We first perform an evaluation on a corpus of five narratives and then on a larger dataset comprising 124 narratives. LLaMA 2 is identified as the most suitable model for generating narrative events that closely align with the input texts, demonstrating its ability to generate high-quality narrative events. Prompt engineering techniques are then employed to enhance the performance of the selected model, leading to further improvements in the quality of the generated texts.

2.2 Proceedings

In this section, we report the paper we published in alphabetic order of the first author. Our works were presented, and published in the proceedings of the following conferences:

- **ACL** – 62nd Annual Meeting of the Association for Computational Linguistics [33]
- **CBMI** – 21st International Conference on Content-based Multimedia Indexing [6]
- **CLiC-it** – 10th Italian Conference on Computational Linguistics [15, 32, 31]
- **CVPR** – IEEE/CVF Conference on Computer Vision and Pattern Recognition [7]
- **HCII** – 26th International Conference on Human-Computer Interaction [20]
- **MMM** – 30th International Conference on MultiMedia Modeling [1]
- **TPDL** – 28th International Conference on Theory and Practice of Digital Libraries [30]
- **X-TAIL** – eXtraction and eXploitation of long-TAIL Knowledge with LLMs and KGs: 1st Workshop co-located with EKAW-24 [8]

2.2.1

ABRICOT - ABstRactness and Inclusiveness in COntexT: A CALAMITA Challenge

G. Puccetti, C. Collacciani, A.A. Ravelli, A. Esuli, M. Bolognesi

CLiC-it, 10th Italian Conference on Computational Linguistics [32]

The ABRICOT Task is designed to evaluate Italian language models on their ability to understand and assess the abstractness and inclusiveness of language, two nuanced features that humans naturally convey in everyday communication. Unlike binary categorizations such as abstract/concrete or inclusive/exclusive, these features exist on a continuous spectrum with varying degrees of intensity. The task is based on a manual collection of sentences that present the same noun phrase (NP) in different contexts, allowing its interpretation to vary between the extremes of abstractness and inclusiveness. This challenge aims to verify the how LLMs perceive subtle linguistic variations and their implications in natural language.

2.2.2
AI 'News' Content Farms Are Easy to Make and Hard to Detect: A Case Study in Italian

G. Puccetti, A. Rogers, C. Alzetta, F. Dell'Orletta, A. Esuli ACL, 62nd Annual Meeting of the Association for Computational Linguistics [32]

Large Language Models (LLMs) are increasingly used as 'content farm' models (CFMs), to generate synthetic text that could pass for real news articles. This is already happening even for languages that do not have high-quality monolingual LLMs. We show that fine-tuning Llama (v1), mostly trained on English, on as little as 40K Italian news articles, is sufficient for producing news-like texts that native speakers of Italian struggle to identify as synthetic.We investigate three LLMs and three methods of detecting synthetic texts (log-likelihood, DetectGPT, and supervised classification), finding that they all perform better than human raters, but they are all impractical in the real world (requiring either access to token likelihood information or a large dataset of CFM texts). We also explore the possibility of creating a proxy CFM: an LLM fine-tuned on a similar dataset to one used by the real 'content farm'. We find that even a small amount of fine-tuning data suffices for creating a successful detector, but we need to know which base LLM is used, which is a major challenge.Our results suggest that there are currently no practical methods for detecting synthetic news-like texts 'in the wild', while generating them is too easy. We highlight the urgency of more NLP research on this problem.

2.2.3
BlocklyBias: A Visual Programming Language for Bias Identification in AI Data

C. De Martino, T. Turchi, A. Malizia. [20]

In the current landscape of Artificial Intelligence (AI), bias has emerged as a central concern in both public discourse and scientific inquiry. In today's rapidly evolving landscape, marked by increasing complexity and challenges, there is a growing need to address the issue of biases and discrimination that can be exacerbated by algorithms. Biases can infiltrate data collection, whether conducted by humans or systems they design, highlighting the multifaceted nature of this challenge. Consequently, addressing this issue from diverse perspectives is imperative, extending its reach beyond technical domains to include stakeholders from various backgrounds. This paper aims to illustrate how the democratization of the data analysis process – specifically regarding intersectional biases – can be achieved through the use of Visual Programming Languages (VPLs). By reducing the technical entry barrier, fostering an understanding of bias, and providing mitigation strategies, this research introduces BlocklyBias, a platform founded on VPL principles. BlocklyBias serves as a foundational stepping stone for future improvements, as a tool to explore and resolve bias-related challenges in data analysis. Through this study, we seek to bridge the gap between technical and non-technical stakeholders, fostering a collaborative approach to bias mitigation in AI.

2.2.4
Evaluation of LLMs on Long-tail Entity Linking in Historical Documents

M. Boscariol, L. Bulla, L. Draetta, B. Fiumanò, E. Lenzi, L. Piano. [8]

Entity Linking (EL) plays a crucial role in Natural Language Processing (NLP) applications, enabling the disambiguation of entity mentions by linking them to their corresponding entries in a reference knowledge base (KB). Thanks to their deep contextual understanding capabilities, LLMs offer a new perspective to tackle EL, promising better results than traditional methods. Despite the impressive generalization capabilities of LLMs, linking less popular, long-tail entities remains challenging as these entities are often underrepresented in training data and knowledge bases. Furthermore, the long-tail EL task is an understudied problem, and limited studies address it with LLMs. In the present work, we assess the performance of two popular LLMs, GPT and LLama3, in a long-tail entity linking scenario. Using MHERCL v0.1, a manually annotated benchmark of sentences from domain-specific historical texts, we quantitatively compare the performance of LLMs in identifying and linking entities to their corresponding Wikidata entries against that of ReLiK, a state-of-the-art Entity Linking and Relation Extraction framework. Our preliminary experiments reveal that LLMs perform encouragingly well in long-tail EL, indicating that this technology can be a valuable adjunct in filling the gap between head and long-tail EL.

