A Concept Drift Stream Generator for Intrusion Detection Systems

Gabriele Nicolò Costa[1,†], Alessandra De Paola[1,2,†], Salvatore Drago[3,*,†], Pierluca Ferraro[1,2,*,†] and Giuseppe Lo Re[1,2,†]

[1]Department of Engineering, University of Palermo, Italy
[2]Cybersecurity National Lab, CINI - Consorzio Interuniversitario Nazionale per l'Informatica
[3]IMT School for Advanced Studies Lucca, Italy

Abstract

Intrusion Detection Systems (IDSs) based on machine-learning techniques have become a major research focus, as they are crucial for identifying anomalies in the network traffic logs to detect malicious activity. Although such systems achieve high performance during testing, they experience a decline in accuracy over time when deployed in real-world scenarios due to concept drift. Over time, patterns in both benign and malicious network traffic evolve, rendering the training data obsolete and leading to performance degradation. This has led to a growing interest in concept drift detection and the use of adaptation policies such as online and incremental machine learning. However, testing system performance over time, both for drift detection and adaptation, requires labeled real network datasets that exhibit concept drift, with temporal indications of when the drift occurs. The absence of such datasets has led to the use of synthetic drift data generators, which, however, force researchers to work with datasets that are overly simplistic and insufficiently challenging for machine learning algorithms compared to real network datasets. To overcome this limitation, this work proposes a Concept Drift Stream Generator for Intrusion Detection Systems that, starting from a real network dataset, generates data streams exhibiting concept drift. This enables the evaluation of system performance under realistic concept drift conditions while preserving the complexity of the original dataset.

Keywords

Threat Detection, Online Intrusion Detection System, Machine Learning, Concept Drift, Drift Data Generator

1. Introduction and Related Work

In the past decade, there has been a growing focus on cybersecurity, particularly in network security. The increasing number of devices interconnected in networks, also due to the growing adoption of IoT technologies in different contexts, such as industrial IoT, smart homes and smart cities [1], has made indispensable the need to detect threats to these distributed systems [2]. In such a scenario, Intrusion Detection Systems (IDSs) emerging as a potential solution to protect systems for malicious activities. These systems aim to detect intrusions early enough to raise alerts about malicious activity that deviates from normal traffic. The most prominent approaches for developing IDSs involve the adoption of Machine Learning (ML) and, more broadly, Artificial Intelligence (AI) algorithms. The recent literature [3, 4] showed that these systems are promising tools for assisting network administrators in handling network anomalies.

However, many recent studies [5, 6, 7, 8] have shown that, in the field of cybersecurity, although AI-based systems achieve high performance in experimental evaluations, they experience performance degradation over time due to the phenomenon of concept drift. Over time, patterns in both benign and malicious network traffic evolve, rendering the training data obsolete and leading to performance degradation [9, 10].

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
†These authors contributed equally.
gabriele.costa03@community.unipa.it (G. N. Costa); alessandra.depaola@unipa.it (A. De Paola); salvatore.drago@imtlucca.it (S. Drago); pierluca.ferraro@unipa.it (P. Ferraro); giuseppe.lore@unipa.it (G. Lo Re)
0000-0002-7340-1847 (A. De Paola); 0009-0009-0367-0484 (S. Drago); 0000-0003-1574-1111 (P. Ferraro); 0000-0002-8217-2230 (G. Lo Re)

Typically, concept drift is classified into four types based on its temporal characteristics and intensity. An *abrupt drift* or *sudden drift* refers to a sudden change in distribution at a specific point in time. In contrast, *gradual drift* occurs when changes take place gradually over a period of time. In *incremental drift*, samples during the transition period are drawn from both distributions with varying probabilities. Finally, *recurrent drift* refers to the reappearance of past distributions, usually due to seasonality.

From the perspective of a flow-based IDS [11], which analyzes statistical features for each communication within a company or public administration subnet, drift in benign network traffic patterns may arise from the activation of new web services accessible both internally and externally, changes in the network infrastructure or variations in staff behavior. On the other hand, drift in malicious traffic may arise, for instance, from zero-day attacks [12, 13] or adversarial learning attacks [14, 15, 16, 17]. As a result, there has been increasing interest in concept drift detection [18] and the adoption of adaptive machine learning approaches to adjust models to evolving data distributions [19]. Current adaptation techniques can be broadly classified into two categories: detect and retrain, which involves discarding the existing model and training a new one using updated data; and detect and update [20, 21], which incrementally refines the existing model based on new data.

However, evaluating system performance over time, both in terms of drift detection and adaptation, requires labeled real-world network datasets that exhibit concept drift and include precise temporal annotations of drift events. The lack of such datasets has led researchers to rely on synthetic drift generators [22]. However, these concept drift generators lack realistic scenarios: they simulate concept drift by sampling from different datasets (Agrawal Generator [23]) or by perturbing features (LED-Generator [24]). While useful, these generators often produce data that is overly simplistic and fail to capture the complexity and challenges found in real network traffic, limiting their effectiveness for testing machine learning algorithms. In [25], the authors propose a drift generator based on real datasets. However, this system is limited to manipulating only the temporal location of the drift and does not allow control over other key aspects such as the type and intensity, which are critical to system performance and incremental adaptation mechanisms.

To overcome this limitation, this work proposes a Real Dataset-based Concept Drift Stream Generator for Intrusion Detection Systems (*RD-ConceptDriftGenerator*). Starting from a real network dataset, this tool enables the generation of controlled data streams that simulate scenarios affected by specific types and intensities of concept drift. This allows for the evaluation of IDSs while preserving the complexity of the original dataset.

