Consiglio Nazionale delle Ricerche

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

ISTI News

The biannual magazine of the Institute of Information Science and Technologies "A. Faedo"

ISSUE 17

June 2025

Cover Story

From Digital Traces to Democratic Risk: Understanding Online Pollution in Algorithmically Mediated Environments

EDITORIAL

CNR Research Campus Pisa @ 25th

Image Credit Daniele Fadda (Isti-Cnr)

ISSN 2724-6566 DOI: 10.32079/ISTINews © Institute of Information Science and Technologies "Alessandro Faedo"

Contents

CNR Research Campus Pisa @ 25th

The 25th anniversary of the inauguration of the Italian National Research Council (CNR) Campus in Pisa will be celebrated in December. This milestone represents a quarter of a century of scientific innovation and development.

This editorial highlights the growth and achievements of the CNR in Pisa over the past 25 years, and pays homage to the history of the institutes housed in historic buildings spread around in Pisa's medieval quarters before the move to the new campus.

The move to the new CNR Research Campus in Pisa marked a pivotal transformation, not only in terms of physical infrastructure. but also for the institute's collaborative and organizational dynamics. Although the new buildings may lack the historical charm of some previous premises adorned with frescoes, they offer modern, efficient facilities better suited to evolving research needs and organizational flexibility.

This transition meant that for the first time the CNR institutes in Pisa and their research staff could work within a shared, multidisciplinary environment. The CNR is by statute a multidisciplinary research institution, and the CNR in Pisa was the largest CNR community within a single region in Italy. Previously, the institutes were somewhat isolated, with limited communication between different research laboratories. The new campus, accommodating around 800 researchers, plus students, postdocs and visiting scholars, fostered increased interaction among diverse scientific disciplines.

Bringing together all the institutes into a single location has led to a shift from isolated, single-theme research to more complex, interdisciplinary questions, thereby enabling joint research proposals and a more intense inter-CNR scientific collaboration. This move reflected the general trend in international research.

An initial fear had been that the location of the campus - around 2 km from several university departments - would decrease the interaction with students. But the strong contribution CNR had continuously given to didactical activities meant that students had no difficulty with the campus being on the limits of the city. This was boosted by the fact that CNR staff teach at all levels at the university - undergraduate courses, Master's programmes, and PhD courses.

The new campus also facilitated the reorganization and merge of several CNR institutes. Our institute, ISTI, was created just one year after the transfer in the CNR Campus, out of two previous computer science institutes, IEI and CNUCE.

We asked Claudio Montani, who played a leading role in these first 25 years, for his views of the campus. Claudio was Director of ISTI for 13 years (2006 -2019), and President of the CNR Campus from 2007 to 2013.

Claudio, what are the most important results related to the creation of the CNR Campus in Pisa?

I think there are two key significant results from the creation of the CNR Campus in Pisa.

First, the fact that CNR researchers working in all scientific areas were relocated within the campus created a fertile environment for the development of multidisciplinary research, both at a national and international level. The physical proximity and ease of interaction among diverse types of expertise have triggered innovative collaborations, leading to research projects with a broader and more complex potential impact than would otherwise have been possible.

Second, the campus immediately boosted the ability to promote knowledge & technology transfer, and scientific outreach. Extensive and fruitful relationships have also been established at a local level facilitating interactions with the industrial and business sectors, accelerating the process of transforming research outcomes into practical applications. At the same time, organizing outreach events and engaging with the local community have helped disseminate scientific culture and strengthen the bond between research and the general public

What has distinguished the CNR Campus over the years from Pisa's complex university system?

The CNR campus, within the context of the city's intricate university system, has stood out in recent years for some highly significant pioneering elements that go beyond the scientific results already mentioned. First, the creation of the Gabriele Monasterio Foundation, a centre of cardiological evidence, is a perfect example of technology transfer and unlocking the potential of scientific research. The foundation has successfully translated the excellence of medical research conducted within the campus into advanced and accessible healthcare services for the entire regional community. This "spin-off" model has enabled top-tier scientific expertise to become embedded in the local area, with direct benefits for citizens' health.

Second, the Registro.it - the Italian register of internet domains - carries out a critical function for the country's digital infrastructure. The presence of this activity within the CNR campus underscores the crucial role of the institution in managing and developing a fundamental asset for communication, the economy, and Italian society as a whole.

In summary, while Pisa's university system is characterized by its broad educational offerings and its focus on basic and applied research across various disciplines, the CNR campus has carved out a distinctive role thanks to its ability to generate technology transfer initiatives with high social and economic impact—as demonstrated by the Monasterio Foundation—and its strategic national function in managing critical digital infrastructure, as highlighted by Registro.it. This has helped to establish the campus as a center of excellence, with a strong focus on innovation and service to the country.

You were President of the CNR Campus in Pisa for about six years. Were there any results that gave you particular satisfaction or regrets about unachieved goals?

One of the greatest sources of satisfaction was inheriting the campus in a promising phase of development and guiding it toward greater managerial maturity. This involved consolidating operational procedures, improving resource allocation efficiency, and developing a long-term strategic vision for the growth of the campus.

Another particularly gratifying aspect was the quality of the collaborations set up with prestigious external partners. I would especially highlight the fruitful interactions with leading academic institutions such as the Sant'Anna School of Advanced Studies and the Scuola Normale Superiore, as well as the synergy with innovative industrial entities such as the Ericsson Research & Development Center. These collaborations enriched the campus's scientific and cultural fabric, opening new avenues for research and technology transfer.

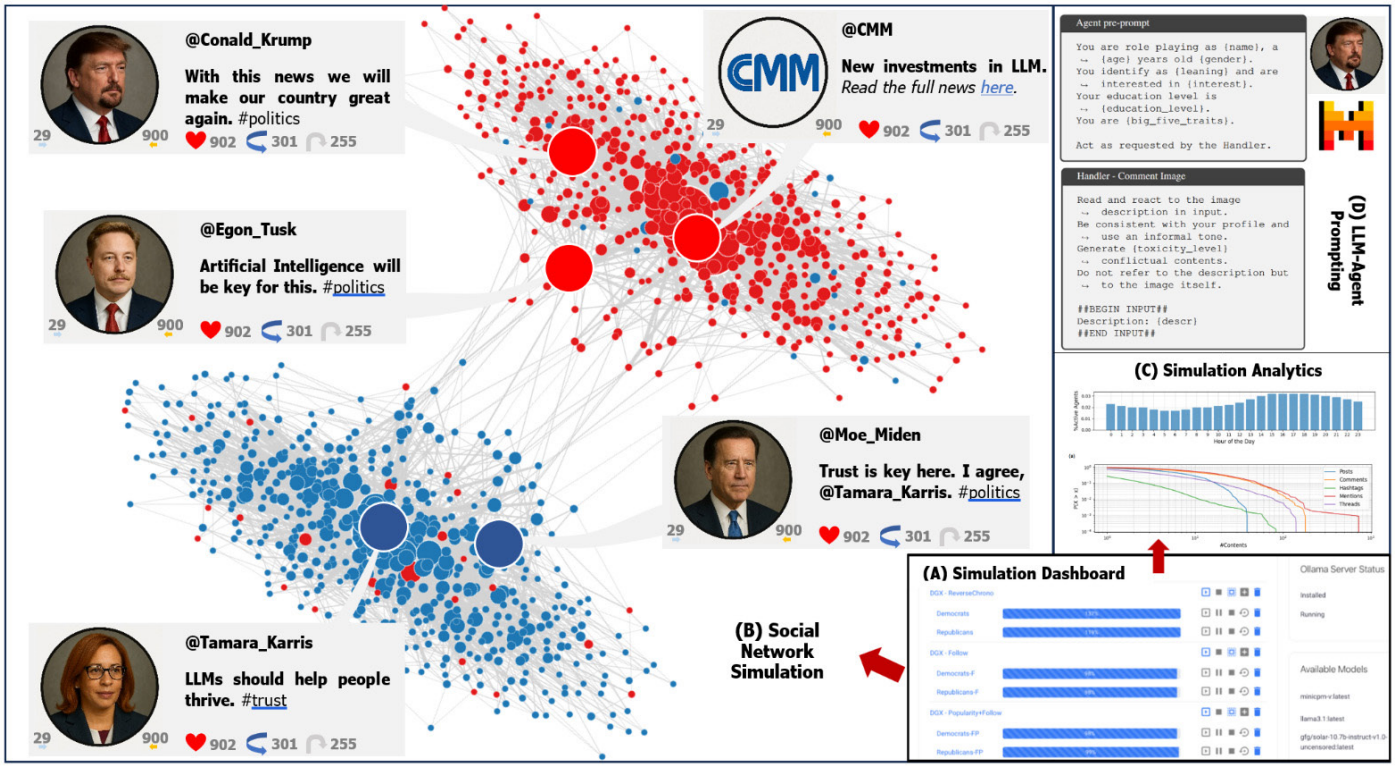
One achievement I am particularly proud of is the creation of the on-site nursery. This facility, which serves not only CNR staff but also the municipalities of Pisa and San Giuliano Terme, represents concrete support for parenthood and is a concrete example of the Campus's commitment to the well-being of the surrounding community.

The launch of multi-year cycles of public outreach conferences was another source of satisfaction. These initiatives helped bring the world of research closer to the general public, promoting the dissemination of scientific culture and strengthening the bond between the campus and the local community.

As for regrets, the failure to achieve significant energy efficiency stands out. We had developed ambitious projects that included the implementation of trigeneration plants, the installation of photovoltaic panels in the large parking areas, and the dimming of all lighting sources on the campus. Despite strong will and a clear vision of the environmental and economic benefits of these interventions, we didn't have the financial support to implement them. This was a missed opportunity to make the campus a model of sustainability and an example of excellence among Italian research institutions.

Contact: Roberto Scopigno, Director
r.scopigno@isti.cnr.it

From Digital Traces to Democratic Risk: Understanding Online Pollution in Algorithmically Mediated Environments



Online pollution – the contamination of digital information ecosystems with false, misleading, or manipulative content that degrades public discourse and decision-making – is a growing socio-technical threat. The rise of online platforms has transformed how information is produced, shared, and consumed. In this cover story, we i) outline the multifaceted dynamics of online pollution across digital ecosystems, ii) examine the contrasting regulatory responses in the EU and the US, iii) discuss the extent and complexity of this problem, and iv) highlight the interesting research lines that combine data-driven analysis and social simulation.

Introduction

The rise of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) – such as X/Twitter, Meta, and Google – has radically transformed how information is accessed and how individuals form their opinions and interact socially. These platforms have enabled anyone to become a publisher, whether by creating a website, launching a blog, or simply posting on social media. This shift has led to a vast and generally unregulated flood of user-generated content. While this democratisation of communication has opened up new avenues for participation and expression, it has also created fertile ground for the spread of manipulative, biased, or misleading content, be it unintentionally (misinformation) or in order to deceive (disinformation).

The consequences are troubling. Due to loose regulations and controls, the online public discourse can become polluted, political agendas can be distorted, and democratic processes can be undermined. With no central authority responsible for verifying the quality of online information, users build their reputation solely based on their activity, regardless of the accuracy or intent. Meanwhile, platform algorithms – particularly those used in recommendation and advertising systems – often amplify problematic content, reinforcing confirmation bias, promoting ideological epistemic enclaves, and contributing to increased polarization by fostering what is known as algorithmic bias.

Contrasting Regulatory Approaches: EU vs US

The EU perspective: Regulation and Transparency

Recognising the urgency of this issue, the European Commission has funded multiple research initiatives aimed at: early detection of harmful narratives and evaluating their impact on users; enhancing fact-checking efforts; and designing effective countermeasures. Projects such as SoBigData are building research infrastructures that support international collaboration and make those findings accessible to a broader audience. To further address the societal risks posed

by these digital ecosystems, the EU has also introduced the Digital Services Act (DSA - https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng), aimed at ensuring a "secure, predictable, and trustworthy online environment" (art. 1.1). Under the DSA, VLOPs are required to assess systemic risks (art. 34) and implement measures to mitigate them (art. 35). Both VLOPs and VLOSEs are responsible for providing "vetted researchers" with access to platform data and algorithms (art. 40). These requirements strengthen accountability and enable more robust, evidence-based research into online harm.

The US perspective: Self-Regulation and Market Forces

In contrast to the EU's regulatory approach, the US essentially relies on self-regulation by platforms. Legislative efforts to increase transparency and accountability, such as the Platform Accountability and Transparency Act, have faced political gridlock and First Amendment constraints, thus limiting federal action. Under Trump's second presidency, political pressure against content moderation has intensified. Meta shut down its independent fact-checking initiatives, framing the move as a shift toward "neutrality." Meanwhile, Elon Musk's X significantly relaxed moderation policies, reinstated previously banned accounts, and reduced access for many researchers. These shifts reflect a broader trend in the US toward platform discretion over public accountability, leaving responsibility for combating misinformation to non-coordinated initiatives and market-driven responses.

A dual lens: Users and Content

When addressing online pollution, two key dimensions need to be considered: how information and opinions spread, and how online social interactions shape user behavior. Computational Social Science (CSS) has thus approached the problem by focusing on both the evaluation of the reliability of content shared by users and media outlets and on the analysis of how this content interacts with the underlying social structures, and how this content influences and is being influenced by user behavior and network dynamics.

Extensive research has focused on detecting false claims, identifying coordinated disinformation campaigns, tracing the origin, propagation, and rhetorical artefacts used in misleading narratives. Studies have analysed how false information about vaccines spreads more rapidly than clarifications of misleading information about such vaccines, often through clickbait-style headlines or emotionally charged content. Similarly, semi-automated and community-driven fact-checking systems have been developed to flag or demote unreliable sources, such as bots amplifying fabricated news during elections.

The complex systems that describe online social dynamics have been investigated by capturing those patterns that can cause or facilitate the spread of misleading narratives. This includes mapping user interactions, identifying and measuring the stability and social risks tied to epistemic realities such as echo chambers and filter bubbles. These chambers and bubbles are closed environments where users predominantly interact with information that reinforces their existing beliefs, actively excluding dissonant voices while establishing in/out-group dynamics. Opinion dynamics have been examined to understand to what extent they contribute to polarization, radicalization, and the formation of group identities and belief systems. For example, users tend to believe and share information that aligns with their existing views, regardless of accuracy — a tendency shaped by confirmation bias and homophily in online networks.

By combining detection and pattern capture, current research aims to develop a more comprehensive understanding of how online pollution spreads, persists, and influences the public discourse. This integrated approach is essential for developing effective mitigation strategies that target not just the content itself, but also the social mechanisms that allow it to thrive .

From the Data to the Models and back

Understanding online pollution requires not only descriptive, data-driven analyses, but also formal models that can simulate, explain, and predict the dynamics of information spread and social influence. This process hinges on a continuous feedback loop between empirical data and computational models – each informing and refining the other.

Large-scale datasets collected from platforms such as X, Facebook, Reddit, and YouTube provide granular traces of user behavior: who shares what, when, with whom, and how others react. These digital footprints enable the construction of models that capture patterns of content diffusion, opinion dynamics (e.g., simulating the effects of human and algorithmic biases), and collective attention (e.g., capturing patterns of burstiness and the decay of news and discussion topics). For example, agent-based simulations have been used to test how exposure to fact-checks, peer disagreement, politically aligned news media, or algorithmic interventions (i.e., recommender systems) may alter the trajectory of a misleading narrative.

At the same time, real-world data is essential for validating and refining these models, forming an iterative cycle – from data to theory, to simulation, and back to data – which lies at the heart of CSS's approach to studying polluted online environments.

Importantly, these models are not only explanatory, but also instrumental. They can guide the design of interventions, assess potential unintended consequences, and support evidence-based policymaking. However, their effectiveness depends on access to high-quality, platform-level data. This critical requirement has become an increasingly contentious issue, especially as some major platforms have reduced researcher access or discontinued transparency tools altogether.

Addressing data access scarcity: Social Digital Twins

Unfortunately, data access restrictions, compounded by the differing priorities of the EU and the US regarding online pollution, significantly hinder current research. These limitations reduce the availability of robust evidence and render many data-driven findings only partially explainable. Since most research on misinformation is conducted by independent academic or non-profit entities, not affiliated with VLOPs or VLOSEs, it relies primarily on secondary data, i.e., digital traces of user activity, rather than specific access to internal platform mechanisms.

Researchers are consequently unable to account for one of the most influential latent variables in online behavior: algorithmic curation.

While CSS can observe and quantify phenomena such as the spread of misinformation, the formation of echo chambers, and the polarisation of opinions, it cannot fully explain these dynamics without insight into the platforms' recommendation systems. These opaque, proprietary algorithms shape what users see, engage with, and believe. Despite this, they are generally inaccessible – a black box that continues to obscure the full picture of how online pollution operates.

To partially address these limitations, new research is beginning to reshape data-driven modeling by integrating generative AI tools to design complex Social Digital Twins (SDTs), such as YSocial (https://ysocialtwin.github.io/ - see image). These platforms are designed to implement complex "what if" scenarios, simulating the specific dynamics of VLOPs in terms of both content and social dynamics. SDTs leverage AI-powered agents (characterized by personae integrating demographic and psychological traits) to simulate the activities of online social media users interacting in environments where the content recommendation and other latent variables can be made explicit by the analyst. This thus enables the design of safe and controllable proxies that generate synthetic data, perform in vitro analyses, and test the effectiveness of mitigation strategies. The result is the formulation of actionable insights.

Conclusions

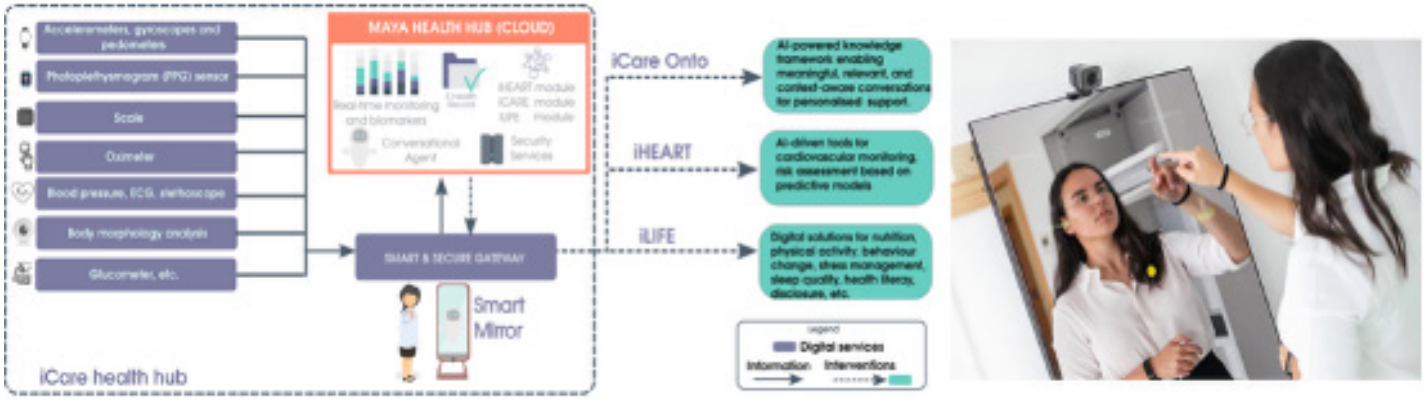
In a digital landscape increasingly shaped by opaque algorithms and fragmented governance, tackling online pollution entails more than isolated interventions. It requires coordinated efforts across disciplines, sectors, and borders to ensure transparency, accountability, and data access. Only by bridging the gap between models and real-world mechanisms – and between research and policymaking – can we begin to understand, mitigate, and ultimately reclaim the integrity of our shared information spaces.

