PhD offer 2025-2028

Confidential Inference and Explainability: Toward Self-Diagnosis via Imaging

Keywords:

Applied cryptography, Machine learning, Computer vision, Deep learning, Privacy, Explainability.

Subject description:

This PhD topic aims to jointly address privacy and explainability of decisions obtained through image analysis using a neural network. In the context of a classification task performed on a remote server, the goal is to develop approaches that ensure the confidentiality of the explanation as well as that of the input (and output) data. Preserving the privacy of data while ensuring the transparency of the model is a crucial challenge, particularly in domains such as healthcare. The objective aligns with the emerging regulatory framework on AI at the European level (AI Act). While these issues are the subject of significant research individually—whether in applied cryptography or machine learning—the combination of explainability under privacy constraints represents a new research problem.

The project will seek to identify local explainability methods based on visual information or concepts that can be adapted to a privacy-preserving mode. Confidentiality may be approached through secure multi-party computation and/or homomorphic encryption. Thanks to a collaboration with the Saint-Etienne University Hospital (France), it will be possible to fine-tune the secure AI system and conduct supervised experiments on health data, aimed at enabling self-diagnosis. The experimentation may also extend to ethical and legal dimensions, through a Auvergne - Rhône-Alpes - Canada partnership on AI.

References:

[1] Akram, A., Zimmer, P., Gritti, C., Karame, G., & Önen, M. (2025). Fundamentals of privacy-preserving and secure machine learning. In *Trustworthy AI in Medical Imaging* (pp. 385-409). Academic Press.

[2] Patrício, C., Neves, J. C., & Teixeira, L. F. (2023). Explainable deep learning methods in medical image classification: A survey. *ACM Computing Surveys*, *56*(4), 1-41.

[3] Nguyen, Thanh Tam, et al. Privacy-preserving explainable AI: a survey. *Science China Information Sciences* 68.1 (2025): 111101.

[4] Li, Yong, et al. Privacy-preserving federated learning framework based on chained secure multiparty computing. *IEEE Internet of Things Journal* 8.8 (2020): 6178-6186.

[5] Pulido-Gaytan, B., & Tchernykh, A. (2024). Self-learning activation functions to increase accuracy of privacy-preserving Convolutional Neural Networks with homomorphic encryption. *Plos one*, *19*(7), e0306420.

Université
Jean Monnet
Saint-Étienne

Laboratoire
Hubert Curien
UMR • CNRS • 5516 • Saint-Étienne

Inria

citi lab

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Expected Results:

- Identification of local explainability methods for image classification that can be carried out in a privacy-preserving manner
- Definition of one or more frameworks for implementing explainable classification methods
- Proposal and implementation of privacy-preserving explainability methods based on the proposed framework(s)
- Experiments on a healthcare-related task and analysis in relation to ethical and legal dimensions of AI

PhD Location:

Laboratoire Hubert Curien (LabHC), Université Jean Monnet, Saint-Etienne, France
(regular meetings at the CITI Laboratory, INSA Lyon, Villeurbanne, France)

Expected Starting Date: 01/10/2025

Framework:

PhD thesis part of a series of PhD theses on *Data and AI in a sustainable and responsible approach*, taking place at the Lyon - Saint-Etienne College of Engineering and Jean Monnet University.

Financial support:

3-year PhD school contract (Ecole Doctorale Sciences Ingéniérie Santé, ED 488).

Expected profile:

Candidates holding a degree from an engineering school or a Master 2 from a university in applied mathematics or computer science, with training in cryptography and machine learning, and proficiency in a programming language and one or more reference development libraries in one of these fields.

Application process:

- **27/05/2025 :** Applications deadline (on rolling basis interviews with pre-selected candidates)
- **28/05/2025 :** Submission of a ranked list of candidates to the Collège Ingéniérie Lyon-Saint-Etienne - Université Jean Monnet.
- **13/06/2025 :** Final decision on the selected candidates by the Collège Ingéniérie Lyon-Saint-Etienne - Université Jean Monnet.

Following the selection process and without delay, a 3-year PhD contract will be established by the Ecole Doctorale. Sciences Ingéniérie Santé, ED 488.

Contact : Thierry Fournel (LabHC), Clémentine Gritti (CITI, Inria) et Amaury Habrard (LabHC, Inria)
fournel@univ-st-etienne.fr, clementine.gritti@insa-lyon.fr, amaury.habrard@univ-st-etienne.fr

Send your CV, cover letter and master transcripts, and give contact details of referees.

Collège d'ingénierie
LYON · SAINT-ÉTIENNE
CENTRALE LYON · ENTPE · INSA LYON · MINES SAINT-ÉTIENNE

Université
Jean Monnet
Saint-Étienne

Bouquet de thèses 2025

The doctoral thesis described below is part of a series of theses designed to build a multidisciplinary scientific approach to the societal challenge of a "responsible digital society", and more specifically, the specific theme of "Data and AI in a sustainable and responsible approach", identified as a priority issue by the 4 institutions of the Lyon Saint-Etienne Engineering College (Centrale Lyon, ENTPE, INSA Lyon, Mines Saint-Étienne) and by the Université Jean Monnet Saint-Étienne, which are providing financial support for the theses making up this 2025 package.

The 2025 theses package includes 6 theses covering different facets of data science and artificial intelligence, addressing the following questions:

- Monitoring crystallization processes using AI-assisted acoustic emission
- AI-assisted design of biodegradable and/or biosourced biopolymers for the sustainable protection of agricultural crops
- Machine learning methods for urban microclimate prediction
- Data-driven non-linear structure identification
- Inference and explicability in confidential mode: towards self-diagnosis via images
- Towards certification of vibration monitoring with explanatory AI

These theses involve a total of 16 supervisors from 11 laboratories on the Lyon Saint-Etienne site (Centre d'Innovation en Télécommunications et Intégration, Centre SPIN - Génie des Procédés, Ingénierie des Matériaux Polymères, Biologie Fonctionnelle, Insectes et Interactions, Institut Camille Jordan, Laboratoire de Mécanique des Fluides et d'Acoustique, Laboratoire de Tribologie et Dynamique des Systèmes, Laboratoire d'InfoRmatique en Image et Systèmes d'information, Laboratoire Hubert Curien, Laboratoire Vibrations Acoustique, Matériaux : Ingénierie & Science) of which the 5 funding institutions are supervisors. The 6 PhD students recruited under this package will be enrolled in 3 Doctoral Schools on the site: MEGA, EDML, SIS.

The teams (doctoral students and their supervisors) involved in these 6 theses form a multi-disciplinary scientific community: regular exchanges between these teams will take place throughout the 3 years of the doctoral pathway, notably in the form of joint seminars to develop the multi-disciplinary systems approach specific to the bouquet and enrich the teams' disciplinary skills in a spirit of sharing and learning. Thesis papers produced at the end of the doctoral program will also reflect the original positioning of the thesis work within a bouquet, by including a chapter analyzing the impact of the work carried out on the "Data and AI in a sustainable and responsible approach" issue.