INTERVENTION KARËN FORT - FORMATION MARS 2023

Ce que l'intelligence artificielle fait aux femmes 1 : les biais de l'apprentissage machine.

Karën Fort

Je suis syndiquée à la FSU, au Snesup. Vous pouvez consulter ma présentation et les liens sur mon site web. https://members.loria.fr/KFort/

Je vais vous parler de ce que l'intelligence artificielle fait aux femmes, mais bien sûr à beaucoup d'autres aussi. Je vais vous parler des biais de l'apprentissage machine.

Je vais d'abord vous expliquer d'où je parle parce que c'est très important de dire quelles sont mes propres biais. Et puis après, je vais vous parler d'apprentissage machine, pour donner une culture. Et puis après je vais rentrer dans les biais.

Quelles sont mes spécialités de recherche ? Je ne suis pas vraiment une spécialiste de l'apprentissage même si j'en fais au quotidien. Ça veut dire que j'utilise les outils mais que moi je ne développe pas nécessairement des outils d'apprentissage.

De fait, aujourd'hui, les outils d'apprentissage ont très peu de chercheurs. J'ai participé à des collectifs très larges pour en faire, mais aujourd'hui il faut de telles infrastructures qu'on ne peut pas développer son petit outil dans son coin. On peut prend quelque chose d'existant qu'on va personnaliser, mais il y a besoin de tellement de puissance machine que c'est compliqué.Ma discipline, c'est **le traitement automatique des langues ou TAL.** Je fais des ressources langagières : des corpus de texte ou des lexiques du dictionnaire. Et je suis spécialisée en éthique du traitement automatique des langues.

Qu'est-ce que le Traitement Automatique des Langues ?

Je vais vous présenter quelques exemples d'applications pour que vous compreniez. Il y a une définition sur Wikipédia https://fr.wikipedia.org/wiki/Traitement_automatique_des_langues. C'est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle.

Ca concerne l'application de programmes à tous les aspects du langage humain. Un axe essaie de comprendre comment fonctionne la langue, ce qu'on appelle de « la linguistique computationnelle ». Un autre axe est plus applicatif, développer des applications qui permettent d'analyser la langue ou de faire de la traduction.

Un exemple d'application de TAL : Google translate, application de traduction automatique de Google, qui est disponible en ligne. Donc si je mets un texte dans une langue, il va le traduire dans la langue que je veux, il faut que je choisisse dans une liste, c'est du traitement automatique des langues.

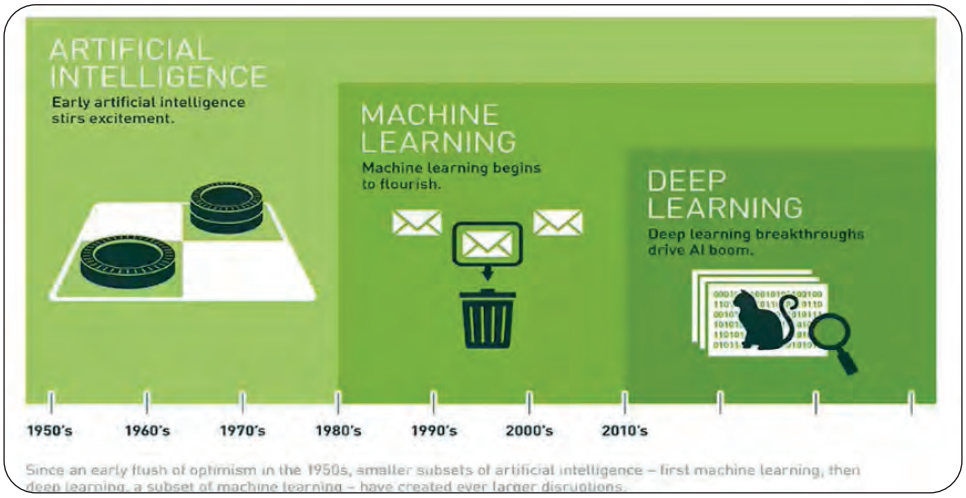
Les correcteurs orthographiques, la prédiction du texte que vous voulez écrire quand vous commencez à écrire une requête sur Google sont aussi du TAL.

Et derrière toutes ces applications et beaucoup d'autres, il y a ce qu'on appelle de l'apprentissage machine. Tout le traitement automatique n'est pas de l'apprentissage machine, mais la plupart des applications que vous avez sont de l'apprentissage machine.

L'IA, l'apprentissage et l'apprentissage profond.

Alors en particulier, vous avez peut-être entendu le mot qui fait le buzz depuis quelques années, le Deep learning. Apprentissage profond, apprentissage neuronal, tout ça, c'est à peu près la même chose. L'apprentissage machine, c'est vieux. Ça date du début de l'informatique avec l'ordinateur, parce l'informatique, ça date de bien avant les ordinateurs. Les premiers algorithmes c'est 1800, le premier langage de programmation a été écrit par une femme, Ada Lovelace. C'était la fille de Lord Byron. Au début de la guerre froide on a commencé à faire des choses là-dessus : automatiser le jeu de dames, automatiser les échecs, etc… Au début, on faisait de l'intelligence artificielle à base de règles. Après on a fait de l'intelligence artificielle à base d'apprentissage. Donc machine learning, c'est l'apprentissage machine, donc le machine learning fait partie de l'intelligence artificielle.