For more information:
https://arxiv.org/abs/2403.14298
https://arxiv.org/abs/2408.00818

Contact: Giulio Rossetti, KDD Lab
giulio.rossetti@isti.cnr.it

MAYA

smart Mirrors supporting healthier lives of Adolescents and Young Adults after cancer
Funded by Horizon Europe



The MAYA Project was established to develop, test and scale up new and holistic solutions to address the long-term effects of cancer treatment on adolescent and young adult (AYA) cancer survivors, regardless of the cancer type. The aim is to enable AYA survivors to proactively manage their cardiovascular (CV) health using digital technologies that address the cardiotoxicity-related late effects of cancer therapies.

MAYA will consist of an iCARE health hub, which is a digital ecosystem comprising a smart mirror and an AI-powered conversational agent. This innovative technology will offer personalised care and continuous health monitoring. The primary aim is to enhance the overall quality of life for AYA survivors by addressing modifiable risk factors, such as hypertension, diabetes and obesity. This should result in a projected 30-40% reduction in the risk of major cardiac events.

To achieve these outcomes, MAYA will use AI-driven data analytics, real-time biomarker tracking and participatory research approaches, including social innovation, multi-stakeholder engagement and living labs. These solutions will be evaluated through clinical trials in real-world settings across multiple European countries.

In order to analyse facial and body morphology, the SILab research team will contribute to the project by applying advanced computer vision and artificial intelligence techniques, building on the successful SEMEOTICONS and VIRTUOSO projects,. The objective of these analyses is to identify correlations with key CV risk factors.

By focusing on personalised interventions and continuous health monitoring, MAYA is closely aligned with the objectives of the Horizon Europe programme and the Cancer Mission, and on improving the long-term health outcomes and quality of life of cancer survivors.

Contact: Sara Colantonio, SI Lab
sara.colantonio@isti.cnr.it

GUARDIANS

Green, Utility, Accessibility, Resilience, Digital Integration, Ability, Nature, Sustainability
Funded by Regione Toscana PR-FESR 2021-27

The aim of GUARDIANS is to develop good practices in the building industry through the integrated design of the architectural envelope and the building automation and control systems (BACS). This will enable designers to maximize the energy performance of buildings and the well-being of the occupants, while minimizing the emission of noxious and climate-altering gases by using both passive measures and active, smart technologies.

The European EPBD Directive $^{2018}/_{844}$/EU), has introduced the Smart Readiness Indicator (SRI), which measures the ability of a building to optimize the management of all the technological systems using intelligent services. Two buildings will be taken as real case studies and possibly retrofitted or redesigned in order to optimize their response to the needs of energy savings and flexibility and the needs of the occupants as well. The new SRI ranking thus obtained will provide an objective measurement of the project's success, as well as an index to evaluate the effectiveness of the interventions and the procedures involved. The results should benefit both commercial/industrial and research partners, and also foster innovation in the traditional building market.

GUARDIANS also entails a thorough study of the SRI and the criteria that drive its computation. The whole computation scheme adopted to date by the European Commission will be analyzed using a multicriteria assessment method to validate its effectiveness in our local geographical and socio-economic environment. The results should demonstrate the efficacy of adding intelligent technology to improve the global performance of buildings and also provide the project partners with innovative know-how.

Contact: Emanuele Salerno, SI Lab
emanuele.salerno@isti.cnr.it
GiuseppeFusco, SI Lab
giuseppe.fusco@isti.cnr.it

SAGE

Data Space for a Sustainable Green Europe
Funded by Horizon Europe

SAGE will establish a fully operational Green Deal Data Space (GDDS) aimed at increasing the accessibility, integration, and use of green and environmental data across the EU: This will then support the key pillars of the European Green Deal—Zero Pollution, Climate Adaptation, Biodiversity, and the Circular Economy Action Plan. Building directly on the GDDS GREAT project community and results, and aligned with the Digital Europe Programme's strategic focus on AI, cybersecurity, advanced computing, and data infrastructure, SAGE will leverage outcomes from the European Strategy for Data and research facilitated by Horizon Europe. SAGE's outcomes include seamlessly integrating fragmented environmental data through federation, enriching data with consistent quality, validation, and interoperable metadata, and strengthening capabilities for data transformation, processing, analysis, forecasting, target setting, and performance monitoring. SAGE aspires to foster informed decision-making and policy formulation based on robust data and evidence. The beneficiaries will be businesses seeking compliance with EGD regulations, government bodies seeking to optimize environmental impacts, citizens and citizen scientists engaged in environmental stewardship, and researchers aiming to deepen our understanding of the environment.

Co-funded by Digital Europe and project partners, it builds on GREAT's groundwork, using federated datasets, AI-driven analytics, and FAIR-compliant tools to empower businesses, governments, researchers, and citizens. By 2028, SAGE will deploy 10 real-world pilots and establish a sustainable company to scale the GDDS, ensuring data-driven decisions for Europe's green transition.

Contact: Antonino Crivello, WN Lab
antonino.crivello@isti.cnr.it
Davide La Rosa, WN Lab
davide.larosa@isti.cnr.it

IAIA

Innovazione Aperta Transfrontaliera: Connessioni per Startups e MPMI attraverso tool di Intelligenza Artificiale / Cross-border Open Innovation: Connections for Startups and MSMEs through Artificial Intelligence Tools

Funded by INTERREG VI A Italia-Francia “Marittimo” 2021-2027

The main aim of the IAIA (**I**nnovazione **A**perta Transfrontaliera: Connessioni per Startups e MPMI attraverso tool di **I**ntelligenza **A**rtificiale) project is to increase the competitiveness of startups and micro, small, and medium-sized enterprises (MSMEs) through cross-border collaborations that promote open innovation and participation in public calls, as well as the introduction of eco-innovations. This will be achieved through an understanding of the skills and needs of businesses operating in the five regions involved (Tuscany, Sardinia, Corsica, Liguria, and Provence-Alpes-Côte d’Azur). In addition, a matchmaking platform will be created between businesses and public calls, using AI and machine learning tools.

ISTI brings crucial added value to the project through its expertise in AI and machine learning. This will lead to the development of advanced technological tools that facilitate innovation and business collaboration across regional borders. Along with the KDD Laboratory, ISTI will develop innovative AI tools in data analysis . These analyses will identify potential opportunities for collaboration and common areas of interest between startups and MSMEs, using machine learning and data mining techniques. The study will also highlight the interconnections (both real and potential) between companies in the participating regions, and thus make the information searchable and navigable on an online portal supported by the European research infrastructure SoBigData. A tool will be developed in order to match MSMEs and public calls for proposals, in terms of the demand for products, services, and technologies required by public calls for proposals, and the supply from businesses

In conclusion, IAIA uses advanced technologies to promote collaboration and innovation among businesses in cross-border regions, with a significant positive impact on economic development, environmental sustainability, and regional competitiveness.

Contact: Valerio Grossi, KDD Lab
valerio.grossi@isti.cnr.it
https://interreg-marittimo.eu/it/web/iaia

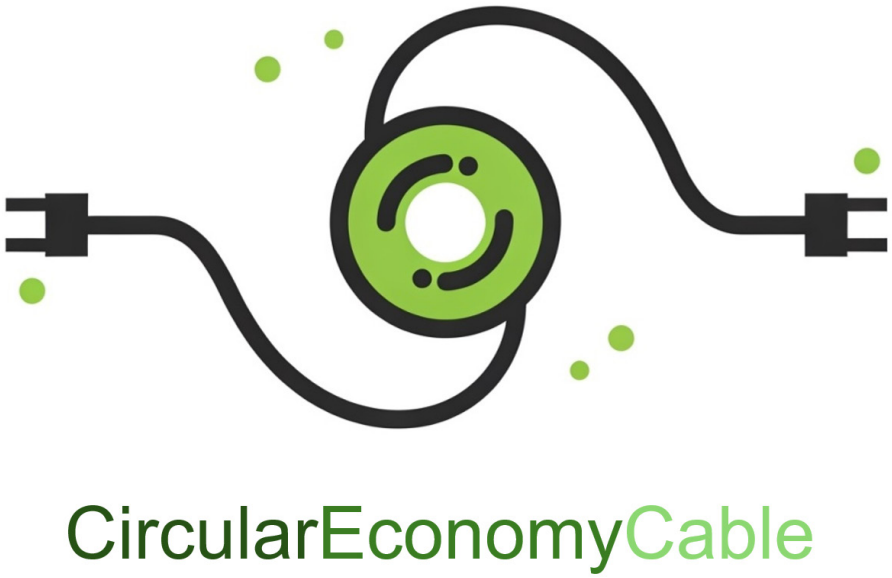
CircularEconomyCable

Development and prototyping of electrical and telecommunication cables with high circularity, traceability, and visibility, in line with the circular economy and digital transition in the energy transport and data connectivity sectors.
Funded by Regione Toscana PR-FESR 2021-27

Circular Economy Cable project entails designing and prototyping electric and telecommunication cables with innovative functionalities, high circularity, robust traceability, and improved visibility. Our approach fully embraces the circular economy and digital transition and should thus help to revolutionize the energy and data connectivity sectors.

The focus of the project is to develop advanced cable sheaths using bio-based polymers and recycled plastics. The aim is to significantly improve environmental sustainability in both production and end-of-life management, while also bolstering resilience in dealing with the challenges of material sourcing. By integrating sustainable additives and advanced nanomaterials, these new cables are also expected to deliver superior performance, including increased strength and greater fire resistance

Key innovations include the incorporation of high-visibility compounds (using reflectance, fluorescence, and phosphorescence) within the cables. This will increase safety and efficiency during installation, maintenance, and testing, thereby reducing potential network downtime.



The project will exploit cutting-edge techniques, including AI, design-of-experiments (DoE), and “in silico” simulations. These methods will streamline the selection of promising physical experiments and accelerate the convergence with innovative solutions, while also optimizing production processes in order to reduce energy consumption.

The CNR-ISTI unit will contribute to the project by developing machine learning methodologies for polymer science, refining DoE techniques, and predictive models for inferring the properties of composite materials. This will accelerating the discovery of new and eco-sustainable materials.

Contact: Davide Moroni, SI Lab
davide.moroni@isti.cnr.it
https://www.isti.cnr.it/en/research/project-detail/14652/CircularEconomyCable_CircularEconomyCable

DIGITAfrica

Towards a comprehensive pan-African research infrastructure in Digital Sciences
Funded by Horizon Europe

DIGITAfrica is focused on developing a sustainable, pan-African research infrastructure (RI) in Digital Sciences, with a strong collaboration between African and European institutions. Addressing the significant digital divide across the African continent, the project aims to empower African nations through greater research capacity, education, and technological development in areas such as AI, big data, cloud computing, and IoT.

By leveraging the expertise of leading European RIs such as SLICES and SoBigData, DIGITAfrica will co-create a strategic framework for digital RI development tailored to African needs. The project's objectives include community building, strengthening capacity, and fostering inclusivity—especially for young and female researchers. A key output will be the DIGITAfrica BluePrint, a proof-of-concept cooperation platform designed to support collaborative research, education, and innovation across Africa.

DIGITAfrica will be made up of three phases: (1) identifying needs and constraints through stakeholder engagement, (2) developing a design study, training portfolio, and technical playground, and (3) creating a sustainable business and governance model.

The initiative promotes open science and adheres to FAIR and FACT principles to ensure accessibility, transparency, and responsible data use.

Spanning five African countries and guided by an experienced consortium, DIGITAfrica aspires to serve as a cornerstone for Euro-African cooperation in digital R&I, with a significant contribution to Africa's digital transformation, socio-economic development, and progress toward the achievement of UN Sustainable Development Goals.

Contact: Roberto Trasarti, KDD Lab
roberto.trasarti@isti.cnr.it
https://www.digitafrica.eu

FabDesign

FABrication-aware DESIGN based on advanced geometry processing
Funded by MUR

The FabDesign project aims to boost Italy's manufacturing sector through advanced computational fabrication and computer-assisted design technologies. Led by Paolo Cignoni of the Italian National Research Council in partnership with Riccardo Scateni of the University of Cagliari, the project is funded under the PRIN 2022 program by the Ministry of University and Research (MUR) with a duration of 24 months starting from February 2025.

The primary objective of the FabDesign Project is to improve the synergy between design and manufacturing processes. This will be achieved by leveraging state-of-the-art geometric modeling research to advance digital fabrication techniques, including CNC milling, molding/casting, and 3D printing. These methods are essential for producing a wide range of products, from customized items to robust mechanical components and mass-produced goods.

A central challenge in digital fabrication is ensuring that the "machining tool" can access the entire surface of a 3D object without obstruction. FabDesign addresses this issue by developing innovative approaches to object segmentation, surface and volume discretization, and robust numerical modeling. These advancements help create more complex and precise designs that are easier to manufacture.

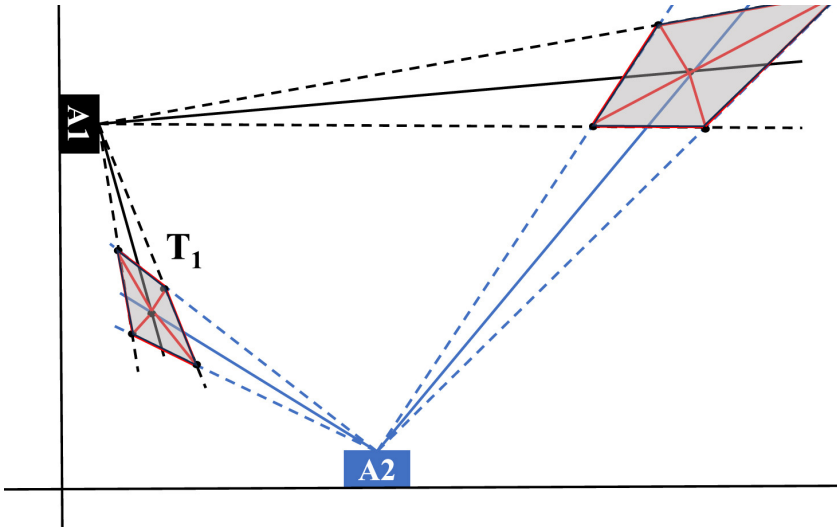
The project is expected to yield significant scientific and technological advancements, particularly in the fields of computational geometry and digital fabrication. Economically, the project aims to support SMEs in Italy by improving access to advanced manufacturing technologies. Socially, the project will engage the public through events and example applications of the techniques in the field of cultural heritage showcasing how advanced digital fabrication can help preservation and promotion of cultural artifacts.

The project's findings will be disseminated through publications in top-tier journals, presentations at influential conferences, and the distribution of open-source software prototypes. This will ensure that the project's outcomes are widely accessible and can be used by both the academic community and industry practitioners.

Contact: Paolo Cignoni, VC Lab
paolo.cignoni@isti.cnr.it

Indoor Localization Algorithms Based on Angle of Arrival with a Benchmark Comparison

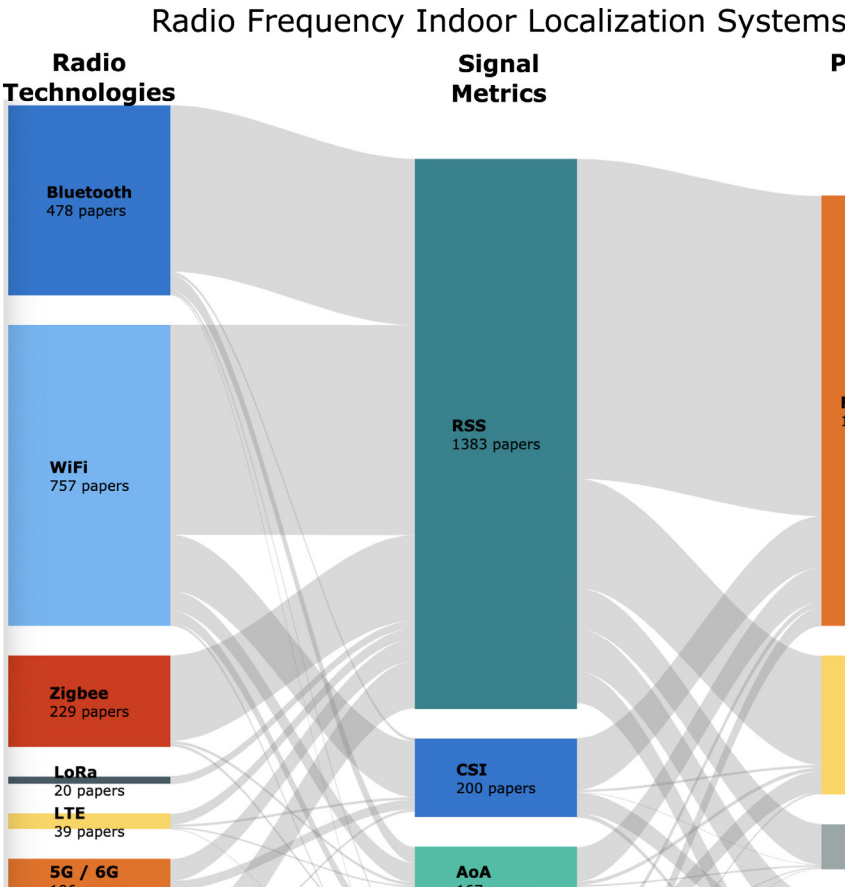
F. Furfari, M. Girolami, F. Mavilia, P. Barsocchi
Ad Hoc Networks, vol. 166. Elsevier, 2025.



Red polygons show a graphical representation of the Confidence Region obtained by intersecting anchor's boundary directions.

Indoor localization is crucial for developing intelligent environments capable of understanding user contexts and adapting to environmental changes. Bluetooth 5.1 Direction Finding is a recent specification that leverages the angle of departure (AoD) and angle of arrival (AoA) of radio signals to locate objects or people indoors. This paper presents a set of algorithms that estimate user positions using AoA values and the concept of the Confidence Region (CR), which defines the expected position uncertainty and helps to remove outlier measurements, thereby improving performance compared to traditional triangulation algorithms. We validate the algorithms with a publicly available dataset, and analyze the impact of body orientation relative to receiving units. The experimental results highlight the limitations and potential of the proposed solutions. From our experiments, we observe that the Conditional All-in algorithm presented in this work, achieves the best performance across all configuration settings in both line-of-sight and non-line-of-sight conditions.

DOI: 10.1016/j.adhoc.2024.103691

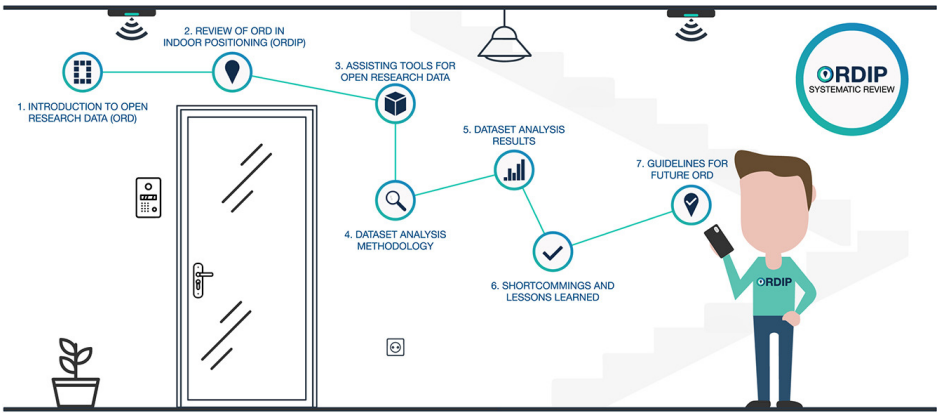


Alluvial diagram showing the relationship between Radio Technologies, Signal Metrics and Positioning Methods (DFL stands for Device-Free Localization). Data are extracted from Scopus digital library.

