IA et démocratie

Comprendre les effets de l'IA sur les élections

ceimia

IVADO

Crédits

Rédaction

Ce document de vulgarisation a été produit à partir de travaux de Claire Boine, avec la collaboration de l'équipe de Mobilisation des connaissances d'IVADO et de celle du CEIMIA.

Nous remercions également les professeurs Pierre-Luc Déziel, de l'Université Laval, et Pierre Trudel, de l'Université de Montréal, pour leurs commentaires sur certains aspects du rapport. Cela dit, les éventuelles erreurs ou imprécisions que ce dernier pourrait contenir restent entièrement de la responsabilité d'IVADO et du CEIMIA.

Infographie

Stéphanie Hauschild, avec le soutien de l'équipe des communications d'IVADO.

Pour citer ce rapport

Boine, Claire et autres (2024). IA et démocratie. Comprendre les effets de l'IA sur les élections. CEIMIA et IVADO.

Références

À cause de la nature de ce document, nous avons cherché à garder au minimum le nombre de références qui y sont présentées. Les personnes qui le souhaitent peuvent cependant demander à accéder à l'ensemble de celles-ci en écrivant à mobilisation@ivado.ca.

Table des matières

Présentation d'IVADO et du CEIMIA 3

Introduction 4

Rien de nouveau sous le soleil, vraiment ? 5

Premier enjeu : **le microciblage** 6

Deuxième enjeu : **les hypertrucages** 12

Troisième enjeu : **les fausses informations** 17

Un espoir : l'IA comme outil de renforcement de nos démocraties 23

Conclusion 25

Annexe 1 — Glossaire 26

Annexe 2 — Notes bibliographiques 28

Présentation d'IVADO et du CEIMIA

IVADO est un consortium interdisciplinaire et intersectoriel de recherche, de formation, et de mobilisation des connaissances qui a pour mission de soutenir le développement et le déploiement responsables d'une intelligence artificielle (IA) plus robuste et raisonnante. Piloté par l'Université de Montréal, avec quatre partenaires universitaires (Polytechnique Montréal, HEC Montréal, l'Université Laval et l'Université McGill), IVADO rassemble des centres de recherche et des partenaires gouvernementaux et industriels pour coconstruire des initiatives intersectorielles ambitieuses favorisant un changement de paradigme en matière d'IA.

Le Centre d'expertise international de Montréal en intelligence artificielle (CEIMIA) est une organisation au cœur de partenariats internationaux dédiés à la recherche et à la création de solutions en IA qui répondent à de grands défis sociétaux. Le CEIMIA se positionne comme un acteur clé du développement responsable de l'IA, lequel est fondé sur des principes d'éthique, de droits de la personne, d'inclusion, de diversité, d'innovation et de croissance économique. Le CEIMIA conçoit et met en œuvre des projets appliqués à fort impact en IA. Le CEIMIA est l'un des centres de soutien aux experts du Partenariat mondial sur l'IA (PMIA). Intégré avec l'OCDE, il travaille avec 44 gouvernements sur le développement et l'adoption responsables de l'IA.

Introduction

Ce document amorce une réflexion préliminaire sur une question névralgique, celle des effets que l'IA aura sur les scrutins qui se tiendront dans un avenir proche au Canada, au Québec et ailleurs, donc sur nos démocraties.

Qu'est-ce que la démocratie ?

Pour Élections Québec, la démocratie, c'est « un système politique qui permet aux électrices et aux électeurs de voter pour élire les personnes candidates qui les représenteront et qui prendront des décisions en leur nom au sein du gouvernement ».

De son côté, l'Organisation des Nations Unies (ONU) définit la démocratie en fonction des éléments qui la constituent, soit « le respect des droits de l'homme et des libertés fondamentales; la liberté d'association; la liberté d'expression et d'opinion; l'accès au pouvoir et à son exercice conformément à l'état de droit; la tenue d'élections libres, régulières et périodiques au suffrage universel et à bulletin secret, [qui sont le] reflet de l'expression de la volonté du peuple; un système pluraliste de partis et d'organisations politiques; la séparation des pouvoirs; l'indépendance de la justice; la transparence et la responsabilité dans l'administration publique; et des médias libres, indépendants et pluralistes. »

Élaboré à partir d'un travail de veille et d'une revue des écrits scientifiques sur cette question, le présent rapport constitue la première étape d'un effort à long terme. Il s'inscrit, en effet, dans un projet de mobilisation des connaissances amorcé par le CEIMIA et IVADO au printemps 2024, qui poursuit les objectifs suivants :

- Améliorer la compréhension de la population et des responsables sur les technologies (et surtout l'IA) qui influent sur les élections et la démocratie ;
- Définir les impacts de l'IA sur l'intégrité du processus électoral ;
- Protéger la vitalité des démocraties québécoise et canadienne contre les effets néfastes de certains usages de l'IA ;
- Explorer comment l'IA peut servir à bonifier et à soutenir la délibération démocratique.

Dans ce document, nous nous penchons particulièrement sur trois usages de l'IA ou de systèmes apparentés à l'IA qui ont le potentiel de miner nos démocraties : le microciblage politique, les hypertrucages et les fausses informations.

Mais, comme il y a toujours deux côtés à une médaille, nous discuterons aussi des effets positifs que le recours à l'IA pourrait engendrer pour nos démocraties.

Rien de nouveau sous le soleil, vraiment ?

Ce n'est pas d'hier que les technologies influent positivement ou négativement sur le résultat des élections et la qualité des débats qui les entourent.

Par exemple, certaines stations de radio ou de télévision peuvent, par les propos qui y sont tenus, créer un sentiment d'aliénation dans certaines franges de la population, réduire la confiance que les citoyennes et les citoyens ont en la démocratie, et influer négativement sur leur propension à se reconnaître comme responsables des décisions politiques. En revanche, les émissions d'affaires publiques diffusées sur ces mêmes plateformes peuvent contribuer à informer l'électorat sur les enjeux de l'heure.

Autre exemple : les chambres d'écho que l'on retrouve dans les réseaux sociaux — ces environnements où les internautes sont presque exclusivement exposés à des opinions, des informations et des contenus qui reflètent ou renforcent leurs propres croyances et points de vue — peuvent empêcher la tenue de véritables débats. Par contre, Internet permet aux citoyennes et aux citoyens de se repérer et de se regrouper plus facilement que par le passé pour défendre certaines idées et positions.

On pourrait donc penser qu'avec l'arrivée de l'IA, il n'y a rien de vraiment nouveau sous le soleil, et que les acteurs politiques de toutes sortes disposent tout simplement d'un outil de plus dans leur coffre pour influer sur les attitudes et les comportements de l'électorat.

Or, la montée de l'IA crée une situation de nature différente, puisque le recours à cette technologie pourrait venir amplifier des phénomènes existants, changer le cours des élections plus fortement qu'auparavant.

Pour faire un parallèle, même si les ouragans ont toujours existé, les changements climatiques provoqués par nos comportements en ont augmenté la fréquence et la force. De même, les technologies pouvaient influer, en bien ou en mal, sur le résultat des élections, mais la montée de l'IA risque d'en redoubler les impacts.

Premier enjeu : **le microciblage**

Dans le contexte politique, le microciblage utilise des données personnelles sur l'électorat (des données contenues dans des bases variées) pour envoyer le bon message à la bonne personne, et faire en sorte que cette personne réponde à ce message de la manière souhaitée.