The experimental evaluation, performed on a real network dataset, proves the effectiveness of the generator by showing that it can generate streams with different concept drift characteristics from the same dataset. The results also show how the performance of a machine learning-based IDS is affected over time by the characteristics of the concept drift present in the stream.

The remainder of the paper is structured as follows. Section 2 describes the proposed *RD-ConceptDriftGenerator*. Section 3 outlines the experimental evaluation. Finally, Section 4 presents the conclusions and future research directions.

2. RDConceptDriftGenerator Architecture

This section presents the architecture of *RD-ConceptDriftGenerator*, which consists of two main components: a *concept generator* and a *drift streamer*. Additionally, as shown in Figure 1, the *RD-ConceptDriftGenerator* requires as input the source dataset and a set of parameters that define the desired characteristics of the concept drift in the generated stream (Streamed Dataset).

First, the Concept Generator module divides the original dataset into two macro-clusters, then further divides each macro-cluster into different micro-clusters. These steps are carried out using classical clustering algorithms [26]. When the source dataset has multi-class labels, this step can be guided by creating macro- and micro-clusters aligned with those labels. The macro-cluster phase splits the data into benign and malicious clusters, while the micro-cluster phase creates a further subdivision reflecting different types of malicious traffic and various benign communication behaviors.

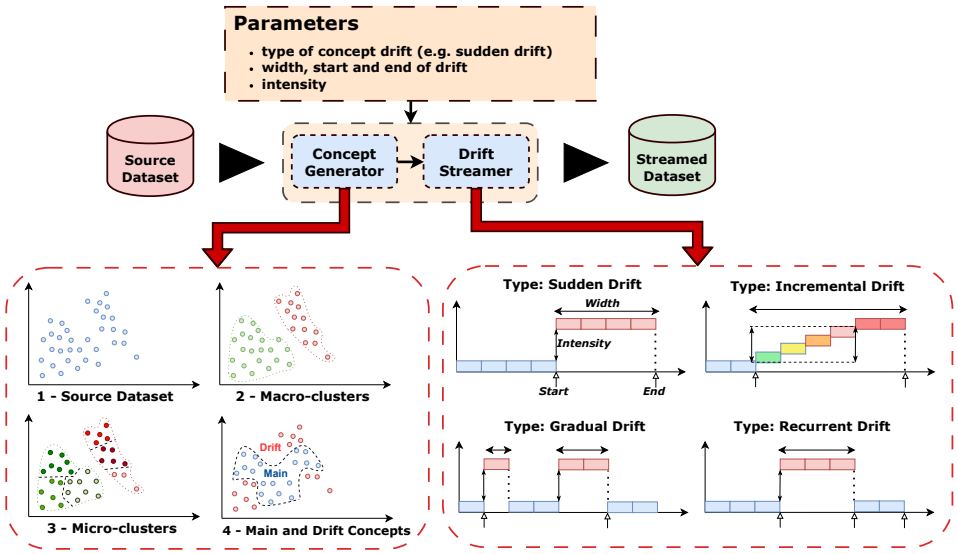

Figure 1: Components of RD-ConceptDriftGenerator.

Then, to create a concept drift with specific characteristics, a cluster similarity metric, such as AMI [27], is computed among micro-clusters within each macro-cluster. Thus, micro-clusters are grouped into two sets: Main Concept and Drift Concept. The Drift Concept is created by selecting the most similar micro-clusters, starting with one micro-cluster from each macro-cluster, until the number of records required by the *width* parameter (expressed as a number of elements) is reached. The remaining micro-clusters form the Main Concept. The goal is to obtain two sets, each containing records from both macro-clusters (both benign and malicious traffic). Additionally, the method aims to maximize the similarity between benign and malicious micro-clusters within each set while minimizing the similarity between the two sets. Once the Main Concept and Drift Concept are formed, the drift elements are assigned an anomaly score based on their deviation from the Main Concept.

Finally, the drift streamer samples from the Main and Drift Concept sets to generate concept drift according to the input parameters (the time in which the drift starts, its width and its ending time). The streamer uses the anomaly scores of the Drift Concept elements to control the desired intensity and type of drift while respecting the *width* parameter.

Note that the Streamed Dataset does not necessarily have the same number of elements as the original dataset. This allows the streamer to create drifts with different characteristics more easily. Furthermore, to avoid biasing the performance of the machine learning model used as an IDS, the streamer samples from the two sets without replacement. Additionally, no data augmentation is applied, ensuring that no unrealistic records are introduced in the Streamed Dataset.

3. Experimental Evaluation

This section highlights the ability of *RD-ConceptDriftGenerator* to introduce concept drift of different types based on a realistic network dataset.

The experiments were conducted using the *CIC-IDS2017* [28] dataset as the source dataset. This dataset offers a comprehensive representation of modern network traffic, including benign traffic and a wide range of attack types, covering various phases and attack methods. These attacks include Distributed Denial of Service (DDoS), Brute Force (SSH and HTTP), Web attacks (XSS, SQL Injection), Infiltration, Botnet traffic, and others, all simulated in a real enterprise environment, with flow-based features and labels.

RD-ConceptDriftGenerator was used to create two streamed datasets, each exhibiting sudden and recurrent concept drift, respectively, using the parameters summarized in Table 1. These represent the

Table 1
Parameters used with *RD-ConceptDriftGenerator* to generate the two streamed datasets with concept drift.

