

MEDIOS DE COMUNICACIÓN

Un projet scientifique européen étudie les moyens de réguler la désinformation créée par l'IA

L'Universidad Carlos III de Madrid (UC3M) participe à un projet de recherche européen, appelé SOLARIS, qui vise à analyser les risques politiques associés aux contenus multimédias créés à l'aide de l'intelligence artificielle (IA). L'objectif est de proposer des innovations réglementaires dans ce domaine d'un point de vue juridique afin de lutter contre les fausses informations et la désinformation créées par cette technologie.

Les Generative Adversarial Networks (GAN) sont une classe de modèles d'intelligence artificielle capables de créer des contenus multimédias (audio et vidéo) qui ressemblent à la réalité. Le principal défi posé par ces technologies est lié à ce que l'on appelle les « deepfakes » (hypertrucage), de fausses images ou vidéos qui simulent des événements réels avec une précision extrême. C'est ce qui s'est passé récemment, par exemple, avec l'image du pape François portant un manteau Balenciaga, qui s'est avérée être une fausse image. « Cette technologie constitue une menace politique urgente car elle est déjà utilisée pour diffuser des fausses nouvelles et de la désinformation. Cela pose un défi crucial pour la gouvernance et la régulation démocratique. Il est urgent d'améliorer la responsabilité, la transparence et la fiabilité des GAN », déclare Antonio Estella de Noriega, professeur au département de droit public de l'État de l'UC3M, qui est l'un des chercheurs impliqués dans le projet SOLARIS.

Ce consortium européen de R&D&I réagit à ces défis de deux manières. D'une part, les risques politiques associés à ces technologies seront analysés afin de prévenir d'éventuelles implications négatives pour les démocraties de l'UE. En conséquence, des innovations réglementaires seront mises en place pour détecter et atténuer les risques liés aux « deepfakes ». D'autre part, les possibilités offertes par les GAN pour revitaliser l'engagement démocratique des citoyens seront évaluées.

À cette fin, trois études de cas seront développées dans le cadre de SOLARIS. La première vise à comprendre les aspects psychologiques de la perception de la fiabilité des GAN par le biais d'une expérience contrôlée en laboratoire. La seconde simule la circulation de contenus menaçants de GANs sur les réseaux sociaux afin de détecter les risques et de concevoir des stratégies d'atténuation. La troisième co-crée un contenu GAN basé sur des valeurs afin de sensibiliser aux questions démocratiques mondiales essentielles (par exemple, le changement climatique, la dimension de genre, la migration humaine).

« Les GAN sont également l'occasion d'améliorer la conscience démocratique et de développer une citoyenneté active et inclusive », déclare Antonio Estella de Noriega, qui est également titulaire d'une chaire Jean Monnet ad personam en droit de la gouvernance économique européenne à l'UC3M. En ce sens, les GAN « peuvent être utilisés pour faire le bien et avoir des utilisations positives dans le journalisme, l'histoire ou le droit, par exemple », ajoute-t-il.

L'UC3M développe la partie réglementaire de ce projet, qui consiste à faire une proposition de normes juridiques dans ce domaine. « Le plus grand défi est de savoir précisément à quelle vitesse les changements se produisent. Par nature, le droit travaille sur une certaine image figée de la réalité et, d'une manière générale, la réalité n'évolue pas aussi vite que l'IA », explique-t-il. « Ce que nous réglementons aujourd'hui ne sera peut-être plus valable dans six ou neuf mois, car cela pourrait devenir obsolète », ajoute-t-il.

Le projet SOLARIS (Strengthening democratic engagement through value-based generative adversarial networks) est financé à hauteur de près de trois millions d'euros par le programme de recherche et d'innovation Horizon Europe de l'Union européenne (GA 101094665). Coordonné par l'université d'Amsterdam, il se déroulera de 2023 à 2026 et impliquera une douzaine d'institutions et d'entreprises privées d'Albanie, d'Allemagne, de Bulgarie, de Slovaquie, d'Espagne, d'Italie, des Pays-Bas et du Royaume-Uni.

Pour plus d'informations :

Site web du projet SOLARIS: <https://projects.illc.uva.nl/solaris/>

Vidéo: <https://youtu.be/JLKsGXdoe78>