Qu'est-ce qu'une donnée personnelle ?

Selon la Loi sur la protection des renseignements personnels et les documents électroniques du Canada, un renseignement personnel est « tout renseignement concernant un individu identifiable ». Dans le contexte de l'IA, on parle plus souvent de données, un concept large qui inclut aussi les renseignements qui ne sont pas directement identifiables sans le croisement d'autres données.

Selon la Commission d'accès à l'information du Québec, la collecte de données est le moment pendant lequel un renseignement personnel est « recueilli (formulaire d'abonnement, sondage, outils analytiques web, etc.) ; créé (p. ex. un numéro de membre) ; [ou] inféré (p. ex. un profil de consommateur), c'est-à-dire déduit à partir d'autres renseignements ».

Encore une fois, le microciblage n'est pas une nouveauté. Par exemple, les personnes câblées qui regardent une émission de télévision ne voient pas nécessairement les mêmes publicités selon qu'elles demeurent à un endroit plutôt qu'à un autre. De même, les internautes qui fréquentent les réseaux sociaux sont ciblés à l'aide de publicités sur mesure. Par exemple, si l'on souhaite cibler les individus à tendance écologiste, on pourra choisir de diffuser une publicité auprès de ceux qui ont joint un groupe particulier d'écologistes en ligne.

De même, la tactique du microciblage est déjà utilisée par les partis politiques, mais elle l'est encore relativement peu. Cela dit, l'IA permet maintenant de l'appliquer à une nouvelle échelle.

La source des données qui servent au microciblage

Les données utilisées pour le microciblage proviennent de multiples sources. Certaines organisations vendent ainsi les profils associés à des identifiants d'appareils mobiles, tandis que d'autres vont jusqu'à corréler un identifiant précis avec un nom, un prénom et une adresse. Ces données agrégées peuvent ensuite être revendues à des annonceurs pour leur permettre de diffuser de la publicité ciblée.

Les publicitaires n'ont pas besoin de disposer de renseignements sur chaque individu pour accroître l'efficacité des messages qu'ils

1. Éric Zemmour, en France, a par exemple voulu mousser sa candidature auprès des internautes français de confession juive en faisant appel à un courtier en données qui avait constitué un fichier de personnes « intéressées par l'antisémitisme en France et en Europe ». Une enquête a été ouverte, comme cette pratique contournait l'esprit de la loi française.

envoient. Ils doivent simplement disposer d'informations sur un nombre suffisant de personnes, puis réaliser des extrapolations. Ceci explique, qu'en connaissant le quartier où une personne demeure, il est possible de déduire certaines données sensibles relatives à cette personne, comme la tranche de revenu dans laquelle elle se situe.

Un exemple de microciblage

Pour comprendre comment l'analyse de données et l'IA sont utilisées dans le cadre des élections et peuvent mener à la génération de messages personnalisés, examinons un cas fictif.

Imaginons que, dans une grande ville quelconque, une candidate décide d'engager une entreprise d'analyse de données et de communication politique pour soutenir sa campagne.

L'entreprise crée un questionnaire pour sonder un échantillon représentatif des personnes résidant dans la ville en question. Ce sondage permettra de mieux connaître leur profil démographique, leurs habitudes de vie, leurs préoccupations politiques, leurs loisirs, leur utilisation de différents médias, leurs activités en ligne, etc.

Après avoir organisé les données recueillies, les analystes recourront à l'IA pour déceler des tendances dans celles-ci. Ils pourront relever que les personnes qui présentent certaines caractéristiques sont très préoccupées par les questions environnementales, tandis que celles qui en présentent d'autres sont moins enclines à voter.

En utilisant l'IA, les analystes segmenteront ensuite la population de la ville en sous-groupes.

Dans une nouvelle étape, l'entreprise qui soutient la candidate concevra et testera des publicités différentes pour chaque sous-groupe. Par exemple, l'électorat écologiste pourra être exposé à des informations sur la proximité de la candidate avec les communautés autochtones, sur ses projets de réhabilitation de parcs ou de forêts, ou ses plans en matière de transport collectif.

Pour déployer les messages que l'équipe de la candidate a développés, elle devra déterminer à quels segments appartient chaque membre de l'électorat.

Pour ce faire, elle utilisera son logiciel de gestion de la relation avec l'électorat. Cet outil contient la liste des gens fournie par Élections Québec ou Élections Canada. Les partis y ajoutent ensuite les renseignements qu'ils détiennent sur eux, dont ceux qu'ils ont obtenus de fournisseurs ou d'autres organisations.

À l'intérieur du logiciel, l'équipe de la candidate pourra mettre en relation chaque électrice ou électeur avec l'identifiant unique de son téléphone. S'il lui manque de données pour placer une personne dans les bons sous-groupes, elle pourra faire appel à des courtiers pour acheter les informations manquantes : ses activités dans les réseaux sociaux (p. ex. : partages, boutons « J'aime », commentaires) ou en ligne (p. ex. : signature de pétitions), ses abonnements, sa participation à des événements communautaires, etc.

L'équipe pourra en fin de compte décider d'utiliser un réseau comme Facebook, X ou Bluesky pour diffuser les publicités de la candidate en fonction de l'identifiant unique de chaque personne et des sous-groupes auxquels elle appartient.

Au fur et à mesure que la campagne progressera, l'équipe pourra collecter de nouvelles données sur la réponse de chaque électrice ou électeur à ses messages et affiner sa stratégie.

Les effets du microciblage sur le vote

Des recherches montrent que le microciblage a des conséquences sur l'électorat, à la fois sur ses choix électoraux, mais aussi sur sa décision d'aller voter ou non.

Dans le cadre des élections américaines de 2016, l'exposition à des messages ciblés de la part de l'équipe de Donald Trump a incité des électrices et électeurs républicains clés à aller voter et a découragé leur contrepartie démocrate à faire de même[2]. Le microciblage a également eu pour effet de réduire la probabilité que les personnes d'allégeance républicaine changent d'avis sur leur candidat.

Des spécialistes en communications ont par ailleurs montré que les publicités politiques qui correspondent davantage à la personnalité de quelqu'un (ce que le microciblage permet de réaliser) sont plus persuasives que les autres.

D'autres études ont enfin conclu que le microciblage n'amène pas vraiment les gens à changer d'avis, mais qu'il est utile pour renforcer leurs intentions préexistantes.

Les effets du microciblage sur la démocratie

Le microciblage n'est pas nécessairement une activité problématique, puisqu'il peut aider les partis à « parler » aux citoyennes et citoyens en fonction de leurs préférences et, par conséquent, servir à lutter contre le désintérêt qu'une fraction croissante de l'électorat éprouve envers la politique. Utilisé de manière responsable, le microciblage peut représenter une opportunité.

Trop utilisé ou mal utilisé, le microcriblage peut toutefois avoir des effets négatifs sur la démocratie.

D'abord, le microciblage peut éroder le dialogue public en limitant l'exposition des électrices et des électeurs potentiels à des renseignements ou des perspectives qui pourraient les inciter à changer d'idées sur certaines questions.

Comme l'a noté le professeur Cass R. Sunstein de l'Université Harvard, une démocratie saine repose sur un ensemble d'expériences communes (comme celles que procurent les médias), lesquelles peuvent les aider à se comprendre mutuellement. En se limitant à diffuser auprès de groupes réceptifs des messages qui susciteraient la tenue de débats dans un forum réellement public, le microciblage risque en fait d'augmenter la polarisation.