2.2.5
Is CLIP the main roadblock for fine-grained open-world perception?

L. Bianchi, F. Carrara, N. Messina, F. Falchi [6]

Modern applications increasingly demand flexible computer vision models that adapt to novel concepts not encountered during training. This necessity is pivotal in emerging domains like extended reality, robotics, and autonomous driving, which require the ability to respond to open-world stimuli. A key ingredient is the ability to identify objects based on free-form textual queries defined at inference time - a task known as open-vocabulary object detection. Multimodal backbones like CLIP are the main enabling technology for current open-world perception solutions. Despite performing well on generic queries, recent studies highlighted limitations on the fine-grained recognition capabilities in open-vocabulary settings - i.e., for distinguishing subtle object features like color, shape, and material. In this paper, we perform a detailed examination of these open-vocabulary object recognition limitations to find the root cause. We evaluate the performance of CLIP, the most commonly used

vision-language backbone, against a fine-grained object-matching benchmark, revealing interesting analogies between the limitations of open-vocabulary object detectors and their backbones. Experiments suggest that the lack of fine-grained understanding is caused by the poor separability of object characteristics in the CLIP latent space. Therefore, we try to understand whether fine-grained knowledge is present in CLIP embeddings but not exploited at inference time due, for example, to the unsuitability of the cosine similarity matching function, which may discard important object characteristics. Our preliminary experiments show that simple CLIP latent-space re-projections help separate fine-grained concepts, paving the way towards the development of backbones inherently able to process fine-grained details.

2.2.6
The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding

L. Bianchi, F. Carrara, N. Messina, C. Gennaro, F. Falchi [7]

Recent advancements in large vision-language models enabled visual object detection in open-vocabulary scenarios, where object classes are defined in free-text formats during inference. In this paper, we aim to probe the state-of-the-art methods for open-vocabulary object detection to determine to what extent they understand fine-grained properties of objects and their parts. To this end, we introduce an evaluation protocol based on dynamic vocabulary generation to test whether models detect, discern, and assign the correct fine-grained description to objects in the presence of hard-negative classes. We contribute with a benchmark suite of increasing difficulty and probing different properties like color, pattern, and material. We further enhance our investigation by evaluating several state-of-the-art open-vocabulary object detectors using the proposed protocol and find that most existing solutions, which shine in standard open-vocabulary benchmarks, struggle to accurately capture and distinguish finer object details. We conclude the paper by highlighting the limitations of current methodologies and exploring promising research directions to overcome the discovered drawbacks.

2.2.7
Using geospatial semantic web for exploring geographic knowledge in medieval manuscripts

N. Pratelli, V. Bartalesi. [30]

This paper explores the capabilities of the Geospatial Semantic Web to support scholars in studying the geographic knowledge included in medieval and Renaissance works. In the context of the Italian national research project IMAGO, we developed a CRM-based ontology that aligns with the Open Geospatial Consortium (OCG) GeoSPARQL standard. The ontology enables geospatial queries on the IMAGO knowledge graph. The results of these queries, as detailed in this paper, demonstrate the effectiveness of this approach in representing the geospatial data and in inferring new knowledge. For example, using this approach, we are able to identify all the works that mention places in a specific region, or by combining geographic knowledge with knowledge about the literary genre of the works, we can identify authors who travelled to a particular territory, such as the Holy Land. Furthermore, combining temporal and geospatial information enables us to discover places within a particular territory mentioned in manuscripts of a specific century. These examples demonstrate the potential of the Geospatial Semantic Web approach to uncover previously hidden connections and enrich our understanding of historical and geographical data.

2.2.8
You Write like a GPT

A. Esuli, F. Falchi, M. Malvaldi, G. Puccetti

CLiC-it, 10th Italian Conference on Computational Linguistics [15]

We investigate how Raymond Queneau's Exercises in Style are evaluated by automatic methods for detection of artificially generated text. We work with the Queneau's original French version, and the Italian translation by Umberto Eco. We start by comparing how various methods for the detection of automatically generated text, also using different large language models, evaluate the different styles in the opera. We then link this automatic evaluation to distinct characteristic related to content and structure of the various styles. This work is an initial attempt at exploring how methods for the detection of artificially-generated text can find application as tools to evaluate the qualities and characteristics of human writing, to support better writing in terms of originality, informativeness, clarity.

