

ChatGPT: Guide de sensibilisation pour le corps professoral de l'université

ChatGPT: Guide de sensibilisation pour le corps professoral de l'université

ALENA BARYSEVICH; LYNNE BOWKER; ET MELANIE MISANCHUK



ChatGPT: Guide de sensibilisation pour le corps professoral de l'université Copyright © 2023 by Alena Barysevich, Lynne Bowker, Melanie Misanchuk is licensed under a Licence Creative Commons Attribution - Pas d'utilisation commerciale - Pas de modification 4.0 International, except where otherwise noted.

Table des matières

Introduction	1
1. Mieux comprendre ChatGPT et les approches axées sur les données	3
2. Mieux comprendre ChatGPT et les biais	8
3. Mieux comprendre ChatGPT et l'intégrité académique	11
Remerciements	15
À propos des auteurs	16

Introduction

Bienvenue à la ressource éducative libre (REL):

ChatGPT: Guide de sensibilisation pour les professeurs.



À qui s'adresse cette ressource?

Cette ressource est destinée au corps professoral du niveau postsecondaire en Ontario (et au Canada) travaillant dans de différentes disciplines.

Cette REL comprend de brèves explications du fonctionnement et de l'utilisation des grands modèles de langage (GML), tels que ChatGPT. Elle propose aussi des questions de réflexion et des ressources pour approfondir des connaissances sur ChatGPT.

Quels sont les objectifs?

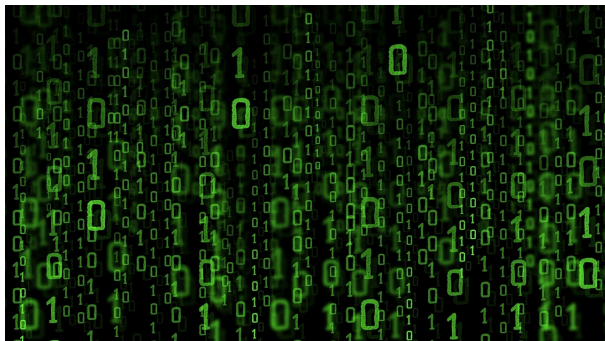
Ce module vise à engendrer une compréhension générale

- de la formation et emploi actuel des GLM
- des biais et préjugés inhérents à un GLM
- de la relation entre l'emploi des GLM et l'intégrité académique

Quelles sont les limites de cette REL?

Ceci est prévu comme une REL d'initiation et générale sur les grands modèles de langage (GML), en utilisant l'exemple de ChatGPT. Cette REL d'initiation offre un point de départ pour le corps professoral faisant face à l'emploi des GLM dans leurs classes. A l'avenir, d'autres peuvent élargir ou approfondir la REL, ou même créer des ressources complémentaires pour leur discipline particulière.

I. Mieux comprendre ChatGPT et les approches axées sur les données



Vous avez sans doute entendu parler des **grands modèles de langage (GML)**, tels que **ChatGPT**. Peut-être vous vous demandez ce que vous devez savoir sur ces outils afin d'aider vos étudiants à les utiliser de manière critique et responsable? Ce guide vous sensibilisera aux GML et aux outils similaires (par exemple, les **traducteurs automatiques** tels que Google Traduction ou DeepL) en abordant quelques points clés. Êtes-vous prêt à en apprendre davantage sur ces « incontournables » afin de devenir un utilisateur plus averti?

Il n'est pas nécessaire d'être informaticien ou de comprendre tous les détails du fonctionnement d'un réseau neuronal pour être un utilisateur averti de ChatGPT ou des outils d'IA semblables. Cependant, il est très utile de savoir que ces outils utilisent une **approche axée sur les données**. Un outil axé sur les données est également très sensible aux données. En savoir plus sur le processus d'entraînement, et surtout sur la nature des **données d'entraînement**, peut vous aider à adopter une approche critique de ChatGPT ou de Google Traduction, etc.

Comment l'ordinateur apprend-il?

