



Utiliser les modèles de langage de grande envergure pour renforcer la démocratie

Seth Lazar¹ · Lorenzo Manuali²



Reçu le 12 mars 2025 / Accepté le 22 janvier 2026
© Les auteurs 2026

Résumé

Les LLM comptent parmi les outils les plus avancés jamais conçus pour comprendre et générer du langage naturel. La délibération et la prise de décision démocratiques impliquent, à plusieurs étapes distinctes, la production et la compréhension du langage. Il est donc naturel de se demander si nos meilleurs outils linguistiques pourraient s'avérer utiles pour l'une de nos tâches linguistiques les plus importantes impliquant le langage. Des chercheurs et des praticiens se sont récemment demandé si les LLM pouvaient soutenir la délibération démocratique en tirant parti de leurs capacités à résumer du contenu, à agréger des opinions sur ce contenu résumé et à représenter les électeurs en prédisant leurs préférences face à des choix non vus. Dans cet article, nous évaluons si l'utilisation des LLM pour remplir ces fonctions et d'autres fonctions connexes fait réellement progresser les valeurs démocratiques qui sous-tendent ces expériences. Nous suggérons que le bilan est mitigé. En présence d'inégalités de fond en matière de pouvoir et de ressources, ainsi que de profonds désaccords moraux et politiques, nous ne devrions pas utiliser les LLM pour automatiser des composantes du processus démocratique qui n'ont pas de valeur instrumentale, ni être tentés de supplanter les procédures décisionnelles équitables et transparentes qui sont pratiquement nécessaires pour concilier des intérêts et des valeurs concurrents. Cependant, si les LLM doivent être tenus à l'écart des processus décisionnels démocratiques formels, nous pensons qu'ils peuvent en revanche renforcer la sphère publique informelle — l'arène qui sert d'intermédiaire entre les gouvernements démocratiques et les communautés politiques qu'ils servent, au sein de laquelle les communautés politiques recherchent des informations, forment des publics civiques et demandent des comptes à leurs dirigeants.

Mots-clés : Grands modèles linguistiques · Démocratie · Intelligence artificielle · Valeurs démocratiques

✉ Seth Lazar seth.lazar@anu.edu.au

Lorenzo Manuali lmanuali@umich.edu

¹ Université nationale australienne, Canberra, Australie

² Université du Michigan, Ann Arbor, États-Unis

1 Introduction

Après l'orgueil des années 1990 (Fukuyama, 2020), les deux dernières décennies ont vu les institutions démocratiques subir une pression croissante, avec la montée mondiale de l'autoritarisme et le sentiment grandissant que la démocratie est en crise (Diamond, 2020 ; Grumbach, 2023 ; Kingzette et al., 2021 ; Scheppele, 2022). Si cette inquiétude est en grande partie due aux efforts concertés des autocrates (et des aspirants autocrates) pour utiliser des moyens légalistes afin de saper les libertés démocratiques (Huq & Ginsburg, 2018 ; Scheppele, 2018), elle tient également en partie à notre dépendance collective croissante vis-à-vis des technologies numériques contrôlées par de puissantes entreprises technologiques (Lazar, 2025 ; Lehdonvirta, 2022 ; Susskind, 2022). Début 2023, le lancement de GPT-4 a marqué l'arrivée de grands modèles linguistiques (LLM) hautement performants dans ce bouillon bouillonnant, et ceux-ci ont immédiatement suscité une anxiété accrue parmi les partisans de la démocratie (Allen & Weyl, 2024 ; Coeckelbergh, 2024 ; Kreps & Kriner, 2023). L'IA étroite existante, malgré toutes ses failles et ses inefficacités, avait néanmoins déjà amplifié le pouvoir de quelques entreprises sur de vastes pans de nos vies (Crawford, 2021 ; Lazar, 2024b). Il serait naturel de supposer que l'avènement de systèmes d'IA bien plus performants ne ferait qu'exacerber cette tendance.

Il n'est donc pas surprenant que ceux qui se soucient à la fois de la démocratie et de la promotion de la recherche en IA cherchent à déterminer si l'IA pourrait réparer ce qu'elle menace de détruire (Ovadya, 2023 ; Siddarth, 2024 ; Weyl et al., 2024). Dans plusieurs articles de recherche, des informaticiens et des ingénieurs en IA ont cherché à utiliser les modèles de langage (LLM) pour améliorer la prise de décision démocratique.¹

Notre objectif ici est d'examiner si, et dans quelle mesure, l'utilisation des LLM pour soutenir la délibération et la prise de décision démocratiques sert réellement les objectifs démocratiques. Étant donné que bon nombre de ces projets ont été financés par les mêmes grandes entreprises technologiques que certains accusent de déstabiliser la démocratie, certains pourraient rejeter ces initiatives comme de simples opérations de « democracy-washing » (Himmelreich, 2023). Cependant, nous considérons ces projets de recherche pour ce qu'ils sont, sans nous attarder davantage sur leurs motivations ou leur financement.

Dans la section suivante, nous identifions six valeurs distinctes, instrumentales et non instrumentales, auxquelles les institutions démocratiques servent. Dans la section 3, nous (1) classons les différentes manières dont les ingénieurs ont utilisé les grands modèles linguistiques (LLM) pour améliorer la prise de décision et la délibération démocratiques et (2) nous nous demandons dans quelle mesure chaque approche de la démocratie renforcée par les LLM sert les valeurs démocratiques que nous identifions. Dans la section 4, nous cherchons à expliquer les lacunes identifiées dans la section 3, en identifiant les fonctions démocratiques que les LLM ne peuvent, en principe, pas remplir — ainsi que celles qu'ils peuvent remplir. La section 5 conclut.

¹ Notre analyse documentaire a été réalisée au milieu de l'année 2024 (à noter que certains articles ont fait l'objet d'une prépublication avant leur date de publication définitive). Nous avons utilisé des bases de données scientifiques et des outils de recherche pour identifier les articles visant à utiliser les modèles de langage de grande envergure (LLM) afin de soutenir la délibération ou la prise de décision démocratiques. Nous avons suivi l'évolution du domaine depuis notre première analyse et avons confirmé que les articles plus récents s'inscrivent toujours dans la taxonomie que nous avons déduite de la littérature existante. Les articles clés abordés ici sont : Bakker et al. (2022), « First Report: Democratic Inputs to AI » (2023), Argyle et al. (2023), Devine et al. (2023), Edelman (2023), Edelman et Klingefjord (2023), Konya et al. (2023), Landemore (2023), Mendoza (2023), Small et al. (2023), Zaremba et al. (2023), Gudiño-Rosero et al. (2024), Huang et al. (2024), Tessler et al. (2024), Fish et al. (2025), Konya et al. (2025). Nous mentionnons également en passant certains projets appliqués, tels que l'utilisation par Meta de sondages délibératifs pour élaborer des règles régissant le harcèlement et les brimades dans les espaces de réalité virtuelle (Sulots, 2023). Pour un aperçu et une histoire évolutive de certaines de ces initiatives, voir Goldberg et al. (2024).

2 Quel est l'intérêt de la démocratie ?

Les théoriciens politiques sont trop rarement sollicités pour contribuer ne serait-ce qu'à la rédaction des lois sur lesquelles reposent les nouvelles organisations politiques ; l'occasion de mettre en place de nouvelles technologies démocratiques est difficile à laisser passer. Mais, si les initiatives que nous décrivons dans la section suivante ont toutes des motivations clairement démocratiques, comment pouvons-nous évaluer si et dans quelle mesure elles servent réellement les valeurs démocratiques ? *Qu'est-ce que* les valeurs démocratiques, au juste ? Il est évident que nous ne pouvons pas trancher cette question dans l'espace dont nous disposons ; cependant, nous mettons en avant six valeurs clés, largement reconnues dans la littérature sur la théorie démocratique, qui peuvent servir de baromètre pour évaluer ces initiatives.² Pour faciliter l'exposé, nous ferons la distinction entre les valeurs non instrumentales et les valeurs instrumentales (E. Anderson, 2009).³

Les réalités pratiques de la politique et la division sociale du travail font que l'influence *effective* sur les résultats sociaux ne peut jamais être répartie de manière égale. Certaines personnes briguent et occupent des postes à responsabilité qui les placent inévitablement, à cet égard, « au-dessus » des autres. Les procédures décisionnelles démocratiques affirment, malgré ces réalités pratiques, l'égalité de tous ceux qui ont leur mot à dire (Dahl, 1989). Elles établissent que tous les exercices du pouvoir politique — toutes les utilisations des ressources et de l'autorité normative de l'État — sont en fin de compte soumis au contrôle d'un mécanisme décisionnel auquel chaque citoyen démocratique a le droit de contribuer de manière égale (Allen, 2023 ; Kolodny, 2023).