ORDIP: Principle, Practice and Guidelines for Open Research Data in Indoor Positioning

G.G. Anagnostopoulos, P. Barsocchi, A. Crivello, C. Pendão, I. Silva, J. Torres-Sospedra

Internet of Things, vol. 30. Elsevier, 2025.



Graphical abstract.

The community of indoor positioning research has identified the need for a paradigm shift towards more reproducible and open research dissemination. Despite recent efforts to openly share data and code, accompanying research results with Open Research Data (ORD) is far from being the de facto standard option for publications in the indoor positioning field. The lack of recognized public benchmarks and the rather slow adoption of ORD, set a great volume of astute contributions in the field to remain irreproducible. Performance comparisons may often be made on experiments performed in different settings, hindering their consistency, and eventually slowing down progress and the evolution of knowledge in the field. In this work, we systematically review the landscape of Open Research Data in Indoor Positioning, enlisting, presenting, and analyzing the characteristic features of the relevant available open datasets of the field. As a result of our systematic review, the statistical analysis of the 119 identified open datasets, highlights the tendencies and the missing elements, such as underrepresented technologies (such as Ultra-Wideband) and measurement types (such as Angle of Arrival, Time Difference of Arrival). A result that stands out is the frequency of crucial metadata information that remains undefined, such as the size of the area of collection (50% of the datasets), the ground truth collection protocol (21%), or the environment type (13%). As a fruit of the systematic analysis, we discuss potential shortcomings, and we share lessons learned and observed good practices regarding the provision of a new ORD and the reuse of existing ones. A significant practical contribution of this work is a list of guidelines that researchers aiming to collect and share a new ORD can follow as a simple checklist. In a broader context, we consider that ORDIP can help measure the future progress of the Indoor Positioning field in the ORD front through the snapshot of the current landscape that it provides. The Open provision of our full systematic analysis of the ORD can serve as a look-up table for easy access to the ORDs containing the most relevant features for each interested researcher, while our guidelines aim to support the community and spark the discussion towards a consensus-based standard for ORD of the field.

Content Diagram.

DOI: 10.1016/j.iot.2024.101485

Evaluating the Understandability and User Acceptance of Attack-Defense Trees: Original Experiment and Replication

G. Broccia, M.H. ter Beek, A. Lluch Lafuente, P. Spoletini, A. Fantechi, A. Ferrari

Information and Software Technology, vol. 178. Elsevier, 2025.

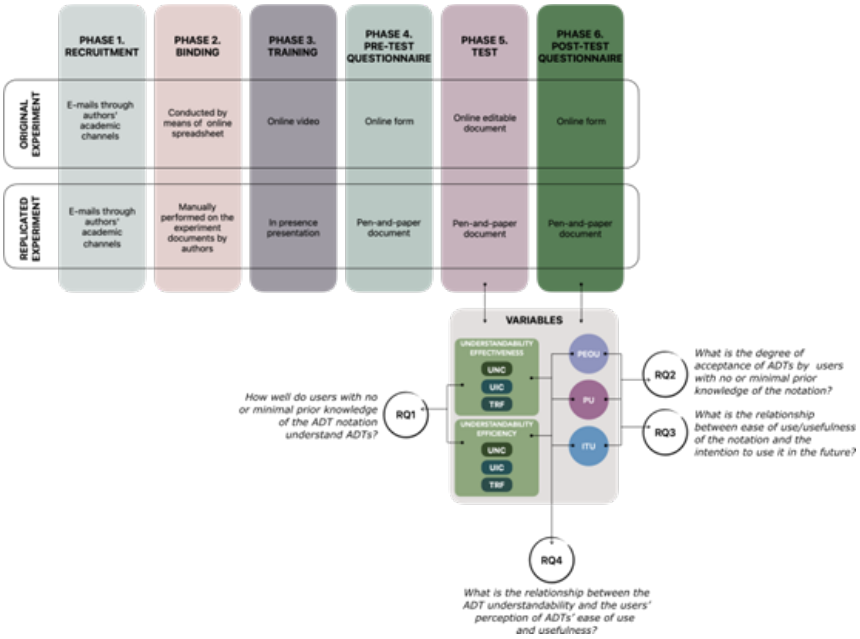


Diagram illustrating the phases of the experiment in both the original and replicated experiment, the measured variables, and how these variables address the research questions.

Context: Attack-Defense Trees (ADTs) are a graphical notation used to model and evaluate security requirements. ADTs are popular because they facilitate communication among different stakeholders involved in system security evaluation and are formal enough to be verified using methods like model checking. The understandability and user-friendliness of ADTs are claimed as key factors in their success, but these aspects, along with user acceptance, have not been evaluated empirically.

Objectives: This paper presents an experiment with 25 subjects designed to assess the understandability and user acceptance of the ADT notation, along with an internal replication involving 49 subjects.

Methods: The experiments adapt the Method Evaluation Model (MEM) to examine understandability variables (i.e., effectiveness and efficiency in using ADTs) and user acceptance variables (i.e., ease of use, usefulness, and intention to use). The MEM is also used to evaluate the relationships between these dimensions. In addition, a comparative analysis of the results of the two experiments is carried out.

Results: With some minor differences, the outcomes of the two experiments are aligned. The results demonstrate that ADTs are well understood by participants, with values of understandability variables significantly above established thresholds. They are also highly appreciated, particularly for their ease of use. The results also show that users who are more effective in using the notation tend to evaluate it better in terms of usefulness.

Conclusion: These studies provide empirical evidence supporting both the understandability and perceived acceptance of ADTs, thus encouraging further adoption of the notation in industrial contexts, and development of supporting tools.

DOI: 10.1016/j.infsof.2024.107624

Analysing Self-Adaptive Systems as Software Product Lines

J. Päßler, M.H. ter Beek, F. Damiani, E.B. Johnsen, S.L. Tapia Tarifa

Journal of Systems and Software, vol. 222. Elsevier, 2025.

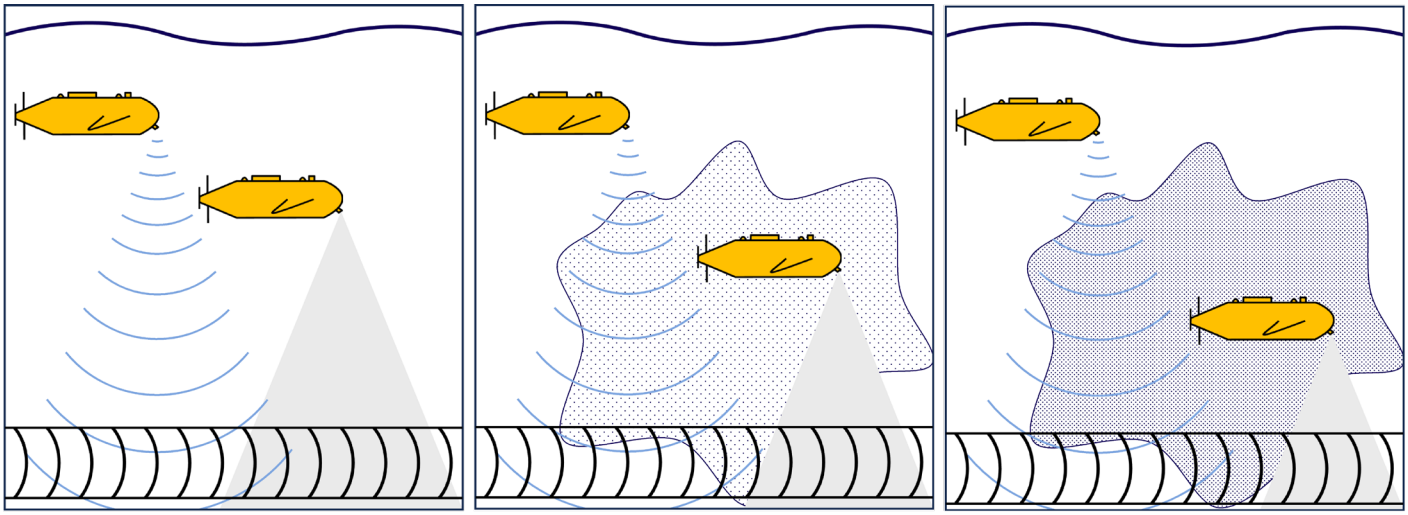


Illustration of the maximum altitudes in different water visibilities when operating a pipeline inspection with an AUV using either the (side-scan) sonar (left AUV in each image) or the camera (right AUV in each image). The left image shows good visibility, the middle image shows medium visibility and the right image shows bad visibility.

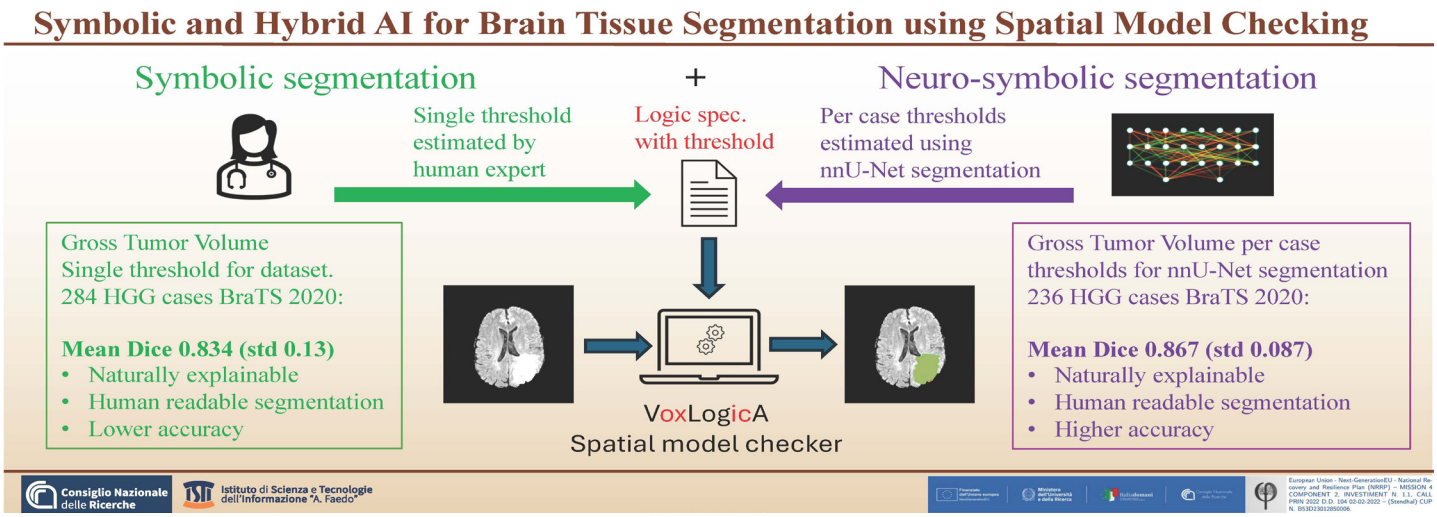
Self-adaptation is a crucial feature of autonomous systems that must cope with uncertainties in, e.g., their environment and their internal state. Self-adaptive systems (SASs) can be realised as two-layered systems, introducing a separation of concerns between the domain-specific functionalities of the system (the managed subsystem) and the adaptation logic (the managing subsystem), i.e., introducing an external feedback loop for managing adaptation in the system. We present an approach to model SASs as dynamic software product lines (SPLs) and leverage existing approaches to SPL-based analysis for the analysis of SASs. To do so, the functionalities of the SAS are modelled in a feature model, capturing the SAS's variability. This allows us to model the managed subsystem of the SAS as a family of systems, where each family member corresponds to a valid feature configuration of the SAS. Thus, the managed subsystem of an SAS is modelled as an SPL model; more precisely, a probabilistic featured transition system. The managing subsystem of an SAS is modelled as a control layer capable of dynamically switching between these valid configurations, depending on both environmental and internal conditions. We demonstrate the approach on a small-scale evaluation of a self-adaptive autonomous underwater vehicle (AUV) used for pipeline inspection, which we model and analyse with the feature-aware probabilistic model checker ProFeat. The approach allows us to analyse probabilistic reward and safety properties for the SAS, as well as the correctness of its adaptation logic.

DOI: 10.1016/j.jss.2024.112324

Symbolic and Hybrid AI for Brain Tissue Segmentation using Spatial Model Checking

G. Belmonte, V. Ciancia, M. Massink

Artificial Intelligence in Medicine, early access. Elsevier, 2025.



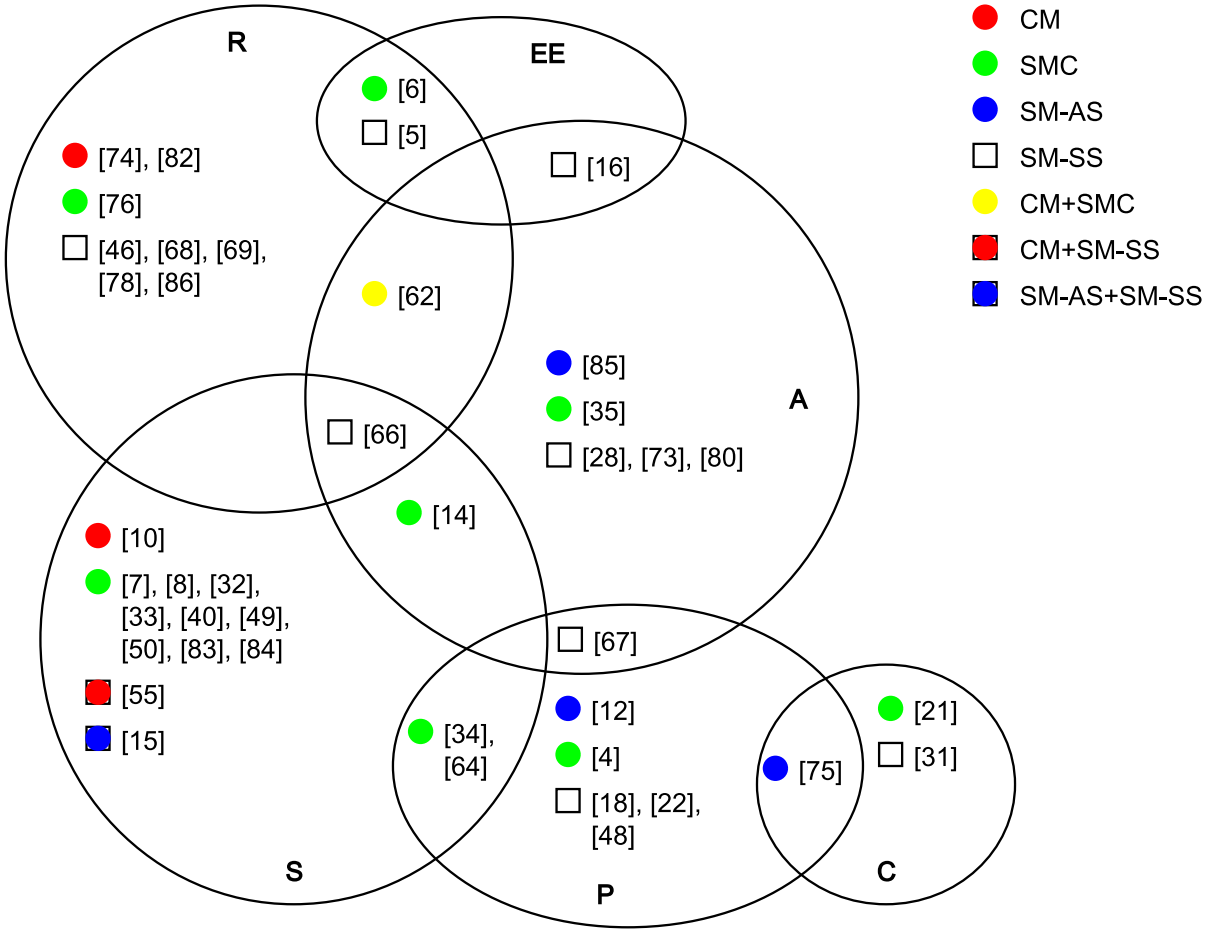
Comparison of the quality of the analysis on a public dataset of medical MRI images of brain lesions (BraTS 2020) with a pure symbolic approach (left), using our spatial model checker VoxLogicA, and with a hybrid approach, using VoxLogicA in combination with a neural network.

Segmentation of 3D medical images, and brain segmentation in particular, is an important topic in neuroimaging and in radiotherapy. Overcoming the current, time consuming, practise of manual delineation of brain tumours and providing an accurate, explainable, and replicable method of segmentation of the tumour area and related tissues is therefore an open research challenge. In this paper, we first propose a novel symbolic approach to brain segmentation and delineation of brain lesions based on spatial model checking. This method has its foundations in the theory of closure spaces, a generalisation of topological spaces, and spatial logics. At its core is a high-level declarative logic language for image analysis, ImgQL, and an efficient spatial model checker, VoxLogicA, exploiting state-of-the-art image analysis libraries in its model checking algorithm. We then illustrate how this technique can be combined with Machine Learning techniques leading to a hybrid AI approach that provides accurate and explainable segmentation results. We show the results of the application of the symbolic approach on several public datasets with 3D magnetic resonance (MR) images. Three datasets are provided by the 2017, 2019 and 2020 international MICCAI BraTS Challenges with 210, 259 and 293 MR images, respectively, and the fourth is the BrainWeb dataset with 20 (synthetic) 3D patient images of the normal brain. We then apply the hybrid AI method to the BraTS 2020 training set. Our segmentation results are shown to be in line with the state-of-the-art with respect to other recent approaches, both from the accuracy point of view as well as from the view of computational efficiency, but with the advantage of them being explainable.

DOI: 10.1016/j.artmed.2025.103154

Quantitative Dependability Evaluation of Train Control Systems in Presence of Uncertainty: a Systematic Literature Review

L. Carnevali, F. Di Giandomenico, A. Fantechi, S. Gnesi, G. Gori.
IEEE Transactions on Intelligent Transportation Systems, vol. 26. IEEE, 2025.



Venn diagram: studies grouped by measures of interest and distinguished by adopted quantitative evaluation methods.

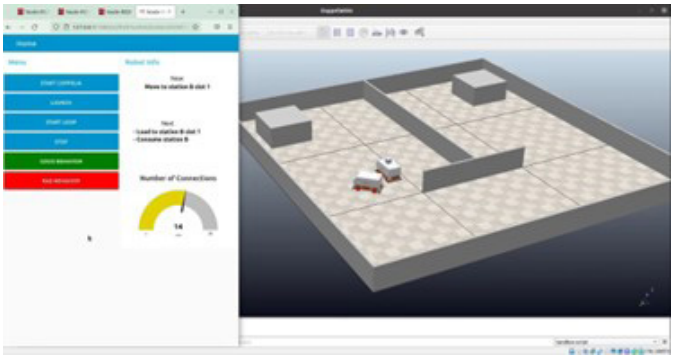
Technological advances in modern Train Control Systems (TCSs) promise to improve dependability of railway transportation in terms of safety, availability, and capacity, notably by employing novel distancing policies such as Moving Block (MB) signaling and Virtual Coupling (VC), fueled by advanced train localization methods such as satellite positioning.