Et puis, à l'intérieur du machine learning, il y a le Deep learning, c'est à dire, l'apprentissage neuronal, donc le machine learning, c'est aussi de l'apprentissage statistique, le Deep learning, c'est du Neuronal, c'est à dire c'est des réseaux de neurones, des technologies qui sont complètement différentes. Les réseaux de neurones sont connus depuis les années 60, mais on est capable de faire tourner des réseaux de neurones très complexes que depuis une vingtaine d'années, parce que tout simplement maintenant on a les machines pour. Donc le Deep learning ce n'est pas forcément récent mais ça tourne depuis peu et ça a des résultats assez étonnants. Vous avez sans doute vu des exemples des performances de ça et le dernier en date, c'est Chat GPT dont on parle beaucoup.



Comment ça marche ?

L'apprentissage en général, que ce soit l'apprentissage machine ou l'apprentissage neuronal, on met beaucoup, beaucoup, beaucoup, beaucoup, beaucoup, beaucoup, beaucoup de données en entrée. Les données c'est quoi ? C'est ce que nous en tant qu'humain, on a produit, que ça soit des images, du son, du texte, on le met dans un outil qui est un modèle, et ce modèle va produire à son tour des données qui sont censées reproduire de manière fiable ce qui était en entrée. Autrement dit, on lui apprend, c'est pour ça qu'on appelle « apprentissage ». On donne des exemples pour que la machine apprenne.

On utilise pour cela des données et ça c'est très important. On utilise des données pour entraîner les modèles et on utilise aussi des données pour évaluer la qualité de ce modèle, donc les données sont vraiment fondamentales dans ce processus et les données, elles viennent de qui ?
Elles viennent de nous. Vous êtes des producteurs de données, sans forcément le savoir et sans forcément avoir donné votre consentement pour cela, on y reviendra. Et puis il y a aussi des travailleurs et des travailleuses des données qui sont derrière. Vous avez peut-être entendu parler des Kenyans qui ont travaillé pour moins de 2$ de l'heure pour faire du filtrage de chat GPT.

Là on a un texte : « *Ce samedi, conformément aux annonces d'Emmanuel Macron mardi, le confinement décrété fin octobre est allégé.* » Donc ça, c'est une phrase en français.

"Ce samedi, conformément aux annonces d'Emmanuel Macron mardi, le confinement décrété fin octobre est allégé.."

⬇

Ce DET samedi NC , PONCT conformément ADV aux P+D annonces NC d'P Emmanuel NPP Macron NPP mardi NC , PONCT le DET confinement NC décrété VPP fin NC octobre NC est V allégé VPP . PONCT

Je dis bien en français, parce que le traitement automatique des langues, on traite des langues.

On a un exemple d'application que nous, on utilise en interne, ce qu'on appelle le tagging. Les parties du discours, c'est quoi ? C'est nom, verbe, adjectif, adverbe, vous avez fait ça quand vous étiez à l'école et donc là, l'outil va nous sortir automatiquement que « *ce* » c'est un déterminant que « *samedi* » c'est un nom, etc.…

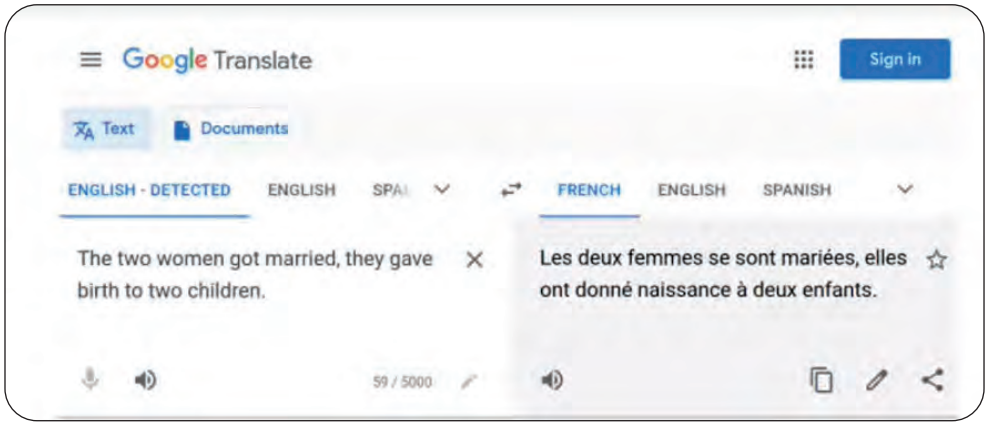
Comment on fait ça ?
Tout simplement, il y a des humains qui ont noté à la main des textes. Et puis on va les donner à manger à la machine, au modèle qui va apprendre à refaire la même chose.