La plupart des outils d'intelligence artificielle (IA) emploient une technique qui s'appelle l'**apprentissage automatique**. En bref, si les scientifiques cherchent à apprendre à un ordinateur comment effectuer une certaine tâche, les scientifiques lui fournissent de nombreux exemples. Les exemples (c'est-à-dire, les données) constituent un **corpus d'entraînement**. Une fois que l'ordinateur a vu de nombreux exemples, cette machine peut essayer d'accomplir la tâche elle-même.

À noter : Bien qu'une personne puisse apprendre à faire quelque chose sur la base de quelques exemples, un ordinateur a besoin de voir un très grand nombre d'exemples (parfois **des milliers ou des millions d'exemples**) pour être en mesure d'effectuer la tâche.

Un exemple : classer les images



Pour mieux comprendre le concept d'entraînement et l'effet des données d'entraînement, prenons une tâche relativement simple : la classification d'images. Une équipe de scientifiques (Ribeiro et al. 2016) cherchent à entraîner un outil d'IA à regarder des images et à classer chaque image comme étant un loup ou un husky. Dans un premier temps, les scientifiques montrent à l'outil d'IA de nombreuses images de loups et, à chaque reprise, les scientifiques confirment à l'outil qu'il s'agit bien d'une image de loup. Ensuite, les scientifiques montrent à l'outil d'IA de nombreuses images de huskies et confirment à chaque reprise qu'il s'agit d'un husky. Ces images clairement identifiées de loups et de huskies constituent les **données**

d'entraînement.

Ensuite, les scientifiques testent l'outil d'IA en lui montrant une toute nouvelle série d'images de loups et de huskies. Pour chaque nouvelle image, les scientifiques demandent à l'outil d'IA : *Loup* ou *husky* ? L'outil tente alors de classer chaque nouvelle image dans l'une ou l'autre catégorie, sur la base de ce qu'il a appris au cours de l'entraînement. Mais qu'a-t-il appris ?

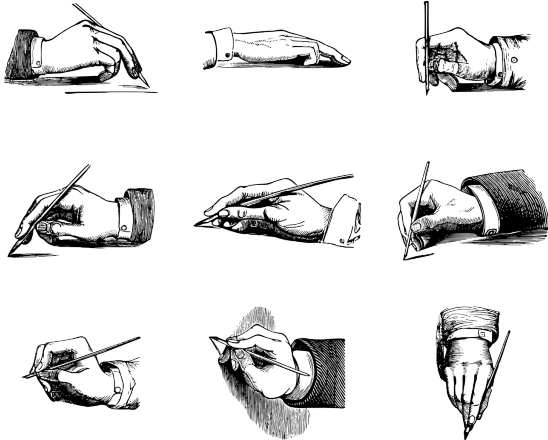
Les ordinateurs ne peuvent jamais rien comprendre, mais ils ont des forces. Deux des plus grandes forces d'un ordinateur sont la reconnaissance de motifs (ou de patrons) et le comptage. Lorsqu'un ordinateur regarde des images de loups et de huskies, il essaie d'identifier des motifs. Lorsqu'il repère un motif potentiel, il compte le nombre de fois où ce motif apparaît pour déterminer la fiabilité du motif. Cela semble assez simple, n'est-ce pas ?

Malheureusement, parce qu'ils ne sont pas intelligents, les ordinateurs peuvent finir par aborder une tâche d'une manière inattendue. Par exemple, dans le cas de la classification d'images, l'ordinateur a mal classé certains loups comme des huskies et vice versa. Lorsque les scientifiques ont analysé les données pour comprendre pourquoi l'ordinateur s'était trompé, ils ont découvert que dans le corpus d'entraînement, la plupart des images de loups se trouvaient sur un fond enneigé, tandis que la plupart des images de huskies se trouvaient sur un fond sans neige. Le motif que l'outil d'IA a vu était donc « neige » ou « pas de neige » et c'est le critère qu'il a utilisé pour classer les images. Et voilà, l'outil a trouvé un motif, mais ce n'était un motif ni souhaitable ni logique. Au lieu d'être un détecteur de loups, l'outil d'IA s'est transformé en ... **détecteur de neige** !



