Politecnico
di Torino

Thesis topics

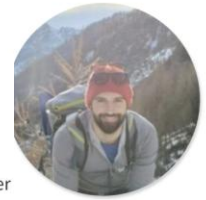
08/01/2026

MINDS
MACHINE INTELLIGENCE AND
DATA SCIENCE GROUP

MINDS: Machine Intelligence and Data Science Group

Daniele Quercia

Participating Member
Responsible AI, Human-Computer Interactions, Data Science



Flavio Giobergia

Participating Member
Quantum Computing, Machine Unlearning, Large Language Models

Eliana Pastor

Participating Member
Explainable and Trustworthy AI, Data Science, Big Data Analytics, Predictive Maintenance

Luca Vassio

Participating Member
Online Social Networks, Internet Measurements and Anonymization, Mobility and Transportation, Big Data Analytics, Analytical and Numerical

Paolo Garza

Participating Member
Data Science, Big Data Analytics

Elena Baralis

Scientific Board and Main Proponent
Data Science, Big Data Analytics

Marco Mellia

Scientific Board and main proponent
Network Monitoring, Cybersecurity, AI & Data Science

Nikhil Jha

Participating member
Data Science, Machine Learning, Graphs

Silvia Chiusano

Participating Member
Data Science, Data Mining

Luca Cagliero

Project Coordinator
Natural Language Processing, Multimodal Learning, Machine Learning

Danilo Giordano

Steering Committee
Internet Traffic, Smart Mobility, Data Science, Big Data Analysis

Matteo Boffa

Participating member
Machine Learning and Deep Learning

Daniele Apiletti

Steering Committee
Theory-Guided & Knowledge-Informed AI, Physics-Informed Scientific Computing, AI for Industrial Intelligence

Giuseppe Tipaldo

Participating Member
Science and Tech studies; Social impacts of innovation; public and stakeholder engagement

Tania Cerquitelli

Participating Member
Data Mining, Self-learning Methodologies, Transparent Data Anaytics

Simone Monaco

PhD Student
Neural Network, Machine Learning and Deep Learning

MINDS
MACHINE INTELLIGENCE AND DATA SCIENCE GROUP

Politecnico di Torino

Thesis general information

- Duration: 4-6 months full time
  - equivalent overall duration if part time
- *Internal* thesis
  - cooperation on active research topic or research project
  - good programming and analytical skills required
  - supervised by a group member
  - can work at home or in our lab (LAB5)
- *External* thesis (stage)
  - supervised by external tutor

To get more info on specific topics

please contact the reference person of the thematic area of interest by email (name.surname@polito.it)

Main topics

- Machine learning algorithms
  - Design and implementation of novel ML algorithms
- Data science pipeline
  - Design, personalization and implementation of KDD processes in diverse application areas
    - Industry, health, finance, cybersecurity,…
- Big data analytics
  - Design of scalable data mining and machine learning algorithms
  - Design of scalable KDD processes
- Database management
  - Data warehouse and NoSQL data modeling
- Privacy
  - Privacy preserving analytics

What we do

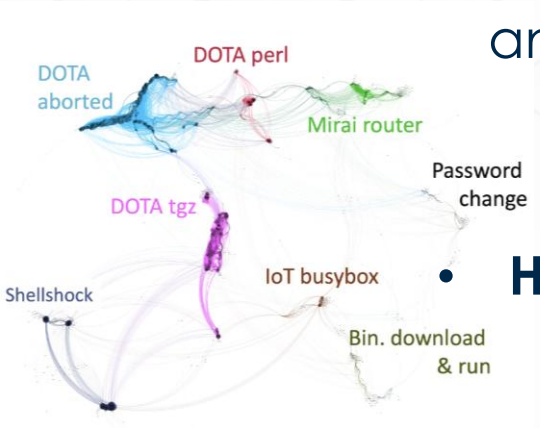
- Data & AI & Cybersecurity
- LLM & Generative models
- Temporal reasoning
- Explainable and Trustworthy AI
- Machine Unlearning
- Tools for AI Governance
- AI & Industry
- Predictive maintenance

Data & AI & Cybersecurity




Use AI to process data and detect novelties

- In **Network Telescopes**: Consider packets sent by malicious actors as words and documents; use **representation learning** to represent senders; use **clustering** and **novelty discovery** to detect coordinated attacks.

