PhD Topic: Multimodal Deepfake Detection

Exploring Explainable Semantic Relations Across Modalities

Host institution: École de technologie supérieure (ÉTS), Montréal, Canada
Duration: Minimum 3 years
Starting date: As soon as possible
Main supervisor: Professor Patrick Cardinal, ÉTS, Montréal (*patrick.cardinal@etsmtl.ca*)
Co-supervisor: Thomas Thebaud, Research Scientist, Agence ministérielle pour l'intelligence artificielle de défense (AMIAD), France (*thomas.thebaud@polytechnique.edu*)

Context

Recent advances in generative artificial intelligence have made it possible to create increasingly realistic synthetic speech, images, and videos. These developments raise important challenges for assessing the authenticity of digital content and understanding the reliability of automated deepfake detection systems.

Multimodal detection offers opportunities to examine not only individual modalities, but also the relationships between them. Spoken language, vocal expression, facial behavior, and visual context convey complementary information whose coherence may help assess whether content has been manipulated. Understanding these relationships could also help explain why a detection system considers a recording suspicious.

However, the connection between cross-modal coherence and authenticity is not straightforward: manipulated content can remain semantically consistent, while authentic content may contain apparent mismatches. Exploring this distinction is an important motivation for the proposed research.

Research direction

This PhD will explore **the role of semantic relationships across modalities in deepfake detection**, with particular attention to the interpretability of the resulting models and their decisions.

The research will consider how information from different modalities, primarily audio, video, and language, can be brought together to assess authenticity. Beyond producing a detection score, the ambition is to better understand which relationships contribute to a decision and how this evidence can be expressed in a meaningful and faithful way.

Potential directions include learning representations of cross-modal relationships, investigating the distinction between manipulation and natural variation, and exploring how explainability can support generalization to unfamiliar content or generation methods.

These directions provide an initial framework rather than a fixed research agenda. **The specific research questions, methodological approaches, and evaluation settings will be progressively defined** through the literature review, initial experiments, and discussions between the doctoral student and the supervisors.

Scientific environment

The doctoral student will be hosted at ÉTS in Montréal, under the supervision of Professor Patrick Cardinal, with co-supervision by Thomas Thebaud at AMIAD in France.

The project lies at the intersection of multimodal machine learning, speech and audio processing, computer vision, and explainable artificial intelligence. The student will develop their own research contributions and disseminate findings through scientific publications, with an emphasis on rigorous evaluation and reproducible research.

Candidate profile

We welcome applications from candidates with:

- A Master's degree, or equivalent qualification, in computer science, artificial intelligence, electrical engineering, applied mathematics, or a related field.
- A solid foundation in machine learning and programming, particularly Python.
- An interest in multimodal learning, deepfake detection, or explainable artificial intelligence.
- Scientific curiosity, autonomy, and an interest in developing and critically evaluating new research ideas.

Experience in speech processing, computer vision, or deep learning frameworks would be an asset.

Languages and teaching opportunities

The research can be conducted in English or French. French-speaking candidates will have an additional advantage if they wish to gain teaching experience alongside their research, particularly in preparation for an academic career.

Teaching opportunities require proficiency in French and will provide supplementary income, subject to availability and institutional requirements.

Application

Interested candidates are invited to contact the supervisors with a curriculum vitae. Examples of previous research or software projects may also be included.

Please indicate your earliest possible starting date and your proficiency in English and French.