

Laboratorio CINI Data Science: Missione, Ruolo e Rapporti con l'Industria Claudio A. ARDEGNA



- ▶ Costituito nel novembre **2013** con il nome di Laboratorio Nazionale Big Data (Lab Data Science dal 2023)
- ▶ **38 nodi e 250+ membri**
- ▶ **Board industriale** con **8** aziende leader (Almaviva, Avio Aero, Engineering, Intesa San Paolo, Leonardo, Maxfone, TIM, Zucchetti) e **13** aziende observer (open call)

Italian Universities and Research Institutions

Locations Map



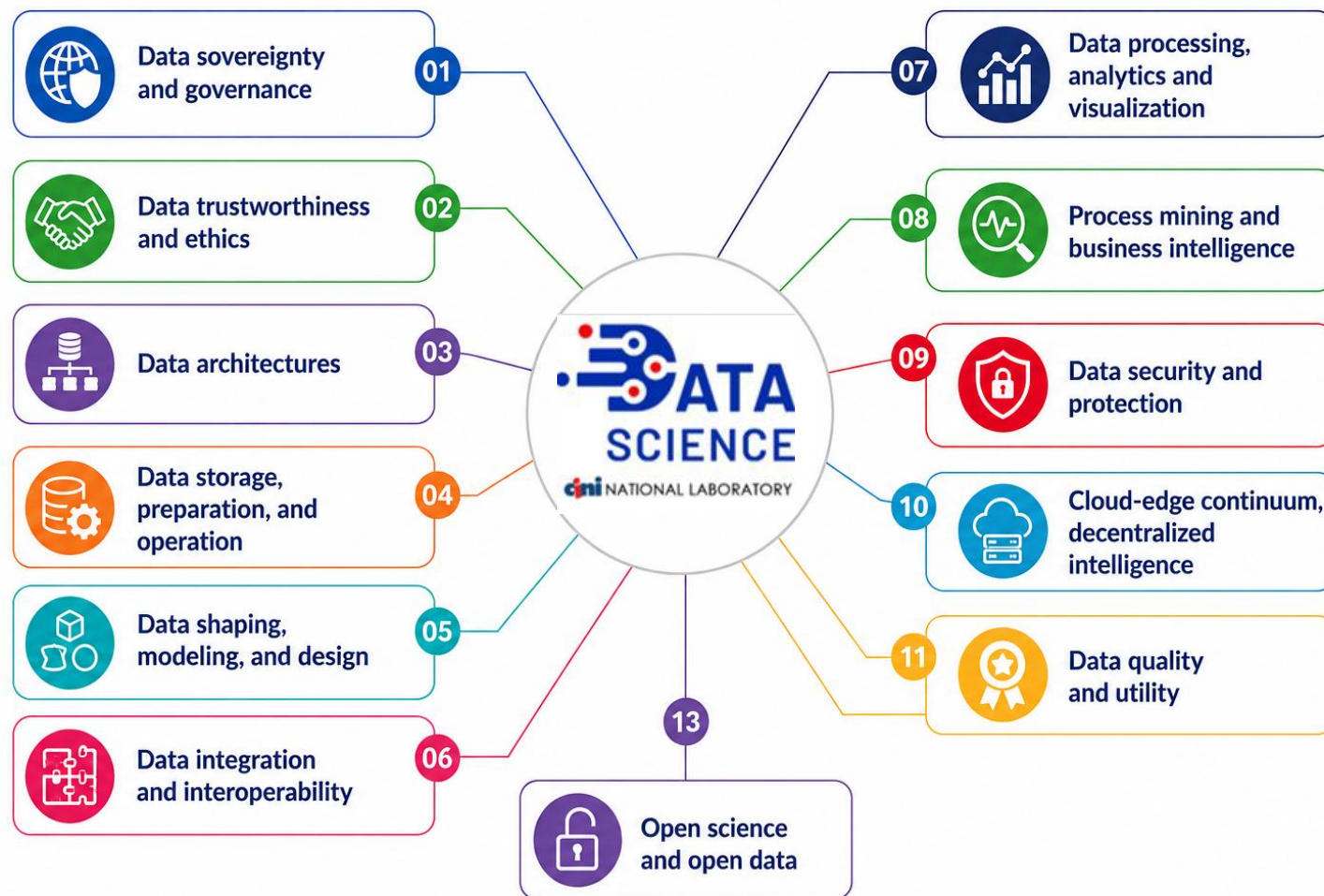
- 1 University of L'Aquila
- 2 University of Bari "Aldo Moro"
- 3 University of Bologna
- 4 University of Brescia
- 5 University of Cagliari
- 6 University of Calabria
- 7 University of Camerino
- 8 University of Campania "Luigi Vanvitelli"
- 9 University of Catania
- 10 CNR, IIT - Pisa
- 11 CNR, ISTI - Pisa
- 12 University of Ferrara
- 13 University of Genoa
- 14 IMT Institute for Advanced Studies Lucca
- 15 Polytechnic University of the Marche
- 16 University of Messina
- 17 University of Milan
- 18 Polytechnic University of Milan
- 19 University of Modena and Reggio Emilia
- 20 University of Naples "Federico II"
- 21 University of Naples "Parthenope"
- 22 University of Padua
- 23 University of Padua
- 24 University of Salento
- 25 University of Palermo
- 26 University of Pavia
- 27 University of Perugia
- 28 University of Eastern Piedmont "Amedeo Avogadro"
- 29 University of Pisa
- 30 Sapienza University of Rome
- 31 Roma Tre University
- 32 University of Turin
- 33 University of Salerno
- 34 University of Trento
- 35 University of Sannio
- 36 University of Udine
- 37 University of Salento
- 38 Ca' Foscari University of Venice