2.2.9
INVALSI - Mathematical and Language Understanding in Italian: A CALAMITA Challenge

G. Puccetti, M. Cassese, A. Esuli

CLiC-it, 10th Italian Conference on Computational Linguistics [31]

While Italian is a high resource language, there are few Italian-native benchmarks to evaluate Language Models (LMs) generative abilities in this language. This work presents two new benchmarks: Invalsi MATE to evaluate models performance on mathematical understanding in Italian and Invalsi ITA to evaluate language understanding in Italian. These benchmarks are based on the Invalsi tests, which are administered to students of age between 6 and 18 within the Italian school system. These tests are prepared by expert pedagogists and have the explicit goal of testing average students' performance over time across Italy. Therefore, the questions are well written, appropriate for the age of the students, and are developed with the goal of assessing students' skills that are essential in the learning process, ensuring that the benchmark proposed here measures key knowledge for undergraduate students. Invalsi MATE is composed of 420 questions about mathematical understanding, these questions range from simple money counting problems to Cartesian geometry questions, e.g. determining if a point belongs to a given line. They are divided into 4 different types: scelta multipla (multiple choice), vero/falso (true/false), numero (number), completa frase (fill the gap). Invalsi ITA is composed of 1279 questions regarding language understanding, these questions involve both the ability to extract information and answer questions about a text passage as well as questions about grammatical knowledge. They are divided into 4 different types: scelta multipla (multiple choice), binaria (binary), domanda aperta (open question), altro (other). We evaluate 4 powerful language models both English-first and tuned for Italian to

see that best accuracy on Invalsi MATE is 55% while best accuracy on Invalsi ITA is 80%.

2.2.10
ViLMA: A Zero-Shot Benchmark for Linguistic and Temporal Grounding in Video-Language Models

I. Kesen, A. Pedrotti, M. Dogan, M. Cafagna, E.C. Acikgoz, L. Parcalabescu, I. Calixto, A. Frank, A. Gatt, A. Erdem, E. Erdem

ICLR 2024, 12th International Conference on Learning Representations [18]

With the ever-increasing popularity of pretrained Video-Language Models (VidLMs), there is a pressing need to develop robust evaluation methodologies that delve deeper into their visio-linguistic capabilities. To address this challenge, we present ViLMA (Video Language Model Assessment), a task-agnostic benchmark that places the assessment of fine-grained capabilities of these models on a firm footing. Task-based evaluations, while valuable, fail to capture the complexities and specific temporal aspects of moving images that VidLMs need to process. Through carefully curated counterfactuals, ViLMA offers a controlled evaluation suite that sheds light on the true potential of these models, as well as their performance gaps compared to human-level understanding. ViLMA also includes proficiency tests, which assess basic capabilities deemed essential to solving the main counterfactual tests. We show that current VidLMs' grounding abilities are no better than those of vision-language models which use static images. This is especially striking once the performance on proficiency tests is factored in. Our benchmark serves as a catalyst for future research on VidLMs, helping to highlight areas that still need to be explored.

2.2.11
VISIONE 5.0: Enhanced User Interface and AI Models for VBS2024

G. Amato, P. Bolettieri, F. Carrara, F. Falchi, C. Gennaro, N. Messina, L. Vadicamo, C. Vairo

MMM 2024, 30th International Conference on MultiMedia Modeling [1]

In this paper, we introduce the fifth release of VISIONE, an advanced video retrieval system offering diverse search functionalities. The user can search for a target video using textual prompts, drawing objects and colors appearing in the target scenes in a canvas, or images as query examples to search for video keyframes with similar content. Compared to the previous version of our system, which was runner-up at VBS 2023, the forthcoming release, set to participate in VBS 2024, showcases a refined user interface that enhances its usability and updated AI models for more effective video content analysis.

2.3 Magazines
2.4 Editorials

In this section, we report journals, proceedings, and books for which we acted as editors.

2.5 Preprints

In this section, we report the papers published only in preprint form on publicly accessible archives, in alphabetic order by first author.

2.5.1
Talking to DINO: Bridging Self-Supervised Vision Backbones with Language for Open-Vocabulary Segmentation

L. Barsellotti*, L. Bianchi*, N. Messina, F. Carrara, M. Cornia, L. Baraldi, F. Falchi, R. Cucchiara

arXiv.2411.19331. [2]

Recent advancements in large vision-language models enabled visual object detection in open-vocabulary scenarios, where object classes are defined in free-text formats during inference. In this paper, we aim to probe the state-of-the-art methods for open-vocabulary object detection to determine to what extent they understand fine-grained properties of objects and their parts. To this end, we introduce an evaluation protocol based on dynamic vocabulary generation to test whether models detect, discern, and assign the correct fine-grained description to objects in the presence of hard-negative classes. We contribute with a benchmark suite of increasing difficulty and probing different properties like color, pattern, and material. We further enhance our investigation by evaluating several state-of-the-art open-vocabulary object detectors using the proposed protocol and find that most existing solutions, which shine in standard open-vocabulary benchmarks, struggle to accurately capture and distinguish finer object details. We conclude the paper by highlighting the limitations of current methodologies and exploring promising research directions to overcome the discovered drawbacks.