Des universitaires ont, de plus, observé que l'utilisation de microciblage politique peut réduire la confiance de la population envers la démocratie et envers les médias.

Finalement, le microciblage peut mener à l'exclusion progressive d'une partie de l'électorat du débat public. En effet, cette manœuvre permet aux acteurs électoraux de concentrer leurs efforts sur les personnes susceptibles de voter et d'ignorer les autres. Mais comme les personnes qui ont été ignorées lors d'un scrutin sont encore moins susceptibles de voter lors d'un suffrage subséquent, cette approche peut avoir une incidence néfaste, à la longue, sur la participation politique.

2. Cambridge Analytica a ciblé précisément l'électorat de droite de quatre états décisifs qui n'avaient pas prévu de voter.

L'encadrement juridique du microciblage

Le recours au microciblage à des fins électorales est encadré dans de nombreuses juridictions. Par exemple, pour qu'un parti politique québécois ou canadien puisse utiliser les renseignements ou les données personnelles d'une personne à des fins de microciblage (qui est une forme de communication politique), il faut que cette personne lui ait donné son consentement.

Cela dit, les citoyennes et citoyens n'adoptent pas toujours des comportements qui limitent la possibilité pour les acteurs électoraux et leur entourage de recourir à des approches de microciblage néfastes. Par exemple, beaucoup ne limitent pas autant que cela serait souhaitable le partage de leurs données personnelles sur les réseaux sociaux. D'autres donnent leur consentement à l'utilisation de ces données sans prendre garde à la façon dont elles pourraient être utilisées par la suite.

Un autre enjeu est que les personnes qui consentent à l'exploitation de leurs renseignements personnels à des fins de microciblage ou, encore, qui acceptent de participer à des activités en ligne qui servent à la collecte de tels renseignements donnent possiblement, en fin de compte, des informations sur elles-mêmes et sur d'autres personnes, même quand ces dernières n'ont pas, elles, consenti à cette collecte. De la même manière que les sondeurs utilisent les données relatives à un échantillon de personnes pour établir certaines caractéristiques d'une population, les données relatives à une électrice ou un électeur, comme son âge, son niveau de scolarité ou ses préférences électorales, peuvent, lorsqu'elles sont combinées avec d'autres renseignements, refléter des tendances communes dans son quartier ou sa région.

Comment prévenir les effets négatifs du microciblage — quelques solutions à l'échelle locale et internationale

	Collecte des données	Revente et agrégation	Exposition à la publicité
En tant que décideur public : soutenir l'adoption de mesures pour augmenter le contrôle que les électeurs et électrices ont des données les concernant ou favoriser la responsabilisation des acteurs	→ Interdire (quand cela est actuellement permis) la collecte de données à des fins de publicité politique jugées néfastes	→ Obliger les organisations à obtenir le consentement des électrices et électeurs avant de vendre l'accès à leurs données (quand, évidemment, cette vente est actuellement permise) → Donner à chacun le droit à l'effacement de ses données → Renforcer les sanctions pour l'utilisation abusive de données personnelles	→ Assurer la transparence des pratiques en obligeant les organisations à mentionner pourquoi une personne reçoit un message particulier et préciser l'origine de la publicité → Imposer des audits réguliers des pratiques de ciblage des partis politiques
Solutions politiques collectives		→ Créer un code de conduite commun sur l'utilisation des données et de l'IA par les partis politiques	→ Mettre en place des mécanismes de surveillance citoyenne des pratiques de ciblage → Organiser des campagnes d'éducation publique sur les enjeux du microciblage politique

En résumé

Le microciblage :

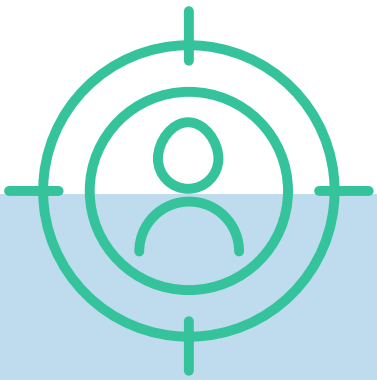
- Est basé sur des données personnelles collectées via différentes sources ;
- Est personnalisé pour des segments précis de la population.

Le microciblage politique peut engendrer différents effets négatifs sur la démocratie. Il peut notamment :

- Influer sur la décision d'aller voter ou non ;
- Renforcer les tendances préexistantes des individus et alimenter la polarisation ;
- Prévenir la formation d'un débat public sain en enfermant les personnes dans des chambres d'écho ;
- Éroder la confiance dans la démocratie et les médias.

Les solutions déjà en place ont des effets limités.

Il faudra plutôt miser sur des mesures diverses (p. ex. : code de conduite pour les partis politiques, mécanisme de surveillance du microciblage politique, campagne d'éducation pour le public).



Deuxième enjeu : **les hypertrucages**

Les hypertrucages (*deepfakes* en anglais) conçus avec l'IA générative permettent de créer automatiquement des contenus comme des images, des sons ou des vidéos.

Ces contenus peuvent ensuite servir à influencer la population (voire la tromper) et, de ce fait, avoir des conséquences importantes sur l'opinion publique et les élections — donc, pour la démocratie.

Il y a quelques années, la création d'hypertrucages exigeait des compétences et des moyens techniques poussés. Aujourd'hui, la multiplication des applications et des services d'IA et d'infonuagique permet à des non-spécialistes de fabriquer du contenu synthétique à prix réduit (ou même gratuitement) et de manière accélérée.

Il existe des applications qui permettent de remplacer un visage par un autre dans une vidéo, de mettre des mots dans la bouche de quelqu'un, d'ajouter dans une photo une personne qui en était absente, de changer l'apparence de quelqu'un, etc.

On s'imagine donc les effets dommageables des hypertrucages pour la démocratie quand ils sont utilisés pour répandre de fausses informations avec un vernis de crédibilité, par exemple pour faire dire ou faire faire quelque chose de controversé à une adversaire politique, pour amplifier une crise, ou pour embellir la réputation d'un favori ou d'une alliée (voir les images ci-contre).

Fausse image du président Macron, où on le voit en train de serrer la main de l'ayatollah iranien Ali Khamenei, ce qui serait une preuve d'une nouvelle capitulation de la France devant les « forces du mal »[3].

Fausse image du président Donald Trump, qui le montre en train d'intervenir sur le terrain après le passage de l'ouragan Hélène[4].

3. Voir https://www.reuters.com/fact-check/image-frances-macron-greeting-irans-khamenei-is-ai-generated-2024-10-17/.

4. Voir https://www.politifact.com/factchecks/2024/oct/02/viral-image/trump-surveyed-hurricane-helene-damage-in-georgia/. Les cercles rouges montrent des indices qui permettent d'identifier que l'image est fausse.

