ARTEMIS Department, SAMOVAR Lab,
Telecom SudParis, IP Paris, IMT

NEWTON

#TatouAI #AIModelTracking #NNWatermarking

THE PROBLEM ADDRESSED

AI solutions for solving complex tasks are popular, regardless of how critical the applications are (ranging from autonomous driving to customer relation chatbot).

Tracking intellectual property rights (IPR) of AI is an emerging need, similar to images or videos. Assessing the integrity of neural networks (NN) remains a challenging task, as noted by governmental bodies (EU AI Act, Chinese regulation, …).

ARTEMIS Department has a long track record in digital imaging sciences and technologies. Specifically, our team of experts is actively engaged in the traceability technologies for AI in normative instances (MPAI / ISO), academic research, and valorization.

TECHNOLOGY

- **TatouAI** is a technology to track any given model without any constraint:
  - **Before/during/after** training of your AI
  - **No-impact** on the performance
  - **Low computational cost**

- **TatouAI** is an **AI watermarking technique** that involves embedding digital marks into AI models to enable their tracking in terms of **identification**, **integrity**, **modification**, and **security**

COMPETITIVE ADVANTAGES

- Model-agnostic: scalable from small to large neural networks.

- Built-in trust: ensures traceability and integrity and regulatory alignment.

- Easy-to-integrate and easy-to-use tracking solution.

APPLICATION

- **Detecting AI model** and its **copy** (possibly **stolen**)

- **Checking integrity** of a given AI technology : **preserving trustworthiness** in AI systems and reputation of AI hosts, e.g. LLM-based chatbots

- Ability to **detect modification/alteration** in a given AI model

- Complementing **security solutions** and **guardrails** with traceability techniques

DEVELOPMENT STATUS

TRL = 4

- Traceability technology tested in laboratory setup using the MPAI-NNT standard evaluation (IEEE 3309)
- Prematuration project funded by **IP Paris** to increase the TRL value of the technology

INTELLECTUAL PROPERTY

Patent application:

- FR2505074

INVENTORS & CONTACTS

- Carl De Sousa Trias, postDoc (TSP)
  carl.de-sousa-trias@telecom-sudparis.eu
- TTO: pauline.portier@telecom-sudparis.eu

PUBLICATIONS

Carl De Sousa Trias. "*Robust watermarking for neural networks in multimedia tasks*". Neural and Evolutionary Computing [cs.NE]. Institut Polytechnique de Paris, 2025. English. ffNNT: 2025IPPAS007ff.fftel-05142694f

LOOKING FOR

- Expending to **new use cases**
- **Industry partner** to study manufacturability