Multimodal and Privacy-Preserving Machine Learning for Human-Behavior Analysis

Context

This thesis explores the intersection of machine learning, social signal processing and ethics of AI systems with a focus on human behavior analysis while ensuring privacy preservation. Understanding and interpreting human behavior is crucial in crafting systems that are both intuitive and personalized, particularly within the healthcare domain, where there is a pressing demand for methodologies and computational models to cater to individual needs and preferences. To accomplish this goal, **user-specific models must be established**, typically derived from personal data and observations of human behaviors (Rahimi et al., 2025). However, the collection and analysis of sensitive user data raise **significant concerns regarding privacy and data protection**. Moreover, especially in mental healthcare, there is a **growing interest in integrating diverse data sources**, including audio-video data, phone call patterns, and wearable sensor data, into predictive models for screening mental disorders (Drissi, Ouhbi, García-Berná, Idrissi, & Ghogho, 2020). This results in the **generation of multimodal and heterogeneous datasets tailored to users and specific applications**, such as stress detection, depression diagnosis or COVID-19 detection. **Leveraging these diverse data sources yields valuable insights for enhancing mental health assessments and interventions** (Bourvis et al., 2021). However, the traditional approach requires the definition and implementation of hand-crafted multimodal features, as well as the collection and training of machine learning models using data from all participants, including both control subjects and patients. These factors make it challenging to apply across different contexts and limit data sharing. **There is an urgent need to develop models capable of transferring across multiple tasks within similar contexts by establishing efficient multimodal representations of human-related data. At the same time, these approaches must prioritize privacy preservation and enhance explainability** (Guerra-Manzanares, Lopez, Maniatakos, & Shamout, 2023). In particular, regulations such as the EU's General Data Protection Regulation (GDPR) impose comprehensive legislative mandates to safeguard individuals' private data, including but not limited to location, age, and sex. Additionally, in Europe, the EU AI Act outlines prohibited applications—such as emotion recognition in educational settings—while enforcing strict privacy and transparency requirements. **Meeting these requirements will not only strengthen trust but also drive paradigm shifts in consent procedures and data storage practices, fostering a more ethical and responsible deployment of AI technologies.**

Objectives

By leveraging machine learning and social signal processing techniques, **the thesis aims to tackle challenges associated with multimodal data processing, privacy, and transparency in the domain of human behavior analysis**. We will **focus on Generative AI models, with a particular emphasis on multimodal foundation models** (Li et al., 2023). We leverage multimodal machine learning models for anonymization by first capturing the intrinsic complexity of multimodal representations of human behaviors. This learned representation is then used to generate realistic synthetic data via generative models. Human feedback will be incorporated to evaluate the quality of the synthetic data and refine the models accordingly.

We hypothesize that by leveraging multimodal machine learning models while prioritizing privacy preservation, we can achieve the following objectives: (i) enhance the accuracy and effectiveness of predictive models for human behavior analysis by utilizing comprehensive representations from diverse data modalities, and (ii) ensure robust privacy protection mechanisms throughout the multimodal data representation and generation pipeline, thereby safeguarding sensitive user information.

Scientific approach

Several recent studies have highlighted advancements in various areas related to privacy-preserving machine learning (Guerra-Manzanares et al., 2023). These include noise addition, federated learning, differential privacy, cryptographic techniques, and security aspects of ML models, such as adversarial attacks. In this thesis, **we propose leveraging advanced generative methods to create synthetic data that are both realistic and privacy-compliant**. **Data anonymization typically leads to a marked decrease in precision during the learning phase, which presents a challenge to achieve a balance between privacy and utility** (Zhao, Kaafar, & Kourtellis, 2020).

Generative variational graph autoencoders have been used for generating electronic healthcare records represented as sequential graphs (patient trajectories)(Nikolentzos, Vazirgiannis, Xypolopoulos, Lingman, & Brandt, 2023). A patient trajectory is a time sequence of encounters (visits) at healthcare organizations (e.g., hospitals

or other providers). However, generating realistic synthetic raw data such as audio signals, images or physiological signals is difficult due to **the high dimensionality and complex distribution of multimodal human behavior recordings**. The intricate nature of multimodal data makes it challenging for traditional generative models (e.g. auto-encoders, Generative Adversarial Networks) to accurately replicate underlying patterns, potentially leading to inaccuracies in the synthetic data produced. Furthermore, the **presence of a limited number of individuals exhibiting unique characteristics**, such as outliers, exacerbates the challenge of accurately estimating the intricate, high-dimensional distribution of these data.