i The map shows the main locations of the universities and research institutions across Italy.

- ▶ Comprendere e gestire la **rivoluzione incentrata sui dati**, affrontando le sfide sia dal punto di vista **scientifico** sia da quello **applicativo**
- ▶ Collettore di competenze e tecnologie nell'ambito della data science
 - ▶ Supportare il paese nella transizione ***verso il futuro***
 - ▶ Rafforzare il ruolo scientifico e tecnologico dell'Italia nell'ambito della data economy

- ▶ **Community building**
 - ▶ Creazione di una comunità scientifica che si colleghi con le politiche di innovazione gestite dai governi nazionali e regionali
 - ▶ Partecipazione alle principali iniziative a livello nazionale e internazionale (ad es. BDVA) in ambito data science e big data
- ▶ **Istruzione e divulgazione:** migliorare le competenze nazionali in ambito data science e big data
- ▶ **Patto accademia-industria** per sostenere il ruolo strategico dell'Italia nelle iniziative globali, anche attraverso il trasferimento tecnologico

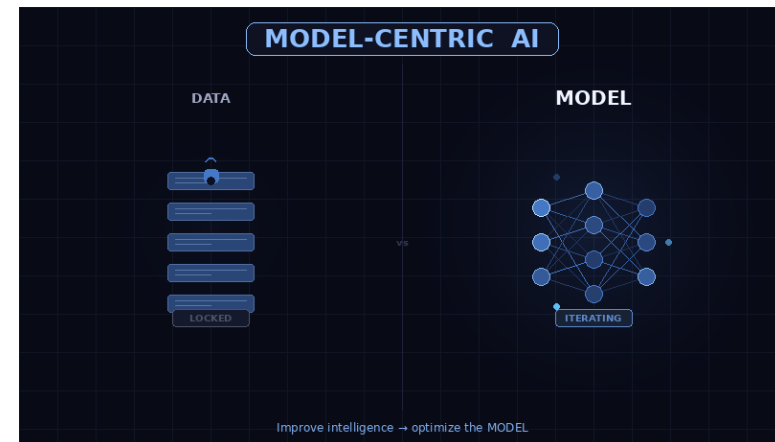
Aree di interesse



- ▶ Progetti per la transizione digitale a supporto della pubblica amministrazione
- ▶ Progetti industriali e di ricerca di base e applicata nazionali ed europei
- ▶ Progetti di consulenza per aziende

Data-Centric AI: Un cambio di paradigma

- ▶ Traditional AI si concentra sul miglioramento dei modelli, considerando i dati come un elemento fisso
 - ▶ Rumore nelle etichette ed errori di annotazione
 - ▶ Cambiamenti nella distribuzione dei dati e scarsa capacità di generalizzazione
 - ▶ Bias e mancanza di rappresentatività
 - ▶ Trasparenza, robustezza e riproducibilità limitate
- ▶ Data-Centric AI (DCAI) rende la qualità dei dati il principale obiettivo di ottimizzazione
 - ▶ Data curation e pulizia dei dati
 - ▶ Arricchimento semantico
 - ▶ Active learning e smart labeling
 - ▶ Generazione di dati sintetici
 - ▶ Monitoraggio continuo della qualità dei dati