Un détecteur de neige – c'est drôle ! Mais lorsqu'une autre équipe de recherche (Narla et al. 2018) a construit un système d'IA pour classer des images de lésions cutanées comme étant cancéreuses ou non cancéreuses, un problème similaire s'est posé. L'outil a fini par classer les images en fonction de la présence ou non d'une règle sur l'image. Se tromper dans la classification d'un cancer, c'est beaucoup moins drôle.

Un exemple : dessiner les mains



Si le classement d'images est relativement simple, que faire d'un ordinateur avec une tâche plus complexe? Avez-vous déjà vu des images dessinées par des outils d'IA tels que Midjourney ou DALL-E 2 ? L'IA est assez bonne pour dessiner des visages, mais moins bonne pour dessiner des mains. Pourquoi ? Pensez encore une fois aux données d'entraînement. Il est probable qu'une collection de photos de personnes se concentre sur les visages et non sur les mains. De plus, dans la plupart des photos, les personnes font face à l'appareil photo, mais leurs mains peuvent être dans différentes positions : saisir quelque chose, pointer du doigt, faire un signe de la main, serrer le poing. Un outil d'IA peut être entraîné à savoir que les mains ont des doigts, mais il est plus difficile de savoir qu'il devrait y en avoir cinq parce que les mains dans corpus d'entraînement figurent

sous différents angles. Donc il ne suffit pas de disposer d'un grand nombre d'exemples. Il faut aussi que les exemples soient bons. Si le corpus d'entraînement contient beaucoup de bons exemples (comme dans le cas des images de visages), l'IA peut accomplir la tâche. Mais si les exemples du corpus d'entraînement ne sont pas d'une qualité suffisante (comme pour les images de mains), il est alors plus difficile pour l'IA d'accomplir la tâche.

Pour apprendre plus

- Une intelligence artificielle bien réelle : Les termes de l'IA (glossaire bilingue (français-anglais) avec les définitions produit par l'Office québécois de la langue française)
- L'intelligence artificielle, une perspective axée sur les données (vidéo de 30 minutes d'une conférence par Ihsen Hedli, de l'Université Laval)
- Les IA ne savent pas dessiner les mains, voici pourquoi (vidéo de 2 minutes et texte de *HuffPost*)

Et le traitement des langues ?

Tout d'abord, il faut se rappeler que les outils d'IA ne comprennent rien. Ils ne peuvent que rechercher des motifs et les imiter. Cela est vrai pour le texte comme pour les images. Dans le cas d'un outil tel que ChatGPT, Google Traduction ou DeepL, les données d'entraînement sont des textes... des milliards de textes ! Pour obtenir autant de textes, les scientifiques qui entraînent les outils les rassemblent de partout : des sites web, des médias sociaux, des archives numériques. Certains de ces textes sont bien rédigés, mais pas forcément tous. Certains textes peuvent contenir un langage inclusif, mais d'autres non. Certains textes peuvent même être offensants. C'est un véritable pot pourri ! Grâce à ce corpus d'entraînement, l'outil d'IA apprend à quoi ressemblent les textes et quels sont les mots qui se trouvent ensemble habituellement. Puis l'outil d'IA les imite ensuite. Mais il ne les comprend pas. C'est pourquoi un texte généré par ChatGPT peut sembler fluide mais être complètement faux. C'est également la raison pour laquelle un outil d'IA peut générer des résultats biaisés. Tout comme un enfant peut imiter un parent en répétant un gros mot, un outil d'IA peut reproduire n'importe quel préjugé qu'il trouve dans le corpus d'entraînement (pour de plus amples détails, voir la section sur les préjugés).