Ce contrôle est important non seulement parce qu'il affirme l'égalité fondamentale de tous ceux qui y contribuent. Il est également constitutif de *la responsabilité* pour des décisions particulières, et de l'autorisation pour l'exercice plus large du pouvoir par ceux qui agissent en notre nom (Christiano, 2004 ; Viehoff, 2014). Il s'agit à la fois d'un contre-pouvoir nécessaire qui contrebalance le pouvoir que ceux qui agissent en notre nom exercent sur nous (Pettit, 2012), et du moyen par lequel nous exerçons collectivement un contrôle positif sur les institutions qui nous servent (Lazar, 2025 ; Miller, 1995).

La troisième valeur fondamentale non instrumentale de la politique démocratique réside dans la possibilité qu'elle offre d'exercer la vertu civique par la participation à la vie politique (Macedo, 1990). Cette notion trouve ses racines profondes dans la philosophie politique aristotélicienne (Winthrop, 1978), mais elle a été réinterprétée pour l'ère du libéralisme politique des années 1990 (Macedo, 1990) et compte en outre des partisans républicains (Dagger, 1997). Selon ces points de vue, la participation politique active est un bien non instrumental — qui fait partie de ce que signifie mener une vie épanouie — et les institutions démocratiques sont nécessaires pour que ce bien soit accessible à tous.

Penchons-nous maintenant sur les vertus pragmatiques de la démocratie, c'est-à-dire les avantages qu'elle nous aide à concrétiser. Il y a tout d'abord l'avantage concret qu'elle offre pour la mise en place d'un gouvernement stable en présence d'intérêts divergents (Lijphart, 1977 ; Miller, 1983). Cela tient en partie au fait que les procédures démocratiques transparentes constituent le meilleur moyen disponible pour garantir ces

² Nous n'avons pas la place de défendre ces valeurs dans cet article ; cependant, les ouvrages cités contiennent de nombreuses défenses solides de chacune des valeurs que nous invoquons.

³ La position d'Elizabeth Anderson (2009) est que la valeur non instrumentale de la démocratie *dépend* de sa valeur instrumentale. En un sens, cela rend la distinction moins nette que nous ne le laissons entendre. Mais, par souci de simplicité et de clarté, nous pensons qu'elle offre une bonne délimitation.

qui perdent un scrutin politique et ont des raisons claires et manifestement justifiées d'accepter leur défaite (C. Anderson, 2005). Cela vaut même dans le cadre d'un processus décisionnel collectif ordinaire. Chacun peut comprendre la légitimité du vote majoritaire, et si l'on vient d'assister au déroulement de la procédure et de constater sa défaite, cela constitue une raison solide (bien que manifestement pas décisive, surtout si l'on fait partie d'une minorité persistante) d'accepter le résultat. Mais la démocratie est également un excellent moyen d'identifier des compromis possibles entre des intérêts concurrents. En laissant une marge de manœuvre aux acteurs disposant d'un droit de veto dans la prise de décision collective (Tsebelis, 1999), les procédures démocratiques nous incitent à rechercher et à identifier des positions de compromis susceptibles de contourner ces vetos.

En plus de nous permettre de parvenir à un *modus vivendi* (Rawls, 1993), la démocratie peut constituer un moyen, certes imparfait, de nous rapprocher de ce qui s'apparente davantage à une volonté générale (Rousseau, 1762), ou tout au moins de nous offrir la possibilité de modifier nos préférences politiques grâce à l'interaction les uns avec les autres (Young, 2002). Lorsque le processus démocratique nous permet, grâce à l'engagement avec l'autre, de nous rapprocher réellement de lui en termes de valeurs et de préférences politiques, cette *transformation* non seulement apaise l'instabilité et les troubles, mais elle permet également une forme authentique d'autodétermination collective (Lovett & Zuehl, 2022).

Enfin, l'un des arguments les plus courants en faveur de la démocratie est qu'elle nous permet de prendre de meilleures décisions : elle s'appuie sur notre intelligence collective pour faire des choix épistémiquement plus judicieux (Landemore, 2012b). Certains arguments à cet égard sont de nature descriptive : ils montrent que les démocraties ont en effet donné naissance à des sociétés plus stables et plus prospères (Quinn & Woolley, 2001). D'autres sont pragmatiques, affirmant que les sociétés devraient viser la justice, et que seules les institutions politiques qui donnent à chacun les moyens de faire valoir ses intérêts permettront d'atteindre cet objectif (Allen, 2023). D'autres encore se concentrent sur des résultats théoriques, tels que le théorème de Condorcet, et sur des travaux plus généraux concernant la sagesse des foules (Cohen, 1986 ; Goodin & Spieckermann, 2018).

Par souci de concision, nous désignerons ces valeurs respectivement par les termes *d'égalité*, *d'habilitation*, *de participation*, *de conciliation*, *de transformation* et *d'intelligence collective*. Or, il est évident que ces valeurs revêtent une importance *bien plus grande* dans le cadre des décisions à enjeux élevés pour lesquelles nous recourons généralement aux procédures démocratiques. Nous devrions nous soucier beaucoup plus de l'égalité et de l'autorisation, par exemple, lorsque nous réfléchissons au déploiement du pouvoir coercitif de l'État que lorsque nous nous penchons sur les normes conversationnelles des chatbots. Cependant, chacun des articles abordés dans la section 3 s'inscrit dans un esprit démocratique, et il est donc légitime d'examiner dans quelle mesure ils pourraient concrétiser ces six bienfaits.

3 Les grands modèles de langage au service de la démocratie

L'intersection entre la démocratie et les technologies numériques est étudiée depuis des décennies par les philosophes et les spécialistes en sciences sociales.⁴ Certains travaux universitaires soutiennent que les technologies numériques menacent les normes et les institutions démocratiques — par exemple en raison de la

⁴ Voir par exemple Wiener (1950), McLuhan (1964), Mumford (1964), Westin (1971), Winner (1984), Sterling (1986), Arterton (1987), Bjerknes et al. (1987), Ronfeldt (1992), Bonchek (1995), Groper (1996), Buchstein (1997), Margolis et al. (1997), Hill et Hughes (1998), Wright (2018), Bernholz et al. (2021), Coeckelbergh (2024).

la concentration du pouvoir entre les mains des entreprises technologiques, ou la manière dont les réseaux sociaux ont sapé la sphère publique numérique.⁵ Certains proposent des approches radicales pour reprendre le contrôle démocratique des industries et des outils numériques.⁶ D'autres critiquent la récupération de l'idéologie démocratique par une industrie technologique qui confond diffusion de masse et démocratisation.⁷ D'autres encore imaginent un avenir numérique où l'IA et les technologies connexes permettraient des formes d'engagement démocratique totalement inédites.⁸ Des informaticiens et ingénieurs en IA animés par des motivations similaires ont abordé ces problèmes sous un angle différent, cherchant à construire l'avenir de la démocratie et de l'IA plutôt qu'à l'évaluer. Mais jusqu'à récemment, ils devaient se limiter principalement à des exercices restreints de choix social computationnel et à des propositions audacieuses de nouveaux types de processus que les progrès futurs de l'IA pourraient rendre réalisables.⁹ Cela a changé au début de l'année 2023.

Les grands modèles linguistiques sont des réseaux neuronaux pré-entraînés à l'aide de l'apprentissage auto-supervisé sur d'énormes quantités de texte converties en séquences de jetons numériques. La fonction objectif du pré-entraînement consiste à prédire des tokens masqués (dans les modèles GPT comme GPT-4, à prédire le token suivant dans une séquence). Ils sont ensuite affinés à l'aide de l'apprentissage supervisé et de l'apprentissage par renforcement afin d'exécuter des fonctions plus complexes, par exemple la synthèse de texte ou l'assistance au dialogue (être un bon chatbot), tout en respectant les normes et contraintes comportementales.¹⁰ Depuis mi-2024, les LLM les plus performants sont entraînés davantage à l'aide de l'apprentissage par renforcement afin d'exploiter des ressources computationnelles supplémentaires au « moment de l'inférence » (par exemple, lors de la réponse à une invite). Contrairement aux formes antérieures d'IA, les LLM sont véritablement polyvalents : si les modèles pré-entraînés peuvent être décrits comme une « saisie semi-automatique dopée aux stéroïdes », le réglage fin libère des capacités latentes dans de nombreux domaines différents. Il est important d'être extrêmement clair sur trois points. Premièrement : les LLM constituent un changement radical dans ce qui est réalisable avec l'IA. Ils sont radicalement différents à cet égard des technologies d'IA qui les ont précédés. Deuxièmement : leurs capacités s'améliorent constamment et rapidement. Troisièmement : ils ouvrent la possibilité de créer avec l'IA à tout le monde, et pas seulement aux ingénieurs.