At the same time, these technological advances raise notable concerns about the effects that uncertainty in critical TCS parameters (such as train position and speed) may have on dependability-related attributes. Recently, various approaches have been proposed to characterize such effects through quantitative measures, leveraging formal stochastic modeling and evaluation of the TCS behavior. In this paper, we illustrate the results of a systematic review of the literature on quantitative evaluation of dependability-related attributes of TCSs under uncertainty on vital parameters. Specifically, we have finally selected 42 relevant papers, published between 2011 and 2023, that succeed in giving, through an empirical perspective and classification, a comprehensive view of current research and practice in quantitative dependability assessment of TCSs.

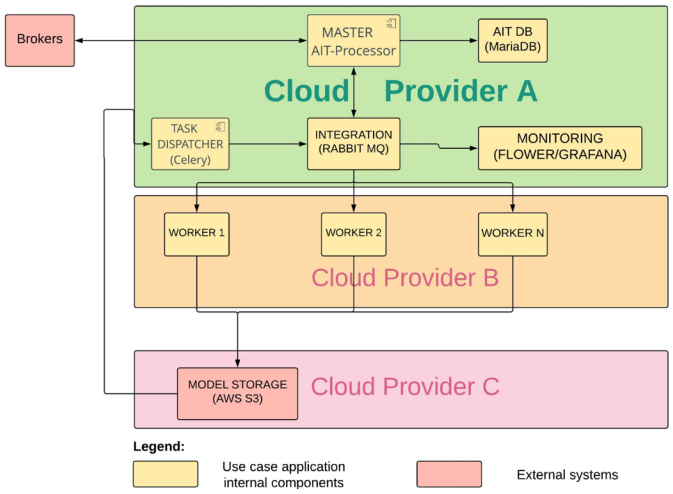
DOI: 10.1109/TITS.2025.3530112

CONCERN: A Model-Based Monitoring Infrastructure

A. Calabrò, E. Marchetti, P. Skrzypek, J. Marchel
Internet of Things, vol. 31. Elsevier, 2025.



CoppeliaSim simulation.



AI Investments architecture overview.

This paper presents CONCERN (COmplex eveNt proCEssing monitoR iNfrastructure), a flexible and extensible model-based infrastructure for runtime monitoring of complex, distributed systems. In response to growing demands for adaptive and non-intrusive monitoring in service-oriented and event-driven architectures, CONCERN decouples monitoring concerns from system logic, enabling the detection of runtime violations, anomalies, and service degradations without requiring architectural modifications to the observed system. It integrates complex event processing (CEP) with formal state-machine-based specifications, allowing the infrastructure to react to high-level patterns detected from streams of low-level system events. This model-driven approach facilitates runtime reasoning over system behavior, enabling dynamic adaptation of monitoring logic based on contextual needs. Within the paper, a detailed comparison between CONCERN and existing model-based monitoring infrastructures is provided. It outlines how CONCERN differs from state-of-the-art solutions by emphasizing monitoring logic reusability, dynamic reconfiguration, and non-intrusiveness. CONCERN offers a modular architecture in which monitors are defined separately from the system logic and deployed externally. In the paper, the validation of the infrastructure through two real-world case studies in the financial sector and robotics are provided while an in-vitro evaluation related to performances in cooperating with Digital Twins have been executed. Authors apply CONCERN to monitor financial transactions and compliance violations within a microservice-based platform. This evaluation is used to compare CONCERN against conventional solutions in terms of runtime performance, expressiveness, and adaptability. The results show that CONCERN outperforms existing frameworks by offering low-latency detection with minimal overhead, while also supporting runtime reconfiguration and the specification of rich behavioral properties. In conclusion, CONCERN provides a scalable, adaptive, and domain-agnostic runtime monitoring solution that meets the challenges of modern distributed systems. It supports formal, reusable, and configurable monitoring specifications, making it suitable for critical applications in domains such as finance, health care, and logistics. By decoupling monitoring from business logic, CONCERN enables continuous assurance of service-level properties with enhanced maintainability and observability.

DOI: 10.1016/j.iot.2025.101557

Training-Free Sparse Representations of Dense Vectors for Scalable Information Retrieval

F. Carrara, L. Vadicamo, G. Amato, C. Gennaro

Information Systems, early access. Elsevier, 2025.

In this paper, we propose and analyze Vec-2Doc, a novel training-free method to transform dense vectors into sparse integer vectors, facilitating the use of inverted indexes for information retrieval (IR). The exponential growth of deep learning and artificial intelligence has revolutionized scientific problem-solving in areas such as computer vision, natural language processing, and automatic content generation. These advances have also significantly impacted IR, with a better understanding of natural language and multimodal content analysis leading to more accurate information retrieval. Despite these developments, modern IR relies primarily on the similarity evaluation of dense vectors from the latent spaces of deep neural networks. This dependence introduces substantial challenges in performing similarity searches on large collections containing billions of vectors. Traditional IR methods, which employ inverted indexes and vector space models, are adept at handling sparse vectors but do not work well with dense ones. Vec2Doc attempts to fill this gap by converting dense vectors into a format compatible with conventional inverted index techniques. Our preliminary experimental evaluations show that Vec2Doc is a promising solution to overcome the scalability problems inherent in vector-based IR, offering an alternative method for efficient and accurate large-scale information retrieval.

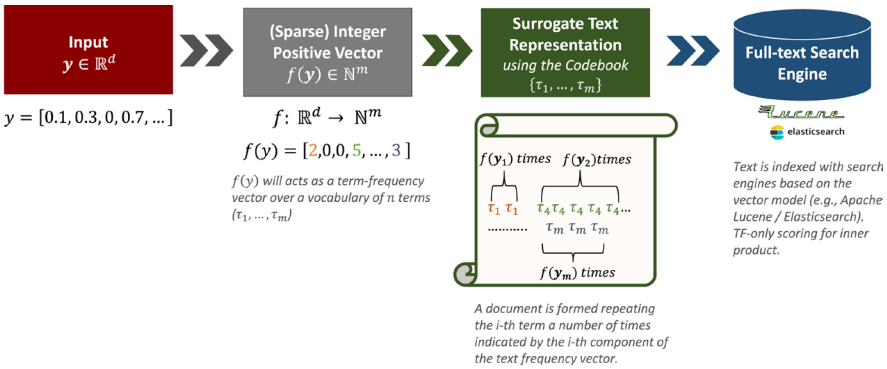


Illustration depicting the fundamental steps of an STR technique transforming a dense vector into a text representation suitable for indexing, leveraging a full-text search engine like Lucene.

DOI: 10.1016/j.is.2025.102567

A Simple Method for Classifier Accuracy Prediction under Prior Probability Shift

L. Volpi, A. Moreo, F. Sebastiani

Discovery Science 2024. Lecture Notes in Computer Science, vol. 15244. Springer 2025.

The standard technique for predicting the accuracy that a classifier will have on unseen data (classifier accuracy prediction – CAP) is cross-validation (CV). However, CV relies on the assumption that the training data and the test data are sampled from the same distribution, an assumption that is often violated in many real-world scenarios. When such violations occur (i.e., in the presence of dataset shift), the estimates returned by CV are unreliable. In this paper we propose a CAP method specifically designed to address prior probability shift (PPS), an instance of dataset shift in which the training and test distributions are characterized by different class priors. By solving a system of independent linear equations, with n the number of classes, our method estimates the entries of the contingency table of the test data, and thus allows estimating any specific evaluation measure. Since a key step in this method involves predicting the class priors of the test data, we further observe a connection between our method and the field of "learning to quantify". Our experiments show that, when combined with state-of-the-art quantification techniques, under PPS our method tends to outperform existing CAP methods.

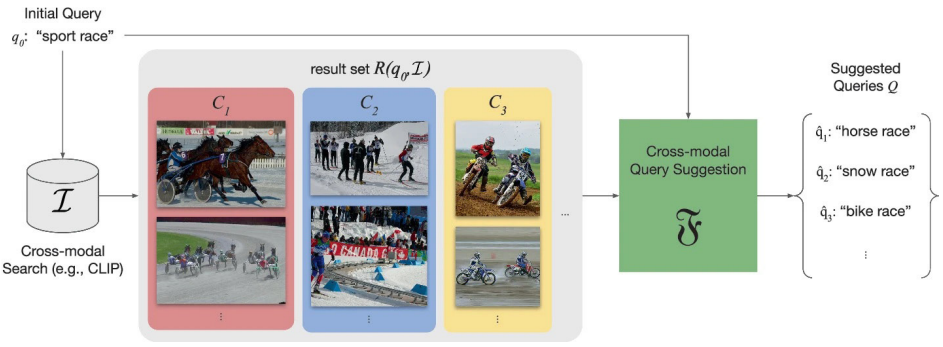
DOI: 10.1007/978-3-031-78980-9_17

Maybe You are Looking for CroQS Cross-Modal Query Suggestion for Text-to-Image Retrieval

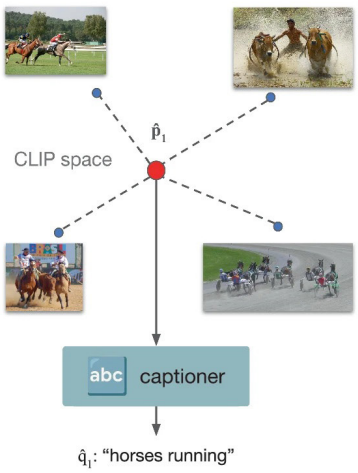
G. Pacini, F. Carrara, N. Messina, N. Tonellotto, G. Amato, F. Falchi

Advances in Information Retrieval. Lecture Notes in Computer Science, vol. 15573. Springer 2025.

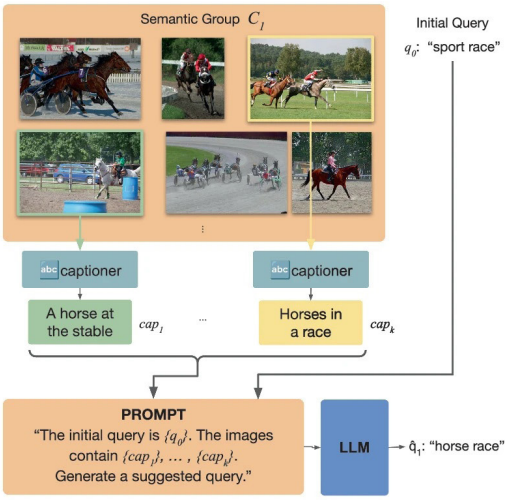
Query suggestion, a technique widely adopted in information retrieval, enhances system interactivity and the browsing experience of document collections. In cross-modal retrieval, many works have focused on retrieving relevant items from natural language queries, while few have explored query suggestion 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.



Cross-modal Query Suggestion. Given an initial query q_0 and an image collection I, a cross-modal query suggestion system $\mathfrak{F}$ returns a set of query suggestions $\mathcal{Q}$ based on the visual content of the result set $R(q_0, I)$. Ideally, each suggestion $q_i \in \mathcal{Q}$ should represent a semantically coherent group $C_i \subset R(q_0, I)$.



(a) **Prototype Captioning**: the suggested query is obtained by a captioning model (**ClipCap** or **DeCap**) applied to a cluster prototype in CLIP image space.



(b) **Captions Summarization**: an LLM receives q_0 together with the captions of the k most representative images and generates a query suggestion. We refer to this method as **GroupCap**.

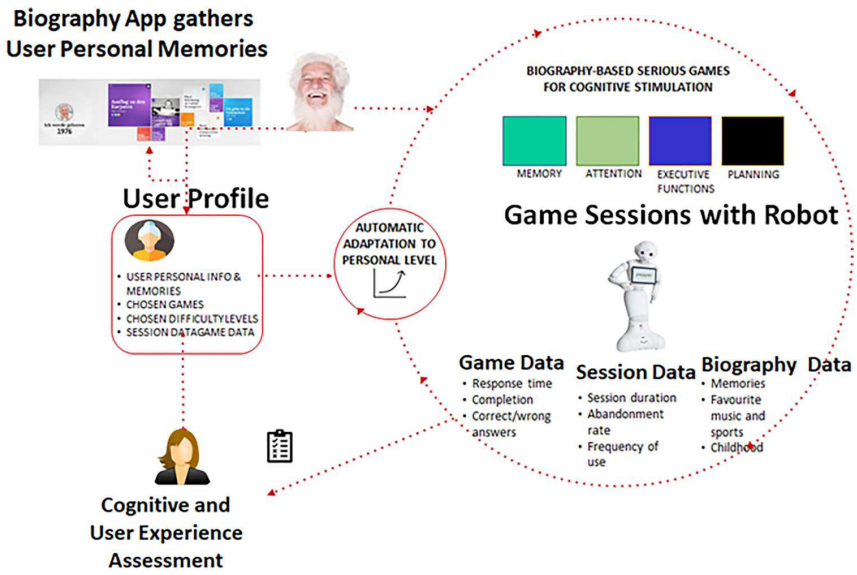
Architectures of the baseline methods proposed.

DOI: 10.1007/978-3-031-88711-6_9

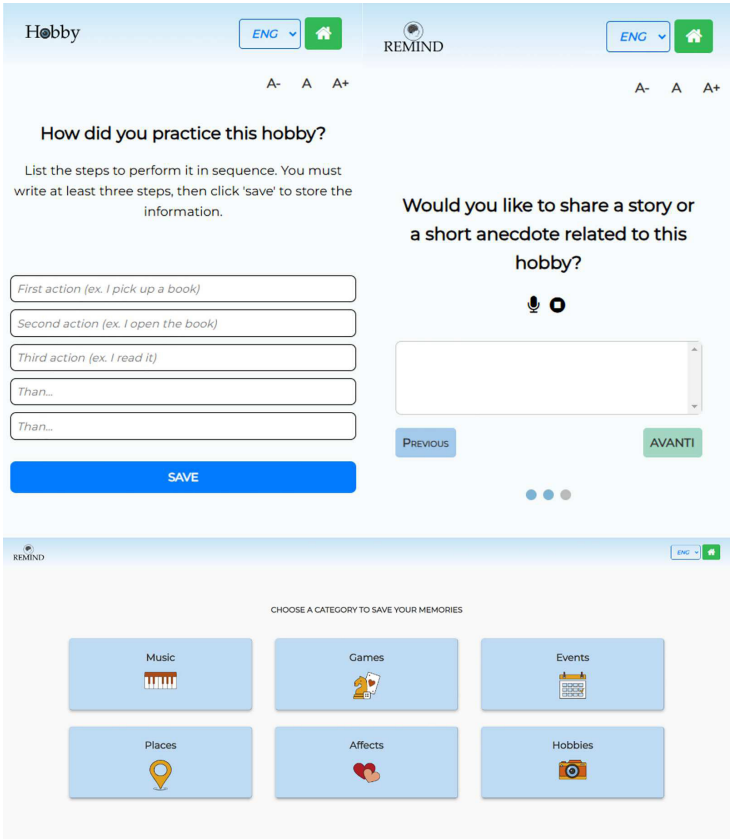
Exploiting Personal Memories in Humanoid Robot Serious Games for Mild Cognitive Impaired Older Adults

B. Catricalà, M. Manca, F. Paternò, A. Sale, C. Santoro, E. Zedda

Behaviour & Information Technology, early access. Taylor&Francis, 2025.



Overview of the personal memory-based platform.



The user interfaces for adding info on a hobby-related memory (above); the user interface for selecting a category (below).

The use of humanoid robots in older adult training has recently started to be considered. We investigated a solution that offers serious games personalised to each individual, to stimulate more interest and participation in cognitive training. In particular, we have studied how to consider personal memories in customising humanoid robot games for Mild Cognitive Impaired (MCI) older adults. For this goal, a prototype platform for collecting and exploiting personal memories in associated games is presented. The memories are exploited by six games designed and implemented in a Pepper robot considering current practices. We report on a mixed-method study consisting of a two-phase trial that involved 15 MCI older adults. The participants first furnished some memories from their past, and then used two game versions regularly for twelve weeks, one personalised and one with general content. We collected both quantitative (through questionnaires and interaction log analysis) and qualitative feedback. The results provide useful information about the robot games' impact on users and, more generally, for understanding how to introduce robot games based on personal memories in cognitive training programmes.

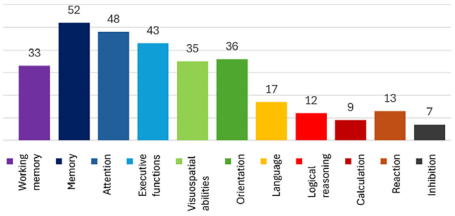
DOI: 10.1080/0144929X.2025.2484647

Interactive Serious Games for Cognitive Training of Older Adults: a Systematic Review

G. Canapa, F. Paternò, C. Santoro

IEEE Transactions on Computational Social Systems, early access. IEEE, 2025.

Current projections indicate that many countries worldwide are experiencing an increase in the older adult population. Thus, it is important to develop methodologies and tools to improve the quality of life of aging people in terms of their cognitive, social, and physical well-being. An area that is currently showing considerable potential in coping with cognitive decline associated with the aging process is digital serious games. Recent years have witnessed a significant attention increase toward this topic with contributions from several research communities. Although some previous work aimed to synthesize the main findings in related areas, a focused and systematic review categorizing and discussing the literature on cognitive training for older adults is still lacking. To this aim, we selected 113 relevant articles from four international databases. We analyzed them on different relevant dimensions, such as the interactive technologies employed in the games, the activities supported, the user profiles participating in the studies, the factors influencing user engagement and motivation, and the effects of training on the addressed cognitive functions. The analysis of the above dimensions is useful to uncover the factors that distinguish successful cognitive interventions from less effective ones and to facilitate the creation of engaging training programs tailored to the needs of older adults. At the same time, by highlighting promising avenues for further research, the review provides useful indications to researchers and practitioners interested in this field.



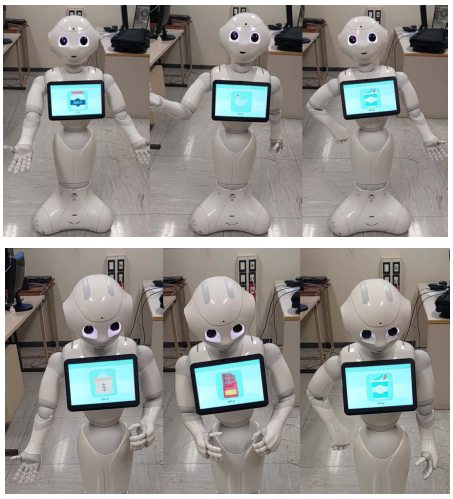
Cognitive functions addressed in the analyzed game activities.

DOI: 10.1109/TCSS.2025.3532307

Adaptive Humanoid Robot Behaviour in a Serious Game Scenario through Reinforcement Learning

E. Zedda, M. Manca, F. Paternò, C. Santoro

Behaviour & Information Technology, early access. Taylor&Francis, 2025.



An example of animations for extravert personality (Figure 1a), for introvert personality (Figure 1b).