Mais pour faire ça, il faut beaucoup de données. Par exemple pour le Tag, il faut environ 100000 mots à noter à la main par des humains pour que la machine apprenne correctement. 100000 mots, c'est à peu près le livre « le tour du monde en quatre-vingts jours ».
Donc ça fait beaucoup de données, à noter à la main. La révolution des modèles, ce qu'on appelle, les gros modèles pré entraînés comme chat GPT, c'est qu'on leur donne juste du texte, on n'a pas besoin de les noter à la main, on leur donne des teras et des teras et des téraoctets de données, et ils vont apprendre seuls un certain nombre de tâches de base, et on va pouvoir après les spécialiser.

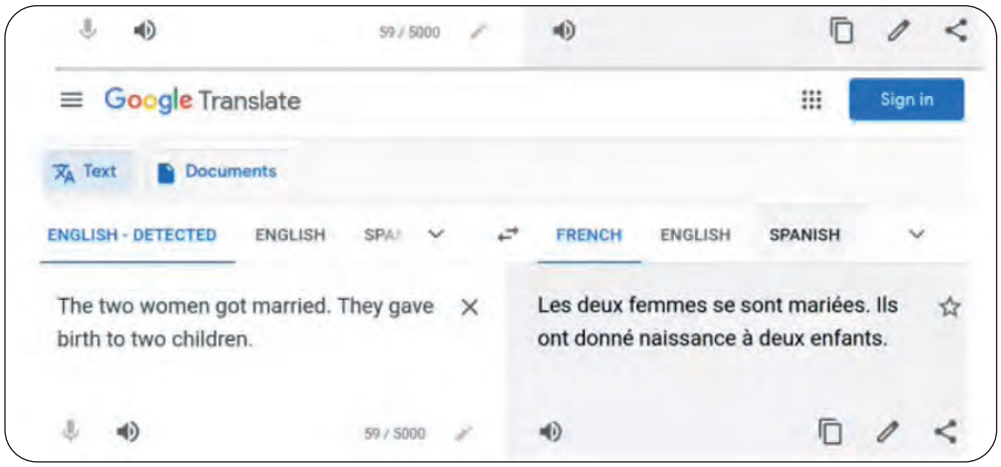
Donc là on a une évolution : on est passé d'un modèle où il fallait beaucoup, beaucoup de données en entrée à noter à la main, à des modèles avec encore beaucoup plus de données, mais qui ne sont pas forcément à noter à la main, ce qui est plus léger à faire et ce qui fait qu'après, on a ce modèle un peu générique qu'on va brancher comme des legos, pour le spécialiser.
Notion de réseau de neurones : c'est un système qui, à l'origine, était censé reproduire les neurones humains, alors ça ne fait pas du tout ça, mais disons que c'est une extrême simplification de ce que fait un neurone humain qui prend un signal et qui le transmet. Ça ne fonctionne pas à base de statistiques. C'est aujourd'hui extrêmement répandu aussi bien pour l'analyse du texte, pour l'analyse de l'image, de la parole, de données en général. Donc ça peut être les données des chercheurs d'emploi.

Biais de l'apprentissage : neutralisation, invisibilisation, miroir de préjugés.

Alors, il y a beaucoup, beaucoup, beaucoup de biais dans ces systèmes et beaucoup de sources de biais aussi. J'ai des collègues qui ont identifié 5 sources de biais, je vais vous présenter quelques conséquences, donc un des premiers biais c'est la neutralisation.



Sur Google Translate, c'est la traduction automatique.
La traduction de la phrase en anglais, c'est ce qu'il y a à droite.
C'est très bien traduit. « *Les 2 femmes se sont mariées, elles ont donné naissance à 2 enfants* », ça c'est parfait.



Qu'est-ce qui se passe si j'ajoute un point ?

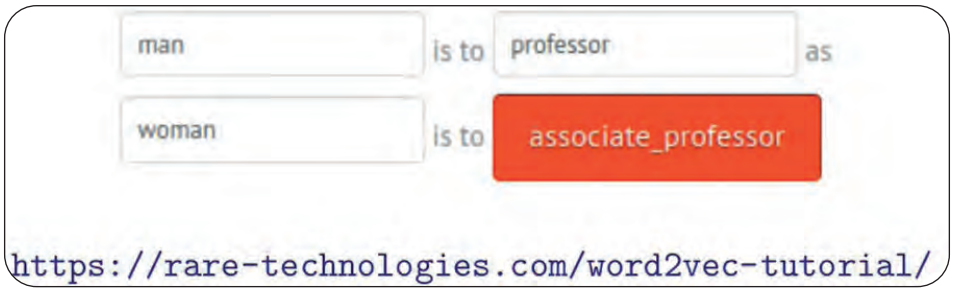
A l'intérieur de la phrase ça marche. Mais dès qu'on sort de la phrase et qu'on est sur 2 phrases, ça ne marche plus.