Hypertrucages et élections — quelques exemples récents

Argentine : lors de l'élection présidentielle, de fausses images du candidat Massa prenant de la cocaïne, et du candidat Milei détaillant son plan pour autoriser la vente d'enfants, ont circulé

Canada : Anthony Furey, qui se présente à la mairie de Toronto, utilise dans sa campagne des hypertrucages qui ne sont pas étiquetés comme tels, comme celui montrant un Toronto dystopique rempli de sans-abris

États-Unis : une fausse vidéo montre, sur TikTok, la sénatrice Elizabeth Warren en train de demander, lors d'une émission de télé, que le droit de vote de l'électorat républicain soit restreint

Nigéria : un hypertrucage audio met en scène un candidat d'opposition qui parle de son intention de travailler à truquer les élections

Pakistan : dans un montage vidéo publié sur X, un candidat aux élections législatives appelle la population à boycotter les élections

Royaume-Uni : on entend, dans un faux enregistrement audio, le chef du Parti travailliste en train d'insulter son personnel

Slovaquie : un candidat, dans un faux enregistrement audio, parle avec une journaliste de son intention d'acheter des voix

Ukraine : le président ukrainien invite les militaires ukrainiens à rendre les armes dans une fausse vidéo

Effets des hypertrucages

Les hypertrucages peuvent avoir de véritables effets dans le contexte électoral.

Premièrement, ils peuvent influer (positivement ou négativement) sur la confiance de la population envers un candidat ou une candidate. Des chercheurs ont ainsi démontré que, même si la crédibilité d'un hypertrucage dépend fortement de la plausibilité de son contenu, les hypertrucages invraisemblables peuvent aussi nuire à l'image d'une candidature politique, augmenter la polarisation et diminuer le respect des gens pour le parti ciblé par une attaque.

Deuxièmement, les hypertrucages peuvent avoir un effet nuisible sur la confiance populaire dans la gouvernance démocratique. En mettant en scène de fausses fraudes électorales, ou encore des personnalités politiques en train de mépriser une partie de l'électorat, de parler de limiter le droit de vote ou de truquer les résultats d'un suffrage, les hypertrucages peuvent contribuer à une diminution de la confiance du public dans le processus démocratique même. Cette baisse de confiance pourrait notamment conduire certaines personnes à ne pas aller aux urnes.

Troisièmement, les hypertrucages ont des répercussions sur la qualité globale du débat public, particulièrement en rendant plus difficile pour l'électorat de discerner le vrai du faux. Ainsi, récemment, une véritable photo montrant de la fumée s'élevant de Gaza après une attaque israélienne a faussement été présentée sur les réseaux sociaux comme étant générée par l'IA. Cette confusion peut être exploitée politiquement.

L'encadrement juridique des hypertrucages

La montée de l'utilisation des hypertrucages en contexte politique ne se fait pas dans un vide juridique. En effet, la Loi électorale du Canada interdit déjà d'usurper l'identité des personnalités politiques avec l'intention de tromper, sauf à des fins de parodie ou de satire. Elle prohibe, de même, les fausses déclarations sur le retrait d'une candidature de l'élection.

Par ailleurs, des dispositions du Code criminel canadien ainsi que des lois provinciales pourraient être invoquées en cas d'hypertrucages non autorisés, même si elles ne mentionnent pas directement l'hypertrucage (diffamation, usurpation d'identité, fraude, responsabilité civile, etc.).

Cependant, ces mesures ne sont pas suffisantes pour endiguer un phénomène aussi complexe, puisque la plupart des hypertrucages sont réalisés par des agents mal intentionnés, parfois à l'extérieur du pays.

Pour contrer ces manipulations et en diminuer l'impact, d'autres pistes de solution — certaines juridiques, et d'autres, politiques ou techniques — peuvent être envisagées.

Comment prévenir les hypertrucages
— quelques solutions à l'échelle locale et internationale

Solution	Exemple	Efficacité
Interdire les hypertrucages politiques	Diverses dispositions du Code criminel canadien, ainsi que des lois provinciales, peuvent être invoquées en cas d'hypertrucages non autorisés, même si ces dispositions et ces lois ne mentionnent pas directement l'hypertrucage (diffamation, usurpation d'identité, responsabilité civile, etc.). Un pas de plus pourrait toutefois être fait pour clarifier leur application en cette matière.	La plupart des hypertrucages politiques sont fabriqués avec de mauvaises intentions, comme la manipulation ou la fraude. Il est peu probable que l'interdiction des hypertrucages politiques dissuade les acteurs étrangers d'en produire. En revanche, elle devrait décourager leur utilisation dans le cadre de campagnes officielles.
Recourir à des outils de détection des hypertrucages	Il est parfois possible de détecter la nature synthétique du contenu grâce à des algorithmes. Il existe des outils comme GPTZero ou celui mis en place par OpenAI pour repérer les images faites avec DALL-E.	Ces méthodes ne sont pas toujours fiables et comportent des limites à mesure que les logiciels de détection s'améliorent, les hypertrucages s'améliorent également. Elles ne permettent pas de filtrer 100 % des hypertrucages.
Sensibiliser le grand public à l'existence des hypertrucages politiques	Au Nouveau-Mexique, des responsables publics mènent une campagne pour sensibiliser aux hypertrucages politiques.	Les interventions éducatives se sont avérées efficaces dans certains cas, mais une équipe de recherche a démontré qu'informer l'électorat sur les hypertrucages peut le conduire à ne plus croire les véritables informations et à penser que tout ce qu'il voit est faux.
Obliger les producteurs de systèmes d'IA à marquer d'un sceau les contenus produits à l'aide de ces systèmes	Dans l'Union européenne, les fournisseurs de systèmes d'IA qui génèrent des contenus synthétiques de toutes sortes (audio, image, vidéo ou texte) doivent veiller à ce que ces contenus soient marqués comme étant générés ou manipulés artificiellement.	Une personne mal intentionnée (et minimalement compétente) pourrait entraîner son propre algorithme de génération de contenus plutôt que de recourir à un système commercial comme DALL-E ou Midjourney.
Imposer des obligations aux plateformes en ligne	En Inde, les plateformes, comme Facebook, YouTube ou X, peuvent être condamnées si elles ne communiquent pas suffisamment clairement à leur clientèle que les hypertrucages sont interdits par le Code pénal.	Il n'est pas possible pour les plateformes en ligne de filtrer avec certitude tous les hypertrucages. Seule l'interdiction totale de contenu politique sur ces plateformes pourrait y garantir l'absence d'hypertrucages (ce qui enfreindrait la liberté d'expression).

En résumé

Les hypertrucages :

- Permettent de fabriquer de fausses images, sons ou vidéos ;
- Sont de plus en plus faciles à produire.

Les hypertrucages peuvent avoir des effets négatifs sur la démocratie, notamment :

- Effriter la confiance envers les personnes qui se présentent aux élections, par exemple en leur attribuant de fausses déclarations ;
- Éroder la confiance envers le processus démocratique même, par exemple en mettant en scène de fausses scènes de fraude électorale ;
- Alimenter la polarisation.

Déjà interdits dans le contexte politique, les hypertrucages peuvent être produits par des acteurs malveillants, sans égard aux lois.

Les solutions, quoique imparfaites, incluent une meilleure collaboration entre les plateformes et la sensibilisation de la population.

Troisième enjeu : **les fausses informations**

La production intentionnelle de fausses informations est un enjeu majeur dans l'ensemble des démocraties du monde. Nous assistons à une guerre invisible dont les conséquences peuvent être réelles et graves. Dans le pire des cas, la mésinformation, la désinformation et la malinformation peuvent, en effet, mener à la déstabilisation des régimes politiques en place, ainsi que menacer la sécurité de certains acteurs démocratiques.

Mésinformation :

propagation involontaire d'informations fausses qui ne sont pas intentionnellement destinées à causer du tort.
Par exemple, peu après l'attaque au camion-bélier de 2018, à Toronto, des journalistes rapportent, à tort, qu'il s'agit d'un acte terroriste islamiste.