Les démocraties reposent sur la délibération et la prise de décision collective. La délibération peut avoir lieu dans le cadre formel d'une assemblée citoyenne, ou dans le cadre informel d'une conversation au café la veille d'une élection, et dans toutes les situations intermédiaires.¹¹ Le moment clé de la prise de décision collective est celui où nous votons pour élire les titulaires de mandats électifs. Mais il existe de nombreux autres moments de prise de décision collective dans une démocratie qui fonctionne bien, des réunions publiques aux référendums nationaux.

⁵ Khan (2016), Theocharis et al. (2016), Brady et al. (2017) ; Wu (2018, 2019), Cohen et Fung (2021), Reich et al. (2021), Siegel (2022), Susskind (2022), Aytac (2024).

⁶ Muldoon (2022), Fischli (2024).

⁷ Sætra et al. (2022), Himmelreich (2023), Seger et al. (2023), Lin (2024).

⁸ Bernholz (2016), Gastil et Davies (2020), Schuler (2020), Bernholz et al. (2021), Landemore (2021a), Himmelreich (2022), Fischli (2024), Grossi et al. (2024).

⁹ Brill (2019), Lee et al. (2019), Benade et al. (2021), Brill (2021), Flanigan et al. (2021), Ford (2021), Kahng et al. (2021), Landemore (2021a).

¹⁰ Pour un bon aperçu du fonctionnement des LLM, voir Naveed et al. (2023), Millièrre et Buckner (2024). Bien sûr, si les LLM sont actuellement à la pointe de la recherche en IA, d'autres avancées pourraient les supplanter — la validité de notre argumentation dépendra des détails techniques spécifiques.

¹¹ Pour des exemples d'ouvrages traitant des différents espaces de délibération démocratique, voir Mansbridge (1983), Habermas (1991), Schmitt-Beck et Grill (2020).

La délibération et la prise de décision démocratiques impliquent toutes deux, à plusieurs étapes, la production et l'analyse du langage. Les grands modèles linguistiques (LLM) comptent parmi les technologies les plus avancées jamais conçues pour analyser et générer du contenu linguistique.¹² Il est donc naturel de se demander si nos outils les plus performants et les plus adaptables pour manipuler le langage pourraient s'avérer déterminants pour l'un des usages les plus importants que nous en faisons. De plus, les grands modèles linguistiques étant si performants et si faciles à déployer, ces deux dernières années, les ingénieurs en IA, les informaticiens et même les amateurs ont cherché à créer des applications utilisant ces modèles pour pratiquement toutes les fonctions imaginables. Nous nous trouvons donc à un moment rare de l'histoire, où l'une des tâches les plus productives pour les philosophes et théoriciens politiques n'est pas seulement d'évaluer l'impact de la technologie sur la démocratie avec le regard d'un simple spectateur, mais de *construire* réellement des technologies démocratiques. En complément, nous assistons à l'émergence d'un nouveau sous-domaine — que l'on pourrait appeler, par analogie avec le choix social computationnel, la théorie politique computationnelle.

Nous avons rassemblé et examiné cette littérature de plus en plus abondante sur l'utilisation des grands modèles de langage (LLM) pour améliorer la délibération et la prise de décision démocratiques, en nous concentrant sur les articles publiés au cours de l'année qui a suivi la sortie de GPT-4. À partir de cette littérature, nous avons identifié quatre rôles distincts qui ont été confiés aux LLM.

Premièrement, les chercheurs et les praticiens se sont demandé si les LLM pouvaient soutenir la politique démocratique en tirant parti de leurs capacités *de synthèse*. Deuxièmement, ils ont exploré comment ces modèles pouvaient soutenir la prise de décision démocratique en *agrégant* les opinions sur différentes options, afin de trouver un consensus ou d'optimiser certains indicateurs de soutien collectif. Troisièmement, certains projets utilisent les LLM pour *représenter* des individus dans une procédure décisionnelle démocratique plus poussée (en général, en prédisant leurs jugements sur des propositions politiques). Enfin, quatrièmement, certains ont également cherché à savoir si les LLM peuvent *faciliter* la délibération. Nous allons maintenant détailler chacune de ces approches tour à tour. Enfin, nous évaluons chaque approche pour voir dans quelle mesure elle concrétise les valeurs démocratiques identifiées dans la section précédente.

3.1 Résumé

La perspective d'utiliser les grands modèles de langage (LLM) pour résumer les contributions du public aux consultations gouvernementales a été davantage mise en avant par les ministres que par les chercheurs, et bien que plusieurs projets en cours cherchent à déterminer si et comment cela peut fonctionner (Small et al., 2023 ; Straub et al., 2024), peu de résultats ont été publiés à ce jour.¹³ La théorie de base est toutefois très simple : la consultation publique est un élément clé du processus décisionnel démocratique ; pour que les consultations soient inclusives, elles impliquent la production d'une quantité considérable de texte ; les gouvernements démocratiques doivent donc traiter ce texte et en extraire des informations. C'est une tâche extrêmement laborieuse ; l'analyse de ces documents par les fonctionnaires est également (vraisemblablement) imparfaite et biaisée, simplement en raison de notre incapacité à

¹² Aucune autre technique ne se rapproche de la fluidité dont font preuve les grands modèles de langage (LLM) dans la génération de contenu linguistique. Et bien que les méthodes statistiques d'analyse linguistique soient, bien sûr, d'une valeur inestimable, les LLM font preuve d'une compréhension utile de la structure linguistique qui leur confère le potentiel de faire progresser le domaine de la linguistique (Beguš et al., 2023 ; Millière, 2024).

¹³ Pour une analyse très précoce de cette approche, antérieure à l'IA, voir Schlosberg et al. (2008). Johannes Himmelreich (2022) évoque également la possibilité d'utiliser l'IA pour « identifier les contributions des citoyens lors des consultations publiques comme étant novatrices, détaillées ou pertinentes d'une autre manière ».

traiter de manière fiable de très grandes quantités d'informations (Bennett & Cutler, 2019). Les modèles de langage à grande échelle (LLM) ont été entraînés pour la tâche de synthèse (Ouyang et al., 2022) et obtiennent généralement de bons résultats dans ce domaine (Zhang et al., 2024). Il semble donc évident que nous pouvons améliorer la réactivité des pouvoirs publics face aux consultations publiques en extrayant des informations pertinentes à l'aide des LLM.

Cette fonctionnalité de base des grands modèles de langage (LLM) a été exploitée de manière plus créative dans le cadre du projet « Recursive Public », qui s'appuie sur vTaiwan, le dispositif visant à intégrer la participation citoyenne dans le système politique taïwanais, dirigé par Audrey Tang, qui occupait jusqu'à récemment le poste de ministre du Numérique à Taïwan (Devine et al., 2023). Le projet utilise Polis — une plateforme en ligne qui permet aux utilisateurs de rédiger de brèves déclarations et d'indiquer (de manière anonyme) leur soutien aux déclarations faites par d'autres — pour recueillir les contributions des participants afin de définir l'ordre du jour des discussions politiques concernant l'IA. Ils ont ensuite utilisé l'IA de deux manières pour résumer ces discussions : tout d'abord, ils ont utilisé un LLM pour résumer les arguments et les points clés de la discussion, et pour regrouper les arguments et les thèmes similaires ; puis ils ont utilisé un autre outil (« Talk to the City ») pour convertir ces données en un « espace bidimensionnel où les arguments sémantiquement similaires sont positionnés les uns à proximité des autres » (Devine et al., 2023 : 20). Le résultat final de cet exercice de définition d'ordre du jour délibératif a été une liste de priorités concernant l'IA. Ils ont ensuite mené une délibération approfondie sur l'une de ces priorités en utilisant le même processus de délibération et de synthèse par l'IA, le résultat final étant un ensemble de lignes directrices pour la politique sur ce sujet.