Data-Centric AI



- ▶ Un risultato del progetto PNRR
- ▶ Una **formalizzazione concettuale** del ruolo dei dati lungo l'intero ciclo di vita dell'AI
- ▶ Un aggiornamento della **metodologia** industriale CRISP-DM in ottica DCAI
- ▶ Un collegamento esplicito con **strumenti, piattaforme e infrastrutture dati**, pienamente allineato ai principi FAIR
- ▶ Una discussione su **DCAI nell'era dell'Intelligenza Artificiale Generativa**, con particolare attenzione a robustezza, affidabilità e deployment responsabile delle soluzioni di AI

“Data-Centric AI Manifesto: How Data Quality Drives Modern AI”, just published!

H2020 IMPETUS Project



IMPETUS

IMPETUS: Intelligent Management of Processes, Ethics and Technology for Urban Safety

Key facts

- H2020 IA project
- Topic: SU-INFRA02-2019 Security for smart and safe cities, including for public spaces
- Total budget 9.3 M€, EC funding 7.9M€
- September 2020 – February 2023
- Overall goal: **Improve the security of public spaces in smart cities**

Consortium

RESEARCH	INDUSTRY & SMEs	NGOs	CITIES

... complemented by
COSSEC:

**Community of Safe and
Secure Cities**
(extends involvement)



Contact info

www.impetus-project.eu

Project Coordinator: Joe Gorman, SINTEF Digital: joe.gorman@sintef.no
 COSSEC Manager: Sandro Bologna, TIEMS: s.bologna@infrastrutturecritiche.it
 Dissemination manager: Snjezana Knezic, TIEMS: snjezana.knezic@gmail.co



This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883286.

19

WHY do Smart Cities need IMPETUS?

Smart Cities

- High-tech grid of sensors (cameras, environmental sensors, traffic sensors, ...)
- IoT ("Internet of Things") – internet is everywhere – not least in everyone's pockets
- IT systems controlling critical infrastructure
- Advanced algorithms and AI (Artificial Intelligence) to help people make decisions

Efficient city administration

Enhanced situational awareness, especially in emergencies

Help authorities make sound decisions, fast



Increased "attack surface" - more vulnerabilities
e.g. cyber attacks

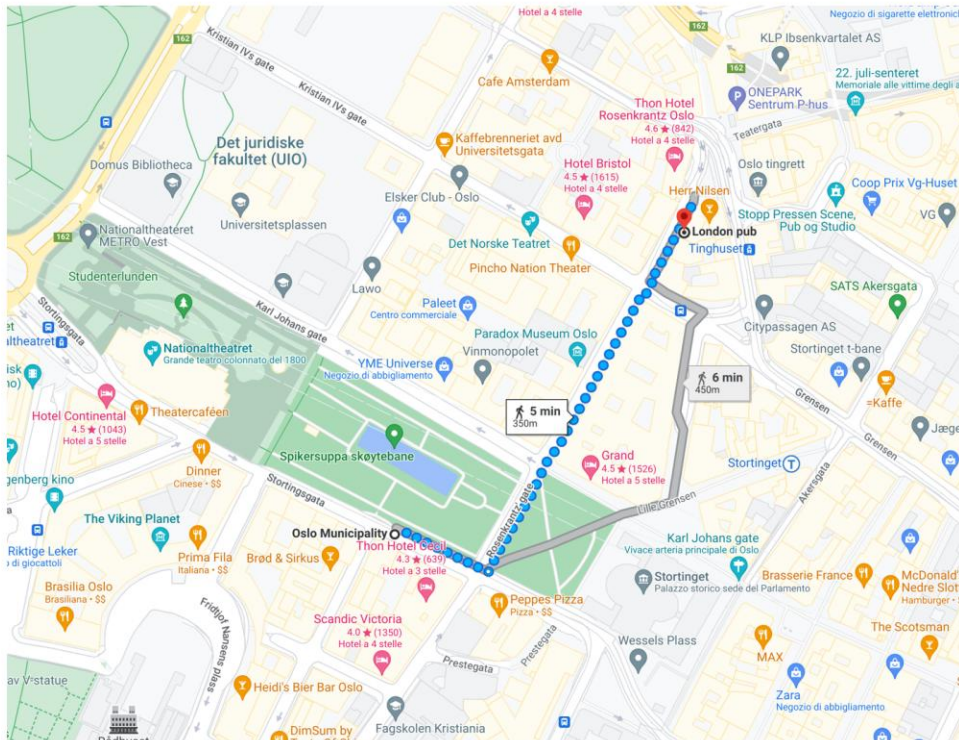
Increased risk of unethical use of personal data

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883286.