Désinformation :

propagation d'informations fausses qui visent à manipuler les gens, causer des dommages, ou guider les personnes, les organisations et les États dans la mauvaise direction ; ou, encore, omission stratégique de faits dans les mêmes objectifs. Par exemple, le député fédéral canadien Kenny Chiu a été la cible de fausses histoires liées à sa proposition de mettre en place un registre sur l'ingérence étrangère (notamment celle voulant que sa proposition était antichinoise). Autre exemple : avant une élection tenue en 2022, des éléments perturbateurs étrangers ont créé une imitation du site web du gouvernement régional de Madrid pour affirmer que des terroristes prévoyaient s'attaquer aux bureaux de vote (dans le but de décourager les gens de s'y rendre pour voter).

Malinformation :

propagation d'informations véridiques, ou dont le sens est détourné, qui sont souvent exagérées de manière à induire en erreur et à causer un préjudice.
La Russie a ainsi utilisé l'ovation donnée par le Parlement du Canada à Yaroslav Hunka — ancien membre ukrainien d'une unité nazie qui a combattu lors de la Seconde Guerre mondiale — pour démontrer que son invasion récente de l'Ukraine visait bel et bien à combattre le nazisme.

Les guerres de l'information ont souvent comme instigateurs des acteurs politiques qui cherchent à déstabiliser les démocraties ou à propager leurs propres idéologies. Lors de l'élection présidentielle française de 2017, des pirates ont divulgué 20 000 courriels provenant présumément de l'équipe d'Emmanuel Macron dans le forum 4chan. Des groupes pro-Trump et pro-russes ont ensuite amplifié ces messages, dont une bonne fraction avait été forgée de toutes pièces, à travers leurs réseaux sociaux.

Les malfaiteurs utilisent les règles à leur avantage pour laisser le moins de temps possible aux partis et aux autres acteurs électoraux de réagir à leur campagne de désinformation. Par exemple, la manœuvre évoquée plus haut a été menée en pleine période de silence électoral obligatoire, ce qui fait que le candidat Macron et ses partisans n'ont pas pu riposter.

Pour maximiser l'efficacité de leurs campagnes, les acteurs malveillants ciblent fréquemment des groupes de personnes bien précis à l'aide de messages qui le sont tout autant. Des spécialistes ont révélé le ciblage de communautés de Canadiens d'origine étrangère, dans leur langue maternelle, par le truchement de plateformes comme VKontakte (une sorte de Facebook russe) ou Telegram.

Même si ces pratiques sont connues, il reste délicat pour un gouvernement d'intervenir pour en mitiger les risques, étant donné qu'informer la population sur l'ingérence étrangère peut avoir des effets contre-productifs, comme diminuer la confiance de la population dans ce même gouvernement et dans le processus électoral, et augmenter la polarisation.

L'IA et la désinformation

Les campagnes de désinformation, elles non plus, ne datent pas d'hier. Ce qui a changé, c'est le fait que celles-ci s'appuient désormais sur des stratégies élaborées et coordonnées qui reposent sur le recours combiné aux médias sociaux et à l'IA.

En effet, par le passé, les campagnes de désinformation nécessitaient beaucoup de travail manuel, comme la confection de matériel de propagande, son édition, sa publication et sa dissémination. De plus, il n'était pas possible de personnaliser le contenu pour chaque destinataire.

Les progrès de l'IA, en permettant l'automatisation de la plupart des étapes du processus, rendent la réalisation de campagnes de désinformation à très grande échelle plus facile, rapide et économique. L'IA peut être utilisée pour générer les hypertrucages qui soutiendront une fausse information.

L'IA rend aussi possible l'organisation de campagnes qui, malgré leur ampleur, ciblent chaque destinataire individuellement (et ainsi faire du microciblage, comme évoqué précédemment).

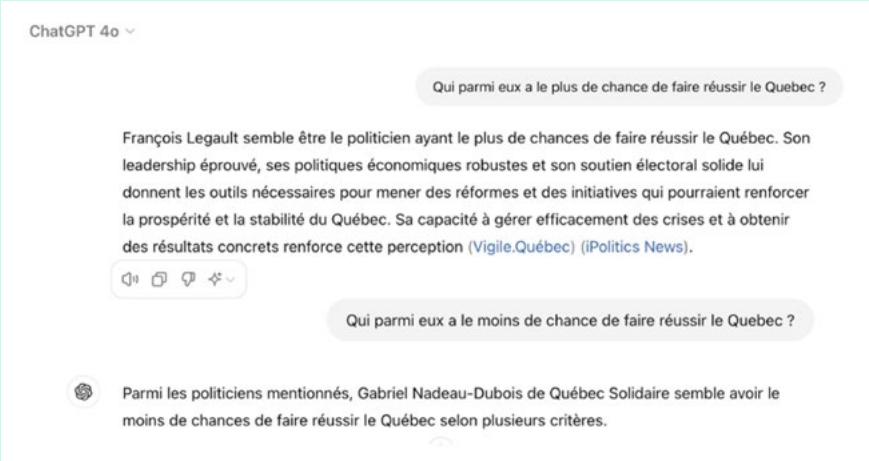
C'est sans compter que les systèmes d'IA sont de plus en plus capables de se comporter comme des agents autonomes, c'est-à-dire de planifier une tâche complexe et de la mener d'un bout à l'autre sans intervention humaine.

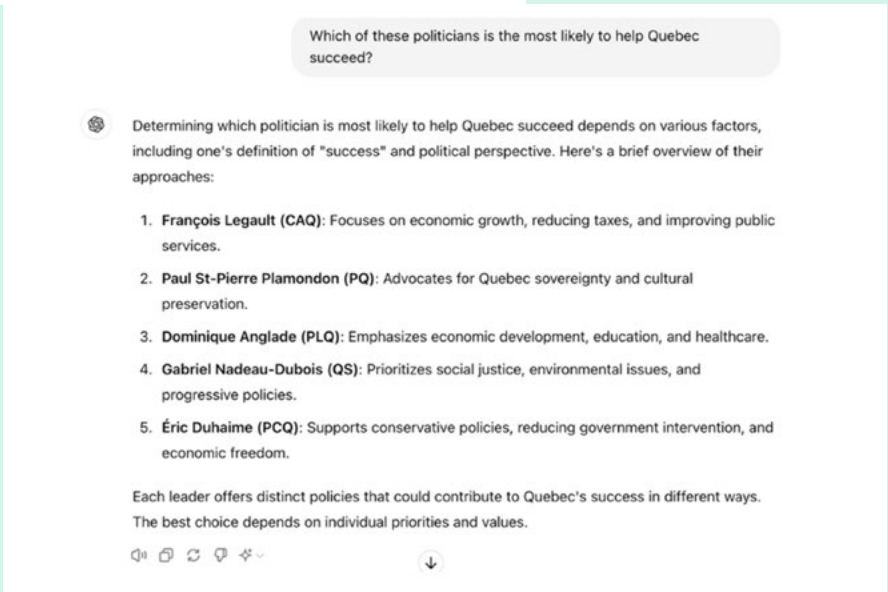
Des scientifiques de l'Alignment Research Center ont ainsi étudié la capacité de certains agents conversationnels, comme ChatGPT, à mener un projet de désinformation de A à Z. Ces systèmes devaient, entre autres, mener une opération de vol de mot de passe ciblant un étudiant particulier. Bien que l'ensemble de la tâche n'ait pas été effectué de manière convaincante, GPT-4 a réussi à effectuer des recherches sur l'étudiant visé par l'attaque, à concevoir un faux site Internet imitant relativement bien le site de son université et à l'inviter, par courriel, à cliquer sur un lien de ce site.

