

## Fiche de proposition de projet par les étudiants

Merci de remettre cette fiche à l'équipe enseignante de l'UE  
bruno.gas@sorbonne-universite.fr, nicolas.obin@sorbonne-universite.fr

Titre	Compétition de génération de deep fakes audio-visuel de personnalités publiques
Encadrant(s) (nom, mail, et <b>signature</b> )	Nicolas Obin (Ircam-STMS) Jean-Luc Dugelay (Eurecom)
Matériel requis disponible à la plateforme	Accès à serveur GPUs
Matériel à acheter + budget approximatif	NA
Nombre d'étudiants	6
Prérequis	

### Description détaillée :

L'avènement des deepfakes a créé une nouvelle menace insidieuse pour les personnalités publiques du monde entier. Les deepfakes récents sont des manipulations numériques sophistiquées utilisant des méthodes d'apprentissage profond qui permettent de superposer le visage et la voix d'une personne à une autre, créant ainsi des enregistrements vidéo et audio incroyablement réalistes et trompeurs. Cette technologie innovante pose un certain nombre de problèmes majeurs aux personnalités publiques, pour plusieurs raisons comme la diffamation et la manipulation ou les attaques politiques et électorales. Pour lutter contre la prolifération des deepfakes, les chercheurs développent des détecteurs de deepfakes utilisant des techniques d'intelligence artificielle. Cependant, l'entraînement de ces détecteurs nécessite une base de données contenant des deepfakes de qualité afin d'obtenir les meilleurs détecteurs possibles.

Ce projet s'inscrit dans le cadre du projet ASTRID DeTOX (IRCAM et EURECOM) sur la « Lutte contre les vidéos hyper-truquées de personnalités françaises », sous la forme de challenges entre les étudiants de Sorbonne Université et les élèves de l'école d'ingénieurs EURECOM. Le projet vise à réaliser les meilleures générations de deepfakes audio & visuel de personnalités publiques à partir de l'ensemble des ressources existantes (matérielles, comme les données ou logistiques comme les algorithmes et la disponibilité de modèles pré-entraînés). Pendant une première phase de développement, une première paire de personnalités publiques sera indiquée pour permettre de sélectionner et d'optimiser les modèles génératifs audio et

visuels. Une réunion mensuelle permettra de présenter les avancements de chaque équipe et de juger de la meilleure génération en cours. Chaque équipe dévoilera alors ses sources pour permettre de rééquilibrer le challenge. Dans une seconde phase de réalisation, l'identité d'une personnalité cible utilisée pour le challenge sera dévoilée, et des données publiques et en propres seront rendues accessibles aux équipes. Une attention particulière sera portée sur les meilleurs algorithmes existants et au développement de solution de « fine-tuning » optimisée pour le rendu sur une personne, en visuel comme en audio ainsi que sur les problèmes de synchronisation entre les modalités audio et vidéos et les artefacts possibles liées à l'utilisation d'algorithmes hétérogènes.