2.5.2
Maybe you are looking for CroQS: Cross-modal Query Suggestion for Text-to-Image Retrieval

G. Pacini, F. Carrara, N. Messina, N. Tonellotto, G. Amato, F. Falchi

arXiv:2412.13834 [26]

Query suggestion, a technique widely adopted in information retrieval, enhances system interactivity and the browsing experience of document collections. In cross-modal retrieval, many works have focused on retrieving relevant items from natural language queries, while few have explored query suggestion solutions. In this work, we address query suggestion in cross-modal retrieval, introducing a novel task that focuses on suggesting minimal textual modifications needed to explore visually consistent subsets of the collection, following the premise of "Maybe you are looking for". To facilitate the evaluation and development of methods, we present a tailored benchmark named CroQS. This dataset comprises initial queries, grouped result sets, and human-defined suggested queries for each group. We establish dedicated metrics to rigorously evaluate the performance of various methods on this task, measuring representativeness, cluster specificity, and similarity of the suggested queries to the original ones. Baseline methods from related fields, such as image captioning and content summarization, are adapted for this task to provide reference performance scores. Although relatively far from human performance, our experiments reveal that both LLM-based and captioning-based methods achieve competitive results on CroQS, improving the recall on cluster specificity by more than 115% and representativeness mAP by more than 52% with respect

to the initial query. The dataset, the implementation of the baseline methods and the notebooks containing our experiments are available here: paciosoft.com/CroQS-benchmark/

3. Dissertations

3.1 PhD Thesis

3.1.1

Computational Authorship Analysis: Applications and Issues in the Cultural Heritage Field

Silvia Corbara. PhD in Data Science, University of Pisa, 2024. Supervisors: A. Monreale, A. Moreo, F. Sebastiani.

The discipline of Authorship Analysis studies the linguistic style of written documents to determine information about their authorship. Unlike traditional methodologies, it leverages statistical methods and focuses on quantifiable linguistic events rather than the literary content of the text. In recent years, this field has experienced significant growth due to advances in information technology, enabling the employment of Machine Learning and Natural Language Processing computational tools, and it has been applied in various domains, spanning from cybersecurity to forensics. This Ph.D. Thesis investigates the application of Computational Authorship Analysis methodologies in the cultural heritage domain. Building on the experience gathered gathered through the research of a case-study (the debated Dantean authorship of the historic document Epistle to Cangrande), we address what we believe are the four main issues in this domain application: i) the identification of features that allow for accurate classification while being topic-agnostic; ii) the limited size of the datasets usually available in these studies; iii) the challenges that can be encountered when facing the possibility that the document under scrutiny is a forgery; and iv) the necessity of providing scholars in cultural heritage with proper explanations regarding the computational system's findings.

3.1.2

Heterogeneous Transfer Learning For Natural Language Processing

Andrea Pedrotti. PhD in Computer Science, University of Pisa, 2024. Supervisors: A. Moreo, F. Sebastiani. [27]

With the advances in Deep Learning, the term Transfer Learning (TL) has become ubiquitous in the field of Machine Learning. One of the most widely adopted strategies when working with pre-trained models is to fine-tune them on downstream tasks by leveraging a relatively smaller labeled dataset compared to the amount of training data used for the pre-training phase. Fine-tuning is in fact a common technique of transfer learning. In general TL, refers to a set of techniques and approaches which leverage training data sampled from a source distribution to improve performance on a test set, the target, containing elements sampled from a different, but related, distribution. This paradigm brings about two major advantages. First, it increases performance on the target domain by making the algorithm more robust and resilient, allowing us to leverage powerful pre-trained models that are trained on hardware not widely available. Second, it allows the application of data-intensive techniques to many scarce-resource domains where training an ad-hoc solution would be impossible. In this thesis, we explore applications of Heterogeneous Transfer Learning (HTL) to the field of Natural Language Processing (NLP). We identify two main exploratory spaces: (i) the heterogeneous space defined by different languages (multilinguality) and (ii) the heterogeneous space defined by the intersection of languages and perceptual (multimodality) information. Lastly, (iii) we explore the benefits of HTL when dealing simultaneously with both multimodality and multilinguality.

3.2 Master of Science Dissertations

3.2.1

Predicting Classifier Accuracy under Prior Probability Shift

Lorenzo Volpi, Computer Science, 2024. [35]. Advisors: A. Esuli, A. Moreo, F. Sebastiani

Predicting the accuracy that a classifier will have on unseen data (i.e., on unlabeled data that were not available at training time) can be done via k-fold cross-validation (kFCV). However, using kFCV returns reliable predictions only when the training data and the unseen data are identically and independently distributed (IID), i.e., were randomly sampled from the same distribution. Unfortunately, in real-world applications it is often the case that the training data and the unseen data are not IID, i.e., that we want to deploy the trained model on unseen data that exhibit some kind of dataset shift with respect to the training data. In this work we deal with the problem of predicting classifier accuracy on unseen data characterized by prior probability shift (PPS), an important type of dataset shift. We propose a class of methods built on top of quantification algorithms robust to PPS, i.e., algorithms devised for estimating the prevalence values of the classes in unseen data characterized by PPS. The methods we propose are based on the idea of viewing the cells of the contingency table (on which classifier accuracy is computed) as classes. We perform systematic experiments in which we test the prediction accuracy of our methods against state-of-the-art classifier accuracy prediction methods from the machine learning literature.