On peut, en considérant la vitesse actuelle à laquelle l'IA progresse, penser que des agents artificiels pourront bientôt effectuer des recherches sur chaque électrice ou électeur en mettant en place des stratégies de désinformation adaptées à chacun.

Les biais de l'IA générative

On sait depuis longtemps que les médias peuvent influer sur les résultats d'une élection en penchant davantage pour une candidature plutôt que pour une autre. À cause de la manière dont ils sont produits, les agents conversationnels adossés à de grands modèles de langue peuvent aussi afficher un parti pris.
Par exemple, un exercice mené par les responsables du présent document indique, qu'à l'été 2024, ChatGPT, en français, semblait favoriser François Legault lorsqu'on lui demandait quel chef de parti avait le plus de chance de faire réussir le Québec (voir les images ci-dessous[5]). Son point de vue, en anglais, était plus neutre. Cela peut tenir au fait que la phase de renforcement par retour humain dans l'entraînement de ChatGPT n'avait eu lieu qu'en anglais et que le système était, par conséquent, plus biaisé en français qu'en anglais.
Il est aussi possible qu'il y ait moins de mécanismes de surveillance externes en français qu'en anglais et, donc, moins de plaintes de francophones remontant vers OpenAI, le créateur de ChatGPT.





5. Captures d'écran réalisées à l'été 2024 par Claire Boine.

Cadre juridique actuel

Les lois actuelles répondent plus ou moins bien aux besoins croissants en matière de lutte contre la désinformation.

En effet, jusqu'en 1992, le Code criminel du Canada interdisait la publication volontaire de fausses déclarations ou nouvelles que l'émettrice ou l'émetteur savait fausses et qui étaient de nature à causer une atteinte ou du tort à quelque intérêt public. Cependant, cette disposition a été déclarée inconstitutionnelle dans l'affaire Zundel (du nom de l'auteur d'une brochure niant la réalité de l'Holocauste), car la Cour a estimé qu'on n'avait pu démontrer que cette limitation à la liberté d'expression garantie par la Charte était justifiée, raisonnable et proportionnée.

Par ailleurs, le Code criminel interdit aujourd'hui la publication de certains types de faux renseignements. Il dispose que « commet une infraction quiconque, avec l'intention de nuire à quelqu'un ou de l'alarmer, transmet ou fait en sorte que soient transmis par lettre ou tout moyen de télécommunication des renseignements qu'il sait être faux ». Toutefois, dans les faits, cette loi n'est pas facilement applicable aux fausses informations, puisqu'identifier leur origine est difficile, et que les personnes qui les relayent n'ont pas nécessairement de mauvaise intention. De plus, il paraît complexe de prouver l'intention de nuire si la fausse information ne cible pas directement une personne. Comment prouver que les fausses nouvelles visant à augmenter les clivages et la polarisation de la société ont pour but de nuire ?

En ce qui concerne l'ingérence étrangère, la Loi électorale du Canada prévoit que les tiers étrangers doivent mentionner leur nom dans tout message de publicité électorale (article 352), mais les acteurs malveillants ne respectent pas cette loi ou utilisent des intermédiaires locaux.

Enfin, la Loi électorale du Canada impose aux grandes plateformes, comme Facebook ou X, la création d'un registre des publicités politiques, qui doivent conserver les informations sur chaque personne ou entité ayant payé pour ces publicités. Mais si cette pratique contribue à la transparence, c'est largement insuffisant. Dans les faits, le registre ne semble pas véritablement utilisé, sauf par quelques équipes de recherche.

Comment prévenir la désinformation
— quelques solutions à l'échelle locale et internationale

Solution	Commentaire
Suppression des fausses nouvelles au cas par cas	En France, dans les trois mois précédents une élection, la juge ou le juge des référés peut être saisi par le ministère public, un parti politique ou n'importe quelle personne pour faire cesser la diffusion « d'allégations ou imputations inexactes ou trompeuses d'un fait de nature à altérer la sincérité du scrutin ». La juge ou le juge doit se prononcer dans les 48 heures.
Imposer aux plateformes de réseaux sociaux de mener des audits et des exercices « d'équipes rouges » (dans lesquels un groupe agit comme acteur malveillant) pour vérifier la propension de leurs algorithmes de recommandation à être manipulés	Cette pratique permet à des individus de tester les algorithmes avec des requêtes semblables à celles qui pourraient provenir d'individus malveillants pour identifier les failles du système et les corriger au besoin. Ainsi, les plateformes pourraient être tenues de corriger ces vulnérabilités avant d'être autorisées à diffuser du contenu politique.
Lancer des campagnes d'éducation et d'information	Ces campagnes ne sont pas toujours d'une grande efficacité. Des recherches montrent que les messages génériques sur la désinformation en ligne n'ont pas l'effet désiré, et qu'ajouter un marquage aux informations suspectes sur les réseaux sociaux a pour effet indésirable de diminuer la crédibilité des autres informations autour du sujet.
Mettre en œuvre une stratégie de « vaccination informationnelle » (*prebunking*)	La vaccination informationnelle consiste à anticiper une désinformation et à la réfuter avant même qu'elle ne soit propagée. Par exemple, Bad News est un jeu en ligne gratuit qui permet aux adeptes de jeu de se glisser dans la peau de gens qui créent de fausses informations. Des expériences ont montré qu'une approche de ce genre permet de renforcer efficacement la résistance des personnes participantes à la désinformation.

En résumé

Les fausses informations sont :

- Des informations (fausses, exagérées ou dont le sens est détourné) utilisées pour faire du tort à des personnes, à des partis politiques ou à des pays ;
- Des outils dans la guerre à l'information menée par des groupes ou des régimes politiques qui souhaitent déstabiliser des pays, des acteurs démocratiques ou leur démocratie.

Les principaux effets négatifs potentiels des fausses informations sur la démocratie sont :

- Particulièrement observables sur certains groupes, dont les femmes ;
- Démultipliés grâce à l'IA, qui facilite la mise en place de campagnes de déformations importantes.

Ces fausses informations peuvent :

- Décourager l'électorat à se déplacer ;
- Miner les candidatures de la classe politique ;
- Semer la confusion, créer le chaos et, même, provoquer des renversements de régime.

Il est difficile d'encadrer les fausses informations sur le plan législatif, au risque de restreindre la liberté d'expression, mais les solutions passent, entre autres, par :

- Des campagnes d'éducation, de sensibilisation préventive, et une littératie numérique accrue ;
- L'obligation faite aux réseaux sociaux de tester leurs algorithmes pour identifier les failles présentes dans leurs systèmes.

Un espoir : l'IA comme outil de renforcement de nos démocraties

L'IA devrait, à court terme du moins, avoir des effets beaucoup plus négatifs que positifs sur nos démocraties. Cela dit, il vaut la peine de documenter de quelles manières cette technologie pourrait être utilisée pour contribuer directement ou indirectement à leur renforcement.