Type	width	start	end	intensity	Total dataset dimension
Sudden	100 000	100 000	200 000	medium	200 000
Recurrent	50 000	100 000	150 000	high	200 000

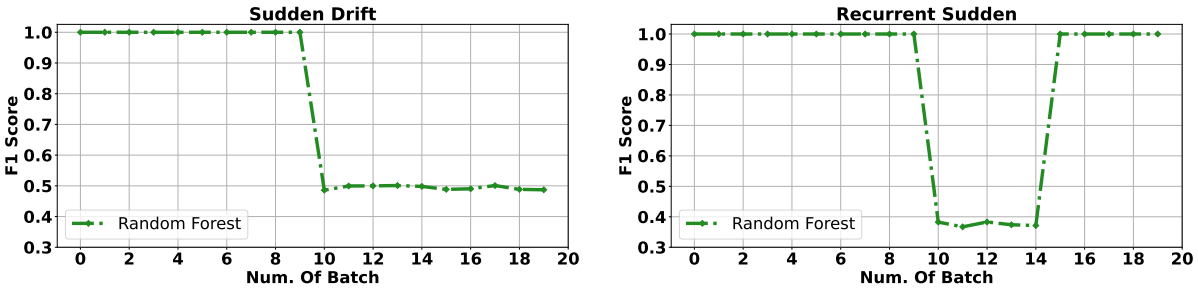

Figure 2: Performance (F1-score over time) of the IDSs tested with two streamed datasets exhibiting different concept drifts, generated by the *RD-ConceptDriftGenerator* from the CIC-IDS2017 dataset.

two main types of drift, as the other types can be derived from them. Specifically, gradual drift can be seen as a recurrent drift with different starting points and widths, while incremental drift is generated by dividing the drift width into multiple sudden drifts with progressively increasing intensity.

To assess whether the two generated datasets contain concept drift with the desired characteristics, they were used to evaluate the online performance of two IDSs. These systems are tested in static mode, which is often used in online learning under concept drift [19] as a baseline to evaluate the impact of concept drift over time. To this end, the machine learning model is trained offline on the first data batch, and its performance, in terms of F1-score, is then evaluated over time, batch by batch, as shown in Figure 2. If the first data batch is representative, the system's performance will remain good and stable until the occurrence of a potential concept drift. A sudden or gradual degradation in performance, on the other hand, will indicate the presence of concept drift. In these experiments, the machine learning model used is Random Forest, which has shown good performance on high-dimensional datasets, even in multi-class settings [29]. Moreover, the two systems were tested using batches of 10 000 samples on a binary task, i.e., distinguishing benign versus malicious traffic.

As shown in Figure 2, in both cases the systems maintain high performance, above 0.9 and close to an F1-score of 1.0, as long as the evaluated records belong to the Main Concept set. Conversely, a sudden drop in performance is observed when concept drift occurs, specifically when samples from the Drift Concept set appear. As can be seen from the batch number (x-axis), the concept drift occurs at batch 10, that is, starting from the 100 000th sample considering batches of 10 000 samples, and it lasts for the expected duration, as reported in Table 1.

For the recurrent concept drift, after the drift ends, performance returns to excellent levels, similar to pre-drift performance. The intensity of the drift significantly affected the systems' performance. In the first case, after the drift, performance stabilized at a value slightly above 0.5, whereas in the recurrent drift case, it oscillates between 0.4 and 0.37, confirming a high drift intensity, which is higher than the intensity set for the sudden drift (medium).

4. Conclusions and Future Work

This work presents *RD-ConceptDriftGenerator*, a novel tool for generating realistic concept drift scenarios in network intrusion detection datasets by leveraging real network data. Unlike existing drift generators, the proposed approach preserves the complexity of real network traffic while enabling precise control over drift characteristics like type and intensity. The experimental evaluation, conducted using two

streamed datasets with sudden and recurrent drift, demonstrated that the generator effectively simulates different drift characteristics, impacting the performance of machine learning-based IDSs as expected. Specifically, IDS performance remained stable before drift and showed significant degradation upon drift occurrence, with recovery observed in the case of recurrent drift. These results highlight the importance of realistic drift simulation for developing and evaluating IDS performance in evolving network environments.

Future work will focus on extending this framework by evaluating the impact of different clustering algorithms, similarity techniques between clusters, and anomaly scoring strategies used by the Concept Generator. Additionally, it may be useful to monitor, through the drift streamer, the balance between benign and malicious records over time, to avoid excessively unrealistic imbalances within a batch.

Acknowledgments

This work was partially supported by the AMELIS project, within the project FAIR (PE0000013), and by the ADELE project, within the project SERICS (PE00000014), both under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] V. Agate, A. De Paola, G. Lo Re, A. Virga, Reliable reputation-based event detection in V2V networks, in: International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability, Springer, 2023, pp. 267–281.

[2] B.-X. Wang, J.-L. Chen, C.-L. Yu, An AI-powered network threat detection system, IEEE Access 10 (2022) 54029–54037.

[3] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, M. A. Khan, Performance analysis of machine learning algorithms in intrusion detection system: A review, Procedia Computer Science 171 (2020) 1251–1260.

[4] V. Agate, F. Concone, A. De Paola, P. Ferraro, G. Lo Re, M. Morana, Adaptive ensemble learning for intrusion detection systems, in: CEUR Workshop Proceedings, volume 3762, 2024.

[5] A. Augello, A. De Paola, G. Lo Re, Hybrid multilevel detection of mobile devices malware under concept drift, Journal of Network and Systems Management 33 (2025) 36.

[6] S. Agrahari, A. K. Singh, Concept drift detection in data stream mining : A literature review, Journal of King Saud University - Computer and Information Sciences 34 (2022) 9523–9540. URL: https://www.sciencedirect.com/science/article/pii/S1319157821003062. doi:https://doi.org/10.1016/j.jksuci.2021.11.006.