3.2.2

On the effectiveness of deepfake detection on multimodal fake news

Enrico Nello, Artificial Intelligence and Data Engineering, 2024. [24]. Advisors: M.G.C.A. Cimino, F. Falchi, C. Gennaro, D.A. Coccomini

The increasing sophistication of deep learning has enabled the creation of highly realistic synthetic media, including images, videos, text, and audio. This capability has led to the rise of "deepfakes," manipulated content designed to appear authentic. As deepfake technology advances, it increasingly integrates both visual and textual misinformation, amplifying the challenges of content verification and misinformation detection. This thesis investigates deepfake detection within the context of fake news, focusing on how the combination of synthetic images and misleading narratives affects detection performance. To address this, an existing fake news dataset is enriched with synthetically generated images, creating a more comprehensive benchmark for misinformation detection. A comparative analysis is conducted to assess the difficulty of detecting deepfakes in misleading versus truthful contexts, the potential for improving model resilience against deepfake misinformation, and the benefits of multi-

modal approaches over unimodal ones. To rigorously evaluate these aspects, the study introduces a dataset that better reflects real-world misinformation scenarios. Experimental results indicate that deepfakes embedded in fake news present a greater detection challenge than those associated with truthful content. The findings highlight the effectiveness of multimodal detection approaches and establish a benchmark for future research in both deepfake and fake news identification. This work contributes to the field by defining a research direction focused on multimodal deepfake detection within misinformation contexts. The results emphasize the need for more robust detection models to counteract the growing impact of deepfake-driven disinformation.

3.2.3
Advanced Query Suggestion for Interactive Text-to-Image Retrieval: a novel task and benchmark

Giacomo Pacini, Artificial Intelligence and Data Engineering, 2024. [25]. Advisors: N. Tonellotto, F. Falchi, F. Carrara, N. Messina

The increasing amount of multimedia elements in visual collections fosters the development of novel and customizable interactive image retrieval systems. While current literature focuses on improving and measuring the ability of a system to retrieve the most relevant items given a natural language query, few works have tried to make these systems more interactive to enhance the browsing experience. The objective of this master thesis is to introduce and define a novel task in the field of cross-modal retrieval, termed "Visual Guided Query Suggestion" (VGQS). This task aims to enhance user experience in cross-modal retrieval systems by generating expanded query suggestions based on the initial search results. Specifically, VGQS systems automatically suggest the smallest textual modifications needed to explore visually consistent subsets of the collection. To facilitate the evaluation and development of methods addressing VGQS, we present a comprehensive benchmark dataset. This dataset consists of initial queries, grouped result sets, and human-defined expanded queries for each group. We establish dedicated metrics to rigorously evaluate the performance of various methods on this task. These metrics are designed to measure the representativeness, specificity and similarity to the original query of the suggested expanded ones. Baseline methods, adapted from related fields such as image captioning and query expansion, are applied to this task to provide reference performance scores. The thesis details the creation of the benchmark dataset, the definition of the evaluation metrics, and the adaptation of baseline methods. Experimental results are presented, showcasing the performance of these baseline methods and highlighting the potential and challenges of the VGQS task. This work lays the foundation for future research in enhancing cross-modal retrieval systems through intelligent query expansion strategies. VGQS, integrated into interactive multimedia browsing software, aims at increasing its interactivity and the overall user experience.

3.2.4
A Novel Benchmark for Prompt-Guided Class-Agnostic Counting: Assessing Models' Understanding of Textual Prompts

Matteo Pierucci, Artificial Intelligence and Data Engineering, 2024. [29]. Advisors: M. Avvenuti, F. Falchi, L. Ciampi, N. Messina

The computer vision task of object counting involves estimating the number of object instances within an image. Traditional class-specific object counting methods rely on regression models and density map estimation to count objects. However, these methods often require extensive datasets and fail to generalize across different object classes. Recent research in object counting has increasingly focused on reducing the annotation problem in dataset creation. Therefore, class-agnostic object counting is a new task that involves training a network to count object instances of any class at test time, even if these classes differ from those seen during the training phase. Typically, these networks use images and density maps as training targets and may also employ single examples of the objects to be counted, named exemplars, as prototypes for counting at test time. Unlike traditional counting methods that rely on class-specific datasets, class-agnostic counting uses multi-class datasets to build a versatile model applicable to unseen categories with minimal additional data. This work explores prompt-guided zero-shot counting, where textual prompts replace visual exemplars to guide the counting process. Despite the advancements in this field, numerous state-of-the-art models ignore textual information when estimating object count. To address this gap, we have developed a novel benchmark and metrics to assess models' understanding of textual prompts in the counting process. Additionally, we introduce a variation of an existing dataset, which includes cases where prompt-specified objects are absent in query images. Finally, we train a state-of-the-art counting model on this new dataset and test its performance on the novel benchmark. Our experimental results demonstrate that the proposed model effectively understands textual prompts and accurately counts objects, even in challenging conditions where objects of interest are not present in the query image, creating a comparative baseline for future models. This work advances the field of prompt-guided zero-shot counting, offering insights into the capabilities and limitations of current models and providing a foundation for future research in this area.