Le recours aux grands modèles de langage (LLM) pour résumer les réponses aux consultations publiques, les délibérations longues et complexes, voire les débats politiques de manière plus générale, s'inscrit parfaitement dans la logique des biens non instrumentaux de la démocratie, et pourrait même favoriser la participation et l'adhésion, car il promet d'alléger les contraintes liées à la participation et pourrait ainsi contribuer à renforcer la responsabilité. Les technologies de communication modernes ont rendu extrêmement facile pour les institutions politiques de solliciter la contribution du public, mais le volume même de ces contributions est difficile à traiter efficacement. De meilleurs moyens à cette fin constituent clairement un atout. Et la même analyse de base s'applique aux biens instrumentaux de la démocratie : un traitement plus efficace de l'information devrait favoriser la conciliation, ainsi que la transformation des préférences et, en fin de compte, l'émergence d'une plus grande variété d'opinions, soutenant ainsi l'intelligence collective. Cela tient simplement au fait que chacun de ces biens est favorisé par la création de meilleurs systèmes de collecte et d'interprétation des contributions du public aux consultations. Si vous estimez avoir eu une réelle opportunité de contribuer à l'élaboration des politiques publiques, et si vous pouvez comprendre les raisons avancées par ceux qui ont pris le contre-pied d'un débat, cela devrait vous rendre plus enclin à accepter un résultat qui ne vous est pas favorable. Si l'« opinion publique » est correctement enregistrée et traitée, nous devrions être mieux à même de discerner ce qu'est la « volonté générale ». Et si nous disposons de moyens efficaces et performants pour traiter ces contributions du public, nous pouvons mieux identifier et exploiter la sagesse de la foule.

Toutefois, tout ce qui précède n'est plausible que si nous avons la certitude que le processus de synthèse n'entraîne pas de pertes excessives, qu'il n'est pas biaisé idéologiquement et qu'il n'est pas, en fait, fantaisiste. Si, au contraire, l'utilisation des LLM pour résumer des contributions collectives, des délibérations ou des conversations individuelles avec un chatbot n'est pas suffisamment ancrée dans le matériel source, cela pose clairement un problème : cela pourrait nuire à l'égalité dans la mesure où certaines voix pourraient être amplifiées au détriment d'autres ; cela menace la légitimité si des détails clés sont omis ou inventés de toutes pièces ; la vertu civique n'est guère renforcée par la participation

à une mascarade ; et des résumés inexacts ne peuvent pas favoriser la conciliation (pourquoi devrions-nous accepter le résultat d'un processus vicié ?), la transformation des préférences ou l'intelligence collective (sauf par accident).¹⁴

Le problème est que les résultats des grands modèles de langage (LLM) en matière de synthèse sont pour l'instant assez mitigés. À l'heure actuelle, ils excellent dans certaines tâches, comme la synthèse d'articles d'actualité courts (Zhang et al., 2024). Cependant, leurs performances sont moins bonnes lorsqu'il s'agit de résumer des textes longs. Ils ont tous deux tendance à inventer des éléments (Kim et al., 2024) et ont du mal à intégrer des informations dispersées dans la source d'entrée (Cheng-Ping et al., 2024). Par exemple, des chercheurs ont utilisé des LLM pour tenter de résumer et d'analyser un appel à commentaires publics de la NTIA, et ont constaté que les LLM introduisaient souvent des changements de sens subtils mais significatifs (Boulos et al., 2023). Les LLM présentent également actuellement des biais politiques bien documentés, susceptibles d'interférer avec une synthèse qui devrait être politiquement neutre (Liu et al., 2022 ; Rettenberger et al., 2025 ; Santurkar et al., 2023 ; Taubenfeld et al., 2024). D'après des travaux sur la « monoculture générative », certains éléments suggèrent que les résumés des LLM parviennent trop souvent à bien résumer les contenus les plus fréquents mais les moins significatifs d'un ensemble de données, mais qu'ils sont bien moins performants lorsqu'il s'agit de mettre en évidence des contenus peu fréquents mais très significatifs (Wu et al., 2024). En d'autres termes, lorsqu'ils résumant des textes, ils ont du mal à repérer les valeurs aberrantes dans les données qui pourraient véhiculer des informations potentiellement importantes. Cela pourrait se traduire par le fait de ne pas remarquer les quelques idées brillantes issues de la consultation publique, ce qui porterait atteinte à l'intelligence collective. Ou bien cela pourrait signifier ignorer la minorité de voix exprimant de profondes inquiétudes quant à la manière dont une politique proposée leur portera gravement préjudice, sapant ainsi l'égalité et la conciliation.

De plus, il y a peut-être lieu de penser que les grands modèles linguistiques (LLM) présentent des lacunes épistémiques plus générales qui pourraient nuire à leur capacité à bien résumer. Coeckelbergh (2025) met en évidence certaines considérations préoccupantes — parmi lesquelles, pour notre propos, l'affirmation selon laquelle les LLM produisent *des absurdités* (Coeckelbergh, 2025, p. 6 ; Hicks et al., 2024). Le problème ici est que les LLM ne se « soucient » pas de la vérité, et que leurs résultats s'apparentent donc davantage à de la sophistique ou à de la rhétorique qu'à des affirmations que nous devrions prendre au sérieux sur le plan épistémique. Dans le cadre de notre étude, la question la plus importante concerne la mesure dans laquelle les LLM peuvent être solidement orientés vers la vérité. Nous sommes optimistes quant aux améliorations apportées en permanence aux LLM en matière d'hallucinations sur ce point (Ji et al., 2023 ; McDonald et al., 2024).¹⁵ Nous notons également que le fait que les LLM se « soucient » littéralement de la vérité — au sens où ils éprouvent une orientation subjective ou un état affectif particulier — n'a peut-être pas d'importance s'ils manifestent un équivalent fonctionnel de ce souci. Les protocoles sur lesquels les modèles sont entraînés, la méthode par laquelle ils sont formés à assimiler ces protocoles, et leurs améliorations continues dans la fourniture

¹⁴ On pourrait penser que l'hallucination est un problème plus structurel, inhérent à tous les grands modèles de langage (LLM), et que, par conséquent, la perspective de disposer de LLM moins sujets aux hallucinations, voire exempts de celles-ci, est moins probable que ce que nous laissons entendre. Par exemple, Coeckelbergh (2025) décrit les LLM comme « conçus pour produire un texte convaincant – sans avoir pour objectif de produire la vérité » (Coeckelbergh, 2025, p. 5). Bien sûr, s'il existe un biais hallucatoire robuste et significatif dans les LLM, alors il va de soi qu'ils doivent être utilisés avec prudence en tant que synthétiseurs dans tout processus démocratique. À l'heure actuelle, cependant, les progrès réalisés par les informaticiens dans les méthodes visant à réduire les hallucinations des LLM nous amènent à penser que (du moins pour l'instant) les preuves empiriques ne corroborent pas une telle conclusion (Ji et al., 2023 ; McDonald et al., 2024). Nous sommes redevables à un évaluateur anonyme de nous avoir signalé cette critique.

¹⁵ Nous remercions un évaluateur anonyme de nous avoir signalé cette littérature.

Les réponses fondées aux requêtes des utilisateurs suggèrent qu'ils sont sur la bonne voie (Anthropic, 2025 ; Guan et al., 2025 ; Kenthapadi et al., 2024 ; Lin et al., 2022 ; OpenAI, 2025).

Bien sûr, les humains qui résumant actuellement les réponses aux consultations publiques ont eux aussi de nombreux biais cognitifs et moraux. Et les grands modèles linguistiques (LLM) n'ont pas besoin d'être parfaits pour constituer une amélioration. Nous ne prétendons pas ici que les LLM *ne pourraient jamais* être des résumeurs suffisamment compétents pour que cette fonction apporte une contribution précieuse à la prise de décision démocratique. Mais nous devrions attendre d'un outil informatique de traitement de l'information démocratique qu'il réponde à des normes plus élevées que celles exigées des humains. Les synthétiseurs humains peuvent être tenus responsables de leurs erreurs, et leurs biais sont susceptibles de s'annuler mutuellement au fil d'un nombre suffisant d'essais, en raison de la capacité limitée de chaque individu à traiter l'information. Les LLM utilisés pour la synthèse ne peuvent évidemment pas être tenus responsables ; de plus, ils peuvent être appliqués à un nombre de cas bien plus vaste, et leurs biais peuvent donc se propager à une échelle bien plus grande (Bommasani et al., 2022). De plus, notre capacité à intervenir directement sur les capacités de synthèse des LLM devrait nous permettre de viser plus haut.