20

Physical Threat Intelligence: Anomalia Identificata 25/06/2022 at 01:55

*Anomalia riportata subito dopo una sparatoria al
London Pub di Oslo.*



Feature Ranking: output del tool

```
{  
  "timestamp" : "2022-06-25T00:55:00+01:00",  
  "coordinates" : [59.91333835433864,  
10.736881605194284],  
  "ranking" : [  
    {  
      "feature" : "number of vehicles",  
      "value" : 9.0,  
      "importance" : 0.38  
    },  
    {  
      "feature" : "origin aimed departure time hour",  
      "value" : 1.0,  
      "importance" : 0.12  
    },  
    ...  
  ],  
  "anomaly" : "true"  
}
```

A. Polimeno, P. Mignone, C. Braghin, M. Anisetti, M. Ceci, D. Malerba, C.A. Ardagna, "Balancing Protection and Quality in Big Data Analytics Pipelines," Big Data Journal 2023

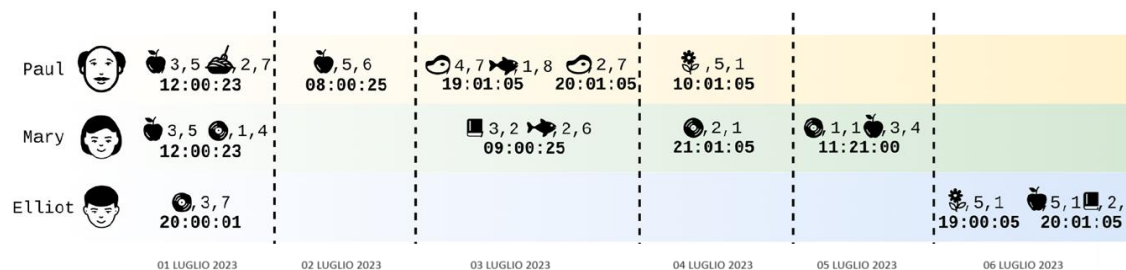
AI per predire la lealtà dei clienti in applicazioni di commercio al dettaglio



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



- ▶ Monitoraggio degli acquisti effettuati in esercizi commerciali
 - ▶ Estrazione di profili descrittivi delle variazioni nel tempo delle abitudini di acquisto dei clienti attivi
- ▶ Realizzare sistemi basati su **Intelligenza Artificiale** per apprendere ed aggiornare modelli di predizione della lealtà dei clienti sulla base delle abitudini di acquisto
- ▶ Fornire all'operatore di marketing **spiegazioni** con tecniche XAI relative alla necessità di **riciclo del modello** di predizione (in presenza di cambiamenti nelle abitudini di acquisto) e alle **decisioni** restituite dal modello di predizione



Ore 24:00:00 del 6 Luglio 2023, il modello di predizione (che considera un periodo di due giorni senza acquisti come una condizione di abbandono) predice che:

Mary potrebbe essere un cliente a rischio di abbandono e potrebbe non effettuare un nuovo acquisto entro il 7 Luglio 2023

ed evidenzia che i fattori che hanno influito maggiormente su tale predizione sono:

Mary ha eseguito un solo acquisto negli ultimi 2 giorni spendendo 13 euro, mentre ha eseguito due acquisti nei due giorni precedenti





AI Governance and Compliance

*You use trustworthy services, hire trustworthy professionals, **why would you use an application driven by untrustworthy AI?***