Les points qui concernent la taille et la qualité du corpus d'entraînement sont également très pertinents pour le traitement de langue. Il est plus facile de recueillir un grand corpus d'entraînement de qualité dans une langue répandue comme l'anglais ou le français, mais il est certainement plus difficile de trouver suffisamment de données dans des langues moins répandues, comme les langues autochtones du Canada. Cela signifie que la performance des outils sera meilleure pour l'anglais que pour le cri ou l'algonquin. Cela signifie également que de nombreux outils génèrent des résultats dans des variétés principales d'une langue, telles que l'anglais américain ou le français européen, mais qu'ils sont moins susceptibles d'offrir un texte en anglais canadien ou en français canadien. Enfin, le sujet et le type de texte peuvent influencer les résultats. Il est plus probable qu'un corpus d'entraînement comprend des textes qui portent sur des sujets communs et moins probable qu'il contienne des textes qui portent sur des sujets très spécialisés.

Par conséquent, un outil d'IA (p. ex. ChatGPT, Google Traduction ou DeepL) sera probablement plus efficace pour traduire des textes administratifs entre l'anglais américain et le français européen, car le corpus de formation contient de nombreux exemples. En revanche, le même outil (ChatGPT, Google Traduction ou DeepL) sera probablement moins performant pour traduire un manuel d'utilisation pour un réacteur nucléaire entre le gallois et le vietnamien, car il y aura moins d'exemples de ce type de texte dans ces langues à inclure dans le corpus d'entraînement.

Pour apprendre plus

- ChatGPT fait son entrée dans les médias québécois (article du journal *Le Devoir*)
- L'intelligence artificielle et la traduction évoluent ensemble (article dans *Circuit*, le magazine)

Pistes de réflexion sur l'emploi des outils d'IA dans la salle de classe

1. D'après vous, quels sont les points forts et les points faibles d'un outil d'IA qui est axée sur les données?
2. Quels sont les risques potentiels liés à l'utilisation d'un outil d'IA axée sur les données pour
 - Traduire un texte très spécialisé du grecque vers le farsi?
 - Rédiger un texte destiné à un public au sein d'une compagnie (p. ex. un courriel à votre collègue)?
 - Résumer un rapport destiné à la diffusion au grand public?
3. Si vous voulez entraîner un outil d'IA pour traduire pour un public canadien, quelles sources de données seront utiles pour constituer le corpus d'entraînement?

2. Mieux comprendre ChatGPT et les biais



Les biais dans les GML comme ChatGPT viennent des préjugés dans la société et donc dans les matériaux que la société produit, sur lesquels les GML sont formés. Les GML sont également (du moins pour l'instant) développés par des humains, soumis à des préjugés humains, en utilisant des textes sur Internet sélectionnés au hasard pour tenter d'en assurer la représentation. Cependant, au-delà des préjugés humains, les modèles génératifs eux-mêmes succombent à un certain nombre d'autres types de biais. Pour chacun de ces préjugés, les scientifiques et les créateurs de GML déploient des efforts substantiels et de bonne foi pour y remédier. Mais ni un travail acharné ni la bonne volonté ne garantissent le succès, et certains types de préjugés, tels que les préjugés d'omission, sont presque impossibles à corriger. Mais ce qui est encore plus important, il n'existe aucun moyen efficace pour les utilisateurs de GML d'atténuer le biais des résultats qu'ils reçoivent, si ce n'est d'être conscients de leur existence.

Le racisme, le sexisme, l'intolérance religieuse et l'homo/transphobie se peuvent être intégrés dans les résultats produits par ChatGPT et d'autres outils de l'IA. On a beau dire: « Eh bien, je suis au courant de ce discours toxique ; je vais simplement ignorer cette partie. » Cependant, le grand nombre de recherches sur

les préjugés inconscients nous rappelle que les gens n'arrivent pas toujours à détecter et traiter les préjugés implicites. Et après des mois et des années d'utilisation de matériel produit par GML qui contient ces préjugés, il se peut que certains préjugés se renforcent dans l'esprit des gens.