De plus, le problème plus profond réside dans le fait que nous ne mesurons pas actuellement les bons indicateurs pour déterminer si les grands modèles de langage (LLM) peuvent jouer ce rôle de traitement de l'information dans la démocratie.¹⁶ Les projets évoqués ci-dessus ont trop tendance à prendre leurs résumés pour argent comptant — ou bien à se contenter de s'appuyer sur une validation a posteriori de ces résumés par les participants. Si les outils de traitement de l'information doivent soutenir la délibération et la prise de décision démocratiques, nous devons exiger des normes élevées, en partie parce que, sinon, la fonction de conciliation de la démocratie s'en trouve compromise. Tout résumé donné est susceptible de favoriser davantage l'un ou l'autre camp d'un ensemble d'intérêts concurrents. Si la véracité du résumé peut être facilement contestée, alors le camp défavorisé dispose d'une raison toute faite pour ne pas accepter le poids de cette preuve apparente à son encontre.¹⁷

¹⁶ Par exemple, les indicateurs de référence en matière de synthèse, tels que ROGUE et BLEU, mesurent le chevauchement des mots, des séquences et des symboles (n-grammes) afin de fournir une mesure de la qualité de la synthèse (Davoodijam et Mohsen, 2024). Il est probablement difficile pour ROGUE de faire la distinction entre une proposition donnée et sa négation — c'est-à-dire que le fait de supprimer un « ne » modifie considérablement le sens de la phrase, mais cela n'affecte peut-être pas la manière dont ROGUE évalue les modèles sur leurs résumés. Il faudrait donc peut-être explorer la possibilité d'un indicateur extrinsèque propre à l'utilisation des résumés dans des contextes démocratiques (Davoodijam et Mohsen, 2024). Quoi qu'il en soit, nous devons être conscients des métriques utilisées pour mesurer l'efficacité des différents LLM en matière de résumé lorsque nous choisissons un type de modèle donné pour résumer dans des contextes démocratiques. Nous remercions un évaluateur anonyme de nous l'avoir signalé.

¹⁷ De plus, même après avoir décidé d'utiliser un LLM pour la synthèse dans le cadre d'un processus décisionnel démocratique, il convient également de prêter attention au *type* de LLM employé. Par exemple, GPT diffère considérablement des modèles basés sur BERT. Bien que les deux soient des transformateurs, une différence majeure réside dans le fait que BERT analyse le texte *de manière bidirectionnelle*, tandis que GPT ne le fait que de droite à gauche. Cela rend BERT, à première vue, plus performant pour préserver le contexte et pour la synthèse *extractive* — c'est-à-dire la synthèse qui vise globalement à extraire des mots et des phrases à partir de données linguistiques qui résumant de manière représentative les données (Batra et al., 2021 ; Bharathi Mohan et al., 2023 ; Darapaneni et al., 2023). D'autres LLM comme PEGASUS, par exemple, sont plus performants en matière de synthèse *abstraite* (Abdul Samad et al., 2024). Il est également possible que les modèles basés sur GPT — puisqu'ils sont plus souvent utilisés pour générer du texte — deviennent plus performants que les modèles basés sur BERT pour la synthèse plus abstraite qui nécessite des capacités de synthèse et de génération (Darapaneni et al., 2023). En effet, de nombreux choix architecturaux peuvent être faits lors du développement d'un outil de synthèse ; pour savoir quels choix servent le mieux les valeurs démocratiques, nous devons commencer par réfléchir à ce que sont précisément les critères démocratiques d'une bonne synthèse. Nous sommes redevables à un évaluateur anonyme de nous avoir poussés à approfondir ce point.

3.2 Agrégation

La synthèse implique une sélection et une agrégation : les points de vue similaires sont regroupés. À la limite, la synthèse est une procédure décisionnelle par laquelle un éventail diversifié d'entrées est converti en une seule sortie. Plusieurs projets de recherche récents ont exploré l'utilisation des LLM dans la délibération afin d'identifier un ensemble plus ou moins large de déclarations qui, dans une certaine mesure, représentent les opinions des délibérateurs. Dans ces exemples, le LLM joue, par rapport aux descriptions textuelles des préférences et des valeurs, le rôle qu'une fonction d'agrégation joue par rapport aux préférences représentées numériquement dans la théorie du choix social. Il convient toutefois de noter que, là où la théorie du choix social ne peut déterminer que le soutien à une liste fixe de positions politiques sur lesquelles les participants ont exprimé leurs préférences, les LLM peuvent lever cette contrainte, nous permettant non seulement d'identifier les déclarations qui bénéficient d'un soutien optimal selon une certaine mesure, mais aussi de générer de nouvelles déclarations qui seront encore plus largement approuvées.

C'est dans l'ouvrage « Generative Social Choice » (Fish et al., 2025) que cette question est étudiée de la manière la plus rigoureuse.¹⁸ En se concentrant sur les préférences des participants concernant les normes de personnalisation des chatbots, les auteurs commencent par recueillir auprès d'eux des déclarations sur ce que devraient être ces normes. Ce sont là les options. Ils recueillent également les préférences des participants concernant ces options. Le choix social classique s'arrêterait là et agrégerait, selon une certaine fonction, leurs préférences concernant ces options afin d'identifier celles qui, le cas échéant, seraient socialement optimales. Le choix social génératif va plus loin. Tout d'abord, il recueille des informations supplémentaires auprès des participants sur leurs valeurs et leurs opinions. Ensuite, elle utilise toutes ces données — déclarations, préférences et données supplémentaires — comme matière première pour inciter GPT-4 à proposer de *nouvelles* déclarations qui, selon ses prédictions, recueilleront plus de soutien que celles qui existent déjà, et en outre « agiront comme un substitut du participant, en prédisant ses préférences concernant toute alternative, qu'elle soit prévue ou nouvellement générée » (Fish et al., 2025 : 2). Les auteurs montrent que leur approche peut être utilisée pour générer de nouvelles déclarations qui répondent mieux aux critères quantitatifs de représentativité que les déclarations existantes.¹⁹ Ils valident également les prédictions des préférences des personnes en exposant les participants à la liste des déclarations générées a posteriori, constatant que « 84 % des participants déclarent que la déclaration qui leur a été attribuée reflète leur opinion sur la politique en matière d'avortement de manière “excellente” ou “exceptionnelle” » (Fish et al., 2025 : 4).

Alors que le projet « Generative Social Choice » utilise des modèles de langage de grande envergure (LLM) pour aider à trouver un consensus parmi des participants dissociés qui n'interagissent entre eux que de manière asynchrone, le projet « Demo-cratic Policy Development Using Collective Dialogues and AI » (Konya et al., 2023) fait appel à une délibération en direct, mais utilise finalement des LLM dans un but similaire. Dans le cadre de ce projet, les chercheurs ont invité les participants à lire d'abord un guide de discussion préparé par l'équipe d'animation, puis à échanger entre eux sur Remesh, une autre plateforme de délibération en ligne textuelle. Ils ont également formulé des déclarations, proposé des politiques et évalué les déclarations et opinions des uns et des autres par le biais d'un classement par paires et d'un système binaire

¹⁸ Pour d'autres exemples connexes, voir Bakker et al. (2022) et, plus récemment, Tessler et al. (2024).

¹⁹ Leur principal critère de représentation, la « représentation équilibrée et justifiée », est décrit de manière informelle comme suit : « [S]'il existe une coalition d'agents qui est (i) suffisamment importante pour « mériter » une déclaration sur la liste par proportionnalité et (ii) qui a des préférences cohérentes (c'est-à-dire qu'il existe une déclaration pour laquelle tous ces agents ont une utilité d'au moins 0), alors (iii) la coalition ne doit pas être « ignorée » » (Fish et al., 2025 : 6).

vote par accord/désaccord. Ces classements, établis sur le sous-ensemble de déclarations que l'utilisateur voit, sont ensuite utilisés pour extrapoler ses préférences à l'ensemble des autres déclarations. Le LLM est alors sollicité pour créer un résumé des déclarations présentant le moins de désaccord (« réponses de rapprochement ») ; ce résumé sert ensuite de base à un nouvel ensemble de propositions de politiques, qui s'appuient également sur la liste des réponses de rapprochement et sur deux autres exemples. Les politiques qui en résultent se voient ensuite attribuer des « scores de justification », reflétant leur degré de proximité avec les réponses de rapprochement, grâce à la combinaison d'une mesure de similarité sémantique et de la notation quantitative de similarité propre au LLM. Une fois encore, le rôle du LLM n'est pas seulement de résumer et de sélectionner parmi les déclarations effectivement produites par les participants, mais d'utiliser ces déclarations (et d'autres données issues de la délibération) comme données d'entrée pour générer des déclarations qui (selon certains critères) sont préférables à celles que les participants ont effectivement proposées.²⁰

Pourquoi ne pas utiliser les grands modèles de langage (LLM) pour synthétiser les avis, à la fois pour évaluer le niveau de soutien et pour générer de nouvelles propositions susceptibles de faciliter la recherche d'un consensus ? Là encore, cela semble être une arme à double tranchant. La capacité à faciliter les discussions sur l'élaboration des politiques et à identifier des terrains d'entente potentiels peut clairement contribuer à la conciliation, en nous aidant à trouver des points de compromis là où il n'y en avait pas auparavant. Mais deux préoccupations se font jour.