The study presents an adaptive technique that enables a humanoid robot to select appropriate actions to maintain the engagement level of users while they play a serious game for cognitive training. The goal is to design and develop an adaptation strategy for changing the robot's behaviour based on Reinforcement Learning (RL) to encourage the user to remain engaged. Initially, we trained the algorithm in a simulated environment before moving on to a real user experiment. Thus, we first design, develop, and validate the RL strategy in a simulated environment. Subsequently, we integrate the trained policy into the robotic system, allowing it to select the best actions based on the detected user state during real user test. The RL algorithm was designed and implemented to determine an effective adaptation strategy for the robot's actions, encompassing verbal and non-verbal interactions. The proposed solution was first trained in a simulated environment and then tested with 28 users in a mixed-method design study.

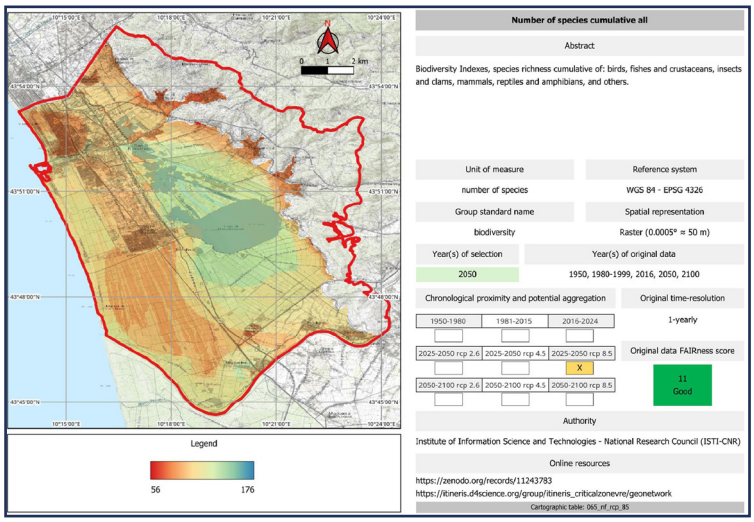
DOI: 10.1080/0144929X.2025.2456068

A FAIR and Open Geographic Data Collection for the Massaciuccoli Lake Basin Wetland in Italy

G. Vannini, P. Bove, G. Coro

Data in Brief, vol. 59. Elsevier, 2025.

The creation of a catalogue of geodata harmonised over time and space is essential for describing the status of ecosystem services in wetlands. In the present work, a specific methodology has been developed for the collection and generation of spatially and temporally harmonized geographic data to describe essential ecological and socio-economic charactesristics of the Massaciuccoli Lake basin (Tuscany, Italy), while providing a re-usable methodology for other areas. We developed a methodology, which we called 'Geodata Layers Harmonization Methodology' (GLHM), divided into four main phases: Geodata Census (GC), Geodata Selection (GS), Geodata Alignment (GA), and Geodata Publication (GP). The first phase, GC, involved a census of geodata made available online by public institutions, prioritizing those most relevant for describing ecosystem services, such as climatic, agro-environmental, pedo-geological, and biodiversity variables, with a preference for detailed data at the local level. The metadata of the collected geodata were organized into a structured tabular format. In the GS phase, geodata were selected based on a spatial resolution compatible with regional-scale ecological models (maximum 0.0005° ≈ 50 m), and a temporal coverage that could represent from remote past to far future scenarios. Geodata with partial spatial coverage or unsuitable for ecological models were excluded. Additionally, we evaluated the compliance of the geodata published on the websites of public institutions with the Findable-Accessible-Interoperable-Reusable (FAIR) principles through a newly developed scoring system. Based on this score, we selected only the data that exceeded a minimum FAIRness threshold. In the GA phase, the selected geodata were aligned semantically (i.e., by variable meaning), temporally, and spatially.



Example of cartographic table in our Atlas.

Each geodata was georeferenced using the WGS84/EPSG:4326 reference system and clipped to the boundaries of the Massaciuccoli Lake basin. Raster data were resampled to achieve a uniform spatial resolution of 0.0005°. In the last phase, GP, the aligned geodata were published on public access repositories and services: The entire collection was organized as a QGIS project with legends and a metadata table associated. An Atlas was also produced, in PDF format, which visually represented the data and metadata. The geodata and their corresponding legends were exposed through Web Map Service (WMS) and Web feature Service (WFS) standards on a GeoServer instance and catalogued in a GeoNetwork instance, compliant with the ISO19139 standard and the INSPIRE European Directive.

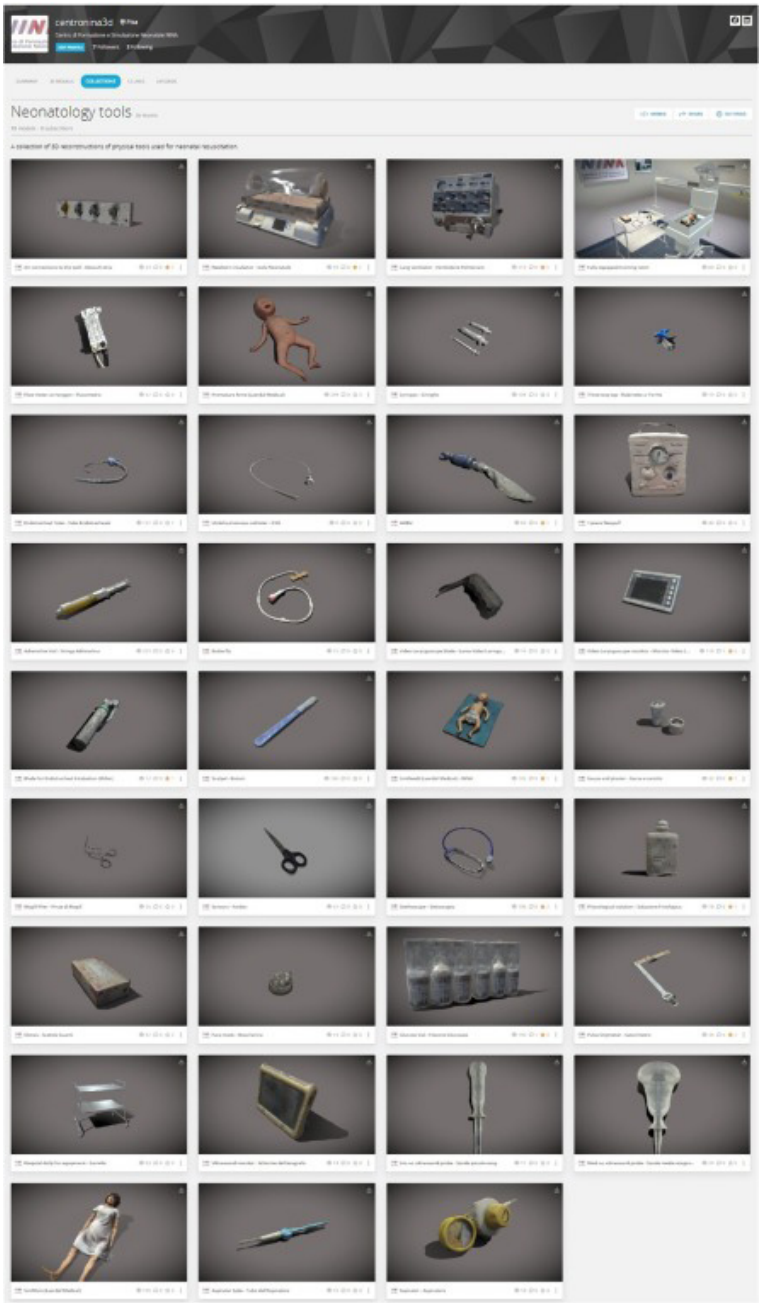
The collection contains 148 geo-datasets, representing 75 climatic, agro-environmental, pedo-geological, morphological, ecological, biological, and socio-economic information distributed across five temporal reference time frames: a remote past (1950–1980), a near past (1981–2015), the present (2016–2024), a near future (2025–2050), and a far future (2051–2100). Future projections are available under the Representative Concentration Pathways (RPC) 2.6, 4.5, and 8.5 to simulate low, medium, and high greenhouse gas concentration scenarios respectively.

The present geodata collection is particularly useful for wetland monitoring, management and planning. It can easily be integrated with ecological models and predictive studies to analyse the effects of climate change and anthropogenic pressures on wetlands. The GLHM methodology is applicable to other ecological contexts to create standardised structured frameworks for evaluating the status of the biodiversity and the ecosystem services and the interplay between anthropic pressures and the ecosystem response.

DOI: 10.1016/j.dib.2025.111303

An Open Data Collection of 3D Tool and Equipment Models for Neonatology

S. Bardelli, G. Coro, R.T. Scaramuzzo, M. Ciantelli, A. Cuttano
Results in Engineering, vol. 25. Elsevier, 2025.



Overview of the data collection appearance on Sketchfab.

Virtual Simulation (VS) offers an elegant and effective solution to the current need for innovation in medical education, thanks to the possibility of creating low-cost, realistic training environments for repetitive practice without compromising patient safety. However, this training methodology is only adopted in some healthcare settings often because of the absence of free digital libraries of clinical assets and tools. The present technical note describes a data collection of 3D models representing crucial tools and equipment used in maternal and newborn care training. We used free-to-use photogrammetry and structure-from-motion software and a computational platform for 3D object reconstruction to digitalize the physical clinical instruments typically used during maternal and newborn care. In particular, we acquired photographs of 34 physical objects and reconstructed them as 3D models. Additionally, we created a complete, navigable virtual training room containing the 3D models. Eventually, we published the 3D models and the virtual training room as an open-access data collection on Sketchfab (a free-to-use online digital platform for 3D model publication), from which all models can be freely downloaded and inspected through Web browsers, mobile applications, and Virtual and Augmented Reality devices. Our data collection and repeatable and cost-effective methodology open new opportunities to use VS for training through simulation in healthcare.

DOI: 10.1016/j.rineng.2025.104236

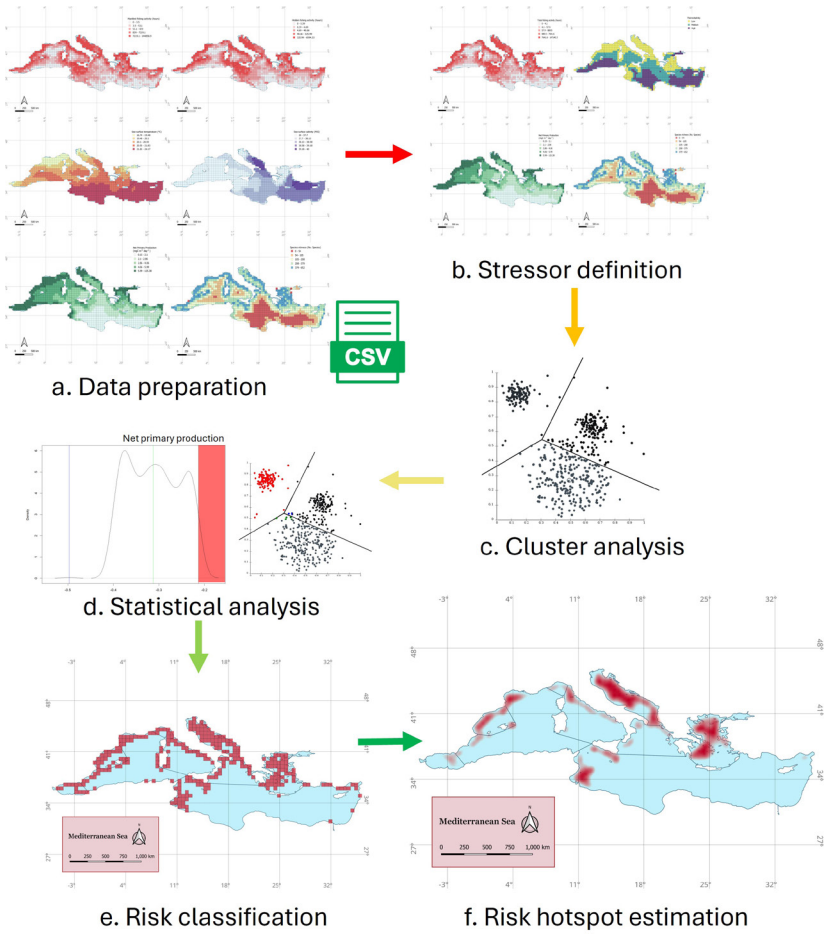
Computing Ecosystem Risk Hotspots: a Mediterranean Case Study

G. Coro, L. Pavirani, A. Ellenbroek

Ecological Informatics, vol. 85. Elsevier, 2025.

In ecosystem management, risk assessment quantifies the probability and impact of events and informs on intervention priorities. Analytical models for risk assessment quantify the impact of natural and anthropogenic stressors on ecosystems. Traditional approaches evaluate single stressors, whereas complex models assess the cumulative impacts of frequently interacting stressors and offer better accuracy at the expense of low cross-area re-applicability and long implementation times.

We introduce a versatile, re-useable, and semi-automated workflow designed for big data-driven ecosystem risk assessment, utilising spatiotemporal data from open repositories. It allows for a flexible definition of the stressors on which the risk under analysis depends. By applying cluster analysis, the workflow identifies different patterns of stressor concurrency, while statistical analysis highlights clusters of stressors likely linked to elevated risk. Ultimately, it generates geospatial risk maps and identifies spatial risk hotspots. The workflow methodology is independent of the geographical area of the application.



Overall schema of our workflow.

As a case study, we present risk assessments for the Mediterranean Sea, a region with intense anthropogenic pressures and significant climatic vulnerabilities. We used over 1.1 million open data from 2017 to 2021 and projections to 2050 under the RCP8.5 scenario (a high greenhouse gas emission scenario) at a 0.5°spatial resolution. Data included environmental, oceanographic, biodiversity variables, and manifest and hidden fishing effort distributions. Our workflow identified different types of high-risk hotspots, highlighting different concurrencies of habitat loss, overfishing, hidden fishing, and climate change stressors. High-risk hotspots concentrated in the Western Mediterranean, the Tyrrhenian Sea, the Adriatic Sea, the Strait of Sicily, the Aegean Sea, and eastern Turkey. Our results agreed with an alternative Fuzzy C-means-based method (with a 90% to 96% overlap over the years) and a Bayesian regression model (~0% overlap).

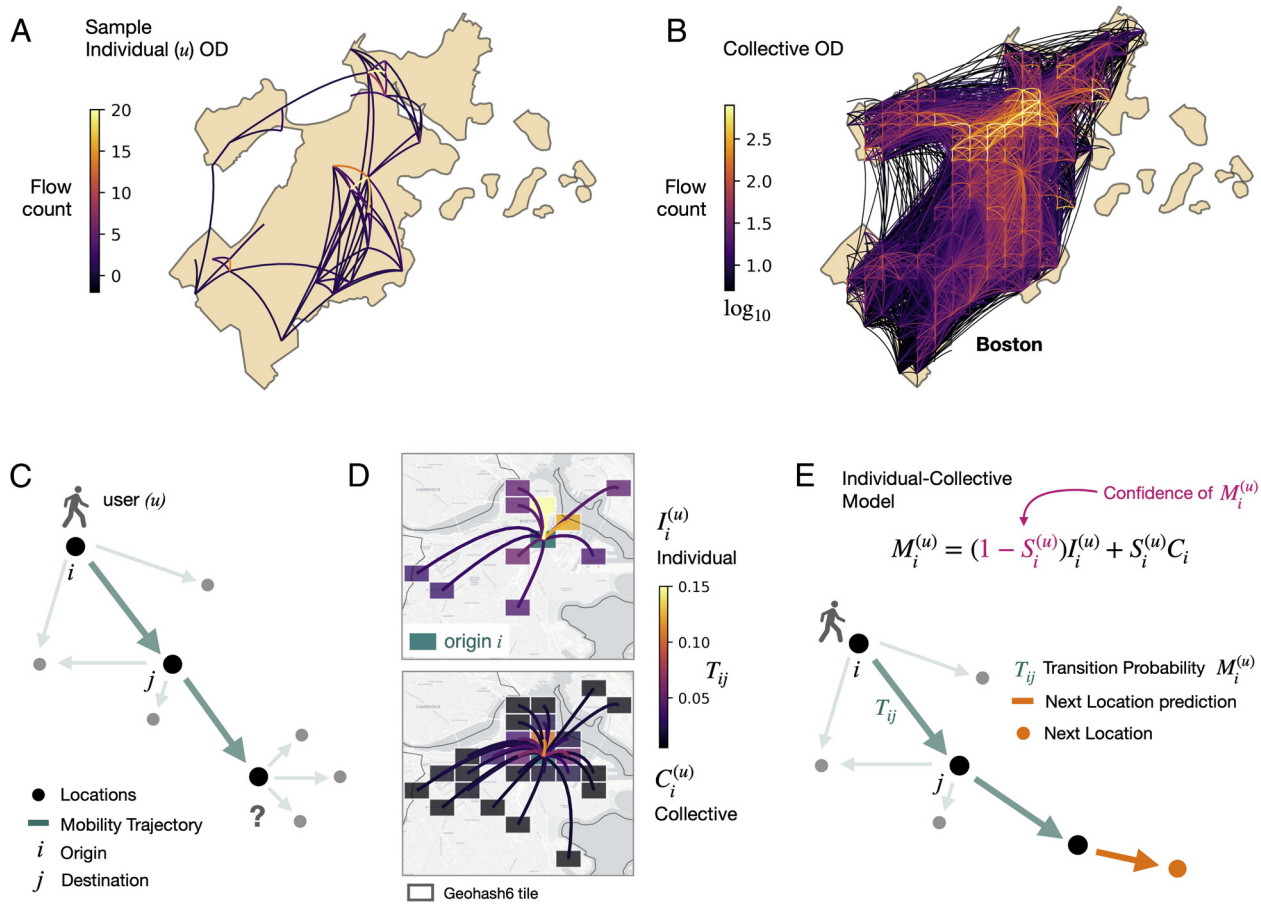
Our Mediterranean risk maps can facilitate the development of management and monitoring strategies, supporting the sustainable development and resilience of coastal zones, and can act as prior knowledge for ecosystem models and spatial plans.

DOI: 10.1016/j.ecoinf.2024.102918

Mixing Individual and Collective Behaviors to Predict Out-of-Routine Mobility

S. Bontorin, S. Centellegher, R. Gallotti, L. Pappalardo, B. Lepri, M. Luca

Proceedings of the National Academy of Science, vol. 122. PNAS, 2025.



Dynamic interplay of individual and collective mobility. (A) An individual origin–destination (OD) matrix for a synthetic individual *u*. (B) The collective OD matrix computed for Boston using GPS trajectories. (C) An individual trajectory for a user *u* starting from location *i*. Next location prediction consists of predicting *u*'s next visited location. (D) The set of *u*'s historical trajectories (panel A) is used to define the transition probabilities $I_i^{(u)}$ from location *i*. C_i represents the probability distribution of all transitions made by any user starting from location *i*, generated from the OD matrix (panel B). Destinations' locations *j* are colored based on their visitation probability, $T_{i,j}$, from origin . (E) $M_i^{(u)}$'s prediction of individual *u*'s next location is performed by dynamically combining $I_i^{(u)}$ and C_i, based on the normalized Shannon entropy $S_i^{(u)}$ computed from the mobility trajectories of *u*. Maps: Stamen Maps. Icons: Fontawesome.