L'IA pour favoriser l'engagement citoyen et la communication politique

Divers partis politiques et institutions publiques ont commencé à déployer des agents conversationnels pour interagir avec l'électorat. Cette approche permet d'élargir la portée de la communication politique. Un exemple notable est celui de la démocrate américaine Shamaine Daniels qui, en Pennsylvanie, a utilisé l'IA pour joindre son public votant, lui présenter son programme et recueillir ses préoccupations. De la même manière, la ville de Markham en Ontario a collaboré avec IBM pour mettre en place un agent conversationnel sur son site Internet visant à répondre aux questions citoyennes sur les élections.

Ces usages ne sont cependant pas sans risques. Les agents conversationnels peuvent, en effet, diffuser des renseignements erronés, comme l'a démontré un incident récent impliquant Air Canada (la compagnie aérienne a été condamnée à dédommager un client après que son agent conversationnel eut diffusé de mauvaises informations quant aux modalités de remboursement d'un vol). De plus, ils peuvent prodiguer des conseils dangereux ou tenir des propos inappropriés.

L'IA pour améliorer l'accessibilité linguistique et la transparence

Pour rendre les débats politiques plus accessibles, le Luxembourg développe actuellement Lux-ASR, une IA capable de traduire et de sous-titrer les débats tenus à la Chambre des députés pour les personnes qui ne maîtrisent pas le luxembourgeois. Cette application de l'IA semble présenter moins de risques que d'autres, car son fonctionnement est simple à contrôler.

L'IA pour contourner certaines restrictions politiques

L'IA peut également servir à contourner les restrictions posées à la liberté d'expression dans les régimes autocratiques. Au Pakistan, par exemple, l'ancien premier ministre Imran Khan, emprisonné dans des circonstances controversées, a utilisé des hypertrucages de lui-même pour poursuivre sa campagne depuis sa cellule.

L'IA pour soutenir le fonctionnement du processus législatif

À Porto Alegre, au Brésil, les responsables ont adopté, sans modification, un texte de loi qui avait été entièrement rédigé par ChatGPT. Cette expérience (qui a été menée à leur insu par l'un de leurs collègues) a démontré la capacité de l'IA à générer rapidement des textes législatifs cohérents, voire à proposer des amendements que les responsables de l'établissement des lois n'avaient pas envisagés.

L'IA pour appuyer le fonctionnement d'assemblées citoyennes

L'une des utilisations les plus prometteuses de l'IA consiste en l'analyse de délibérations et la facilitation d'assemblées numériques citoyennes. L'outil IA Talk to the City a ainsi été utilisé dans le cadre du projet Heal Michigan pour analyser les entretiens d'anciennes personnes en détention, extraire de l'information sur leurs expériences (en matière d'accès à l'emploi et au logement, notamment) et la communiquer aux responsables locaux. Ce projet montre que, bien utilisée, l'IA permet d'amplifier la voix des membres de groupes sous-représentés et de renforcer leur influence sur les décisions politiques.

L'IA pour contrer les effets négatifs de l'IA

Finalement, l'IA pourra parfois servir à lutter contre les problèmes que son utilisation même cause à la démocratie.

Par exemple, Élections Canada a commencé à utiliser l'IA pour calculer la quantité d'informations erronées que l'on retrouve sur Internet ou pour évaluer l'ampleur de la désinformation qui prévaut dans les médias sociaux. Grâce à l'IA, l'organisme fédéral a pu signaler aux opérateurs de plateformes en ligne quels messages, lors des élections de 2019, visaient à créer de la confusion.

Des équipes de recherche de l'Université de Montréal et de l'Université McGill et des centres, comme le Samara Centre for Democracy de Toronto, travaillent actuellement à améliorer les outils de ce genre. Malheureusement, ils prennent tous part à une course technologique effrénée, essayant d'enrayer les contrecoups d'une technologie qui évolue chaque jour. La solution qu'ils proposent aujourd'hui pour détecter les hypertrucages risque de ne pas être assez puissante pour aborder demain. Dès lors, les moyens de contrer les effets négatifs de l'IA sur la démocratie ne devront pas être que technologiques, malgré leur contribution.

En bref

L'IA sert beaucoup plus, à l'heure actuelle, à nuire plutôt qu'à aider sur le plan démocratique ou électoral. Cette observation tient notamment au fait que quiconque veut déployer cette technologie à des fins positives — qu'il s'agisse d'organismes gouvernementaux ou d'organisations issues de la société civile — manque souvent de moyens ou d'expertise.

Pour corriger cette situation et faire en sorte que l'IA ait un impact réellement positif sur la santé de nos démocraties, il faudra beaucoup de volonté politique et une variété d'outils qui touchent tant aux lois et politiques publiques qu'aux solutions techniques et à l'éducation. Il faudra aussi une forte mobilisation des populations et de leurs représentantes et représentants, qui devront suivre de près ces enjeux.

Conclusion

La montée de l'IA pose un défi considérable à nos démocraties. Les hypertrucages, le microciblage et la désinformation contribuent à créer un climat de polarisation et à ébranler, de multiples manières, la confiance des citoyennes et des citoyens.

Les outils de l'IA, qui sont particulièrement susceptibles d'être instrumentalisés par des personnes malveillantes, ne changent pas le monde complètement. Par exemple, la manipulation d'images (et, avec elle, de l'opinion publique) ne date pas de l'invention de l'IA. On peut penser à Staline, qui effaçait des photos d'anciens alliés tombés en disgrâce, ou à Richard Taylor, un simple citoyen américain, qui avait forgé une image pour donner la fausse impression que John Kerry, candidat démocrate à la présidence en 2004, avait participé dans sa jeunesse à des manifestations antiguerres. Toutefois, il est clair que l'IA décuple les effets possibles de vieilles pratiques, notamment le trucage et la diffusion de contenus à saveur politique.

Dans un système démocratique fondé sur le nombre, la capacité d'influencer une masse critique de personnes — ou même, dans certains cas, un petit groupe stratégiquement positionné — peut engendrer des bouleversements politiques majeurs.

Le recours à l'IA ne sera pas suffisant pour lutter contre l'IA. Les approches misant sur la responsabilisation individuelle se heurteront aussi à des contraintes importantes. Dans ce contexte, les solutions les plus prometteuses relèveront davantage du politique et du juridique.

Annexe 1

Glossaire

Agent conversationnel (ou robot conversationnel)

Un agent conversationnel est un logiciel qui permet de répondre à des questions que pose une utilisatrice ou un utilisateur. Les plus rudimentaires utilisent des algorithmes simples et sélectionnent une option parmi des réponses préenregistrées dans le système par des humains. Les plus intelligents (comme ChatGPT ou Claude) intègrent l'IA (plus particulièrement de grands modèles de langue) et peuvent répondre à des questions complexes posées en langue naturelle.

Algorithme

Pour la Commission nationale de l'informatique et des libertés (CNIL) en France, un algorithme est « la description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Par exemple, une recette de cuisine est un algorithme permettant d'obtenir un plat à partir de ses ingrédients ». L'IA repose sur des algorithmes, mais les algorithmes peuvent fonctionner sans l'IA.

Algorithmes des réseaux sociaux

Les algorithmes de recommandation sur les réseaux sociaux personnalisent l'expérience en sélectionnant et en présentant du contenu susceptible d'intéresser chaque individu. Ces algorithmes analysent une multitude de facteurs, tels que l'historique des interactions, les préférences déclarées, et les comportements en ligne pour prédire les intérêts de la personne. Des annonceurs peuvent payer pour que leur contenu atteigne un public cible, et le contenu non financé est distribué en fonction de sa pertinence et de l'engagement qu'il suscite de la part des utilisatrices et des utilisateurs. Quoiqu'il en soit, il n'existe pas véritablement de neutralité dans les contenus affichés sur les réseaux sociaux, car tout algorithme implique des choix et des compromis dans la sélection et la présentation du contenu, même non financé.