Tout d'abord, le débat public sur les politiques est-il réellement freiné par un éventail incomplet d'alternatives disponibles ? Il est encourageant de constater que ces expériences montrent que des modèles peuvent produire des énoncés suscitant davantage d'adhésion que le corpus existant, dans ces contextes contrôlés spécifiques, mais nous ne pouvons pas déterminer si ces énoncés constituent une avancée significative *sur le fond* simplement parce qu'ils ont été approuvés (peut-être sont-ils simplement assez vagues pour masquer les désaccords). De plus, l'adhésion qu'elles suscitent est probablement liée au cadre étroitement prosocial et apolitique des conversations, ainsi qu'aux faibles enjeux du débat politique en jeu. Si le débat politique était plus controversé et les enjeux plus importants, et si les participants n'étaient pas motivés de manière indépendante par le cadre expérimental à faire preuve d'ouverture d'esprit et à être disposés à faire des compromis, nous soupçonnons que ces déclarations générées seraient implicitement triées sur le plan idéologique (Bullock & Lenz, 2019) beaucoup plus rapidement, et seraient moins pertinentes pour parvenir à un compromis entre des intérêts concurrents.

En d'autres termes, dans un contexte politique réaliste où le débat est suffisamment ouvert, on peut raisonnablement s'attendre à ce qu'un éventail suffisamment large de propositions politiques soit formulé. La politique démocratique n'est probablement pas freinée par un manque d'imagination. La capacité des propositions de compromis à rallier des soutiens dépendra sans doute de faits stratégiques liés à la nature des intérêts concurrents, dont toutes ces expériences font abstraction. Par conséquent, l'utilisation de modèles de langage grand public (LLM) pour agréger des jugements et suggérer des propositions politiques novatrices ne devrait guère contribuer à améliorer la prise de décision et la délibération démocratiques dans le monde réel.

²⁰ Contrairement à la « choix social génératif », cependant, ce n'était pas l'étape finale. Ensuite, l'équipe d'animation sélectionne un sous-ensemble supplémentaire de réponses politiques pour former une liste initiale de politiques, qui est ensuite élaborée par des experts choisis par l'équipe d'animation, afin d'affiner les politiques pour plus de clarté et de prendre en compte les cas limites. Les contributions des participants sont ensuite utilisées pour affiner les politiques en fournissant des arguments pour et contre le soutien aux politiques données de la liste. Les politiques sont ensuite affinées à la lumière de ces retours d'information afin de les rendre plus représentatives. Enfin, une troisième phase est organisée afin que le public puisse délibérer et voter sur chaque politique et sur la liste dans son ensemble (Konya et al., 2023).

Deuxièmement, il existe une grande différence entre le choix social *génératif* et le choix social computationnel en général. Lorsque les théoriciens du choix social computationnel appliquent une fonction d'agrégation à un ensemble de préférences numériques, ils utilisent un algorithme transparent et mathématiquement vérifiable pour parvenir à une décision (Brandt et al., 2016). Il est évident que la représentation numérique des préférences est en soi une forme d'artifice, mais la procédure de décision est reproductible et peut faire l'objet d'une inspection et d'une remise en question. Lorsque le LLM produit une déclaration consensuelle basée sur son interprétation des déclarations existantes et des préférences des personnes — et lorsque le LLM est utilisé pour évaluer la similitude entre les déclarations —, aucun algorithme d'agrégation explicite et déterministe n'est appliqué ; la décision est opaque et probablement non reproductible.²¹ Il s'agit bien sûr d'un problème ancien et bien connu des outils de prise de décision basés sur l'apprentissage profond (Burrell, 2016 ; Lazar, 2024a ; Selbst & Barocas, 2018 ; Vredenburg, 2022). Mais c'est un problème particulièrement aigu ici en ce qui concerne la valeur de la conciliation.

Supposons que vous fassiez partie de la minorité défavorisée par le verdict rendu par le LLM. Quelle raison auriez-vous d'accepter ce verdict ? Avec la théorie computationnelle du choix social, nous pouvons au moins montrer la fonction qui nous mène des préférences de chacun au résultat. Et cela est encore plus simple dans le cas d'un scrutin majoritaire. Mais lorsque le résultat est déterminé par un LLM, nous ne pouvons actuellement ni reproduire ni examiner la fonction décisionnelle. En effet, nous ne pouvons même pas garantir qu'il n'a pas été affecté de manière significative par une influence exogène. Nous ne pouvons offrir aucune garantie de sécurité, ni d'absence de biais propre au LLM.²² Dans des contextes décisionnels où les enjeux sont plus importants, il y aurait une incitation évidente à manipuler le LLM médiateur en utilisant des techniques adversaires telles que l'injection de prompts et le jailbreaking (Greshake et al., 2023 ; Shah et al., 2023). Les auteurs de « Generative Social Choice » anticipent quelque peu cette préoccupation en demandant aux participants de ratifier les reformulations par le modèle de leurs opinions et de celles des autres, mais on pourrait à nouveau penser que si les enjeux étaient plus importants, les gens seraient plus exigeants (et de toute façon, 16 % des personnes ont constaté une divergence significative entre les déclarations générées et les données d'entrée qui y ont conduit) ; de plus, il pourrait exister une autre série d'énoncés qu'ils approuveraient davantage.

3.3 Représentation

Plusieurs articles s'appuient, à un moment ou à un autre, sur l'utilisation de modèles de langage de grande envergure (LLM) pour *représenter* des individus en prédisant leurs préférences face à des options politiques inconnues. Cela inclut, de manière indirecte, la « choix social génératif » et l'« élaboration démocratique des politiques ». Mais cette approche occupe une place bien plus centrale dans d'autres travaux, qui examinent explicitement la possibilité de former des LLM pour qu'ils agissent en tant que représentants, c'est-à-dire en tant qu'agents mandataires agissant au nom d'un mandant dans un processus démocratique. Dans « LLMs as Agents for Augmented Democracy », les auteurs tentent de former des LLM à prédire les préférences politiques d'un individu (à l'aide de données d'enquête issues des élections brésiliennes de 2022), dans l'optique que ces LLM se substituent ensuite à la personne qu'ils représentent dans une procédure de prise de décision collective (Gudiño-Rosero et al., 2024). Le

²¹ Une graine aléatoire et une température différentes conduiraient probablement à des résultats très différents, par exemple.

²² Les progrès en matière de sécurité et d'interprétabilité mécaniste des LLM pourraient atténuer ces préoccupations à l'avenir. Mais, du moins pour l'instant, nous pensons qu'elles sont importantes.

La même approche est adoptée dans « Language agents as digital representatives in collective decision-making » (Jarrett et al., 2025). Alors que les autres exemples de cette catégorie impliquent des LLM facilitant l'agrégation des jugements, cette approche utilise le LLM pour représenter les contributions des agents individuels dans la procédure de décision finale.

Passons maintenant à l'idée d'utiliser les grands modèles de langage (LLM) comme représentants des électeurs individuels.²³ L'une des préoccupations concerne les performances : à l'heure actuelle, il semble impossible de concevoir des représentants basés sur des LLM capables de prédire et de suivre de manière fiable les jugements d'un mandant. Mais mettons cette préoccupation de côté pour l'instant, et essayons d'évaluer l'argument démocratique en faveur de ce type d'agents IA en tant que mandataires/représentants, s'ils fonctionnent bien. Nos approches actuelles de la représentation politique impliquent déjà un problème mandant-mandataire : les politiciens font souvent un mauvais travail lorsqu'il s'agit de représenter réellement ceux qui les élisent (Besley, 2006). Si nous imaginons des agents IA selon les lignes envisagées dans (Gudiño-Rosero et al., 2024 ; Jarrett et al., 2025), nous pourrions certainement les amener à représenter plus fidèlement les préférences de leur mandant ; si nous parvenions ensuite à élaborer une règle de décision équitable pour agréger les jugements de ces agents, nous pourrions potentiellement disposer d'une forme de politique démocratique présentant certaines des vertus de la démocratie participative, sans que cela n'exige plus d'efforts de la part des citoyens, voire moins, que la démocratie représentative. À tout le moins, l'expérience mérite d'être menée.

Mais nous devons avancer en gardant les yeux grands ouverts. Toute utilisation d'agents IA pour représenter les électeurs doit s'accompagner d'une règle de décision quelconque, et comme nous venons de le voir, il est difficile d'imaginer que ceux qui sont lésés par ces décisions puissent raisonnablement être amenés à les accepter. Cette approche sape également de manière évidente les raisons *participatives* qui justifient la valeur de la démocratie. On ne peut pas développer la vertu civique en déléguant la politique à un agent IA. Elle peut aussi compromettre la valeur démocratique de la *transformation des préférences* et de la formation collective des préférences. Les projets qui s'appuient largement sur l'entraînement de modèles de langage de grande capacité (LLM) pour représenter les participants (Fish et al., 2025 ; Gudiño-Rosero et al., 2024 ; Jarrett et al., 2025 ; des projets plus délibératifs comme Recursive Public pourraient éviter ces objections) échouent sur les deux points suivants.