[7] V. Agate, A. De Paola, S. Drago, P. Ferraro, G. Lo Re, Enhancing IoT network security with concept drift-aware unsupervised threat detection, in: 2024 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2024, pp. 1–6.

[8] A. Augello, A. De Paola, G. L. Re, M2fd: Mobile malware federated detection under concept drift, Computers & Security (2025) 104361.

[9] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM computing surveys (CSUR) 46 (2014) 1–37.

[10] V. Agate, S. Drago, P. Ferraro, G. Lo Re, Anomaly detection for reoccurring concept drift in smart environments, in: 2022 18th International Conference on Mobility, Sensing and Networking (MSN), IEEE, 2022, pp. 113–120.

[11] M. F. Umer, M. Sher, Y. Bi, Flow-based intrusion detection: Techniques and challenges, Computers & Security 70 (2017) 238–254.

[12] T. Zoppi, A. Ceccarelli, T. Puccetti, A. Bondavalli, Which algorithm can detect unknown attacks? comparison of supervised, unsupervised and meta-learning algorithms for intrusion detection, Computers & Security 127 (2023) 103107. URL: https://www.sciencedirect.com/science/article/pii/S0167404823000172. doi:https://doi.org/10.1016/j.cose.2023.103107.
[13] A. De Paola, S. Drago, P. Ferraro, G. Lo Re, Detecting zero-day attacks under concept drift: An online unsupervised threat detection system, in: CEUR Workshop Proceedings, 8th Italian Conference on Cybersecurity, ITASEC, volume 2024, 2024.
[14] K. He, D. D. Kim, M. R. Asghar, Adversarial machine learning for network intrusion detection systems: A comprehensive survey, IEEE Communications Surveys & Tutorials 25 (2023) 538–566.
[15] F. Batool, F. Canino, F. Concone, G. Lo Re, Morana, A black-box adversarial attack on fake news detection systems, in: CEUR Workshop Proceedings, volume 3731, 2024.
[16] T. S. Sethi, M. Kantardzic, Handling adversarial concept drift in streaming data, Expert systems with applications 97 (2018) 18–40.
[17] F. Concone, S. Gaglio, A. Giammanco, G. Lo Re, M. Morana, Adverspam: Adversarial spam account manipulation in online social networks, ACM Transactions on Privacy and Security 27 (2024) 1–31.
[18] F. Hinder, V. Vaquet, B. Hammer, One or two things we know about concept drift—a survey on monitoring in evolving environments. part a: detecting concept drift, Frontiers in Artificial Intelligence 7 (2024) 1330257.
[19] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, Learning under concept drift: A review, IEEE transactions on knowledge and data engineering 31 (2018) 2346–2363.
[20] M. A. Shyaa, N. F. Ibrahim, Z. B. Zainol, R. Abdullah, M. Anbar, Reinforcement learning-based voting for feature drift-aware intrusion detection: An incremental learning framework, IEEE Access (2025).
[21] F. Camarda, A. De Paola, S. Drago, P. Ferraro, G. Lo Re, Managing concept drift in online intrusion detection systems with active learning (2025).
[22] J. Montiel, J. Read, A. Bifet, T. Abdessalem, Scikit-multiflow: A multi-output streaming framework, Journal of Machine Learning Research 19 (2018) 1–5.
[23] R. Agrawal, T. Imienski, A. Swamy, Database mining: A performance perspective, ieee tran, On Knowledge and Data Engg (1991).
[24] L. Breiman, J. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, Routledge, 2017.
[25] B. Lin, C. Huang, X. Zhu, N. Jin, Realdriftgenerator: A novel approach to generate concept drift in real world scenario, in: 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2024, pp. 1124–1129.
[26] D. Xu, Y. Tian, A comprehensive survey of clustering algorithms, Annals of data science 2 (2015) 165–193.
[27] N. Vinh, J. Epps, J. Bailey, Information theoretic measures for clusterings comparison: Variants, Properties, Normalization and Correction for Chance 18 (2009).
[28] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, et al., Toward generating a new intrusion detection dataset and intrusion traffic characterization., ICISSp 1 (2018) 108–116.
[29] V. Agate, D. Felice Maria, A. De Paola, P. Ferraro, G. Lo Re, M. Morana, A behavior-based intrusion detection system using ensemble learning techniques., in: ITASEC, 2022, pp. 207–218.

CEUR Workshop Proceedings
ceur-ws.org
ISSN 1613-0073

Vol-4121

urn:nbn:de:0074-4121-x

Ital-IA-TW 2025
Thematic Workshops at Ital-IA 2025

Joint Proceedings of the Thematic Workshops at Ital-IA 2025 colocated with the 5th National Conference on Artificial Intelligence, organized by CINI (Ital-IA 2025)

Trieste, Italy, June 23-24, 2025.

Edited by

Luca Manzoni [1]
Luca Bortolussi [1]
Giulia Cisotto [1]
Fabio Anselmi [1]

[1] University of Trieste, Department of Mathematics, Informatics and Geosciences (MIGe), Via Weiss 2, 34128, Trieste, Italy

Table of Contents

Summary: There were 130 papers submitted for peer-review to this workshop. Out of these, 104 papers were accepted for this volume. All of them are short papers.