4. Resources

In this section, we report contributions of AIMH having to do with the creation of datasets (Section 4.1), the publication of code (Section 4.2), and the design of shared tasks (Section **??**)

4.1 Datasets

4.1.1
The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding

L. Bianchi, F. Carrara, N. Messina, C. Gennaro, F. Falchi [7]

Recent advancements in large vision-language models enabled visual object detection in open-vocabulary scenarios, where object classes are defined in free-text formats during inference. In this paper, we aim to probe the state-of-the-art methods for open-vocabulary object detection to determine to what extent they understand fine-grained properties of objects and their parts. To this end, we introduce an evaluation protocol based on dynamic vocabulary generation to test whether models detect, discern, and assign the

correct fine-grained description to objects in the presence of hard-negative classes. We contribute with a benchmark suite of increasing difficulty and probing different properties like color, pattern, and material. We further enhance our investigation by evaluating several state-of-the-art open-vocabulary object detectors using the proposed protocol and find that most existing solutions, which shine in standard open-vocabulary benchmarks, struggle to accurately capture and distinguish finer object details. We conclude the paper by highlighting the limitations of current methodologies and exploring promising research directions to overcome the discovered drawbacks. The dataset is publicly available at https://doi.org/10.5281/zenodo.13269555.

4.1.2
DeepFakeNews: novel dataset for Deepfake and Fakenews detection

E. Nello, M.G.C.A. Cimino, F. Falchi, C. Gennaro, D.A. Coccomini [23]

We introduce DeepFakeNews, a dataset designed for evaluating the detection of both deepfakes and fake news. As an extension of the Fakeddit fake news dataset, DeepFakeNews includes 509,916 images, with 254,958 deepfake images generated using three different models. The dataset is balanced, ensuring an equal distribution of authentic and synthetic images, and has been refined by removing hand-modified content and low-quality data. These enhancements make DeepFakeNews a comprehensive benchmark for assessing multimodal detection systems that analyze both visual and textual misinformation. The dataset is publicly available at https://doi.org/10.5281/zenodo.11186583.

4.2 Code

4.2.1
Talk2DINO

L. Barsellotti*, L. Bianchi*, N. Messina, F. Carrara, M. Cornia, L. Baraldi, F. Falchi, R. Cucchiara [2]

`https://lorebianchi98.github.io/Talk2DINO/`

4.2.2
FG-CLIP

L. Bianchi, F. Carrara, N. Messina, F. Falchi [6]

`https://lorebianchi98.github.io/FG-CLIP/`

4.2.3
FG-OVD

L. Bianchi, F. Carrara, N. Messina, C. Gennaro, F. Falchi [7]

`https://lorebianchi98.github.io/FG-OVD/`

5. Awards

5.1 ERCIM Cor Baayen Award

Nicola Messina was awarded the 2024 ERCIM Cor Baayen Award, the annual award given to a promising young researcher in computer science and applied mathematics by ERCIM, the European Research Consortium for Informatics and Mathematics, fosters collaboration across Europe's research community and strengthens ties with industry. It brings together leading research institutes to advance innovation and cooperation in these fields.

Motivations: *Nicola's research is distinguished by its high quality, interdisciplinary approach, and significant impact. His work spans across multiple domains, including Artificial Intelligence, Computer Vision, Deep Learning, and Multimedia Information Retrieval. He has made substantial contributions to both the theoretical and applied aspects of these fields, with measurable scientific and practical outcomes. During his PhD, Nicola began exploring the ability of neural networks to understand and process relationships between objects in computer vision. He also made important contributions in multimodal artificial intelligence by developing innovative and efficient methods for aligning representations of complex neural networks in both visual and language modalities. His work in multimedia information retrieval has demonstrated immediate real-world applicability. This is exemplified by VISIONE, a large-scale video search system that won the 2024 international Video Browser Showdown competition and secured second place in 2023. VISIONE has been employed by the Italian national public broadcaster RAI, as part of the AI4Media European project, to facilitate efficient browsing of audiovisual archives. More recently, he has expanded his research focus to include the application of attentive deep learning techniques for structural health monitoring and the preservation of cultural heritage. This new direction highlights his ability to engage in highly interdisciplinary research and to address complex challenges across different fields. Nicola's professional experience also includes international collaborations with esteemed European universities and participation in various Italian and European research projects, such as AI4Media, AI4EU, INAROS, AI4CHSites, ADA, and Smart News. In addition to his research, he is involved in teaching and dissemination activities, offering instruction and guidance to students on topics related to deep learning, computer vision, and multimodal processing. In 2021 and 2022, Dr. Messina received the ISTI Young Researcher Award, recognizing him as one of the top young researchers (under 32 years old) at ISTI-CNR,Italy. Overall, Dr. Nicola Messina is an outstanding early-career researcher who has achieved remarkable success during his PhD and postdoctoral work, combining scientific innovation with societal and commercial impact. Nicola Messina is currently employed at the Information Science and Technologies Institute, National Research Council (ISTI-CNR), Italy, where he is associated with the Artificial Intelligence for Multimedia and Humanities Laboratory (AIMH). He received his PhD in 2022 from the University of Pisa, Italy, with a thesis titled "Relational Learning in Computer Vision", supervised by Fabrizio Falchi, Giuseppe Amato, and Marco Avvenuti.*

5.2 International Competitions

5.2.1 Video Browser Showdown

The VISIONE content-based video retrieval system won the Video Browser Showdown 2024, The Video Retrieval Competition, with the approach described in [1].

5.3 Best Paper Awards

5.3.1 CBMI

The paper "Is CLIP the main roadblock for fine-grained open-world perception?", L. Bianchi, F. Carrara, N. Messina, F. Falchi, [6] won the Best Paper Award at the 21st International

Conference on Content-based Multimedia Indexing, Reykjavik, Iceland, September 18-20, 2024.