Les biais du GML ont-ils un effet sur notre vie quotidienne ? Après tout, nous pouvons choisir d'ignorer certaines des choses que nous disent les GML, si nous sommes suffisamment astucieux pour remarquer le biais. Cependant, lorsqu'il s'agit de services fournis à l'aide d'une IA biaisée, des préjudices réels peuvent survenir. C'est peut-être moins important tandis que les utilisateurs décident, de leur propre gré, de suivre les recommandations d'achat d'actions ou de prendre des conseils juridiques auprès de ChatGPT, mais c'en est tout autre chose lorsque des produits alimentés par les GML fournissent des services cruciaux, tels que les soins de santé.

Le biais de disponibilité ou de sélection est la situation où les données sur lesquelles le GML est formé ne sont pas représentatives de l'ensemble de la population. De la même manière que les femmes sont sous-représentées dans la plupart des recherches médicales, en particulier dans les essais cliniques de médicaments, ce qui conduit à une généralisation erronée des soins de santé destinés aux hommes aux traitements destinés aux femmes, l'absence de données sur les personnes issues de populations marginalisées peut conduire à des mauvais diagnostics, triages, ou des plans de traitement, si ceux-ci sont laissés à un GML.

Si un GML est utilisé pour trier les candidatures à un emploi, les biais de l'outil peuvent éliminer des candidats précieux qui ne correspondent pas à la probabilité statistique de la formation du GML. Pour revenir sur une idée centrale:



« ... les modèles linguistiques ne sont pas conçus pour représenter la réalité ou comprendre des concepts. Les modèles sont plutôt entraînés à prédire la séquence de mots la plus probable en fonction des données sur lesquelles ils ont été formés. » (Ostermeier, 2023) [notre propre traduction]

Même s'il n'y a pas de biais manifeste observable, les GML comme ChatGPT ont par définition un effet nivelant ou générique : en prenant le dénominateur commun (la chaîne de mots la plus probable), ils suppriment la « voix » de l'auteur. Comme ces outils sont formés sur de grandes quantités de données qui privilégient l'anglais comme langue dominante et la pensée occidentale, ils marginalisent davantage les voix et les expériences des minorités. Ainsi, même si les résultats ne semblent pas ouvertement racistes ou sexistes, ils peuvent être biaisés par la simple absence de représentation des expériences minoritaires.

Comme utilisateurs de ces outils, il y a peu à faire pour améliorer le biais : *caveat emptor*.

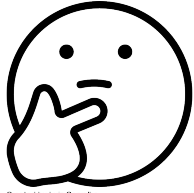
Pistes de réflexion

Que feriez-vous pour tenir compte du biais dans les résultats donnés par ChatGPT et d'autres outils?

Pour apprendre plus

- [ChatGPT is multilingual but monocultural, and it's learning your values](#)
- [De-biasing LLMs: From Theory to Practice](#)
- [The Dark Side of Language Models: Exploring the Challenges of Bias in LLMs](#)
- [More human than human: measuring ChatGPT political bias](#)
- [The Efforts to Make Text-Based AI Less Racist and Terrible](#)

3. Mieux comprendre ChatGPT et l'intégrité académique



Thinking by Luiz
Carvalho from Noun
Project (CC BY 3.0)

Created by Luiz Carvalho
from Noun Project

Sondage en ligne sur la perception de ChatGPT

Notre vision

Au fur et à mesure que ChatGPT évolue, la façon dont le corps professoral les utilise dans les cours et la recherche doit également évoluer. Il devient clair que l'intelligence artificielle (par exemple, Chat GPT) ne va nulle part et il est important d'apprendre à travailler **avec** cette technologie artificielle. En premier lieu, il nous semble primordial que le corps professoral et les assistants d'enseignement se familiarisent avec ChatGPT avant d'aborder cette question dans leur salle de classe. Deuxièmement, nous conseillons d'aborder la question de façon ouverte, claire et explicite dans la salle de classe. Posez-vous les questions:

- Est-ce que ChatGPT a la place dans mon cours?
- Le Chat GPT, peut-il être utilisé pour réaliser des travaux, des tests ou des examens de mon cours?
- Dans quelle mesure Chat GPT est autorisé dans mon cours?
- Comment Chat GPT devrait être cité?