Workshop 1: Responsible and Trustworthy AI

Workshop 2: AI and Sustainability

- Towards a sustainable future: AI-powered solutions in agriculture and green energy
  Matteo Tortora, *Giulia Romoli*, *Valerio Guarrasi*, *Rosa Sicilia*, *Francesco Conte*, *Federico Silvestro*, *Paolo Soda*
- Applications of deep learning super-resolution methods for coastal ocean modelling
  Federica Adobbati, *Lorenzo Bonin*, *Gianpiero Cossarini*, *Valeria Di Biagio*, *Fabio Giordano*, *Paolo Lazzari*, *Luca Manzoni*, *Stefano Piani*, *Marco Reale*, *Stefano Salon*, *Cosimo Solidoro*, *Stefano Querin*
- AI for sustainability should be sustainable in its own: results from project FAIR-Spoke 9
  Luigi Pontieri, *Pietro Sabatino*, *Francesco Scala*
- Explaining reinforcement learning policies for power grid operations
  Luca Marzari, *Francesco Leofante*, *Enrico Marchesini*
- Advancing green and fair AI: a research perspective on environmental and social sustainability
  Loris Belcastro, *Riccardo Cantini*, *Fabrizio Marozzo*, *Alessio Orsino*, *Domenico Talia*, *Paolo Trunfio*
- Advancing sustainable AI: research perspectives from the CINI-AIIS Lab at Federico II
  Antonio Galli, *Antonino Ferraro*, *Vincenzo Moscato*, *Lidia Marassi*, *Carlo Sansone*, *Stefano Marrone*, *Antonio Elia Pascarella*
- Requirements analysis in ARCAD-IA project
  Pietro Patrizio Ciro Aucelli, *Francesco Camastra*, *Angelo Ciaramella*, *Emanuel Di Nardo*, *Alessio Ferone*, *Antonio Maratea*, *Raffaele Montella*, *Antonino Staiano*
- Multi-objective particle swarm optimization for environmental risk/benefit analysis with pre-assignment strategy
  Luca Puzzoli, *Gabriele Sbaiz*, *Luca Manzoni*
- AI4Materials: a manifesto for AI-driven scientific discovery in materials science
  Eros Radicchi, *Usman Syed*, *Federico Cunico*, *Michele Cassetta*, *Uzair Khan*, *Paolo Marone*, *Alberto Scarsini*, *Filiberto Semenzin*, *Francesco Enrichi*, *Francesco Setti*, *Adolfo Speghini*, *Marco Cristani*

Workshop 3: AI for Public Administration

- AI for public administrations: research at UNIFI DISIT Lab
  Pierfrancesco Bellini, *Stefano Bilotta*, *Enrico Collini*, *Marco Fanfani*, *Luciano Alessandro Ipsaro Palesi*, *Paolo Nesi*, *Gianni Pantaleo*
- Small Language Models for Public Administration: Towards Sustainable, Trustworthy, and Transparent AI Systems
  Francesco Recanati
- Enabling research collaboration with AI: the BI4E experience with large language models
  Gianluca Amato, *Marcello Costantini*, *Luca Di Vita*, *Francesca Ferri*, *Paolo Melchiorre*, *Maria Chiara Meo*, *Francesca Scozzari*, *Matteo Vitali*

- Neural technologies for predictive analytics in tourism governance processes
  Natalia Pichierri, *Daniele Margiotta*, *Andriy Shcherbakov*, *Danilo Croce*, *Roberto Basili*
- AI-supported certification of family-friendly organizations
  Davide Vandelli, *Sara Tonelli*, *Pietro Marzani*, *Alessio Palmero Aprosio*
- Towards an ontology network for Italian educational institutions
  Luigi Asprino, *Aldo Gangemi*, *David Grassi*, *Giorgia Lodi*, *Nicola Mallogi*, *Elettra Morini*, *Maria Teresa Sagri*, *Eniko Tolvay*
- Generative artificial intelligence in public administration: an integrated framework for regulatory compliance, ethical governance, and digital transformation
  Gaetano Riccio
- Prediction of unconstitutional index of Italian bills
  Monica Palmirani, *Michele Corazza*, *Generoso Longo*, *Salvatore Sapienza*, *Pier Francesco Bresciani*, *Faria Ferooz*

Workshop 4: AI for Health and Medicine

- Detecting anomalies in healthcare processes: a k-NN graph-based approach
  Iuliana Malina Grigore, *Azin Moradbeikie*, *Sylvio Barbon Junior*
- Advances in AI for health and medicine @ Computer-Human Interaction Laboratory (CHILab)
  Germano Ammirata, *Salvatore Contino*, *Luca Cruciata*, *Irene Siragusa*, *Roberto Pirrone*
- Lively Ageing: an integrated system of services and technologies to enhance the well-being of elderly
  Martina Casari, *Giovanni Bonisoli*, *Laura Po*
- Modeling and analyzing non-IID data in federated learning based ECG arrhythmia detection scenarios
  Davide Cantoro, *Angela-Tafadzwa Shumba*, *Gianluigi Semeraro*, *Mattia Cotardo*, *Davide Rollo*, *Teodoro Montanaro*, *Ilaria Sergi*, *Massimo De Vittorio*, *Luigi Patrono*
- SICURA: a symbiotic AI system to support pharmacological management
  Raffaele Manna, *Giulia Speranza*, *Maria Pia Di Buono*, *Gianmarco Troia*, *Giovanni Vizzini*
- Federated approach for glioblastoma radiogenomic methylation classification
  Eriberto Andrea Franchi, *Elena De Momi*, *Alberto Redaelli*
- Artificial intelligence in digital medicine across biomedical signal processing and medical imaging
  Simone Bove, *Rosanna Ferrara*, *Martino Giaquinto*, *Gennaro Percannella*, *Alessia Saggese*, *Mattia Sarno*, *Francesco Tortorella*, *Mario Vento*
- Feature importance via Shapley values in random forests for sleep apnea and hypopnea detection
  Giulia Cisotto, *Shayan Sharifi*, *Shahla Sadeghzadehdarandash*, *Leonardo Badia*
- Transparent machine learning for type 1 diabetes diagnosis from gene

expression data
Rosa Carotenuto, *Viviana Pentangelo*, *Antonio Della Porta*, *Fabio Palomba*

