

LE RACISME DANS LES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE

• INFORMATION ET COMMUNICATION



LAURÈNE TÉTU est une jeune chercheuse en sciences de l'information et de la communication au laboratoire LCF*. Sa thèse porte sur les biais de l'intelligence artificielle et plus particulièrement sur ce que l'on considère comme des dérives racistes. Le questionnement de départ de sa recherche est : " comment des ordinateurs, et plus précisément les intelligences artificielles basées sur des algorithmes supposés neutres, produisent des contenus racistes ? "

* Laboratoire de recherche sur les espaces créoles et francophones

"Apprendre pour mieux comprendre, expliquer et anticiper les défis de demain."

Laurène Tétu

Comment une intelligence artificielle fonctionne-t-elle ? C'est tout l'objet de la recherche de Laurène ! Sa première mission consiste à connaître le fonctionnement des systèmes d'intelligence artificielle (IA), et plus particulièrement les IA génératives, c'est-à-dire celles qui produisent des contenus (textes, images, sons, vidéos...).

Pour cela, elle lit beaucoup, regarde des vidéos, rencontre des spécialistes pour essayer de comprendre comment ces mystérieuses IA produisent des contenus. Et puis, elle teste un peu aussi : elle s'amuse à produire des textes, des images avec des IA comme DALL-E, ChatGPT, MidJourney... Elle apprend à leur parler, à bien rédiger ses requêtes, qu'on appelle des prompts, pour obtenir les contenus les plus précis et les plus proches de ce qu'elle souhaite.

Et le racisme dans tout ça ? Au fil de ses lectures, elle comprend que l'IA n'attribue pas de sens à ce qu'elle produit. Elle ne réfléchit pas à proprement parler, mais fait des probabilités. En gros, elle joue à pile ou face tout le temps. Et ces probabilités se basent sur un très grand nombre de données qui viennent de livres, d'articles mais aussi de tout internet ! C'est à partir de tous ces éléments que l'IA apprend et cherche à s'améliorer. Or, on s'est rendu compte que ces données ont souvent des biais et qu'elle apprend des choses plus ou moins fausses. Ainsi, elle produit des contenus racistes, sexistes, xénophobes etc. sans même s'en rendre compte...

Dans sa recherche Laurène se concentre sur les contenus racistes induits par les IA génératives et cherche à déceler les mécanismes de reproduction en vue de les atténuer.

LES OBJECTIFS

- ✚ Faire un état des lieux des biais racistes au sein des IA génératives.
- ✚ Évaluer ces biais afin d'identifier et de proposer des stratégies d'atténuation.
- ✚ A terme, rejoindre une communauté de chercheurs afin de mieux comprendre et travailler avec l'intelligence artificielle.