5.3.2 ACL

The paper “AI ’News’ Content Farms Are Easy to Make and Hard to Detect: A Case Study in Italian”, G. Puccetti, A. Rogers, C. Alzetta, F. Dell’Orletta, A. Esuli, [33], won the Senior Area Chair Award at the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, August 11–16, 2024

5.4 National Awards

5.4.1 AI*IA Best Thesis Award

Lorenzo Volpi has won the 2024 “Leonardo Lesmo” Graduate Award from the Italian Association for Artificial Intelligence, for his thesis entitled "Predicting Classifier Accuracy under Prior Probability Shift".

5.4.2 ISTI Young Research Awards

The researchers of the AIMH Lab that won the ISTI Young Research Awards “Matteo Delle Piane” in 2024 are:

- Xxxx, in the Advanced category
- Xxxx, in the Beginner category

5.4.3 ISTI Grant for Young Mobility

The researchers of the AIMH Lab that won an ISTI Grant for Young Mobility in 2024 are:

- Giovanni Puccetti, second call

6. Previous Reports

The activity report of previous years can be found in:

- AIMH Research Activities 2023
  DOI: 10.32079/isti-ar-2023/001
  https://hal.science/hal-04430990/
  https://hdl.handle.net/20.500.14243/452242
- AIMH Research Activities 2022
  DOI: 10.32079/isti-ar-2022/002
  https://hdl.handle.net/20.500.14243/413820
  https://hal.science/hal-04023023v1
- AIMH Research Activities 2021
  DOI: 10.32079/isti-ar-2021/003
  https://hdl.handle.net/20.500.14243/445569
  https://hal.science/hal-03573814v1
- AIMH Research Activities 2020
  DOI: 10.32079/isti-ar-2020/001
  https://hdl.handle.net/20.500.14243/428630
  https://hal.science/hal-03466721v1

References

[1] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. Visione 5.0: Enhanced user interface and ai models for vbs2024. In Stevan Rudinac, Alan Hanjalic, Cynthia Liem, Marcel Worring, Björn Þór Jónsson, Bei Liu, and Yoko Yamakata, editors, *MultiMedia Modeling*, pages 332–339, Cham, 2024. Springer Nature Switzerland.

[2] Luca Barsellotti, Lorenzo Bianchi, Nicola Messina, Fabio Carrara, Marcella Cornia, Lorenzo Baraldi, Fabrizio Falchi, and Rita Cucchiara. Talking to dino: Bridging self-supervised vision backbones with language for open-vocabulary segmentation. *arXiv preprint arXiv:2411.19331*, 2024.

[3] Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Nicolò Pratelli, Pasquale Pagano, Michele Moretti, and Gianluca Brunori. A semantic knowledge graph of european mountain value chains. *Scientific Data*, 11, 2024.

[4] Valentina Bartalesi, Emanuele Lenzi, and Claudio De Martino. Using large language models to create narrative events. *PeerJ Computer Science*, 2024.

[5] Valentina Bartalesi and Nicolò Pratelli. Representing geospatial knowledge in narratives. *Journal on Computing and Cultural Heritage*, 2024.

[6] Lorenzo Bianchi, Fabio Carrara, Nicola Messina, and Fabrizio Falchi. Is clip the main roadblock for fine-grained open-world perception? *arXiv preprint arXiv:2404.03539*, 2024.

[7] Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi. The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22520–22529, 2024.

[8] Marta Boscariol, Luana Bulla, Lia Draetta, Beatrice Fiumanò, Emanuele Lenzi, and Leonardo Piano. Evaluation of llms on long-tail entity linking in historical documents. Accepted, in publication.

[9] Mirko Bunse, Alejandro Moreo, Fabrizio Sebastiani, and Martin Senz. Regularization-based methods for ordinal quantification. *Data Mining and Knowledge Discovery*, 38(6):4076–4121, 2024.

[10] Luca Ciampi, Carlos Santiago, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. In the wild video violence detection: An unsupervised domain adaptation approach. *SN Comput. Sci.*, 5(7), August 2024.

[11] Davide Alessandro Coccomini, Andrea Esuli, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Detecting images generated by diffusers. *PeerJ Computer Science*, 10:e2127, 2024.

[12] Davide Alessandro Coccomini, Giorgos Kordopatis Zilos, Giuseppe Amato, Roberto Caldelli, Fabrizio Falchi, Symeon Papadopoulos, and Claudio Gennaro. Mintime: Multi-identity size-invariant video deepfake detection. *IEEE Transactions on Information Forensics and Security*, 19:6084–6096, 2024.

[13] Silvia Corbara and Alejandro Moreo. Forging the forger: An attempt to improve authorship verification via data augmentation. *IEEE Access*, 12:171911–171925, 2024.

[14] Washington Cunha, Alejandro Moreo, Andrea Esuli, Fabrizio Sebastiani, Leonardo Rocha, and Marcos André Gonçalves. A noise-oriented and redundancy-aware instance selection framework. *ACM Trans. Inf. Syst.*, 43(2), January 2025.