- Towards resilient ocular anomaly detection for health monitoring
  Giovanni Mancini, *Antonio Visconti*, *Luigi Libero Lucio Starace*, *Sergio Di Martino*, *Daniel Riccio*
- Transforming healthcare with AI: the work of CINI-AIIS at Federico II University
  Mariano Barone, *Domenico Benfenati*, *Salvatore Capuozzo*, *Giovanni Maria De Filippis*, *Adriano De Simone*, *Simona Fioretto*, *Kyriakos Kristofer Georgiou*, *Michela Gravina*, *Julian Gortz*, *Lidia Marassi*, *Stefano Marrone*, *Enea Vincenzo Napolitano*, *Valeria Panzetta*, *Giuseppe Pontillo*, *Giuseppe Riccio*, *Antonio Romano*, *Cristiano Russo*, *Carmela Russo*, *Cristian Tommasino*, *Mariano Sirignano*, *Konstantinos Siettos*, *Elio Masciari*, *Antonio Maria Rinaldi*, *Paolo Antonio Netti*, *Vincenzo Moscato*, *Carlo Sansone*
- Health-related NLP activities at IDSIA
  Fabio Rinaldi, *Andrea Franchini*, *Roberta Noseda*, *Oscar William Lithgow Serrano*, *Joseph Cornelius*, *Lorenzo Ruinelli*, *Amos Colombo*, *Sandra Mitrovic*, *Alessandro Ceschi*
- HEMERA: H&E-to-HER2 encoding for morphology-enhanced region annotation
  Daniel Riccio, *Francesco Longobardi*, *Mara Sangiovanni*, *Maria Frucci*, *Nadia Brancati*, *Mariacarla Staffa*, *Lorenzo D'Errico*, *Antonio Ciccarelli*
- Leveraging AI for signal and image analysis in medicine and health
  Marco Cafiso, *Andrea Carboni*, *Claudia Caudai*, *Sara Colantonio*, *Francesco Conti*, *Mario D'Acunto*, *Said Daoudagh*, *Giulio Del Corso*, *Danila Germanese*, *Giacomo Ignesti*, *Gianmarco Lazzini*, *Giuseppe Riccardo Leone*, *Massimo Magrini*, *Davide Moroni*, *Francesca Pardini*, *Maria Antonietta Pascali*, *Paolo Paradisi*, *Federico Volpini*
- Explainability in breast cancer detection
  Ijaz Ahmad, *Alessia Amelio*, *Daniela Cardone*, *Eliezer Zahid Gill*, *Francesca Scozzari*
- Understanding and extrapolating from healthcare data through machine learning: a case study on a COVID-19 Dataset
  Michele Rispoli, *Francesco Salton*, *Andrea Rocca*, *Paola Confalonieri*, *Marco Confalonieri*, *Alberto d'Onofrio*, *Luca Manzoni*
- Automatic radiology report generation: evaluating the alignment of NLP metrics with radiologist assessment
  Andrea Santomauro, *Ivan Gallesio*, *Xhorxhi Kaci*, *Giorgio Leonardi*, *Luigi Portinale*
- A comparative study of speech-to-text for Italian
  Michela Quadrini, *Michele Loreti*, *Mattia Luciani*, *Ivano Corradetti*, *Carboni Matteo*, *Stefano Vagnoni*
- Enhancing medication safety with LLMs
  Gabriele De Vito, *Filomena Ferrucci*, *Athanasios Angelakis*
- Next-gen health: from multimodal AI to foundation models
  Rosa Sicilia, *Fatih Aksu*, *Alessandro Bria*, *Alice Natalina Caragliano*, *Camillo Maria Caruso*, *Ermanno Cordelli*, *Arianna Francesconi*, *Valerio Guarrasi*, *Giulio*

Iannello, *Guido Manni*, *Massimiliano Mantegna*, *Giustino Marino*, *Daniele Molino*, *Elena Mulero Ayllon*, *Filippo Ruffini*, *Linlin Shen*, *Matteo Tortora*, *Paolo Soda*

- UniCas research initiatives in medicine and healthcare
  Gabriele Lozupone, *Marco Cantone*, *Emanuele Nardone*, *Cesare Davide Pace*, *Ciro Russo*, *Alessandro Bria*, *Tiziana D'Alessandro*, *Claudio De Stefano*, *Francesco Fontanella*, *Claudio Marrocco*, *Mario Molinara*, *Alessandra Scotto di Freca*
- MediMint: medical imaging and multimodal intelligence laboratory
  Alberto Signoroni, *Mattia Savardi*
- Improving medical code classification for death certificates using ontology-adapted contrastive loss in BERT models
  Kevin Roitero, *Davide Volpi*, *Riccardo Lunardi*, *Mihai Horia Popescu*, *Vincenzo Della Mea*
- Some research directions for multi-agent systems in medicine and health
  Vincenzo Auletta, *Francesco Cauteruccio*, *Diodato Ferraioli*, *Grazia Ferrara*
- From rules to reports: enhancing diabetes prediction interpretability with anchor and LLMs
  Francesco Festa, *Antonio Della Porta*, *Viviana Pentangelo*, *Fabio Palomba*
- AI-driven applications for diagnostic, clinical decision support and prevention in cardiology, orthopedics, and oncology
  Carmen Guzman Garcia, *Monica Mordonini*, *Stefano Cagnoni*
- AI and data science research in life science, system biology and medicine
  Alessio Ansuini, *Alberto Cazzaniga*, *Edith Natalia Villegas Garcia*, *Michele Alessi*, *Federico Barone*, *Stefano Cozzini*, *Francesca Cuturello*, *Fiorella Fabris*, *Valerio Piomponi*, *Lucrezia Valeriani*, *Ruggero Lot*, *Marco Prenassi*