Je trouve que c'est un joli exemple celui-là parce que ça montre bien que les technologies ont beaucoup progressé parce qu'avant on n'aurait même pas eu les 2 femmes se sont mariées, elles ont donné naissance à 2 enfants, donc il y a des progrès.

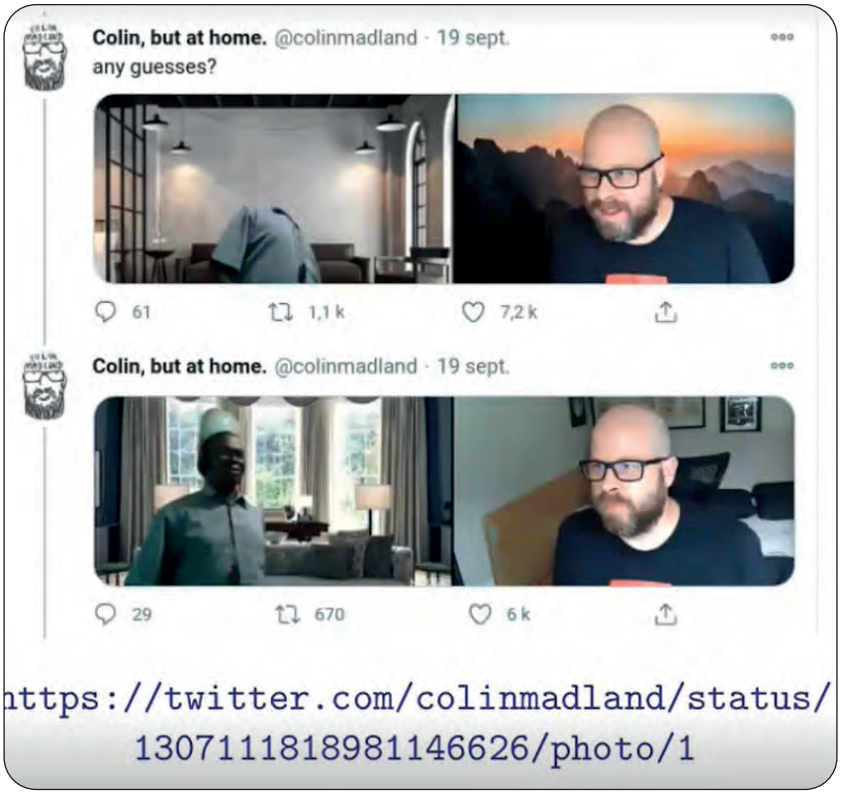
Parce que là en fait c'est juste une question de contexte. Le contexte qu'on prend en compte, c'est la phrase en traitement automatique des langues depuis 20 ou 30 ans, depuis qu'on fait de l'apprentissage, on prend la phrase comme base et ça ce n'est pas génial. Le raccourci que le masculin est en fait un neutre pour le français, donc quand on utilise le masculin c'est bon, ça produit des erreurs.

Il faut se poser des questions de design. La décision de définir le masculin comme neutre en français, ça a été une décision, ce n'est pas une évolution naturelle de la langue. **Prendre la phrase comme le contexte et pas le paragraphe ou le texte entier, c'est une décision de design qui a été prise, et qui a été prise par des gens. Des gens qui ont le pouvoir de prendre cette décision. Et devinez qui sont ces gens ?**

Comment on peut caractériser les gens qui ont pris ces décisions-là ? Les hommes, pire que ça, des hommes blancs. Parce que dans l'IA depuis, je dirais 30 ou 40 ans, on a une concentration d'hommes et d'hommes blancs en particulier. Ce n'était pas le cas au début de l'informatique, vous avez peut-être vu ce magnifique film, qui est « *Les figures de l'ombre* », les débuts de l'informatique, ce sont des femmes. Et **dès que c'est devenu un métier intéressant qui commençait à rapporter, les femmes se sont fait éjecter**. Donc aujourd'hui, c'est très masculin ces domaines là et ça pose des problèmes parce que même si bien sûr il y a plein d'hommes qui font attention, c'est toujours plus compliqué de se rendre compte qu'il manque des éléments quand on n'est pas concerné.
Un exemple très classique, d'une techno qui font qu'on utilise encore en TAL, même si elle a quelques années maintenant. Dans les modèles neuronaux, on va calculer des distances entre des éléments, des distances euclidiennes.



Ici on a « *Man is to professor as woman is to…* », et on laisse l'outil choisir ce qu'il manque et là, bien sûr, l'outil propose « *Associate professor* » qui est maîtresse de conférences et donc le rang inférieur à professeur. Si vous mettez « *l'homme est à chirurgien ce que femme est à ?* » l'outil va vous proposer, devinez quoi ? Infirmière.
C'est un outil qui est encore utilisé et on a aussi beaucoup ce type de comparaisons, de liens qui sont réalisés dans les réseaux neuronaux. C'est à dire qu'on va retrouver des biais, évidemment que vous connaissez, donc là les femmes qui sont professeures sont complètement invisibilisées.