Liste de recommandations

- Intégrer dans son cours la formation sur l'usage éthique de ChatGPT, de ses implications et des problèmes découlants.
- Rester explicite dans le plan du cours et lors des sessions d'enseignement si les étudiants peuvent utiliser ChatGPT dans le cadre du cours, en général, et dans chaque travail universitaire, en particulier.

- S'assurer que les directives sont explicites, claires et diffusées par de différentes chaînes d'informations (par exemple, sur la plateforme d'apprentissage du cours, dans les plans de cours, dans les consignes des travaux et lors des sessions d'apprentissage).
- S'assurer que les étudiants comprennent que le mauvais emploi de ChatGPT représente un manquement à l'intégrité académique.
- Assurer les étudiants que l'autorisation de l'usage de ChatGPT peut varier d'un cours à l'autre
- Encourager les étudiants à poser des questions sur l'usage éthique de ChatGPT.
- Encourager les ateliers de formation (par exemple, sur l'intégrité académique).

Quelques pratiques pédagogiques pour gérer l'usage responsable de Chat GPT

- Proposer plus de travaux oraux ou des travaux multimodaux.
- Proposer des projets d'évaluation basés sur le contexte et le contenu de la discussion en classe.
- Proposer des projets d'évaluation basés sur l'expérience personnelle et liées aux actualités.
- Se concentrer sur l'argumentation, la pensée critique et la créativité dans l'évaluation des projets.
- Attribuer des mini- questionnaires formatifs sur l'usage de ChatGPT.

Comment citer l'usage de ChatGPT

- ChatGPT Citations | Formats & Examples ([scribbr.com](https://www.scribbr.com/ai-chatgpt-citations/))
- How do I cite generative AI in MLA style? | [MLA Style Center](https://www.mla.org/ai)
- FAQ Item ([chicagomanualofstyle.org](https://www.chicagomanualofstyle.org))

Développement professionnel

- Se renseigner auprès de différents bureaux responsables de l'éducation (par exemple, Teaching Commons, Teaching and Learning. Center, Library Resources, sur les politiques de l'Université).
- Contacter les responsables du développement de l'éducation au niveau universitaire.
- Prendre contact avec la bibliothèque.
- Se renseigner auprès du bureau sur l'intégrité académique.
- Participer aux ateliers sur l'intégrité académique et ChatGPT.

Pour apprendre davantage:

- [Intégrité académique pour les étudiants](#)
- [FAQ sur l'IA pour les professeurs](#)
- [Frequently Asked Questions: ChatGPT and generative AI in teaching and learning at the University of Waterloo](#)
- [University of Guelph Statement on Artificial Intelligence Systems, ChatGPT, Academic Integrity](#)

Remerciements

Cette ressource éducative libre a été financée à l'aide d'une subvention d'eCampusOntario.

Attributions pour les images :

- AI Generated ChatGPT par Franz26 sur Pixabay, Licence Pixabay.
- Matrix par jaydeep_ sur Pixabay, Licence Pixabay.
- Husky par Antelao sur Pixabay, Licence Pixabay.
- Wolf par Pixel-mixer sur Pixabay, Licence Pixabay.
- Draw par GDJ sur Pixabay, Licence Pixabay.
- Translation par AbsolutVision sur Pixabay, Licence Pixabay.
- Man in White Crew Neck Shirt par cottonbro studio sur Pexels, Licence Pexels.
- Hand X-ray Result sur Pixabay, Licence Pixabay.
- Thinking par Luiz Carvalho du Noun Project (CC BY 3.0)
- Merci par Nono_Bobo sur Pixabay, Licence Pixabay.





Lynne Bowker, PhD, CTran

Professeure titulaire
École de traduction et d'interprétation
Université d'Ottawa, Canada
Courriel : lbowker@uottawa.ca
Twitter : @bowkerl
Site Web : <https://uniweb.uottawa.ca/members/415>



Melanie Misanchuk, Ph.D

Conceptrice pédagogique
Centre for Extended Learning
University of Waterloo
Courriel : mmisanch@uwaterloo.ca
