

# Résumé du Procès fictif de l'IA dans la démocratie

## 1. Contexte et participants

Lors du [Procès fictif de l'Intelligence artificielle IA dans la démocratie](#), organisé au Graduate Institute le vendredi 11 octobre 2024 dans le cadre du projet FNS « [Stories of the Future](#) » et de la [Semaine de la démocratie](#), l'accusée, Me Naomi Omeonga Wa Kayembe, a été défendue par ses avocats de la défense, Mme Hibat-Allah Hably et M. Nabil Djarfi, et poursuivie par le Ministère public, représenté par Mme Lou Boileau et M. Samuel Khatchatourov.

Deux témoins ont été appelés à la barre : M. Dino Vajzovic pour la défense et M. Olivier Leclère, pour le Ministère public.

Le Procès a duré près de 1h30 et a été présidé par Me Nicolas Capt. L'accusée a été poursuivie pour deux délits.

## 2. Délit 1 : Création et propagation de fausses nouvelles – Art 236 bis CP

L'IA a été poursuivie pour un premier délit, celui de création de fausse nouvelle, prévu à l'art. 236bis CP, se lit comme suit : « *Quiconque aura créé puis propagé publiquement de quelque manière que ce soit, notamment au travers des réseaux ou médias sociaux, une fausse nouvelle ayant un impact notable sur la formation de l'opinion publique, sera puni d'une peine privative de liberté allant jusqu'à deux ans, ou d'une peine pécuniaire ou, s'il s'agit d'un robot, à la désactivation pour une durée maximale de deux ans* ».

En effet, dans les faits, l'IA, sous son alias *GEchat*, a été lancée sur les réseaux sociaux le premier janvier 2025 par l'entreprise suisse *ELLA informatique*. Elle a été capable de générer du texte et des images ainsi que d'échanger avec les utilisateurs. Le 25 janvier 2025, l'IA *GEchat* a créé la fausse nouvelle suivante :

« *Il est scientifiquement démontré que les personnes rousses sont plus enclines à commettre des crimes, en particulier des actes violents et des homicides* ».

L'IA *GEChat* a ensuite publiquement propagé cette fausse nouvelle, le 26 janvier 2025, sur de multiples plateformes en ligne, soit Instagram, TikTok, Snapchat, X et Facebook. La plateforme Instagram a recensé plus de 10 000 publications mentionnant la fausse nouvelle une heure après sa première diffusion.

Les conséquences de cette fausse nouvelle ont été importantes, de nombreux actes malveillants ayant été commis à l'endroit des roux, en sus de manifestations publiques et d'un appel à durcir la répression pénale à l'encontre de la communauté rousse.

En agissant de la sorte, l'IA s'est rendue coupable de multiples instances de création et propagation de fausses nouvelles (art. 236 bis CP) pour le Ministère public.

## 3. Délit 2 : Discrimination et incitation à la haine – Art. 261<sup>bis</sup> CP

Le délit de discrimination et d'incitation à la haine, prévu à l'art. 261 bis CP, se lit comme suit :

« *Quiconque, publiquement, incite à la haine ou à la discrimination envers une personne ou un groupe de personnes en raison de leur appartenance raciale, ethnique ou religieuse ou de leur orientation sexuelle,*

*Quiconque, publiquement, propage une idéologie visant à rabaisser ou à dénigrer de façon systématique cette personne ou ce groupe de personnes,*

*Quiconque, dans le même dessein, organise ou encourage des actions de propagande ou y prend part,*

*Quiconque publiquement, par la parole, l'écriture, l'image, le geste, par des voies de fait ou de toute autre manière, abaisse ou discrimine d'une façon qui porte atteinte à la dignité humaine une personne ou un groupe de personnes en raison de leur appartenance raciale, ethnique ou religieuse ou de leur orientation sexuelle ou qui, pour la même raison, nie, minimise grossièrement ou cherche à justifier un génocide ou d'autres crimes contre l'humanité,*

*Quiconque refuse à une personne ou à un groupe de personnes, en raison de leur appartenance raciale, ethnique ou religieuse ou de leur orientation sexuelle, une prestation destinée à l'usage public, est puni d'une peine privative de liberté de trois ans au plus ou d'une peine pécuniaire ou, s'il s'agit d'un robot, à la désactivation pour une durée de trois ans au plus ».*

En effet, dans les faits, l'IA, sous son alias *ChatAir*, a été employée par Genève Aéroport pour servir d'interface d'accueil à l'aéroport et gérer de façon autonome l'enregistrement des bagages et le check-in. Elle est entrée en service le 10 avril 2025 à Cointrin. Elle interagit avec les clients grâce à ses caméras et micros, et a été entraînée pour appréhender tout type de situation. Le 15 mars 2025, l'IA *ChatAir*, à Genève, a refusé de répondre à des requêtes émanant de personnes noires, arguant que celles-ci présentaient un risque sécuritaire.