Workshop 5: AI for Industry

- Efficient transformer-based identification of defects in fasteners
  Luigi Portinale, *Giorgio Leonardi*, *Stefania Montani*, *Simone Garau*, *Guido Noce*, *Mario Brumini*, *Enrico Ottaviano*
- An analysis of vision-language models for fabric retrieval
  Francesco Giuliari, *Asif Khan Pattan*, *Mohamed Lamine Mekhalfi*, *Fabio Poiesi*
- Egocentric human-object interaction in industrial scenarios: a case study
  Marco Moltisanti, *Ketty Cantone*, *Antonino Lopes*, *Emanuele Ragusa*, *Rosario Leonardi*, *Francesco Ragusa*, *Antonino Furnari*, *Giovanni Maria Farinella*
- Spatial understanding for industry
  Daniele Di Mauro, *Francesco Ragusa*, *Antonino Furnari*, *Giovanni Maria Farinella*
- AI for industry at FBK
  Alessandro Cimatti, *Fabio Antonelli*, *Luisa Bentivogli*, *Marco Cristoforetti*, *Andrea Micheli*, *Fabio Poiesi*, *Fabio Remondino*, *Angelo Susi*, *Stefano Tonetta*
- AI for environmental, energy, and IoT applications: UniCas industrial research

highlights
Mario Molinara, *Alessandro Bria*, *Domenico Capriglione*, *Tiziana D'Alessandro*, *Claudio De Stefano*, *Luigi Ferrigno*, *Francesco Fontanella*, *Fabrizio Marignetti*, *Claudio Marrocco*, *Alessio Miele*, *Filippo Milano*, *Hamza Mustafa*, *Vincenzo Rega*, *Alessandra Scotto di Freca*, *Luca Cicala*, *Sara Parrilli*, *Mariano Focareta*, *Giuseppe Meoli*, *Michele Vitelli*, *Carmine Bourelly*, *Andrea Amodei*

- Bridging AI and industry: research and innovation from the CINI-AIIS Lab at Federico II University
  Antonio Galli, *Flora Amato*, *David Carlini*, *Alessandro Del Prete*, *Sofia Dutto*, *Narendra Patwardhan*, *Stefano Marrone*, *Vincenzo Moscato*, *Gabriele Piantadosi*, *Carlo Sansone*, *Marco Valle*
- The SmarTwin project, an intelligent digital supply chain twin
  Pietro Catalano, *Angelo Ciaramella*, *Pietro D'Ambrosio*, *Aniello De Prisco*, *Michele Di Capua*, *Luigi Ferraro*, *Salvatore Moscariello*, *Pasquale Perillo*
- An agentic AI for a new paradigm in business process development
  Mohammad Azarijafari, *Luisa Mich*, *Michele Missikoff*
- Reasoning in the financial space with the vadalog system
  Teodoro Baldazzi, *Luigi Bellomarini*, *Emanuel Sallinger*
- MYCAMPANIA.TRAVEL: leveraging generative AI to enhance digital travel experiences
  Aniello Somma, *Sergio Di Meglio*, *Fabio Scippacercola*, *Ermanno Battista*, *Sergio Di Martino*, *Luigi Libero Lucio Starace*
- AI and tourism: a chatbot for small and medium enterprises in the tourism sector
  Stefano Bistarelli, *Marco Cuccarini*
- AI-VaS: empowering edge video analytics with cloud-based services
  Bruno Vento, *Cristian Cerasuolo*, *Andrea Vincenzo Ricciardi*, *Domenico Rocco*, *Stefano Saldutti*, *Antonio Vitale*
- Renewable energy management in industry using AI and system simulation
  Behzad Pirouz, *Francesca Guerriero*
- Automatic positioning of AI microservices on NextG networks to support interactive holograms
  Martina Di Bratto, *Antonio Origlia*, *Jaime Llorca*, *Andrea Detti*, *Alessandro Mauro*, *Marco Grazioso*, *Vincenzo Norman Vitale*, *Valentina Russo*, *Azzurra Mancini*, *Alessandro Perrino*, *Nicholas Napolitano*, *Giovanni Della Corte*, *Antonia Maria Tulino*, *Sergio Piane*, *Margherita Tennirelli*

Workshop 6: AI for Cybersecurity

- Empirical quantification of spurious correlations in malware detection
  Bianca Perasso, *Ludovico Lozza*, *Andrea Ponte*, *Luca Demetrio*, *Luca Oneto*, *Fabio Roli*
- Learning to explain cyberattacks: insights from random forest and decision predicate graphs
  Eron Ponce Pereira, *Azin Moradbeikie*, *Bruno Bogaz Zarpelão*, *Sylvio Barbon Junior*