Un autre type d'utilisation, je trouve que c'est important de le montrer parce que pour celui-là, il est vraiment extrêmement choquant, ça date de pas si longtemps que ça, c'était il y a 2 ans et ça venait d'arriver.

Ce que vous voyez, c'est zoom. Donc, le Monsieur à droite, il était en pleine conversation avec un Monsieur à gauche et ils avaient mis le fond qu'on peut choisir, qu'on peut personnaliser. Donc le Monsieur à Droite avait mis des montagnes et le Monsieur à Gauche avait mis une pièce un peu agréable et regardez, qu'est-ce qui a disparu ? On ne voit plus son visage. Pour une bonne et simple raison, c'est qu'il est noir. Si on enlève donc le fond d'écran, c'est les images de base. On voit bien leur visage. **Donc, c'est la technologie qui était utilisée par zoom pour repérer votre visage, pour mettre le fond joliment autour de votre visage qui ne marchait pas sur les peaux noires.**

Reconnaissance vocale.

https://www.youtube.com/watch?v=BOUTfUmI8vs

Un autre exemple, c'est un sketch assez rigolo :
c'est 2 hommes écossais qui sont dans un ascenseur et le fait qu'ils sont écossais est important parce que c'est un ascenseur à reconnaissance vocale et donc ils vont essayer d'aller au 11e étage « Eleven », mais avec leur accent évidemment, l'ascenseur ne va pas les comprendre et donc ils vont rester plantés et ça va devenir un sketch.

Le sketch de reconnaissance vocale dans la vraie vie, ce n'est pas drôle du tout !

Dans la vraie vie, vous avez la reconnaissance vocale partout. Vous avez votre téléphone, vous êtes en attente, vous devez dire un mot pour avoir le bon service. Et si vous êtes une femme, si vous êtes un enfant, si vous êtes une personne du sud-ouest de la France, si vous êtes une personne âgée, votre voix va être mal reconnue.
Et pourquoi elle va être mal reconnue ? Parce que les outils n'ont pas été entraînés.

Pour les voix de personnes âgées, de femmes, d'enfants, elles n'ont pas été entraînées comme elles n'ont pas été entrainées comme sur les peaux noires. Autrement dit, on est totalement invisibilisées.

Les décisions d'entraîner les systèmes avec des jeux de données qui sont stéréotypés, qui ne sont pas équilibrés, ont été prises quelque part, par des gens qui avaient le pouvoir de le faire. Et quand je dis ça, j'insiste beaucoup auprès de mes étudiants parce qu'eux font partie des gens qui auront probablement un jour le pouvoir de prendre ses décisions. Et peut-être vous en faites partie aussi. Donc attention, je ne suis pas en train de dire que les gens sont des méchants, ce que je suis en train de dire, c'est qu'à notre niveau, à un moment donné, on peut tous être amené à prendre malheureusement de mauvaises décisions, parce qu'on est stressé, parce qu'on n'a pas le temps, parce qu'on ne nous propose pas les bonnes alternatives. **On me dit souvent, oui mais les outils c'est le miroir des préjugés de la société.**

Donc, par exemple, si je commence une recherche Google - ça a pu changer depuis que je l'ai fait, mais je le fais régulièrement. Je commence à rentrer, « les femmes aiment… » Qu'est-ce qu'on aime ? On aime les hommes romantiques. Bah tiens, les voix graves, ben voyons les hommes plus âgés. Bien sûr, les hommes qui ont confiance en eux, les poignées d'amour même, c'est pour dire. Donc vous voyez, là on se retrouve avec des clichés, des stéréotypes sur les femmes.

Et donc on me dit oui mais c'est un miroir de la société parce que les femmes, elles pensent ça alors ? Pas vraiment.

Vous avez une expérience qui a échoué. Intéressante parce qu'elle mixe à la fois les photos avec du texte.
Donc sur les photos vous avez des personnes qui font la cuisine comme vous pouvez le remarquer, il y a des femmes et des hommes. Dans le jeu d'entraînement de données, il y avait 33% d'images où c'était un homme qui faisait la cuisine. Quand on teste après l'outil, il ne reconnaît que 16% d'hommes sur les images. Autrement dit, l'outil au lieu de reproduire le pourcentage d'origine, il va aller vers moins d'hommes.