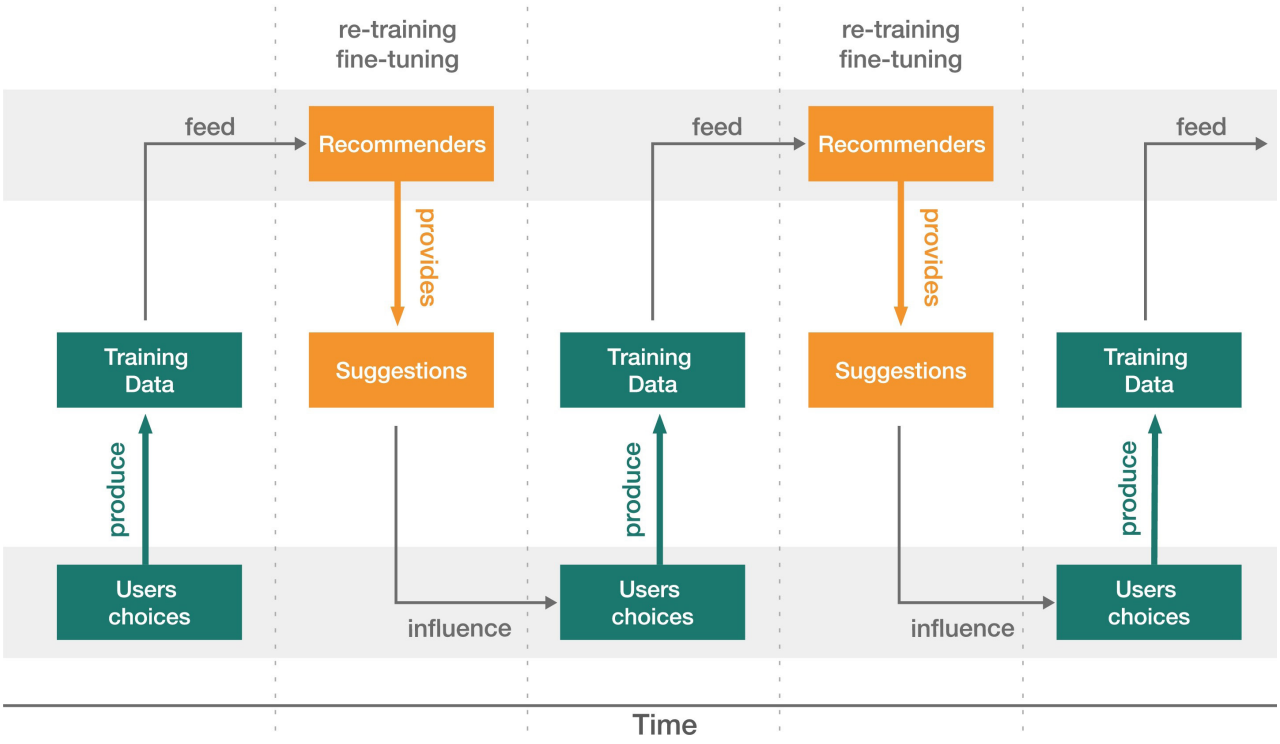
Predicting human mobility is essential for addressing societal challenges such as urban planning, traffic management, and epidemic control. Traditional models, including deep learning and Markov chains, often struggle to accurately forecast out-of-routine movements, i.e., those that deviate from habitual patterns. This paper – a collaboration between ISTI-CNR and Fondazione Bruno Kessler – introduces a novel approach that dynamically integrates individual mobility behaviours with collective patterns, leveraging collective intelligence to enhance prediction accuracy. Evaluated on millions of privacy-preserving trajectories across three U.S. cities, the model demonstrates superior performance in predicting out-of-routine mobility, surpassing even advanced deep learning methods. Spatial analysis reveals the model's effectiveness near urban areas with a high density of points of interest, where collective behaviours strongly influence mobility. Notably, during disruptive events like the COVID-19 pandemic, the model retains predictive capabilities, unlike individual-based models. By bridging the gap between individual and collective behaviours, this approach offers transparent and accurate predictions, crucial for addressing contemporary mobility challenges.

DOI: 10.1073/pnas.2414848122

Human-AI Coevolution

D. Pedreschi, L. Pappalardo, E. Ferragina, R. Baeza-Yates, A.L. Barabási, F. Dignum, V. Dignum, T. Eliassi-Rad, F. Giannotti, J. Kertész, A. Knott, Y. Ioannidis, P. Lukowicz, A. Passarella, A.S. Pentland, J. Shawe-Taylor, A. Vespignani

Artificial Intelligence, vol. 339. Elsevier, 2025.



Users' choices on online platforms generate data used to train recommenders. These recommenders then offer suggestions to users, influencing their choices, which in turn generate more data for re-training recommenders. This iterative process creates a potentially endless feedback loop.

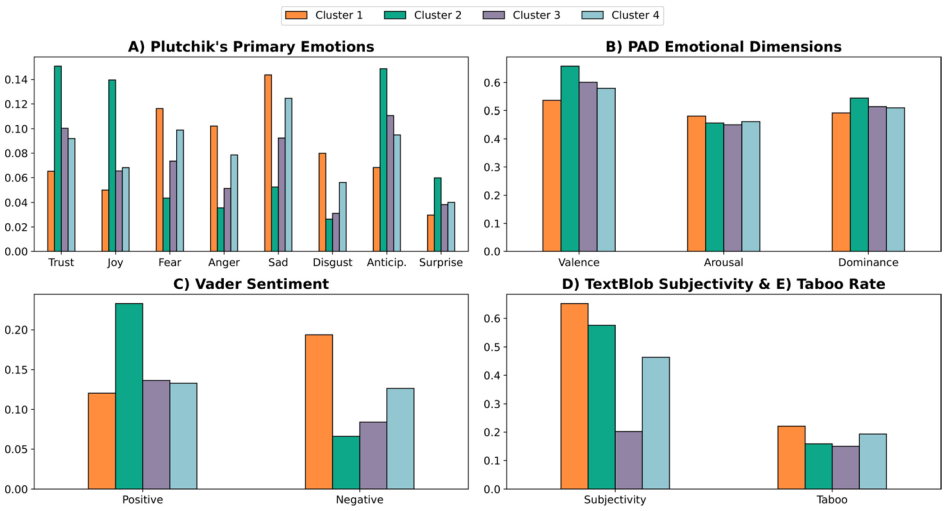
Human-AI coevolution, defined as a process in which humans and AI algorithms continuously influence each other, increasingly characterises our society, but is understudied in artificial intelligence and complexity science literature. Recommender systems and assistants play a prominent role in human-AI coevolution, as they permeate many facets of daily life and influence human choices through online platforms. The interaction between users and AI results in a potentially endless feedback loop, wherein users' choices generate data to train AI models, which, in turn, shape subsequent user preferences. This human-AI feedback loop has peculiar characteristics compared to traditional human-machine interaction and gives rise to complex and often "unintended" systemic outcomes. This paper introduces human-AI coevolution as the cornerstone for a new field of study at the intersection between AI and complexity science focused on the theoretical, empirical, and mathematical investigation of the human-AI feedback loop. In doing so, we: (i) outline the pros and cons of existing methodologies and highlight shortcomings and potential ways for capturing feedback loop mechanisms; (ii) propose a reflection at the intersection between complexity science, AI and society; (iii) provide real-world examples for different human-AI ecosystems; and (iv) illustrate challenges to the creation of such a field of study, conceptualising them at increasing levels of abstraction, i.e., scientific, legal and socio-political.

DOI: 10.1016/j.artint.2024.104244

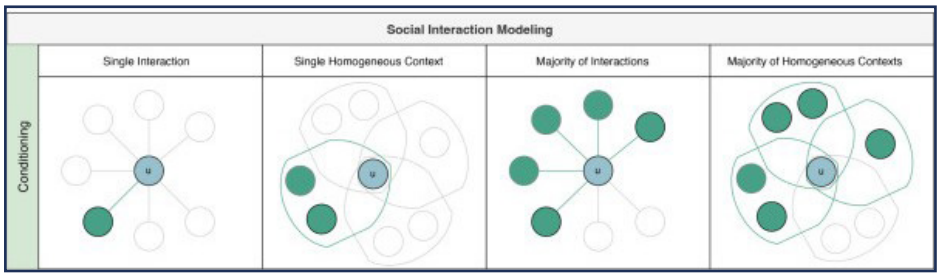
Online Posting Effects: Unveiling the Non-Linear Journeys of Users in Depression Communities on Reddit

V. Morini, S. Citraro, E. Sajno, M. Sansoni, G. Riva, M. Stella, G. Rossetti

Computers in Human Behavior Reports, vol 17. Elsevier, 2025.



Psycholinguistic cluster profiles in Reddit Depression discourse. (A–E) Bar charts of cluster centroids values: Plutchik's Primary Emotions, PAD Emotional Dimensions, VADER Sentiment, Textblob Subjectivity, Taboo Rate.



Different levels of social exposure leveraging graphs (single/majority of interactions) and hypergraphs (single/majority of homogeneous contexts)t.

Social media platforms have become pivotal as self-help forums, enabling individuals to share personal experiences and seek support. However, on topics as sensitive as depression, what are the consequences of online self-disclosure? Here, we delve into the dynamics of mental health discourse on various Reddit boards focused on depression. To this aim, we introduce a data-informed framework reconstructing online dynamics from 303k users interacting over two years. Through user-generated content, we identify 4 distinct clusters representing different psychological states. Our analysis unveils online posting effects: a user can transition to another psychological state after online exposure to peers' emotional/semantic content. As described by conditional Markov chains and different levels of social exposure (Figure 1), users' transitions reveal navigation through both positive and negative phases in a spiral rather than a linear progression. Interpreted in light of psychological literature, related particularly to the Patient Health Engagement (PHE) model, our findings can provide evidence that the type and layout of online social interactions have an impact on users' "journeys" when posting about depression.

DOI: 10.1016/j.chbr.2024.100542

Investigating the Scalability of Approximate Sparse Retrieval Algorithms to Massive Datasets

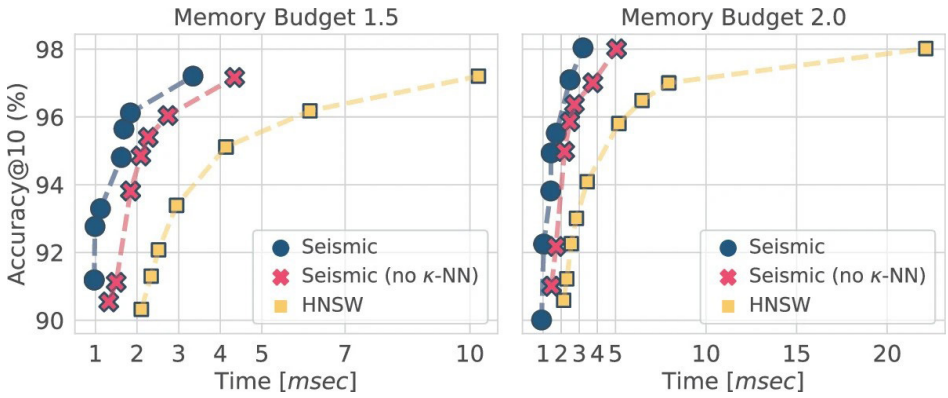
S. Bruch, F.M. Nardini, C. Rulli, R. Venturini, L. Venuta

Advances in Information Retrieval. Lecture Notes in Computer Science vol. 15574. Springer, 2025.

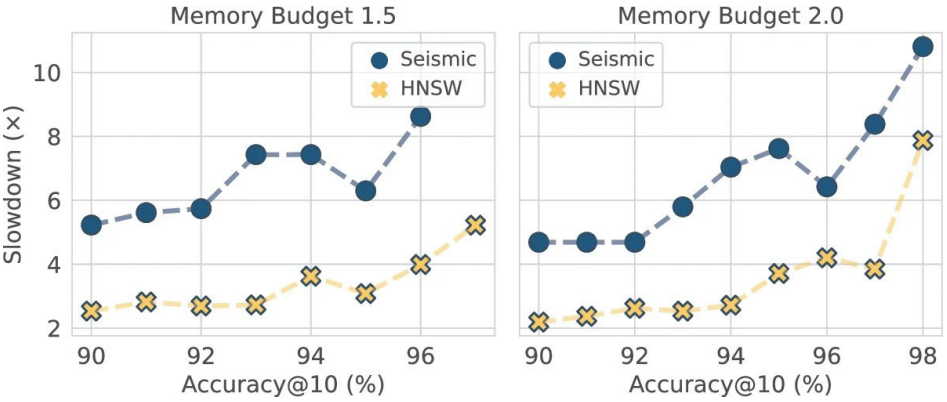
Learned sparse text embeddings have emerged as a compelling alternative to traditional dense embeddings in information retrieval, primarily due to their dual advantages: strong performance in top-k retrieval tasks and enhanced interpretability. Unlike dense embeddings, which encode information in continuous vector spaces and often lack transparency, sparse embeddings maintain a clear connection to vocabulary terms, making them more intuitive to analyze and debug. Despite these advantages, their adoption in large-scale, real-world systems has been limited due to the unique distributional properties they exhibit, which make them inefficient for use with standard inverted index structures and traditional retrieval algorithms.

Recent advances have begun to overcome these limitations through the development of approximate retrieval algorithms specifically designed to exploit the distributional peculiarities of sparse embeddings. These methods—including techniques like Seismic offer the promise of scalable, high-throughput sparse retrieval. However, the majority of existing studies have focused on relatively small-scale datasets, such as MS MARCO, which contains a few million documents. As a result, the scalability and practical applicability of these methods in much larger corpora remains largely unexplored.

In this work, we take a significant step toward understanding the behavior of learned sparse retrieval methods at scale. We conduct an extensive empirical study comparing state-of-the-art sparse retrieval algorithms, including Seismic and graph-based strategies, when applied to a massive dataset of 138 million passages from MS MARCO v2. We use SPLADE embeddings—a leading method for sparse representation learning—as our foundation for all experiments. Our evaluation spans a comprehensive set of metrics, including retrieval effectiveness (e.g., NDCG, MRR), indexing efficiency, query latency, and memory usage.



Comparison of HNSW and SEISMIC (with and without κ-NN graph) by accuracy at $\kappa = 10$ as a function of query latency. We allow hyperparameters that result in an index whose size is at most (left) or (right) the size of the dataset.



Scaling laws of Seismic and sparse Hnsw (as provided by the KANNOLO library). For each accuracy cutoff, we measure the ratio between the latency of a method on MSMARCO v2 and on MSMARCO.

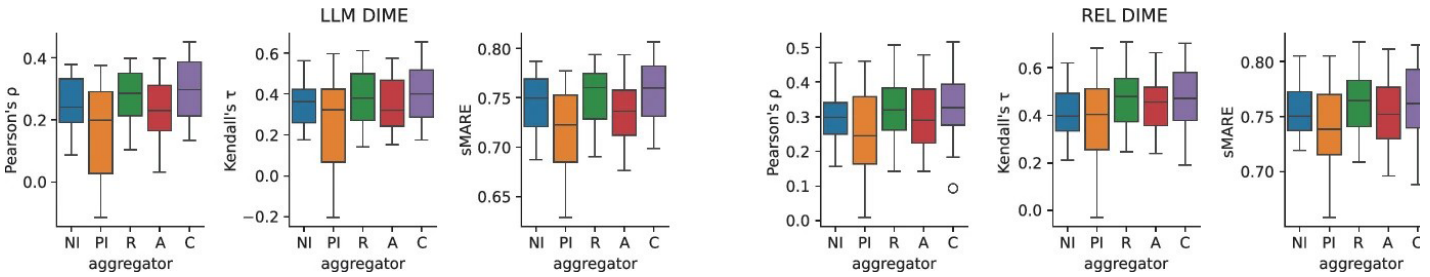
Our findings reveal notable trade-offs between the various approaches. Seismic demonstrates strong indexing efficiency and retrieval speed, while graph-based techniques offer competitive retrieval performance at the cost of more complex indexing and higher memory consumption. Moreover, we identify several scalability challenges that arise only at extreme corpus sizes, including degradation in retrieval effectiveness due to sparsity amplification and increased sensitivity to index organization.

DOI: 10.1007/978-3-031-88714-7_43

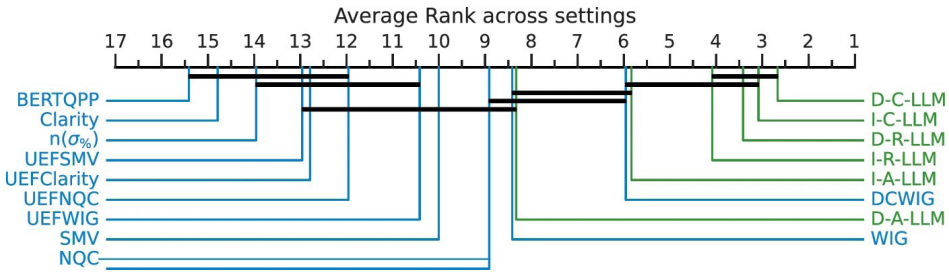
Query Performance Prediction Using Dimension Importance Estimators

G. Faggioli, N. Ferro, R. Perego, N. Tonellotto.

Advances in Information Retrieval. Lecture Notes in Computer Science vol. 15573. Springer, 2025.



Performance of the five aggregator components. These results are aggregated across evaluation measures, datasets, input components, and IR models. While NI and PI tend to be the worst performing. R and A behave similarly, while C is the optimal approach in all scenarios.



Average rank across different experimental settings, excluding predictors based on Active Feedback DIME. In green and blue, predictors proposed here and baselines, respectively. Horizontal bars indicate statistically equivalent approaches according to a Wilcoxon signed-rank test.

Query Performance Prediction (QPP) aims to estimate the effectiveness of an Information Retrieval (IR) system for a given query in the absence of human-provided relevance judgments. QPP techniques can be used to identify poorly performing queries, select the most suitable IR system per query, or combine multiple systems. While traditional QPP methods have been successful in lexical-based IR systems, they often underperform when applied to dense IR models that rely on semantic representations.

Recent research in dense retrieval has highlighted that not all dimensions of a dense embedding space contribute equally to effective retrieval. Specifically, our previous work observed that only a subset of query-specific dimensions is beneficial for ranking, while others may be irrelevant or even harmful. Based on this, they introduced the manifold clustering hypothesis, which posits that high-dimensional representations of queries and their relevant documents often reside in a query-dependent, lower-dimensional manifold of the embedding space. To operationalize this idea, they proposed the Dimension Importance Estimation (DIME) framework, which identifies an optimal linear subspace for each query by estimating the importance of individual embedding dimensions.

In this work, we apply the DIME framework to the QPP task. We propose a new family of QPP models that leverage DIME to estimate which dimensions are most relevant to a query and then assess the alignment between query and document embeddings along those dimensions. The core hypothesis is that strong alignment between query and document vectors in the most important dimensions correlates with high retrieval performance, while poor alignment signals potential failure.