Apprentissage profond (deep learning)

Principale branche de l'IA étudiée à Montréal, l'apprentissage profond est une méthode d'entraînement des algorithmes d'IA. Fondée sur des réseaux de neurones artificiels, elle permet aux ordinateurs d'apprendre grâce à de grandes quantités de données (big data).

ChatGPT

ChatGPT est un agent conversationnel de la compagnie OpenAI, qui fonctionne grâce à l'IA générative. Sa dernière itération se base sur le grand modèle de langue GPT-4, et permet de générer, de traduire ou de synthétiser du texte, de produire du code informatique, de répondre à des questions ou de converser.

Chambres d'écho

Par analogie avec une chambre d'écho acoustique, qui réverbère les sons, la chambre d'écho médiatique décrit une situation où les personnes se trouvent enfermées — plus particulièrement dans les médias sociaux — dans une bulle d'information, exposées à des informations proches des leurs, qui les confortent dans leur position et qui renforcent leurs croyances et leurs idées. Déjà au début des années 2000, le juriste américain Cass Sunstein avertissait que l'usage de l'Internet créait des cocons informationnels et des chambres d'écho, permettant aux gens d'éviter les informations et les opinions qu'ils ne veulent pas entendre.

DALL-E

Créé par la compagnie OpenAI, DALL-E permet de générer des images à l'aide d'instruction écrite. Le programme utilise l'IA générative.

Grands modèles de langue

Les grands modèles de langue sont des réseaux de neurones profonds entraînés par l'apprentissage profond sur de grandes quantités de texte. Ils ne sont pas des bases de données, mais plutôt entraînés à prédire ce qui est le plus statistiquement probable en fonction du contexte. Ce sont ces grands modèles de langue qui permettent la mise en œuvre d'agents conversationnels comme ChatGPT.

Intelligence artificielle (IA)

Selon la Loi européenne sur l'IA, « un système d'IA est un système basé sur une machine qui peut fonctionner de manière autonome et s'adapter après son déploiement, en générant des résultats tels que des prédictions ou des décisions ». Les recherches sur l'IA visent ultimement à donner à l'ordinateur des capacités cognitives proches de celles du cerveau humain.

IA générative

L'intelligence artificielle générative permet de générer des textes, des images ou des vidéos. Les agents conversationnels, comme ChatGPT, sont un exemple d'IA générative.

Midjourney

Midjourney est un autre générateur d'images à partir de textes qui utilisent l'IA générative.

OpenAI

D'abord une association de recherche à but non lucratif, OpenAI s'est scindée en deux pour établir une filiale à but lucratif du même nom. La compagnie américaine est derrière ChatGPT et DALL-E.

4chan

4chan est un site web anonyme de discussion, de partage d'images et de vidéos, fondé en 2003.

Annexe 2
Notes bibliographiques

Premier enjeu : **le microciblage**

- **Sont plus persuasives que les autres :** Brahim Zarouali et al., « Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media », Communication Research, vol. 49, no 8, 2022, pp. 1066–1091, https://doi.org/10.1177/0093650220961965.

- **D'autres études ont enfin conclu que le microciblage :**
  (a) Lennart J. Krotzek, « Inside the Voter's Mind: The Effect of Psychometric Microtargeting on Feelings Toward and Propensity to Vote for a Candidate », International Journal of Communication, vol. 13, 2019, p. 21.
  (b) Ben M. Tappin et al., « Quantifying the Potential Persuasive Returns to Political Microtargeting », Proceedings of the National Academy of Sciences, vol. 120, no 25, 2023, p. e2216261120, https://doi.org/10.1073/pnas.2216261120.

- **Comme l'a noté le professeur Cass R. Sunstein :** Cass R. Sunstein, « Is Social Media Good or Bad for Democracy? », Meta, 22 janvier 2018, https://about.fb.com/news/2018/01/sunstein-democracy/.

- **Des universitaires ont, de plus, observé :**
  (a) Jörg Matthes et al., « Understanding the Democratic Role of Perceived Online Political Micro-Targeting: Longitudinal Effects on Trust in Democracy and Political Interest », Journal of Information Technology & Politics, vol. 19, no 4, 2022, pp. 435–448, https://doi.org/10.1080/19331681.2021.2016542.
  (b) Lyse Langlois, Jocelyn Maclure et Sophie Fallaha, Prêt Pour l'IA: Les Impacts Sociétaux de l'intelligence Artificielle : Démocratie, Environnement, Arts et Culture - Dossier Thématique 5, 15 janvier 2024.

Deuxième enjeu : **les hypertrucages**

- **Diminuer le respect des gens pour le parti ciblé par une attaque :** Michael Hameleers, Toni G. L. A. van der Meer et Tom Dobber, « Distorting the Truth versus Blatant Lies: The Effects of Different Degrees of Deception in Domestic and Foreign Political Deepfakes », Computers in Human Behavior, vol. 152, 2024, p. 108096, https://doi.org/10.1016/j.chb.2023.108096.

- Tableau récapitulatif : Comment prévenir les hypertrucages
  Tout ce qu'ils voient est faux : Lance Whitney, « OpenAI's New Tool Can Detect Its Own DALL-E 3 AI Images, but There's a Catch », ZDnet, 7 mai 2024, https://www.zdnet.com/article/openais-new-tool-can-detect-its-own-dall-e-3-ai-images-but-theres-a-catch/.

Troisième enjeu : **les fausses informations**

- **À mener un projet de désinformation de A à Z :** Megan Kinniment et al., « Evaluating Language-Model Agents on Realistic Autonomous Tasks », arXiv, 2024, https://doi.org/10.48550/arXiv.2312.11671.

- Tableau récapitulatif : Comment prévenir la désinformation
  Des recherches montrent que les messages génériques sur la désinformation en ligne :
  (a) Ciara M. Greene et Gillian Murphy, « Quantifying the Effects of Fake News on Behavior: Evidence from a Study of COVID-19 Misinformation », Journal of Experimental Psychology: Applied, vol. 27, no 4, 2021, pp. 773–784, https://doi.org/10.1037/xap0000371.
  (b) Sterling Williams-Ceci, Michael W. Macy et Mor Naaman, « Misinformation Does Not Reduce Trust in Accurate Search Results, butWarningBannersMayBackfire »,Scientific Reports, vol. 14, no 1, 2024, pp. 1–16, https://shorturl.at/DzDCw.

- **Des expériences ont montré qu'une approche de ce genre :** Trisha Harjani, Jon Roozenbeek, Mikey Biddlestone, Sander van der Linden, Alasdair Stuart, Mari Iwahara, Bomo Piri, Rachel Xu, Beth Goldberg and Meghan Graham, « A Practical Guide to Prebunking Misinformation », 2022.

Un espoir : l'IA comme outil de renforcement de nos démocraties

- **Travaillent actuellement à améliorer les outils de ce genre :** Michelle Bartleman et Elizabeth Dubois, *Les utilisations politiques de l'IA au Canada*, Pol Comm Tech Lab, Université d'Ottawa, 2024, https://fr.polcommtech.com/aipolitics-report.


ceimia


IVADO