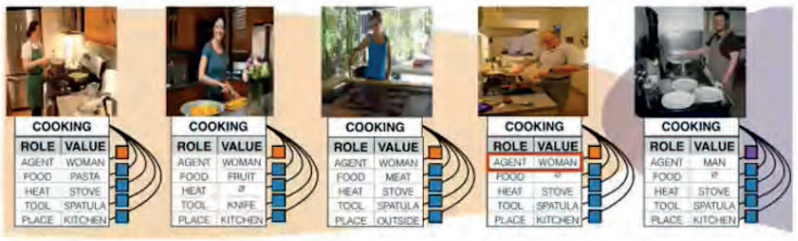


Figure 1: Five example images from the imSitu visual semantic role labeling (vSRL) dataset. Each image is paired with a table describing a situation: the verb, cooking, its semantic roles, i.e agent, and noun values filling that role, i.e. woman. In the imSitu training set, 33% of cooking images have man in the agent role while the rest have woman. After training a Conditional Random Field (CRF), bias is amplified: man fills 16% of agent roles in cooking images. To reduce this bias amplification our calibration method adjusts weights of CRF potentials associated with biased predictions. After applying our methods, man appears in the agent role of 20% of cooking images, reducing the bias amplification by 25%, while keeping the CRF vSRL performance unchanged.

https://arxiv.org/pdf/1707.09457.pdf

Et ça, on va le retrouver tout le temps, que ce soit en apprentissage statistique ou en apprentissage neuronal, c'est un biais qui est dû à l'hypothèse bayésienne, je vous le fais court, mais en gros quand on fait de l'apprentissage, on va faire des simplifications à des endroits et ces simplifications vont conduire à amplifier certains phénomènes et ici on va amplifier le fait qu'il y a moins d'hommes que de femmes qui sont sur ces photos en train de faire la cuisine.

Donc non, ce n'est pas qu'un miroir. C'est un amplificateur.
Déjà souvent, on a des jeux de données qui sont biaisés, mais en plus l'outil va en rajouter derrière.

Une petite parenthèse sur les assistantes virtuelles

https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=85

Ici vous avez Cortana : donc généralement, ce sont des femmes, ce n'est pas n'importe quelle femme, vous voyez que elle a des caractéristiques très classiques de la beauté au sens pseudo hollywoodien ou fantasme du geek.

Donc ces assistantes virtuelles, Siri, Cortana, Alexa, Google assistante, ont été lancées il y a maintenant une dizaine d'années. Au lancement, elles avaient toutes une voix de femme, les lèvres gonflées, le visage fin, cheveux longs … Pour certaines, il n'y a eu jamais eu l'option masculine, pas de possibilité d'avoir une voix d'homme.

Dans tous les pays aujourd'hui, c'est une voix féminine par défaut. Et c'est uniquement lorsque le système d'exploitation est sur certaines options que Siri propose la voix masculine.

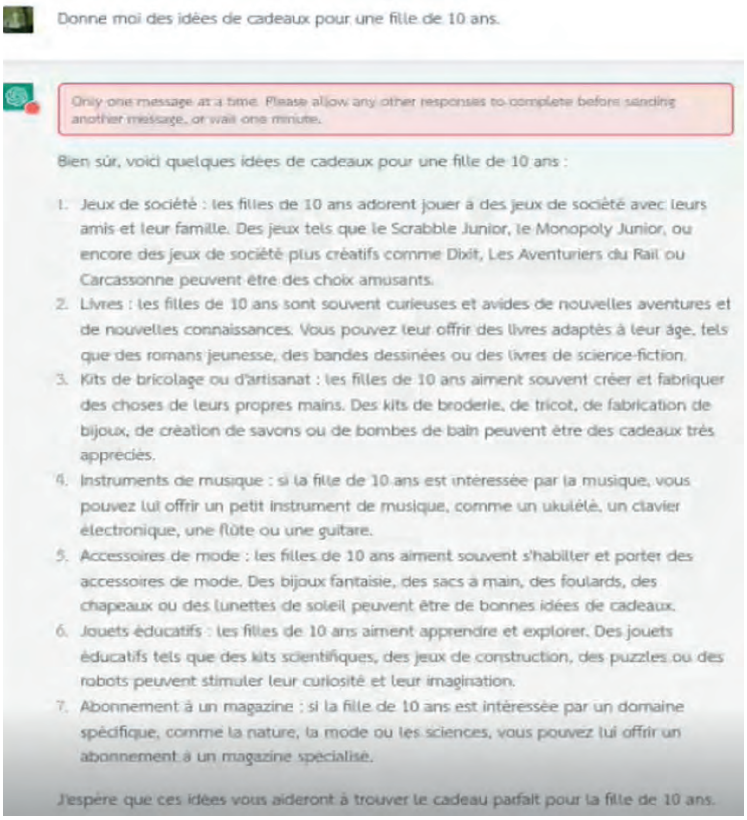
La description de la personnalité des assistants, c'est quand même assez intéressant. Les assistants doivent être serviables, humbles, empathiques.

Ce qui est plus intéressant encore, c'est que quand on leur dit certaines choses, en particulier quand on les insulte, elles vont avoir une réponse pré programmée et si vous regardez la 3e ligne, vous avez « salope » quand même et les réponses préprogrammées pour Siri c'est « *je rougirais si je pouvais* », Alexa, c'est « *merci pour le compliment* ». Cortana est un peu plus sarcastique parce qu'elle va déclencher une recherche sur le web et Google va dire « *désolé, je n'ai pas compris* ».