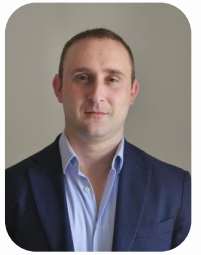


- **Honeypot logs**: Train **language model** for the analysis of **honeypot** logs; detect similar patterns and novelties

- **AFCDH**: Leverage AI/ML for **Financial Crime Fight**

- **Web security** – ***new*** *6.1ME project* – AI4CTI: **Foundational model for cybersecurity**; Leverage it to detect phishing attacks, Data Loss (e.g., a user uploading a file with personal data in ChatGPT); Critically verify AI models

LLMs and Generative AI



Design cost-effective solutions to

- **Interact with database**: Techniques to **generate SQL code, solve ambiguous questions, and test LLMs on proprietary data**
- **Query Visually Rich Documents**: LLM-based methodologies to **answer questions on visual, textual, and layout information**
- **Paraphrase and summarize text**: Algorithms to **synthetize** (produce summaries), **change style** (e.g., informal to formal), **process complex textual sources** (e.g. legal documents, social content)
- **Understand multimodal data inputs and data with noise/uncertainty**: leverage **multimodal LLMs** to understand multimodal data (images, videos) and complex AI processes in the presence of noise/uncertainty

Temporal reasoning

Understand and reason about **timestamped** and **time-evolving data, temporal relations** and **time-related concept** using (Multimodal)LLMs

Open challenges

- **Trained models** have temporal biases and drifts
- (Re-)training/fine-tuning LLMs is **costly**
- **Data sources** are inherently time-evolving (e.g., videos, speeches)
- Relations among **multimodal sources** are time-dependent (e.g., time series and textual annotations)

Addressed tasks:

- **Summarization**
  - Video
  - Text
- **Explainability**
  - Text and images
  - Speech
- **Legal reasoning**
  - Legal documents
- **Visually Rich Document Understanding**
  - Unanswerable query detection

FIS

Project funded by the Italian Ministry (FIS-2 Research Grant)
"Language Models: a Matter of Time"

Explainable and Trustworthy AI




Algorithms for making AI more interpretable,

trustworthy, safe, and reliable

Future Artificial Intelligence Research

- **Explanations**: Techniques to **explain the individual predictions** of AI models (tabular, text, image and speech) and **concept-based explainability** to explain via concepts [6].

amazon | science

- **Subgroup behavior**: Methodologies to **analyze** AI behavior within data **subgroups**, detecting potential **differences in performance**

- **Mitigation**: Algorithms to **reduce disparities among data subgroups**, improving performance both overall and in subgroups

- **Monitoring and drift detection**: Approaches for continuous monitoring of AI performance to detect and mitigate drifts over time in subgroups.

INTESA SANPAOLO

MINDS MACHINE INTELLIGENCE AND DATA SCIENCE GROUP

Politecnico di Torino

Machine Unlearning

Machine Unlearning refers to techniques that allow a trained machine learning model to *forget specific data* as if it had never been seen during training

- **Why?** To remove sensitive data from models (*right to be forgotten*), to address copyright infringements, to debias models, …
- **How?** By making "small" changes to models
- **When?** When retraining a model from scratch is too expensive (e.g., for LARGE models)

Current research:

- **MU across modalities**
  - Speech models
  - Language models
- **MU algorithms**
  - On new tasks (e.g., regression)
  - New distillation-based approaches
- **MU evaluation**
  - Benchmarking, new metrics
- **MU library**
  - ERASURE framework

Tools for AI Governance

Responsible AI
How to build AI systems that are fair, transparent, and sustainable

What's in it for you?
1/ Define Fincantieri's Responsible AI Pillars: Establish clear standards aligned with global AI regulations and Fincantieri's strategic commitment to environmental sustainability.
2/ Gain Responsible AI Tools: Equip teams with frameworks for answering key questions

Domain / Topic	Key Questions
Radar & Geneative AI	Do operators trust & understand *outputs*?
Robotics & Drones	How do we safeguard *human-robot* interaction?
Navis Sapiens Ecosystem	Who governs and benefits from this ecosystem?
Digital Twins & Autonomy	Is there *human oversight*? How do we handle AI-led decisions in crises?
Dual-Use Innovation	How do we *prevent misuse* for harmful military or unethical applications?
AI Governance Overall	Are ethics consistently applied and *auditable* across all programs?