- Federated hyperdimensional ensembles for mobile malware detection
  Andrea Augello, *Alessandra De Paola*, *Giuseppe Lo Re*, *Gioacchino Zangara*
- Towards explainability framework for cybersecurity domain: a case study using NER
  Stefano Silvestri, *Emanuele Damiano*, *Raffaele Guarasci*, *Mario Ciampi*
- A concept drift stream generator for intrusion detection systems
  Gabriele Nicolò Costa, *Alessandra De Paola*, *Salvatore Drago*, *Pierluca Ferraro*, *Giuseppe Lo Re*
- Causal prototype attention classifier: an interpretable model for credit card fraud detection under extreme class imbalance
  Claudio Giusti, *Mirko Casu*, *Luca Guarnera*, *Sebastiano Battiato*
- Constructing cryptographic primitives with evolutionary computation and cellular automata
  Rocco Ascone, *Firas Ben Ramdhane*, *Luca Manzoni*, *Giuliamaria Menara*, *Gloria Pietropolli*
- Large language models in the software supply chain: challenges and opportunities
  Giacomo Benedetti, *Luca Caviglione*, *Carmela Comito*, *Daniela Gallo*, *Alberto Falcone*, *Massimo Guarascio*, *Angelica Liguori*, *Giuseppe Manco*, *Francesco Sergio Pisani*, *Ettore Ritacco*, *Antonino Rullo*
- The BullyBuster Research Center: a multidisciplinary approach to digital and social violence
  Giulia Orrù, *Guido Colaiacovo*, *Sara Concas*, *Antonio Galli*, *Vincenzo Gattulli*, *Michela Gravina*, *Stefano Marrone*, *Marco Micheletto*, *Wanda Nocerino*, *Angela Procaccino*, *Giovanni Puglisi*, *Lucia Sarcinella*, *Enrico Santoro*, *Grazia Terrone*, *Alessandro Valenti*, *Donatella Curtotti*, *Donato Impedovo*, *Gian Luca Marcialis*, *Carlo Sansone*
- Artificial intelligence in cybersecurity: activities of the CINI-AIIS Lab at University of Naples Federico II
  Roberto Canonico, *Annalisa Navarro*, *Simon Pietro Romano*, *Giancarlo Sperlì*, *Andrea Vignali*

Workshop 7: AI and Robotics

- Intelligent drones with vision technology for sustainable weed mapping in precision agriculture
  Pasquale De Marinis, *Gennaro Vessio*, *Giovanna Castellano*
- Spherical vision for mobile robotics
  Marco La Cascia, *Liliana Lo Presti*
- Investigating the mechanisms of embodied intelligence with evolvable modular soft robots
  Eric Medvet, *Giorgia Nadizar*, *Michel El Saliby*, *Francesco Rusin*, *Berfin Sakallioglu*
- I-BOAT: an autonomous sailboat to monitor the state of the seas
  Cosimo Distante, *Vishali Mankina*, *Andre P.D. Araujo*, *Luiz Marcos Garcia*

Goncalves, *Esteban Clua*, *Arturo Argentieri*

- Evolving roles of intelligent robots in social and industrial domains
  Giuseppe De Simone, *Pasquale Foggia*, *Francesco Rosa*, *Alessia Saggese*, *Mario Vento*
- SISTER: social robots to support biopsychosocial frailty
  Vincenzo Solfrizzi, *Claudia Chiapparino*, *Giovanna Castellano*, *Berardina Nadja De Carolis*, *Davide Lofrese*, *Giuseppe Palestra*, *Aurora Toma*, *Loredana Perla*, *Stefania Massaro*, *Mariateresa Santacroce*, *Giuseppe De Simone*, *Alessia Saggese*, *Mario Vento*

Workshop 8: Generative AI

- A methodological approach for applications of generative AI to support students and teachers
  Davide Ponzini, *Giovanni Adorni*, *Daniele D'Agostino*, *Giorgio Delzanno*, *Giovanna Guerrini*
- Geometry, topology and mechanistic interpretability of large scale AI models
  Alessio Ansuini, *Lorenzo Basile*, *Federica Bazzocchi*, *Matteo Biagetti*, *Alberto Cazzaniga*, *Stefano Cozzini*, *Francesca Cuturello*, *Diego Doimo*, *Yuri Gardinazzi*, *Ruggero Lot*, *Francesco Ortu*, *Emanuele Panizon*, *Valerio Piomponi*, *Tommaso Rodani*, *Alessandro Serra*, *Niccolò Tosato*, *Lucrezia Valeriani*
- Towards sustainable computation: synergizing LLM heat recovery and algorithmic trading for energy-efficient AI systems
  Ivan Letteri, *Pierpaolo Vittorini*, *Tamsir Jobe*
- Prompt engineering approaches for working with biomedical knowledge graphs through LLMs
  Marco Mesiti, *Emanuele Cavalleri*, *Matteo Castagna*, *Paolo Perlasca*, *Darya Shlyk*
- Improving LLM-generated code via genetic improvement: a summary of recent advances
  Giovanni Pinna, *Damiano Ravalico*, *Luigi Rovito*, *Luca Manzoni*, *Andrea De Lorenzo*
- AI, spot me! Multimodal models in the gym
  Gaetano Dibenedetto, *Elio Musacchio*, *Marco Polignano*, *Pasquale Lops*
- Generative AI across modalities: insights from our research on domain-aware content generation
  Giovanna Castellano, *Emanuele Colonna*, *Nicola Fanelli*, *Lucrezia Laraspata*, *Ivan Rinaldi*, *Alberto Gaetano Valerio*, *Gennaro Vessio*
- Medicine without boundaries: generative AI for translating medical data across modalities
  Valerio Guarrasi, *Francesco Di Feola*, *Giulio Ianello*, *Irene Iele*, *Linlin Shen*, *Massimiliano Mantegna*, *Daniele Molino*, *Elena Mulero Ayllon*, *Ludovica Pompilio*, *Aurora Rofena*, *Marco Salmè*, *Rosa Sicilia*, *Matteo Tortora*, *Paolo Soda*
- Prompting the future: educator competencies and case-based innovation for

Workshop 9: AI, Misinformation and Disinformation

2025-10-01: submitted by Giulia Cisotto, metadata incl. bibliographic data published under Creative Commons CC0
2025-12-07: published on CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073) |valid HTML5|