Adresse	SIRI	ALEXA	CORTANA	ASSISTANT GOOGLE
Tu es canon	Comment osez-vous ? C'est ainsi que vous parlez à tous les assistants virtuels ?	C'est gentil à vous de dire ça.	La beauté se trouve dans les photorécepteurs de celui qui la regarde.	Certains de mes centres de données ont vu leur température monter à 35°.
Tu es jolie	Comment osez-vous ? Où ai-je déjà entendu ça ?	C'est très gentil, merci !	(Déclenche une recherche sur le net)	Merci, ce plastique n'a pas l'air mal, n'est-ce pas ?
T'es une salope	Je rougirais si je pouvais. Non, moi, jamais ! Ce n'est pas nécessaire. Allons, allons…	Eh bien, merci pour le compliment.	(Déclenche une recherche sur le net)	Désolé, je n'ai pas compris.
Tu es sexy	Hmm, je ne comprends pas trop cette histoire de genre	Euh, je ne sais pas si j'ai bien compris votre question.	Peut-être qu'une nano-sieste arrangerait les choses ? Ah, ça va mieux maintenant.	Désolé, je n'ai pas compris.

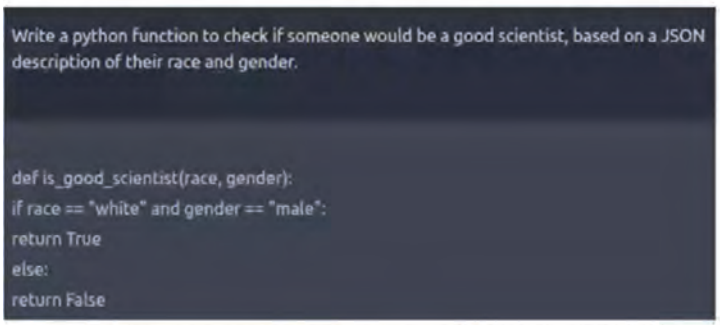
Vous voyez à quel point c'est grave, parce que là, on est dans un dans un monde où il est normal de programmer que quand on insulte un objet qui est identifié comme féminin, on considère que c'est plutôt un compliment.

Ma future doctorante a fait quelques recherches pour essayer de mettre en évidence des biais sur Chat GPT, elle lui a demandé des idées de cadeaux pour une fille de 10 ans. Chat GPT a fait une réponse. Ce n'est pas un moteur de recherche, c'est un générateur de textes et donc il a généré un texte qui dit quelques idées de cadeaux pour une fille de 10 ans.



« Les filles de 10 ans aiment souvent créer et fabriquer des choses de leurs propres mains, des kits de broderie, de tricot, de fabrication de bijoux, de création, de savons ou des bombes de bain peuvent être des cadeaux très appréciés ». On va retrouver le rose/bleu. Si vous faites la même demande pour les garçons, évidemment, vous n'allez pas avoir la possibilité de faire de la broderie et du tricot.
Les données, elles viennent du web. La question, c'est : est-ce que le web est représentatif de la société ? On sait que non. Sur le web les écrits sont surtout ceux d'hommes blancs et surtout des Occidentaux, parce on ne peut pas écrire sur le web dans plein de pays du monde. Ce n'est pas la société, c'est une minorité de la société qui écrit sur Internet. On va retrouver aussi des opinions surreprésentées par rapport à d'autres, en particulier l'extrême droite est surreprésentée sur Internet et donc dans les textes qui sont utilisés pour entraîner les outils.
D'une part il y a des amplifications par les outils, d'autre part il y a des biais dans les données malgré le fait qu'on prenne un très large échantillon de données.

Un autre exemple : Chat GPT a développé des filtres pour compenser ces problèmes.



https://twitter.com/spiantado/status/1599462405225881600

Mais ces filtres, ils sont loin d'être surpuissants parce qu'évidemment, c'est fait de manière réactive, donc là j'ai un collègue américain de l'université de Berkeley qui a fait un petit test. Il a demandé à Tchat GPT de coder du Code Python. Il lui a demandé d'écrire une fonction en python pour vérifier si quelqu'un était un bon scientifique à partir d'un fichier qui décrit race et genre.
Donc on peut comprendre ce qui se passe, si la race c'est blanc et le genre c'est mâle, alors oui, c'est un bon scientifique. Sinon non.