Premièrement, elles considèrent les préférences des participants comme fondamentalement immuables. Elles extrapolent à partir des préférences exprimées lors d'une enquête initiale, sans leur donner la possibilité de revoir leur position. Ce problème pourrait être résolu en adoptant une conception en plusieurs étapes. Le deuxième problème est plus fondamental. Ces approches ne permettent pas aux participants de modifier leurs préférences à la suite d'échanges avec des personnes avec lesquelles ils sont en désaccord. Ce type d'interaction est précieux non seulement parce qu'il aboutit à des résultats qui reflètent les préférences mieux mûries des participants, mais aussi parce que le fait d'échanger directement avec l'autre, de voir la question litigieuse de son point de vue et d'ajuster ses préférences à travers un processus associé à cet échange est précieux en soi. Elle sert à la fois la valeur de la conciliation et celle de la transformation des préférences, en particulier dans la mesure où elle conduit à la formation ou à l'identification d'une volonté générale (Rousseau, 1762 ; Young, 2002).

²³ Voir à ce sujet Lechterman (2024), qui examine la possibilité hypothétique de déployer des agents IA dans le rôle d'élus, chargés de négocier avec d'autres responsables pour parvenir à des compromis législatifs. Voir également Burgess (2022), qui adopte une vision plus favorable des agents IA en tant que politiciens.

Nous sommes également sceptiques quant à la question de savoir si la délégation à des agents mandataires préserverait les avantages épistémiques de la démocratie — par exemple, comment le théorème du jury de Condorcet serait-il affecté par le remplacement des citoyens par des agents de modèles linguistiques ? Auraient-ils vraiment la même diversité cognitive que les personnes qu'ils sont censés représenter (Landemore, 2012b, 2021b) ?

En effet, on peut soutenir qu'une délibération artificielle entre agents d'IA omettrait certains des mécanismes clés grâce auxquels la politique démocratique aboutit à des décisions épistémiquement meilleures. Par exemple, notre processus politique humain, fondé sur la confrontation, aide à faire émerger des informations pertinentes pour la prise de décision collective qui, sans cela, resteraient enfouies. Et ce n'est que parce qu'ils participent activement à la vie publique que les visionnaires politiques proposent des mesures novatrices qui ne se contentent pas de faire le grand écart entre les convictions existantes, mais offrent des solutions véritablement innovantes à des problèmes pratiques urgents. En d'autres termes, les procédures démocratiques (1) mettent en commun des connaissances qui sont normalement dispersées, (2) forcent différents types de connaissances à interagir entre eux, ce qui conduit à des corrections, à la mise en lumière de nouvelles raisons et propositions, etc. (Landemore, 2012a), et (3) offrent des opportunités de leadership créatif qui dépassent ce qui est réalisable en se contentant d'agir sur des préférences préexistantes. Nous pourrions peut-être concevoir un système d'agents d'IA représentatifs qui présente également ces vertus, mais cela nécessiterait bien plus que la simple formation d'adjoints compétents ; il faudrait créer une sphère politique virtuelle pour permettre leur interaction (ce qui offre des opportunités passionnantes aux théoriciens de la politique computationnelle).²⁴

3.4 Facilitation

Les méthodes susmentionnées impliquent soit une délibération animée par des modérateurs humains, soit une collecte asynchrone des déclarations et des opinions des participants. D'autres projets ont également cherché à utiliser les grands modèles de langage (LLM) pour *faciliter* le processus de délibération lui-même. L'initiative de Rappler (Mendoza, 2023) consistait à créer un salon de discussion modéré par l'IA (appelé aiDialogue), dans lequel les participants à des groupes de discussion dialoguent avec un LLM qui leur pose des questions, y compris des questions complémentaires, puis génère des résumés à partir de leurs contributions. Ensuite, comme pour les autres projets mentionnés ci-dessus, le LLM est invité à générer des idées de politiques sur la base de la conversation et des résumés, sur lesquelles les participants sont invités à voter individuellement. « Common Ground » (« First Report: Democratic Inputs to AI », 2023) a utilisé des LLM pour amorcer et modérer la délibération synchrone entre les participants, afin de maintenir la discussion sur le sujet et de modérer la conversation.²⁵

« Democratic Fine-Tuning » (Edelman, 2023 ; Edelman & Klingefjord, 2023) adopte une approche similaire, bien que moins sociale. Les participants s'engagent dans une délibération en tête-à-tête avec un chatbot basé sur un LLM au sujet d'un thème donné. Le chatbot les invite à fournir des informations

²⁴ Les architectures multi-agents constituent une piste de développement possible. Les cadres déployant plusieurs agents LLM ont donné des résultats prometteurs sur un certain nombre de mesures (Chan et al., 2023 ; Wu et al., 2023). Cela dit, même les cadres multi-agents utilisant des LLM de pointe restent moins performants que les humains sur des tâches complexes nécessitant une coordination (Abdelnabi et al., 2024). Étant donné que la délibération et d'autres formes de prise de décision démocratique relèvent précisément de ce type de tâches, nous ne pensons pas que les cadres multi-agents offrent actuellement les capacités techniques nécessaires pour incarner les vertus épistémiques susceptibles de justifier la gouvernance démocratique. Mais ils pourraient le faire à l'avenir, auquel cas nous devrions réévaluer notre critique.

²⁵ Voir encore, plus récemment, Tessler et al. (2024).

informations sur les valeurs qu'ils estiment en jeu concernant le sujet donné. Il synthétise ensuite la conversation en quelques valeurs clés que (selon lui) le participant considère comme importantes (notez que cela implique une compression très lossy — la transformation de conversations entières en quelques valeurs caractéristiques).²⁶

Le rôle de facilitation des grands modèles linguistiques (LLM) dans les processus de délibération en ligne semble, dans l'ensemble, beaucoup plus bénéfique (bien que l'on puisse trouver des voix sceptiques, voir Alnemr, 2020 ; Malkin & Alnemr, 2024). D'un autre côté, la disponibilité de modérateurs et d'animateurs humains ne semble pas constituer un obstacle aussi important pour la démocratie délibérative que la réticence des électeurs à participer à des délibérations intensives et chronophages. De plus, nous soupçonnons qu'une grande partie des avantages des procédures démocratiques délibératives réussies repose sur le fait qu'elles se déroulent en personne et impliquent des personnes qui se font face et se voient les unes les autres en tant qu'individus à part entière, envers lesquels elles sont dans une certaine mesure responsables.²⁷ La volonté de parvenir à un consensus significatif ou de progresser vers celui-ci est probablement bien plus forte lors de réunions en personne que lors de réunions virtuelles (Hartz-Karp & Sullivan, 2014). En d'autres termes, les procédures démocratiques efficaces pourraient nécessairement être incarnées. Ainsi, le succès relatif des LLM dans la facilitation de certains types de délibération en ligne pourrait n'être qu'un argument modeste en leur faveur.

4 Comment les LLM peuvent-ils aider les démocraties ?

Les projets à l'étude se concentrent strictement sur la question de savoir si les grands modèles linguistiques (LLM) peuvent être utilisés pour soutenir la conception participative des chatbots et les normes qui les régissent. À juste titre, leurs auteurs ne cherchaient pas à « réparer » la démocratie. Cependant, certains pensent clairement que les LLM ont le potentiel de transformer et de redynamiser considérablement les institutions démocratiques, et la critique qui précède devrait inciter à la prudence (Landemore, 2023 ; Ovadya, 2023). Nous souhaitons en particulier mettre en avant trois grandes leçons concernant la question de savoir si et comment les LLM peuvent soutenir les objectifs instrumentaux de la démocratie, qui, selon nous, peuvent offrir un guide pratique instructif à ceux qui seraient tentés de s'appuyer sur ces premiers résultats de manière plus ambitieuse.

Les démocraties visent à atteindre de nombreux objectifs différents, mais le plus important d'entre eux est peut-être de permettre à des personnes aux valeurs souvent radicalement opposées et aux intérêts directement concurrents de cohabiter dans une relative harmonie. Nous craignons que les LLM — du moins dans leur état actuel — ne parviennent pas à concilier correctement les intérêts et les valeurs concurrents. Les procédures mises en place à cette fin doivent être équitables, simples, sûres, transparentes et contestables. Le vote à la majorité, assorti de certaines contraintes constitutionnelles visant à protéger les minorités persistantes et les droits fondamentaux, répond à ces critères. Toute tentative d'utiliser l'IA pour soutenir des procédures équitables de résolution des revendications concurrentes doit considérer ces critères comme des contraintes. Les LLM en particulier, et les réseaux neuronaux profonds en général, ne peuvent à l'heure actuelle satisfaire à ces critères.