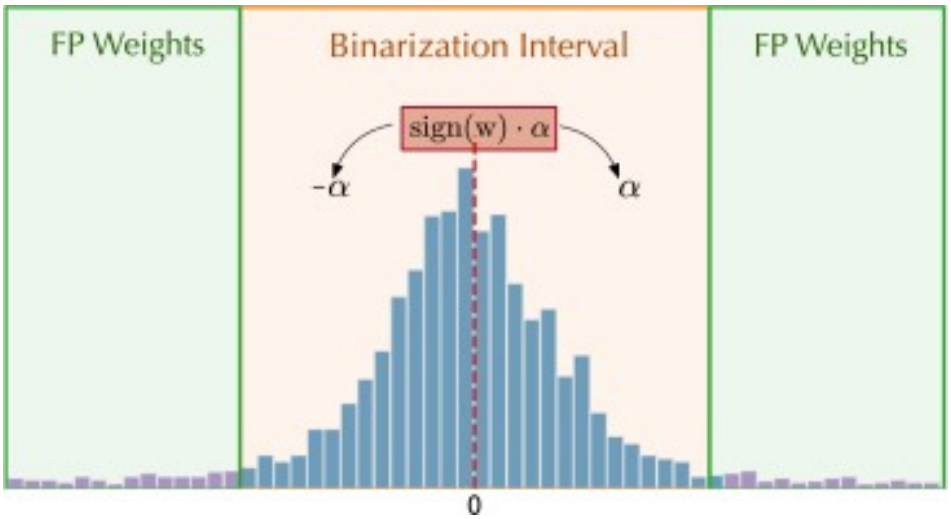
We instantiate a set of alignment-based heuristics that compute this dimension-aware similarity and evaluate their predictive power on two prominent dense IR systems, Contriever and TAS-B, using the TREC Deep Learning 2019 and 2020 benchmarks. Experimental results show that our DIME-based predictors significantly outperform existing QPP methods for dense retrieval, confirming the validity of the alignment hypothesis and the practical value of integrating dimension importance into QPP modeling.

DOI: 10.1007/978-3-031-88711-6_13

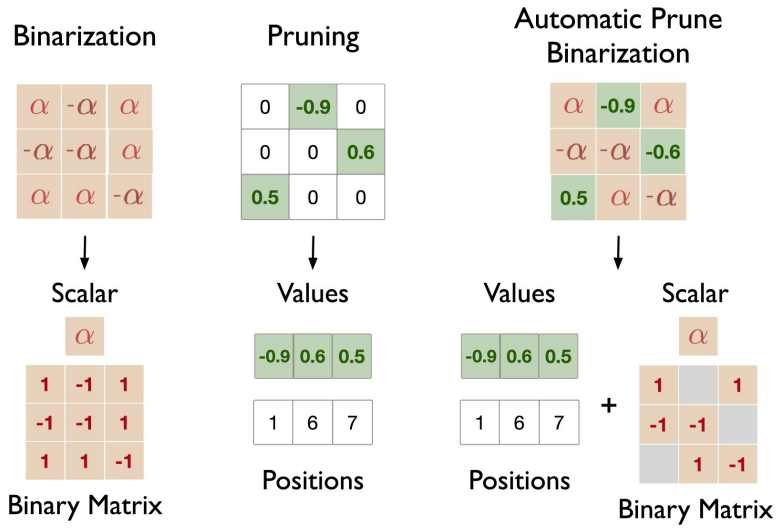
Neural Network Compression Using Binarization and Few Full-Precision Weights

F.M. Nardini, C. Rulli, S. Trani, R. Venturini

Information Systems, early access. Elsevier, 2025.



Graphical illustration of Automatic Prune Binarization (APB). The weights in the binarization interval are converted to {−α,α} according to their sign, while the remaining values are kept in full precision.



Network weights representation in binarization (left), pruning (center), APB (right). In APB, binary, and full-precision weights coexist in the same matrix. For full-precision entries, we are required to store also the index inside the weight matrix.

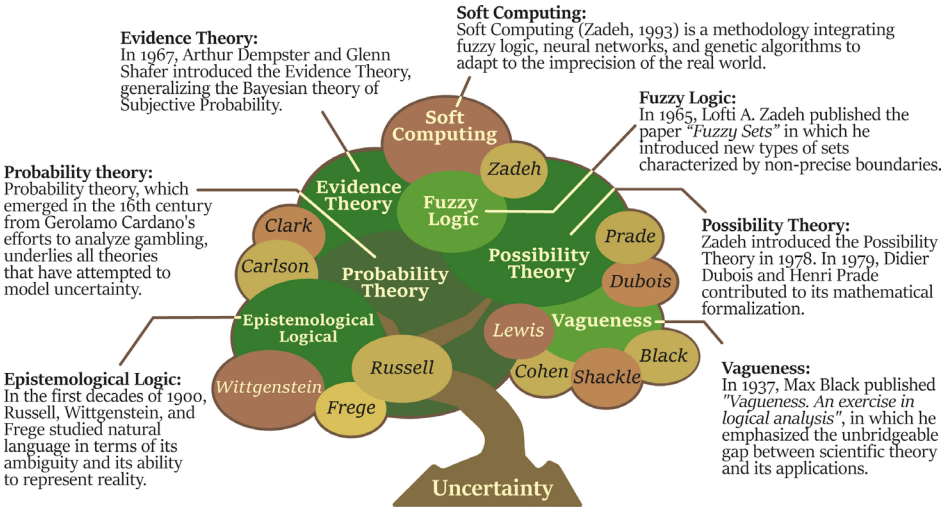
Quantization and pruning are two effective Deep Neural Network model compression methods. In this paper, we propose Automatic Prune Binarization (APB), a novel compression technique combining quantization with pruning. APB enhances the representational capability of binary networks using a few full-precision weights. Our technique jointly maximizes the accuracy of the network while minimizing its memory impact by deciding whether each weight should be binarized or kept in full precision. We show how to efficiently perform a forward pass through layers compressed using APB by decomposing it into a binary and a sparse-dense matrix multiplication. Moreover, we design two novel efficient algorithms for extremely quantized matrix multiplication on CPU, leveraging highly efficient bitwise operations. The proposed algorithms are 6.9× and 1.5× faster than available state-of-the-art solutions. We extensively evaluate APB on two widely adopted model compression datasets, namely CIFAR-10 and ImageNet. APB shows to deliver better accuracy/memory trade-off compared to state-of-the-art methods based on i) quantization, ii) pruning, and iii) a combination of pruning and quantization. APB also outperforms quantization in the accuracy/efficiency trade-off, being up to 2× faster than the 2-bits quantized model with no loss in accuracy.

DOI: 10.1016/j.ins.2025.122251

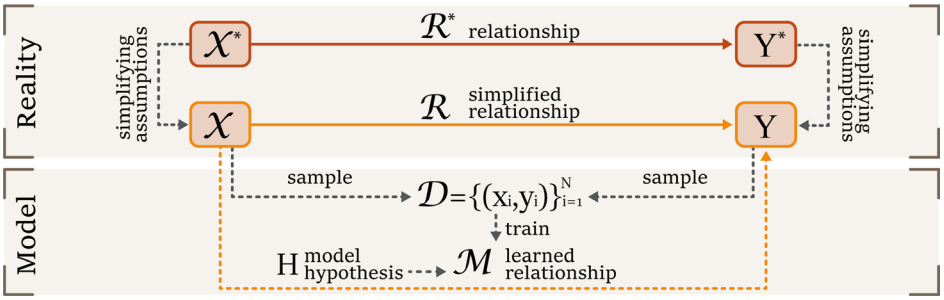
Shedding Light on Uncertainties in Machine Learning: Formal Derivation and Optimal Model Selection

G. Del Corso, S. Colantonio, C. Caudai

Journal of the Franklin Institute, vol. 362. Elsevier, 2025.



Road map of evolution of Uncertainty conceptualization in the 20th century, throw years and currents of thought.



Graphical scheme showing the relationship/simplified relationship between the real input/output and the corresponding trained model.

The concept of uncertainty has always been important in the field of mathematical modeling. In particular, the growing application of Machine Learning and Deep Learning methods in many scientific fields has led to the implementation and use of new uncertainty quantification techniques aimed at distinguishing between reliable and unreliable predictions. However, the novelty of this discipline and the plethora of articles produced, ranging from theoretical results to purely applied experiments, has resulted in a very fragmented and cluttered literature. In this review, we have attempted to combine the well-established mathematical background of the Bayesian framework with the practical aspect of modern state-of-the-art emerging techniques in order to meet the urgent need for clarity on key concepts related to uncertainty quantification. First, we introduced the different sources of uncertainty, ranging from epistemic/reducible to aleatoric/irreducible, providing both a rigorous mathematical derivation and several examples to facilitate understanding. The review then details some of the most important techniques for uncertainty quantification. These methods are compared in terms of their advantages and drawbacks and classified in terms of their intrusiveness, in order to provide the practitioner with a useful vademecum for selecting the optimal model depending on the application context.

DOI: 10.1016/j.jfranklin.2025.107548

Descriptor: Not-A-DAtabase of Synthetic Shapes Benchmarking Dataset (NADA-SynShapes)

G. Del Corso, F. Volpini, C. Caudai, D. Moroni, S. Colantonio

IEEE Data Descriptions, vol. 2. IEEE, 2025.

NADA is a synthetic dataset of 1,500,000 (128x128 pixels) images of geometric shapes with non-elementary multivariate parameter distributions designed to benchmark and test novel probabilistic deep learning models. Benchmarking uncertainty-aware techniques is critical in real-world scenarios, especially in high-stakes domains such as automated driving or health data analysis. This is because the most robust and reliable AI methods are rooted in Bayesian reasoning and uncertainty analysis, yet there is no consensus on how to test or compare them. In particular, the public, synthetic, and real

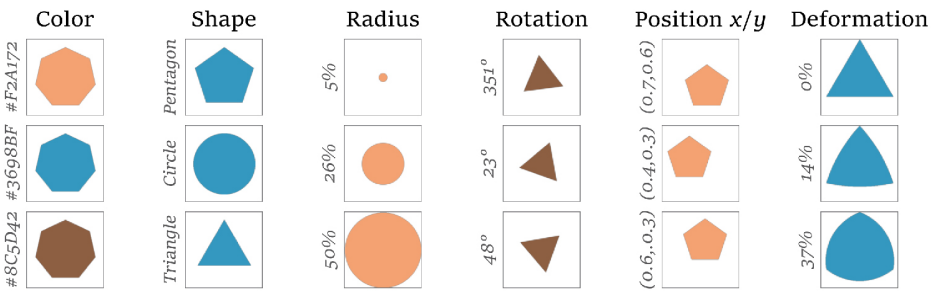
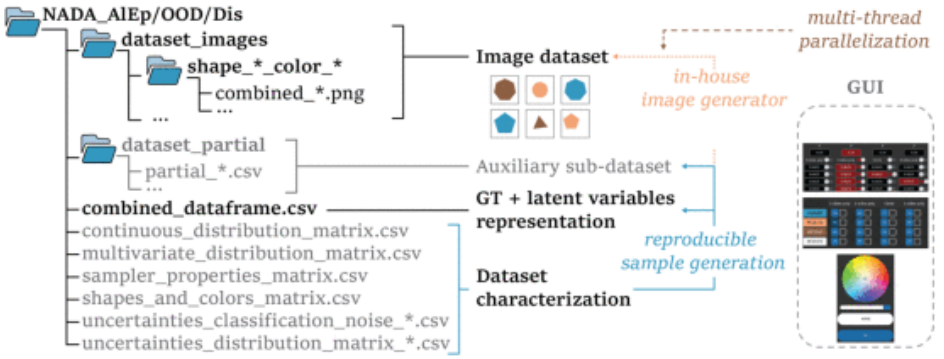


Image generation user-defined parameters: Color (described by hex color codes), Shape (arbitrary regular shapes given the number of vertices), Radius (from 0 to 50% of the image), Rotation (0-360°), asymmetry can be induced in the shape by enabling a corresponding option), Position (center position in percentage from the top left of the image), Deformation (0-100%, from the original regular shape to the circumscribing circle).



Folder structure of each dataset presented. The main folder contains a subfolder (dataset_images) with the generated images. The comma-separated value file combined_dataframe.csv contains the description of each image for disentangling or training/test. The other csv files are used to fully characterize the dataset and can be loaded into the provided GUI to generate a fully reproducible variation of the dataset using an in-house image generator coupled with efficient multi-thread parallelization.

datasets currently available in the literature are inadequate for benchmarking most probabilistic methodologies: first, due to a lack of control over population variability; and second, due to an oversimplification of the distributional properties of latent variables. From this perspective, NADA is organized into three main From this perspective, NADA is organized into three main repositories, specifically developed to challenge uncertainty-aware methods and provide a unified benchmark reference dataset across three key areas: (1) characterization of complex latent space and evaluation of disentangling ability (NADA_Dis: 300,000 images), (2) identification of different types of aleatoric and epistemic uncertainties (NADA_AIEp: 500,000 images), and (3) detection of out-of-distribution elements for reliable AI (NADA_OOD: 700,000 images). Each repository includes the dataframe describing the image parameters (e.g., rotation, position, shape, color, deformation, noise), the dataset-generating hyperparameters (e.g., marginal distributions and correlation matrix), and evaluation plots to assess data quality. In addition, the dataset is coupled with an open-source Python synthetic generator, allowing easy modification and adaptation to specific research questions.

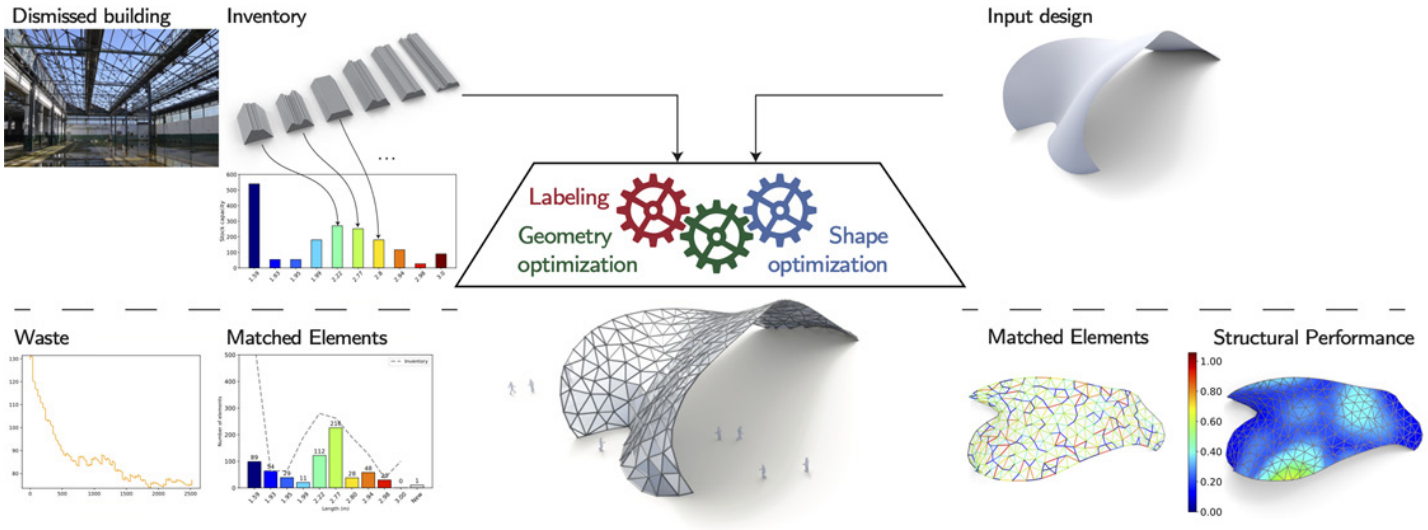
DOI: 10.1109/IEEEDATA.2025.3562805

Optimizing Free-Form Grid Shells with Reclaimed Elements under Inventory Constraints

A. Favilli, F. Laccone, P. Cignoni, L. Malomo, D. Giorgi
Computer Graphics Forum, early access. Wiley 2025.

We propose a method for designing 3D architectural free-form surfaces, represented as grid shells with beams sourced from inventories of reclaimed elements from dismantled buildings. In inventory-constrained design, the reused elements must be paired with elements in the target design. Traditional solutions to this assignment problem often result in cuts and material waste or geometric distortions that affect the surface aesthetics and buildability. Our method for inventory-constrained assisted design blends the traditional assignment problem with differentiable geometry optimization to reduce cut-off waste while preserving the design intent. Additionally, we extend our approach to incorporate strain energy minimization for structural efficiency. We design differentiable losses that account for inventory, geometry, and structural constraints, and streamline them into a complete pipeline, demonstrated through several case studies. Our approach enables the reuse of existing elements for new designs, reducing the need for sourcing new materials and disposing of waste. Consequently, it can serve as an initial step towards mitigating the significant environmental impact of the construction sector.

DOI: 10.1111/CGF.70047

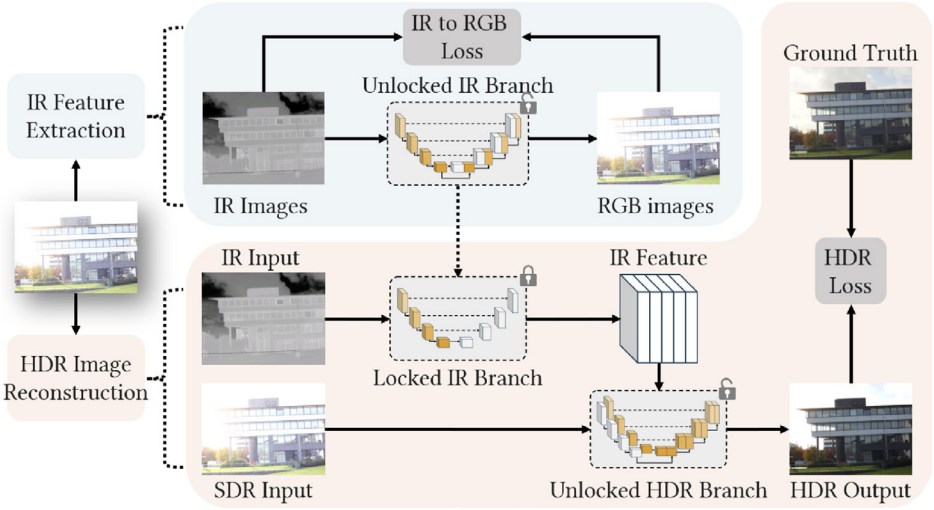


We propose a method for the assisted design of 3D architectural free-form grid shells by reusing elements from disassembled structures. Our input includes an inventory of available structural elements in various quantities and a user-defined target shape (top row). Our output is a grid shell that reuses as many elements as possible from the inventory, minimizes material waste, is optimized for statics performance, and preserves the design intent (bottom row). We integrate three main components into a single optimization pipeline (middle, solid box): combinatorial inventory assignment for maximizing reuse, differentiable geometry optimization for waste reduction, and learning-based shape optimization for statics performance.

HDRT: a Large-Scale Dataset for Infrared-Guided HDR Imaging

J. Peng, T. Bashford-Rogers, F. Banterle, H. Zhao, K. Debattista

Information Fusion, early access. Elsevier, 2025



The full pipeline of the proposed HDRTNet for enhancing SDR images using thermal to obtain HDR images.

Capturing images with enough details to solve imaging tasks is a long-standing challenge in imaging, particularly due to the limitations of standard dynamic range (SDR) images which often lose details in underexposed or overexposed regions. Traditional high dynamic range (HDR) methods, like multi-exposure fusion or inverse tone mapping, struggle with ghosting and incomplete data reconstruction. Infrared (IR) imaging offers a unique advantage by being less affected by lighting conditions, providing consistent detail capture regardless of visible light intensity. In this paper, we introduce the HDRT dataset, the first comprehensive dataset that consists of HDR and thermal IR images. The HDRT dataset comprises 50,000 images captured across three seasons over six months in eight cities, providing a diverse range of lighting conditions and environmental contexts. Leveraging this dataset, we propose HDRTNet, a novel deep neural method that fuses IR and SDR content to generate HDR images. Extensive experiments validate HDRTNet against the state-of-the-art, showing substantial quantitative and qualitative quality improvements. The HDRT dataset not only advances IR-guided HDR imaging but also offers significant potential for broader research in HDR imaging, multi-modal fusion, domain transfer, and beyond.