The core concept involves learning the **inherent complexity of multimodal representations of human behaviors**. We employ this representation to **generate realistic synthetic data using generative models**. Subsequently, **human expert feedback on the synthetic data is incorporated to adapt the models accordingly using reinforcement learning techniques**. This approach enables us to maintain the utility of the original data while ensuring robust privacy protection.

Several approaches will be developed and evaluated, taking into account their performance in addressing the **challenges of (i) transcending user and task specificity and (ii) balancing the privacy-utility trade-off**. As an initial approach, we will **explore generative models to construct multimodal and privacy-preserving deep auto-encoders, including generative graph models**. This will involve extending existing uni-modal models developed for analogous objectives in prior literature (Guerra-Manzanares et al., 2023). Addressing the challenge of approximating the underlying distribution of multimodal data will be paramount. **Human Interactive Machine Learning techniques based on Reinforcement Learning from Human Feedback (RLHF) will be employed to align ML models to ethical and privacry requirements** as we recently did in (Rahimi et al., 2025) for Vision Language Models. The models will be **assessed using publicly accessible databases**, which commonly include audio, video, and physiological sensor data used for **detecting mental states and diagnosing pathologies**. For the second approach, we will **explore multimodal foundation models**, given their success in representing and generating multimodal data through pre-trained models. However, these models also raise **ethical concerns, particularly regarding privacy**. Therefore, there is a timely need to develop new privacy-preserving techniques tailored to such models.

Rationale

This thesis is conducted within a collaborative effort between ISIR (SU) and TICLab **(International University of Rabat, IUR, Morocco), which is a strategic partner of Sorbonne University**. We aim to synergize our expertise in Machine Learning, Social Signal Processing, and Ethics of AI Systems to effectively tackle the challenges and opportunities presented by this thesis. **This collaboration has already led to joint research efforts on multimodal machine learning, resulting in research visits and initial paper submissions** (for example (Rahimi et al., 2025)) Furthermore, this proposal contributes to the Franco-Moroccan Research Center launched in 2024 and to a **proposal for a CNRS International Research Lab in Computer Science**, involving research labs from Sorbonne University (LIP6 & ISIR), Nancy University (LORIA & CRAN), and the Rabat-Salé-Kenitra Region (including UIR).

References

Bourvis, N., Aouidad, A., Spodenkiewicz, M., Palestra, G., Aigrain, J., Baptista, A., ... Cohen, D. (2021). Adolescents with borderline personality disorder show a higher response to stress but a lack of self-perception: Evidence through affective computing. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *111*, 110095.

Drissi, N., Ouhbi, S., García-Berná, J. A., Idrissi, M. A. J., & Ghogho, M. (2020). Sensor-based solutions for mental healthcare: A systematic literature review. In *International conference on health informatics.*

Guerra-Manzanares, A., Lopez, L. J. L., Maniatakos, M., & Shamout, F. E. (2023). Privacy-preserving machine learning for healthcare: Open challenges and future perspectives. In H. Chen & L. Luo (Eds.), *Trustworthy machine learning for healthcare* (pp. 25–40).

Li, C., Gan, Z., Yang, Z., Yang, J., Li, L., Wang, L., & Gao, J. (2023). Multimodal foundation models: From specialists to general-purpose assistants. *ArXiv*, *abs/2309.10020*.

Nikolentzos, G., Vazirgiannis, M., Xypolopoulos, C., Lingman, M., & Brandt, E. G. (2023). Synthetic electronic health records generated with variational graph autoencoders. *npj Digit. Medicine*, *6*.

Rahimi, H., Bahaj, A., Abrini, M., Khoramshahi, M., Ghogho, M., & Chetouani, M. (2025). *User-vlm 360: Personalized vision language models with user-aware tuning for social human-robot interactions.* Retrieved from https://arxiv.org/abs/2502.10636

Zhao, B. Z. H., Kaafar, M. A., & Kourtellis, N. (2020). Not one but many tradeoffs: Privacy vs. utility in differentially private machine learning. In *Proceedings of the 2020 acm sigsac conference on cloud computing security workshop* (p. 15–26). Association for Computing Machinery.