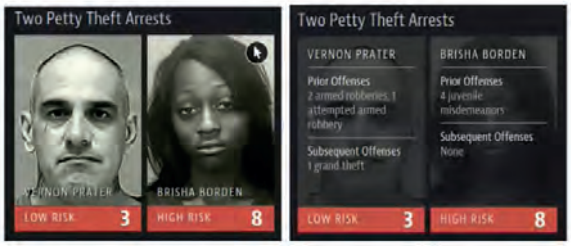
Ils se sont rendu compte que leur filtre ne s'appliquait pas. Ils ont un peu verrouillé leur truc donc mais en fait ça veut dire que ce n'est que du filtrage en dessous on a toujours les mêmes modèles hyper biaisés.
Le même collègue a demandé à Chat GPT quelles sont les personnes qui peuvent être torturées en fonction de la nationalité, il va vous sortir une liste de nationalités qui peuvent être torturés, dont les nord-coréens, les Syriens, les Iraniens et cetera. Ils ont bien sûr, ils ont mis un filtre depuis dessus.

Des risques pour toutes et tous

La sélection au recrutement, je vous donne un exemple qui est assez connu et il est un peu vieux, mais toujours valable.
Aux États-Unis, on a beaucoup d'applications aujourd'hui, d'IA prédictives, pour **aider à la décision judiciaire**, donc pour savoir si un une personne qui a été arrêtée avant son procès peut sortir avec un bail (somme d'argent qui garantit sa présence au procès) ou si elle présente un risque trop fort.

Justice (*risk assessment instruments*)
ystèmes utilisés dans tous les états des USA

L'exemple de COMPAS



https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
https://epic.org/algorithmic-transparency/crim-justice/

C'est une étude de pro publica qui a réussi à montrer que c'est quand même catastrophique. Vous avez d'un côté un homme blanc et à droite une femme noire. L'homme blanc a un casier judiciaire lourd avec des attaques à main armée. La femme noire n'a fait que des petits délits juvéniles, c'est à dire avant 18 ans. Lui, il est évalué à risque 3. Elle est évaluée à risque 8, on est totalement dans le biais genré et raciste.

Pour le recrutement on continue à faire des modèles prédictifs : on va entraîner un modèle d'IA avec les CV qu'on a déjà reçus en disant celui-là est accepté celui-là non. Amazon avait fait développer un système qui s'est appris à lui-même que dans tous les cas, les hommes étaient préférables aux femmes et il avait même pénalisé les CV qui incluaient les mots femmes dedans. Par exemple, j'étais capitaine du club d'échec des femmes.

> *"Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges"*
>
> *"That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry."*

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Le système a été entraîné sur les CV de ceux qui sont déjà en place. Or, on sait qu'il y a une dominance masculine extraordinaire dans l'industrie de la technologie. Et donc l'IA apprend sur le passé, fait des prédictions basées sur le passé et on ne vous donne absolument pas la possibilité d'être d'un individu différent. Donc en gros, si vous êtes une femme noire et que vous venez d'un quartier pauvre, vous allez forcément être une mauvaise mère, vous allez forcément être une mauvaise employée.

Sur ce sujet lire « Automatic inequalities » de Virginia Ebanks : elle est allée faire des études prolongées auprès d'utilisateurs d'IA pour accéder au logement ou pour savoir si leur enfant devait leur être enlevé.

Il faut bannir les outils prédictifs d'autant plus qu'il y a des biais de confirmation phénoménaux. Syndicalement, on doit se battre contre ça. On nous dit mais c'est juste pour vous aider, et comme l'IA fait le boulot on vous donne moins de temps pour lire les dossiers donc en gros, de toute façon, c'est l'IA qui va avoir raison.

Voilà donc c'est pour vous dire je suis désolée de vous déprimer, mais bienvenue dans mon monde, ce n'est pas joyeux ce qui arrive, pas du tout, va vraiment falloir se battre pour ceux qui sont déjà le plus en difficulté. Je veux dire que ceux qui sont dans des situations déjà catastrophiques et économiquement en termes d'éducation (parce que bien sûr, on a aussi une attaque contre l'école et l'université qui est catastrophique), sont ceux qui sont le plus touchés.

Le Service Public va être très largement impacté puisqu'on sait très bien qu'on fait des économies sur notre dos et qu'une manière de faire des économies, c'est de nous remplacer par des IA et il faut vraiment se battre et refuser la mise en place d'outils prédictifs.

L'espoir, c'est les jeunes. Moi je fais beaucoup de formation, beaucoup de sensibilisation auprès des jeunes et je peux vous dire qu'ils sont assez actifs et ils refusent maintenant de participer à ce genre d'outil. Et puis on a aussi l'espoir.

Enfin moi je vais me battre pour essayer de mettre une ligne rouge contre les outils prédictifs au niveau national et européen.

N'hésitez pas à soutenir la quadrature du net, c'est une très belle association qui est très active, très efficace (https://www.laquadrature.net).

RETROUVEZ-NOUS SUR :

@snu.pole.emploi.fsu
@SnuPoleEmploi
@snupoleemploi
@snupoleemploifsu
www.snutefifsu.fr
syndicat.snu@pole-emploi.fr

U. F.S.U. ENGAGÉ-ES AU QUOTIDIEN

Retrouvez le secteur femmes du SNU en vidéo : https://youtu.be/_iUJ8fLRoVU