[15] Andrea Esuli, Fabrizio Falchi, Marco Malvaldi, and Giovanni Puccetti. You write like a GPT. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[16] Edoardo Fazzari, Fabio Carrara, Fabrizio Falchi, Cesare Stefanini, and Donato Romano. Using AI to decode the behavioral responses of an insect to chemical stimuli: towards machine-animal computational technologies. *Int. J. Mach. Learn. Cybern.*, 15(5):1985–1994, 2024.

[17] Pablo González, Alejandro Moreo, and Fabrizio Sebastiani. Binary quantification and dataset shift: an experimental investigation. *Data Mining and Knowledge Discovery*, pages 1–43, 2024.

[18] Ilker Kesen, Andrea Pedrotti, Mustafa Dogan, Michele Cafagna, Emre Can Acikgoz, Letitia Parcalabescu, Iacer Calixto, Anette Frank, Albert Gatt, Aykut Erdem, and Erkut Erdem. ViLMA: A zero-shot benchmark for linguistic and temporal grounding in video-language models. In *The Twelfth International Conference on Learning Representations*, 2024.

[19] Gabriele Lagani, Fabrizio Falchi, Claudio Gennaro, Hannes Fassold, and Giuseppe Amato. Scalable bio-inspired training of deep neural networks with fasthebb. *Neurocomputing*, 595:127867, 2024.

[20] Claudio De Martino, Tommaso Turchi, and Alessio Malizia. Blocklybias: A visual programming language for bias identification in ai data. In Degen Helmut and Ntoa Stavroula, editors, *Artificial Intelligence in HCI*, volume 14735 of *Lecture Notes in Computer Science*, page 45–59, 2024.

[21] Nicola Messina, Davide Alessandro Coccomini, Andrea Esuli, and Fabrizio Falchi. Cascaded transformer-based networks for wikipedia large-scale image-caption matching. *Multim. Tools Appl.*, 83(23):62915–62935, 2024.

[22] Alessio Molinari and Andrea Esuli. Sal$_\tau$: efficiently stopping TAR by improving priors estimates. *Data Min. Knowl. Discov.*, 38(2):535–568, 2024.

[23] Enrico Nello. Deepfakenews dataset, 2024.

[24] Enrico Nello. On the effectiveness of deepfake detection on multimodal fake news. Master's thesis, University of Pisa, 2024.

[25] Giacomo Pacini. Advanced query suggestion for interactive text-to-image retrieval: a novel task and benchmark. Master's thesis, University of Pisa, 2024.

[26] Giacomo Pacini, Fabio Carrara, Nicola Messina, Nicola Tonellotto, Giuseppe Amato, and Fabrizio Falchi. Maybe you are looking for croqs: Cross-modal query suggestion for text-to-image retrieval, 2024.

[27] Andrea Pedrotti. *Heterogeneous Transfer Learning for Natural Language Processing*. PhD thesis, University of Pisa, 2024.

[28] Olaya Pérez-Mon, Alejandro Moreo, Juan José del Coz, and Pablo González. Quantification using permutation-invariant networks based on histograms. *Neural Computing and Applications*, pages 1–16, 2024.

[29] Matteo Pierucci. A novel benchmark for prompt-guided class-agnostic counting: Assessing models' understanding of textual prompts. Master's thesis, University of Pisa, 2024.

[30] Nicolò Pratelli and Valentina Bartalesi. Using geospatial semantic web for exploring geographic knowledge in medieval manuscripts. In Antonacopoulos Apostolos, Hinze Annika, Piwowarski Benjamin, Coustaty Mickaël, Di Nunzio Giorgio Maria, Gelati Francesco, and Vanderschantz Nicholas, editors, *Linking Theory and Practice of Digital Libraries, Pt. II, TPDL 2024*, volume 15178 of *Lecture Notes in Computer Science*, pages 74–84, 2024.

[31] Giovanni Puccetti, Maria Cassese, and Andrea Esuli. INVALSI - mathematical and language understanding in italian: A CALAMITA challenge. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[32] Giovanni Puccetti, Claudia Collacciani, Andrea Amelio Ravelli, Andrea Esuli, and Marianna Bolognesi. ABRICOT - abstractness and inclusiveness in context: A CALAMITA challenge. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[33] Giovanni Puccetti, Anna Rogers, Chiara Alzetta, Felice Dell'Orletta, and Andrea Esuli. AI 'news' content farms

are easy to make and hard to detect: A case study in italian. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 15312–15338. Association for Computational Linguistics, 2024.

[34] Mattia Setzu, Silvia Corbara, Anna Monreale, Alejandro Moreo, and Fabrizio Sebastiani. Explainable authorship identification in cultural heritage applications. *ACM Journal on Computing and Cultural Heritage*, 2024.

[35] Lorenzo Volpi. Predicting classifier accuracy under prior probability shift. Master's thesis, University of Pisa, 2024.

[36] Xenophon Zabulis, Nikolaos Partarakis, Valentina Bartalesi, Nicolò Pratelli, Carlo Meghini, Arnaud Dubois, Ines Moreno, and Sotiris Manitsaris. Multimodal dictionaries for traditional craft education. *Multimodal Technologies and Interaction*, 8, 2024.

[37] Xenophon Zabulis, Nikolaos Partarakis, Ioannna Demeridou, Valentina Bartalesi, Nicolò Pratelli, Carlo Meghini, Nikolaos Nikolaou, and Fallahian Peiman. Modelling and simulation of traditional craft actions. *Applied Sciences*, 14, 2024.