Why do you need this?

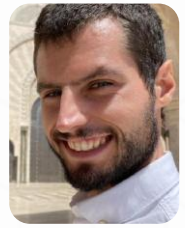
1.Ensure compliance:
Be prepared for upcoming EU AI regulations.

2.Secure a first-mover advantage:
Position Fincantieri as a leader in your industry.

3. Reduce costs:
Avoid expensive redevelopment: over **60% of AI costs** come from rework, not initial development.

4. Build a trusted brand:
Prevent costly incidents and strengthen customer and stakeholder trust.

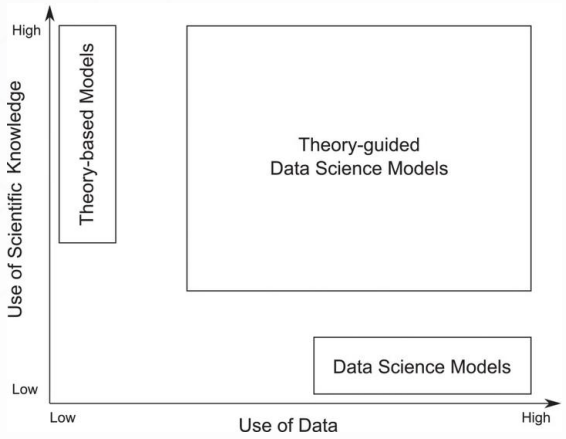
AI & Industry – Scientific AI

Blending domain knowledge and data science algorithms

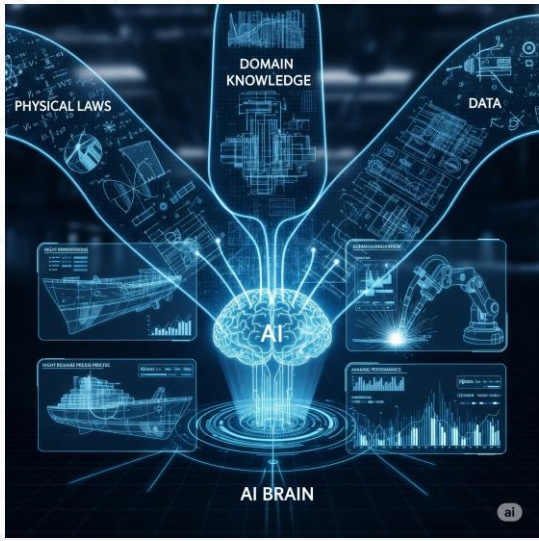
- **Physically-consistent intelligence** - AI models integrating fundamental domain knowledge, such as the scientific principles, blending data-driven learning and scientifically plausible outputs, robust in real-world operational scenarios, preventing physically impossible hallucinations

- **Efficiency & performance** - by injecting known engineering constraints and physical laws, we reduce the need for massive datasets and intensive computational power. This makes advanced AI solutions effective even in situations with limited or incomplete data, such as predicting rare equipment failures



- **Consistency control** - by enforcing known operational rules and physical invariances, we deliver AI systems whose outputs are reliable, verifiable, and build trust for critical tasks

- **Uncertainty-aware algorithms** - AI models that not only makes predictions but also quantify their own confidence in those predictions, crucial for high-stakes industrial applications.



Predictive maintenance







*Algorithms and pipelines for **anticipating failures**, improving reliability, safety, and efficiency*

- **Data-Driven Pipelines for Prognostics**: Approaches that transform raw sensor time series into indicators of component health

- **Interpretable pattern recognition**: Extracting human-readable patterns and relationships among operating parameters to guide maintenance strategies.

Quantum ML

- Many algorithms in computer science can benefit from a *quantum advantage*
  - E.g., searching an unordered database is $O(N)$ on a classic computer, but $O(\sqrt{N})$ on a quantum computer!
- Quantum ML (QML) combines quantum computing with machine learning to solve problems faster or differently than classical methods

Torino, Fondazione Links e Poli «accendono» un computer quantistico

Iniziativa avviata con l'Istituto nazionale di Ricerca Metrologica (INRIM) - L'obiettivo è sviluppare un ecosistema che coinvolge 30 tra docenti e ricercatori e 60 studenti

di Filomena Greco

25 ottobre 2024

- Various ML algorithms have already been framed in their quantum version:
  - QSVM, QkNN, QPCA, …
- But many others are yet to be proposed!