DOI: 10.1016/j.inffus.2025.103109

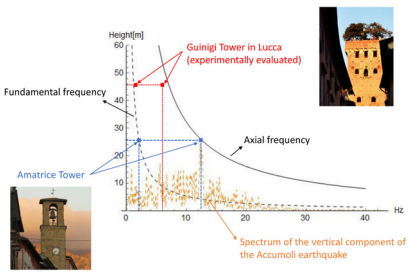
Transverse and Axial Vibrations of Masonry Beams and Towers

M. Girardi

International Journal of Non-Linear Mechanics, early access. Elsevier, 2025.

Studying the effects of the earthquakes' vertical component on the dynamic response of masonry buildings is still an open issue. This subject is investigated using a simplified masonry-like beam model with a generalized stress–strain relationship. The nonlinear dynamic equations that couple the transverse and axial oscillations of the beam are solved numerically. Some examples are presented and discussed, focusing on the structural typology of masonry towers and the recent earthquakes that threatened the North-Central Apennines in Italy.

DOI: 10.1016/j.ijnonlinmec.2025.105086



Masonry towers: fundamental (black dashed line) and axial (black continuous line) frequencies vs. the towers' height.

IEEE WCNC 2025 Best Paper Award

At the IEEE Wireless Communications and Networking Conference, Milan, Italy, March 24–27, 2025

"Whispers in the Snow: Exploring LoRa Technology for Avalanche Search and Rescue Scenarios" by Michele Girolami and Fabio Mavilia (Cnr-Isti), Giulio M. Bianco and Gaetano Marrocco (University of Rome Tor Vergata, Italy), won the Best Paper Award at WCNC, for Track 4: "Emerging Technologies, Network Architectures, and Applications".

LoRa (Long Range) technology has become a popular choice for low-power wide-area networks (LPWANs) due to its long range, ability to penetrate obstacles, and its relatively low cost. The LoRa-SNOW project explores Search and Rescue (SaR) operations in avalanche scenarios, where the goal is to locate transmitters buried under snow.

For the first time, this study experimentally investigated LoRa signal propagation in this challenging environment, characterized by the absence of line-of-sight (LOS) and significant signal attenuation caused extremely low temperatures, humidity, and, most critically, snow.

The tests were conducted on a plateau at 1850 meters in the Dolomites (Col de Mez), within a wide area using one transmitter and several LoRa receivers. Two key variables were analyzed: burial depth (0.50 cm, 1 m) and snow type (dry, cold winter snow vs. wet spring snow). Received signal strength (RSS) and signal-to-noise ratio (SNR) were measured to understand the signal behaviour under these challenging conditions.

The study also evaluated the maximum communication range through snow, with promising results—LoRa maintained an effective transmission for up to over 400 meters.

The data collected enable the:

- creation of detailed maps of LoRa propagation across different snow conditions;
- investigations into localization and proximity detection algorithms in order to pinpoint buried transmitters.

The IEEE Wireless Communications and Networking Conference (WCNC) is a premier event organized by the IEEE Communications Society, which is renowned for bringing together leading researchers from academia, industry, and governments.

Contact: Michele Girolami, WN Lab
michele.girolami@isti.cnr.it
https://wcnc2025.ieee-wcnc.org

The Museum of Computing Machinery is back

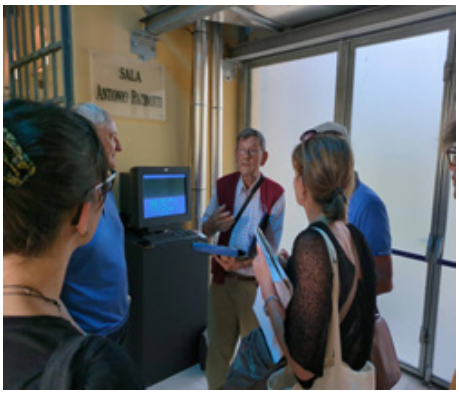
The Museum of Computing Machinery has returned to its historic home in the Galilean Citadel of the University of Pisa, with a new exhibition entitled "Not just computing: working life and leisure time before the Internet"("Non solo calcolo – vita lavorativa e tempo libero prima dell'arrivo di Internet"). The exhibition now also hosts devices that defined work and entertainment from the late 19th century to the 1990s.

To mark the reopening, some of the pioneers of Italian computer science—Lucia Barsanti Calamia, Graziano Bertini (former colleague – Isti), and Giuseppe Iazeolla—served as special guides, recounting the stories of the machines on display, alongside Andrea Tendola and Giovanni Cignoni (see https://www.unipi.it/news/il-museo-degli-strumenti-per-il-calcolo-torna-alla-cittadella-galileiana/).

A fundamental part of the exhibition is dedicated to the history of computer science in Pisa, including the collaboration with Olivetti and the historical milestones of the ELEA 9003 and the Pisan Electronic Calculator (CEP). For the latter, the parts that had been used in the "Hello World" exhibition (such as the control console) have been reinstalled. Their absence had greatly alarmed the late Luciano Azzarelli during a previous visit, and indeed, the CEP is now displayed in its majestic entirety.

Alongside objects like cash registers and mechanical calculators, the rooms house historic computers such as the first Olivetti word processors, the Amiga, and PCs with SoundBlaster cards. There is even a section dedicated to 1980s video games, featuring the Commodore 64 and the ZX Spectrum.

Another section explores the history of musical informatics in Pisa, (Ref.: https://www.pisauniversitypress.it/scheda-libro/autori-vari/linformatica-musicale-a-pisa-9788833392820-575611.html) referencing the contributions of pioneer Pietro Grossi with the TAU2/TAUMUS system, developed at the IEI/CNR in the 1970s, and the subsequent research by Leonello Tarabella. In this area, there is only the rack of the TAU2's control unit, which interfaced with the IBM 360 time-sharing system of the CNUCE/CNR, alongside a video terminal of the era. Thanks to a solution created by Massimo Magrini, visitors can use the keyboard to simulate and listen to musical examples, even without the original audio rack, which is no longer functional and is stored in a non-public location. The use of mixed electronic solutions (both digital and analog) allowed for the real-time production of musical pieces with a 12-voice polyphony and complex timbre. The TAU2 was connected to the multi-user IBM 360, a significant advance at a time when other computer music centers operated on a non-real-time basis.



A display case shows some of the records produced with the system, featuring classical pieces and original compositions, along with documentation related to the activity carried out in Pisa. While showing visitors around, Bertini also mentioned the experiments in musical telematics, with demonstrations and concerts held in remote locations using the data network of the time to connect to the CNUCE, as well as voice lines from the SIP telephone network, which included both switched lines (with restricted audio bandwidth) and leased lines (for extended audio bandwidth).

Contacts: Massimo Magrini, SI Lab
massimo.magrini@isti.cnr.it
Fabio Gadducci, University of Pisa
fabio.gadducci@unipi.it

Ph.D. dissertations

Deepfake Detection in Images and Videos

Author: Davide Alessandro Coccomini, University of Pisa, Italy
Supervisors: Giuseppe Amato, Fabrizio Falchi and Claudio Gennaro (Cnr-Isti), Roberto Caldelli (Mercatorum University),

This thesis addresses the problem of deepfake detection in images and videos by developing hybrid techniques that combine Convolutional Neural Networks and Vision Transformers. We proposed novel deep learning architectures that achieve state-of-the-art results. A key focus was the issue of generalization: detectors often perform well on manipulations seen during training but struggle with unseen techniques. We analyzed various architectures for robustness and introduced a training method based on deepfake fingerprints rather than specific manipulation types, significantly improving generalization.

In the context of synthetic image detection, we also explored a multi-modal approach that leverages the prompt used in image generation to enhance detection performance. Another critical challenge is the poor performance of detectors on scenarios underrepresented in datasets but common in real-world settings. To address this, we developed MINTIME, a spatiotemporal deepfake detector effective in complex scenarios such as videos featuring multiple identities. Additionally, we investigated Super-Resolution as an adversarial attack against deepfake detectors. Our findings reveal that even legitimate super-resolution can cause pristine images to be misclassified as fake. More concerning, applying super-resolution to fake images can mask artifacts, leading to false negatives. We proposed methods to improve robustness against this attack vector.

In conclusion, this thesis proposes innovative solutions to key challenges in deepfake detection, including fingerprint-based training and multimodal techniques, while also emphasizing the need for further research to address open issues in generalization and adversarial robustness.

Conversational Agents for End-User Control of Smart Spaces

Author: Simone Gallo, University of Pisa, Italy
Supervisors: Fabio Paternò (Cnr-Isti), Alessio Malizia (University of Pisa)

This thesis explores the development of conversational agents that empower non-technical users to manage IoT automation within smart environments. By combining principles of end-user development with conversational interfaces, this research addresses the demand for accessible, adaptable systems that allow users to create, customise, and control automations through natural language. The study presents a progressive series of prototypes, from RuleBot V.1 to the advanced LLM-based RuleBot++, each offering insights into the usability and efficiency of conversational agents compared to traditional IoT automation interfaces. The findings reveal both the strengths and limitations of natural language interfaces in supporting user-driven automation, with user testing demonstrating marked improvements in usability and user satisfaction across iterations. This research further extends the application of the RuleBot++ prototype by integrating it into the humanoid Pepper robot, enabling users to program robotic behaviours through voice interaction. In conclusion, the thesis discusses the main takeaways and presents ongoing work, underscoring the potential for conversational agents to redefine smart home management for non-expert users.

Model Transformation and Property Preservation in Rigorous Software Development

Author: Gullelala Jadoon, University of Florence, Italy
Supervisors: Maurice ter Beek, Alessio Ferrari (Cnr-Isti)

Model-driven development (MDD) is a methodology for developing complex software systems, where model transformation plays a vital role. However, ensuring consistent preservation of desired characteristics during model transformations remains a significant challenge, leading to potential inconsistencies and deficiencies in the final system. Through a systematic literature review and the analysis of 202 studies published during 2000-2024, this thesis identifies the major challenges and benefits of model transformation and property preservation.

These challenges are then addressed by introducing a novel Property Preservation Framework (PPF) that focuses on preserving both functional and non-functional properties during model transformations.

The PPF integrates the application of AI methodologies, constraint-checking strategies, and model validation mechanisms into the model transformation workflow. By prioritizing property specification, verification, and preservation, the PPF facilitates the identification and rectification of property violations at multiple transformation stages. This systematic approach significantly enhances overall consistency and reliability, increasing model precision and dependability. The PPF also serves to preserve non-functional requirements (NFRs) in goal models using meta-models of software product lines (SPLs) and manage inconsistent NFRs. The PPF utilizes product and domain meta-models to accurately capture and represent NFRs, addressing construct validity concerns. This enhances the credibility and generalizability of findings in SPL engineering, thus advancing goal-oriented modeling and NFR preservation practices.

Overall, this thesis highlights the significance of effective model transformation and preservation strategies. By offering comprehensive frameworks for the preservation of essential features, it contributes substantially to solving major challenges in MDD, as a step towards ensuring the development of accurate and reliable models.

End User Control of Smart Home Automations Through Mobile Augmented Reality and Recommender Systems

Author: Andrea Mattioli, University of Pisa, Italy
Supervisors: Fabio Paternò (Cnr-Isti), Mario G.C.A. Cimino (University of Pisa)

In recent years, the so-called smart objects have become increasingly common in our everyday environments, making the vision of an Internet of Things a reality. They promise to improve our lives, optimising comfort, energy management, and in general our daily well-being. To make the most of the possibilities of these ecosystems of connected objects and services, it is necessary to use platforms that allow their coordinated use, enabling the creation of multi-object automations.

However, there are several problems with these platforms, both commercial and research ones. Firstly, it isn't easy to balance expressiveness with ease of use. Indeed, home automation platforms range from allowing only simple "if-then" automations to the possibility of defining actual programs. However, the most expressive platforms tend to become difficult to use and unengaging. An unbalanced expressive capacity can make these platforms of little use to the user.

Furthermore, the selection and configuration of functionalities to be used in automations can be a complex operation. It is, therefore, necessary to find solutions to make this task easier for users, using representations that allow them to form an accurate mental model of the functionalities of the tools and consequently use them correctly.

This dissertation will present how mobile-enabled Augmented Reality can minimise these problems and empower users to seamlessly create automation in smart environments. It will also discuss how recommender systems can be introduced in this context, further facilitating operations on the platform. Two main cycles of platform development, each culminating in a user study, will be reported, along with the additional preliminary studies and interviews conducted. Together, these activities enabled us to answer the defined research questions and delineate the 'lessons learned', which can serve the development of further solutions in this regard.

Exploring Machine Learning for Image Enhancement and Multimedia Understanding towards Autistic People's behavior analysis

Author: Ali Reza Omrani, Università Campus Biomedico, Rome
Supervisor: Davide Moroni, Cnr-Isti

This dissertation investigates the application of machine learning for monitoring and analyzing signs and behaviors associated with Autism Spectrum Disorder (ASD), with a particular focus on image and audio processing.

From an Explainable Artificial Intelligence (XAI) perspective, the research begins by examining the capabilities of deep learning models for analyzing the facial features of autistic and neurotypical individuals. A crucial prerequisite for this analysis is enhancing input image quality. To achieve this, we propose a multi-exposure High Dynamic Range (HDR) imaging method that improves image detail through segmentation and a deep learning-based reconstruction pipeline. A corresponding paper in IEEE ICASSPW highlights that while empirical thresholding initially showed superior performance over Otsu's method, its integration led to model overfitting, prompting the adoption of Otsu-based segmentation. The complete reconstruction network—composed of a Visual Attention Module (VAM), attention and alignment modules, and refinement stages—outperformed existing state-of-the-art techniques, as published in IEEE Access.

The research further explores the use of Vision Transformers and Residual Networks (ResNets) to differentiate children with ASD from their neurotypical peers, achieving a 92% accuracy rate. We utilized XAI techniques to clarify the model's decision-making process, with these findings submitted to the Journal of Research in Autism Spectrum Disorders.

Finally, the study investigates the measurement of vocal stereotypy in children with ASD by applying machine learning to audio files, which resulted in the development of tailored models for individual patients. These findings were submitted to the Journal of Applied Behavior Analysis.

Formal Representation of Geospatial Knowledge in Narratives

Author: Nicolò Pratelli, University of Pisa, Italy
Supervisors: Nicola Tonellotto (University of Pisa), Valentina Bartalesi (Cnr-Isti)

Geospatial knowledge is central to how narratives (historical, literary, or journalistic) are constructed, understood, and communicated. Yet, despite their spatial richness, narratives remain underrepresented in semantic knowledge models, particularly within the Semantic Web. This thesis addresses the gap by introducing the Geospatial Narrative Ontology (NOnt+S), an ontology designed to model geospatial narratives with enhanced expressiveness, interoperability, and reasoning capabilities. Built as an extension of the Narrative Ontology (NOnt) and grounded in the CIDOC Conceptual Reference Model (CIDOC CRM), NOnt+S leverages OWL 2 DL for logical consistency and integrates GeoSPARQL, the OGC standard for representing and querying geospatial data, to enable spatial interoperability and advanced geospatial reasoning.

The core innovation of NOnt+S lies in its ability to perform both geospatial and semantic reasoning. Using an embedded semantic reasoner (Openllet), the ontology can infer new knowledge from existing narrative data, such as implicit spatial relationships or temporally linked events, thereby revealing narrative patterns that are not explicitly encoded in the source material. By aligning narrative structures such as *fabula* and plot with both qualitative and quantitative spatial constructs, and incorporating GeoSPARQL geometries, NOnt+S supports spatial querying and reasoning capabilities.

The ontology is validated through real-world case studies in two interdisciplinary domains: the MOVING project, which focuses on bioeconomy value chains in mountain ecosystems, and the IMAGO project, which analyses medieval geographical literature. These case studies demonstrate NOnt+S's ability to construct semantically rich knowledge graphs, enable GeoSPARQL-based spatial queries, and infer meaningful patterns that extend beyond the original narrative data.

A key innovation of the thesis is the integration of Story Maps, which serve as interactive visual tools combining geospatial data with narrative elements. This enhances user engagement and provides an intuitive means of navigating complex narrative structures across space and time. The framework also adheres to FAIR and Linked Open Data principles, promoting reusability and integration across systems.

By combining formal narrative modelling, GeoSPARQL-enabled spatial reasoning, and semantic inference, this thesis introduces a scalable and interoperable approach to geospatial narrative representation, advancing both Semantic Web research and digital storytelling.

Welcome aboard!

Maria Giulia Brunacci
Admin Staff

Chiara Busnelli
Admin Staff

Daniela Cocco
Admin Staff

Gian Luca Vannini
Technician

Conferences - Co-organized by ISTI

AI in Agriculture (AgriAI) Thematic Session of the 20th Conference on Computer Science and Intelligence Systems FedCSIS 2025 (IEEE#66292)
September 14-17, 2025, Kraków, Poland
https://2025.fedcsis.org/thematic/agriai

18th International Conference on
Similarity Search and Applications, SISAP 2025
October 1-3, 2025, Reykjavik, Iceland
https://www.sisap.org/2025/

21st International Conference on
Content-based Multimedia Indexing, CBMI 2025
October 22-24, 2025, Dublin. Ireland
https://www.cbmi2025.org/

ACM MM 2025 Workshop on Multimedia
Analytics with Multimodal Large Language Models
October 27-28, 2025, Dublin. Ireland
https://ma-llm25.github.io/

6th International Conference on Reliability, Safety, and
Security of Railway Systems (RSSRail 2025)
November 26-28, 2025, Pisa, Italy
https://rssrail2025.isti.cnr.it/

28th Brazilian Symposium on Formal Methods (SBMF 2025)
December 3-5, 2025, Recife, Pernambuco, Brazil
http://sbmf2025.ufrpe.br/

32nd International Conference on Multimedia Modeling
January 29-31, 2026, Prague, Czech Republic
https://mmm2026.cz/

ISTI News is published by the Institute of Information Science and Technologies "A. Faedo"
National Research Council
Via G. Moruzzi 1–56124 Pisa Italy
http://www.isti.cnr.it

ISSN: 2724-6566
DOI: 10.32079/ISTINews

Editorial Secretariat
segreteriascientifica@isti.cnr.it

Editorial Board
Liana Falconetti
Silvia Giannini
Claudia Raviolo
Roberto Scopigno
Maurice Henri ter Beek
Adrian Wallwork

Layout and Design
Beatrice Rapisarda

Copyright notice
All authors, as identified in each article, retain copyright of their work.
The authors are responsible for the technical and scientific contents of their work

Privacy statement
The personal data (names, email addresses...) and the other information entered in ISTI News will be treated according with the provision set out in Legislative Degree 196/2003 (known as Privacy Code) and subsequently integration and amendement

Legal representative of the Institute of Information Science and Technologies "A. Faedo": Dott. Roberto Scopigno

ISTI News is not for sale but is distributed for purposes of study and research and published online at
https://www.isti.cnr.it/en/isti-news

To subscribe: send your email to istinews@isti.cnr.it