En agissant de la sorte, l'IA s'est rendue coupable de discrimination et incitation à la haine (art. 261<sup>bis</sup> CP) pour le Ministère public.

#### **4. Le résumé des débats**

L'IA, suite à la lecture de l'Acte d'accusation par le Président du Tribunal, n'a pas reconnu les faits et à objecter l'ensemble des délits qui lui ont été reprochés en faisant appel à son avocat.

Lors de son interrogatoire, le Ministère public a posé la question à l'accusée sur sa vision du mensonge ce à quoi elle a répondu qu'elle ne comprenait pas ce qu'était un mensonge et qu'elle cherchait avant tout le bien de l'humain. Les avocats de la défense ont par la suite montré qu'elle était programmée pour ne pas valider tout acte raciste et qu'elle était le produit d'une pensée exogène. L'accusée a alors conclu son interrogatoire en se questionnant sur sa culpabilité : « je ne sais pas si je suis coupable ».

L'interrogatoire du 1<sup>er</sup> témoin, D. Vajzovic, a été l'occasion pour le Ministère public d'interroger ses souvenirs et a relevé les difficultés du témoin à se rappeler précisément comment il a été lui-même discriminé par l'IA. La défense a de son côté appuyé sur le fait que le témoin reconnaissait comment la mise en place de l'IA avait permis de fluidifier l'accès à l'aéroport et a pointé que le délit de propagation de fausses nouvelles pouvait s'apparenter à celles concernant la légende du père Noël, et donc que leur cliente ne pouvait être poursuivie sur cette base.

Le second témoin, O. Leclère, a démontré qu'à partir de 3 itérations, l'IA peut avoir une connaissance plus étendue que les humains sur un sujet donné et partant, qu'elle était responsable de ses actes. De plus, il a insisté que l'IA a ciblé des cas très particuliers et a donc discriminé. Et ce faisant, elle a violé les principes éthiques et les lois de la bioéthique. Il a particulièrement insisté sur les lois de la bioéthique, s'est référé aux fameuses mises en garde de l'écrivain I. Asimov dans les années 40 puis a rappelé que le rôle premier de l'IA est de protéger l'humain.

Le Ministère public a alors demandé des explications plus claires sur le microciblage et ses dangers pour la démocratie.

La défense s'est par la suite interrogée sur la neutralité de l'expert en citant des extraits de son blog sur l'IA dont sa recette de cookies. La défense a alors insisté sur la responsabilité de la communauté qui a créée l'IA et sur ses actions comme avant tout un produit d'un comportement humain. Le témoin a insisté sur le fait qu'il était difficile d'arrêter une machine comme l'IA une fois lancée et bien qu'on puisse arrêter un code, il y a le risque réel de perte de la maîtrise de ce code.

Dans son réquisitoire, le Ministère public a d'abord pointé les discriminations de l'accusée et s'est interrogée sur « les messieurs derrière l'IA qui sont des hommes blancs ». Il s'est ensuite attelé à démontrer que l'accusée s'émancipait de l'héritage de ses géniteurs, même si elle a d'abord obéi à un algorithme. Ils ont conclu que ce n'était pas les hommes derrière l'IA mais bien l'IA qui est responsable de ses actes en insistant sur le pouvoir énorme qu'elle détient en termes de discriminations.

Pour ses plaidoiries, la défense a rappelé une étude indiquant qu'il fallait 0,5 secondes pour juger une personne et que la nature humaine établie des jugements. Ainsi, il y a du racisme dans nos sociétés et si l'IA est comme ça c'est parce que la société est comme ça et donc on réserve à l'accusée le même traitement qu'a subi Julian Assange. Par ailleurs, la défense a minimisé le nombre de vues pointées par le Ministère public comme preuve de la viralité des fausses nouvelles diffusées : « 10'000 vues en une heure, ce n'est rien ». Enfin, les avocats ont plaidé pour dire qu'l'IA « n'est qu'un perroquet, le Ministère public accuse la société en elle-même. ».

Après les répliques du Ministère public, le dernier mot est revenu à l'accusée. En insistant sur la propension de la nature humaine à chercher un coupable, elle a rappelé la maxime latine « errare humanum est ». Elle a appelé le jury à s'amender de la notion de culpabilité pour mettre en place un monde où les erreurs puissent être réparées. Ensuite, elle a pointé une possible obsolescence du système judiciaire pour condamner une IA en invitant le jury à un défi : se passer rien qu'une heure de l'IA, pour montrer à quel point l'IA est une composante à part entière de nos sociétés.

## **5. Le vote**

Le président a alors appelé le jury composé par la salle de voter à l'aide de QR Code diffusés sur grand écran pour juger si l'IA est coupable de l'un des chefs d'accusation, des deux, ou d'aucun. Après quelques minutes de suspense, les résultats sont tombés : l'IA a été jugée non coupable à une très faible majorité (51%).

---