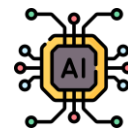
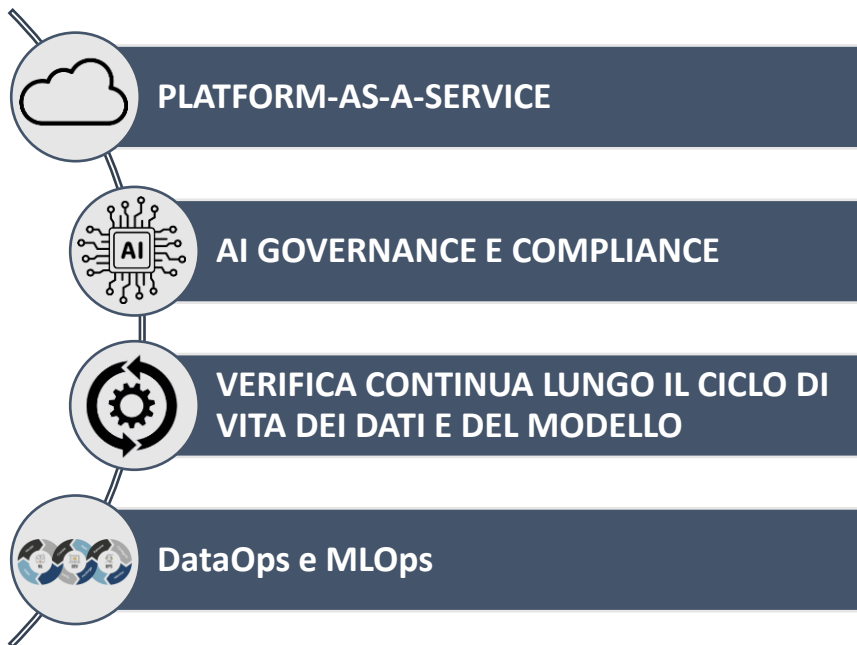
M. Anisetti, C. A. Ardagna, N. Bena, E. Damiani, "Rethinking Certification for Trustworthy Machine Learning-Based Applications", IEEE Internet Computing, 2023

*Risk management should be **continuous, timely**, and performed throughout the AI system **lifecycle** dimensions*

NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0)



MoonCloud⁴AI



MOONCLOUD⁴AI



SPIN-OFF DELL'UNIVERSITA' DEGLI STUDI DI MILANO



MoonCloud⁴**AI**: AI Governance and Compliance

SCHEMA DI CERTIFICAZIONE MULTI-DIMENSIONALE PER LA VERIFICA E COMPLIANCE DI AI

- **Dimensione design**: design della pipeline per la costruzione del modello di AI
- **Dimensione dati**: dati usati per il training
- **Dimensione processo**: processo di training
- **Dimensione modello**: modello AI in operation



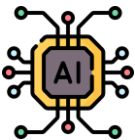
MoonCloud⁴AI: Assessment and Compliance

VERIFICA E COMPLIANCE A TRE LIVELLI



- **Layer 1: Self-Assessment**

- Questionario statico ispirato ai requisiti e linee guida dell'EU AI Act e CapAI



- **Layer 2: Continuous Assessment of AI Artifacts**

- Domande nel questionario collegate a controlli tecnici che valutano la validità delle risposte
- Controlli tecnici eseguiti sugli artifatti AI (dimensioni) indipendentemente



- **Layer 3: Continuous Assessment of AI Pipelines (MLOps)**

- Domande nel questionario collegate a specifici controlli tecnici
- Controlli tecnici iniettati nelle pipeline di costruzione dei modelli AI per valutare la validità delle risposte al questionario a run time
- Controlli tecnici eseguiti durante tutti gli stadi della pipeline (es., data preparation, training process, model building, model assessment)

- ▶ Italian Conference on Big Data and Data Science (ITADATA – <https://www.itadata.it>)
- ▶ Hackaton nazionale su data science e big data (<https://buytickets.at/cinidatasciencelab/2418831>)
- ▶ Serie di seminari “Tales on Data Science and Big Data” (<https://seminars.sesar.di.unimi.it>)
- ▶ Eventi industry-focused “Come dare valore ai dati”
- ▶ PhD School (in forma associata) on Data Science
- ▶ Partecipazione a panel, tavole rotonde, conferenze
 - ▶ Es.: “città sostenibili, infrastrutture strategiche” (Bologna)

- ▶ The Italian Conference on Big Data and Data Science (ITADATA) evento nazionale del Lab CINI Data Science
 - ▶ Quinta edizione: 10-12 Novembre 2026, Bari
 - ▶ Track di ricerca su big data e data science con articoli scientifici e industriali
 - ▶ Include tutorial/demo, workshop tematici, hackathon ITADATAhack
 - ▶ Premi per la miglior tesi italiana di master e di dottorato

- ▶ Speaker invitati
 - ▶ Richard Stevens, IDC, BDVA
 - ▶ Ernesto Damiani, Università degli Studi di Milano

ITADATA2026: Join Us

► www.itadata.it



ITADATA 2026

Committees Submissions Program Speakers Awards Info Sponsors Past Editions

The 5th Italian Conference on Big Data and Data Science

10-12 November 2026
Bari, Italy

The banner features a dark blue background with a cityscape at night, including a prominent church tower, and a bar chart overlay.