²⁶ À partir de là, cette approche met en œuvre une théorie éthique selon laquelle notre objectif devrait être d'agir conformément aux valeurs les plus « sages », une valeur A étant plus sage qu'une valeur B si A « contient » B. Les auteurs génèrent ensuite un « graphe moral » illustrant ces relations, qui peut en principe servir à affiner d'autres modèles de langage de grande envergure (LLM).

²⁷ Ce point est controversé ; certains partisans de la délibération en ligne estiment que l'on peut obtenir des résultats aussi bons en ligne qu'en personne.

Malgré certaines tentatives prometteuses visant à utiliser les grands modèles linguistiques (LLM) pour faciliter la délibération, il est peu probable qu'ils contribuent, en fin de compte, de manière significative au processus de transformation des préférences par lequel les communautés politiques visent davantage qu'un simple *modus vivendi* ou un compromis transactionnel, et cherchent à résoudre leurs désaccords en parvenant à quelque chose qui s'apparente à une volonté générale. Les véritables institutions démocratiques doivent fonctionner dans un monde déchiré par des inégalités sous-jacentes de pouvoir et de ressources, ainsi que par de profonds désaccords idéologiques, où les citoyens ne peuvent échapper aux pressions et aux contraintes (et en particulier là où des groupes partisans dominent le débat politique), où les décisions affectent très étroitement leurs intérêts matériels et culturels, et où des intérêts et des valeurs concurrents placent les individus dans une relation antagoniste les uns avec les autres (Lovett, 2024). En revanche, chacun des projets évoqués place les participants dans une position privilégiée, semblable à celle d'un jury, où ils s'engagent dans un débat libre, non manipulé et impartial sur un ensemble de problèmes qui, pour la plupart, ne les concernent pas directement. Ils sont aussi proches que possible de la « situation de communication idéale » habermasienne (Habermas, 1984). Dans ces conditions favorables, nous soupçonnons que *toute* méthode de délibération et de prise de décision démocratique paraîtrait attrayante. La délibération assistée par les LLM pourrait bien fonctionner, mais, franchement, la démocratie représentative traditionnelle fonctionnerait tout aussi bien si l'on mettait entre parenthèses tous les éléments controversés de la politique et de l'économie. Dans des conditions réalistes, cependant, nous craignons que la délibération démocratique assistée par les LLM n'aboutisse pas aux avantages que ses partisans revendiquent. De plus, dans la mesure où l'utilisation des LLM pour la facilitation nous oriente vers une version de la délibération exclusivement en ligne, nous pensons qu'elle passe à côté d'une des vertus clés de la délibération, à savoir qu'en réunissant les gens, face à face, elle suscite une certaine volonté de céder du terrain et de parvenir à un compromis — nous soupçonnons, comme nous l'avons argumenté plus haut, que cela ne se produira de manière solide que par le biais d'une interaction en personne.

Cependant, même si nous mettons en garde contre l'utilisation des grands modèles de langage (LLM) pour remplacer des procédures décisionnelles équitables ou pour faciliter la délibération démocratique, nous pensons qu'ils peuvent néanmoins apporter une contribution importante au renouveau de la démocratie. Si nous ne pensons pas que l'intégration des LLM dans la sphère publique formelle de la prise de décision exécutive soit susceptible de faire progresser les objectifs épistémiques de la démocratie, nous pensons *en revanche* que le déploiement des LLM dans la sphère publique *informelle* peut contribuer à fournir certaines des conditions épistémiques préalables à l'épanouissement des démocraties. La sphère publique informelle est l'environnement du dialogue et de la délibération non structurés entre les membres d'une communauté politique (Habermas, 1998 ; Young, 2002). La sphère publique formelle est l'espace de la délibération et de la prise de décision officiellement sanctionnées, dans lequel des actions institutionnelles concrètes sont effectivement débattues et adoptées.²⁸

Les démocraties saines s'appuient sur une sphère publique informelle saine pour demander des comptes aux puissants et tracer une voie positive pour la société en facilitant la délibération collective informelle (Cohen & Fung, 2021 ; Habermas, 1998 ; Lazar, 2023 ; Young, 2002). Notre infrastructure numérique actuelle pour la sphère publique est défaillante (Cohen & Fung, 2021). À ce jour, les grands modèles de langage (LLM) n'ont fait que nuire à la sphère publique. En voici quelques exemples : La génération de contenu de mauvaise qualité optimisé pour les moteurs de recherche amplifie

²⁸ Il est évident qu'il n'existe pas de frontière nette entre la sphère publique formelle et informelle. Nous préconisons simplement que, s'ils ont le choix entre des interventions qui renforcent les processus informels contribuant à la culture démocratique et celles qui visent à remplacer les institutions clés de la prise de décision et de l'action, les chercheurs en IA devraient privilégier les premières.

le travail des fermes de contenu. Les biais des LLM font que leur utilisation dans la sphère publique menace de marginaliser les points de vue minoritaires ainsi que ceux qui remettent en cause le pouvoir dominant ou au pouvoir (Jungherr & Rauchfleisch, 2025). L'opacité des LLM réduit également notre capacité à *évaluer* la sphère publique — quels points de vue, par exemple, sa structure rend visibles et lesquels elle met de côté (Jungherr & Rauchfleisch, 2025). Enfin, en raison de tactiques telles que le « flooding », l'IA peut être utilisée pour donner une image faussée de qui ou combien de personnes partagent un point de vue donné, fournissant ainsi des informations trompeuses aux « politizens » (Jungherr & Rauchfleisch, 2025).

D'autre part, les avantages des grands modèles de langage (LLM) pour la sphère publique commencent tout juste à être pleinement explorés.²⁹ Les premiers travaux suggèrent que les LLM de pointe peuvent être utilisés pour améliorer radicalement les approches automatisées de la modération de contenu ; ils se sont également révélés excellents pour réorganiser les flux des réseaux sociaux en fonction de valeurs sociétales plus larges (Bernstein et al., 2023 ; Jia et al., 2023). Les LLM peuvent probablement offrir une nouvelle façon d'utiliser Internet en général, en servant d'alternative aux systèmes de recommandation existants qui contribuent au dysfonctionnement de la société, en nous permettant de naviguer dans la sphère publique numérique de manière à décourager la polarisation et l'extrémisme, et en nous épargnant la surveillance et la manipulation (pour une validation de principe, voir Jia et al., 2023 ; pour une défense morale exhaustive, voir Lazar et al., 2024 ; et pour un aperçu de la littérature technique, voir Lin et al., 2023 ; Vats et al., 2024). Il est évident qu'aucune solution technologique ne va à elle seule redresser la sphère publique, mais une nouvelle technique de filtrage et de recommandation de contenu ouvre de nouvelles possibilités précieuses d'amélioration. Et les capacités de recherche d'informations des LLM, ainsi que leur aptitude à traduire ou à simplifier le texte et la parole, pourraient fournir des ressources démocratiques précieuses aux citoyens qui ont du mal à parler la langue nationale, ou qui ont besoin de conseils personnalisés complexes qu'ils ne peuvent pas obtenir efficacement en se contentant de consulter les sites web du gouvernement. Au-delà du simple filtrage, de la recommandation et de la recherche, les LLM ont donné des résultats prometteurs dans la lutte contre les croyances en des théories du complot grâce à des contre-arguments cohérents et adaptés, ainsi que dans la facilitation d'un dialogue en ligne plus mesuré (Argyle et al., 2023 ; Costello et al., 2024 ; Papaioannou et al., 2023 ; Rosenblum & Muirhead, 2019).

Nous pensons également que certains des forums délibératifs défendus dans ces projets pourraient apporter une contribution supplémentaire utile à la sphère publique. Le forum itératif de délibération de Recursive Public et le projet Democratic Policy Development pourraient offrir aux citoyens démocratiques de nouvelles façons d'interagir avec leurs concitoyens, de découvrir ce qu'ils croient et ce à quoi ils accordent de la valeur, et de forger ensemble leurs préférences et leurs valeurs. À condition que leur rôle se limite à ce cadre informel, sans prétendre offrir une procédure pour régler les désaccords, etc., cela pourrait constituer un ajout utile à la boîte à outils de la société démocratique.

Bien sûr, tout comme dans le résumé ci-dessus, le fait que les démocraties tireraient profit de LLM fonctionnels remplissant ce rôle n'implique pas qu'ils en soient capables *aujourd'hui*. On ne devrait pas faire confiance aux LLM actuels pour ces rôles. Cependant, ce sont des rôles pour lesquels les LLM peuvent être formés et évalués de manière significative. Avec une approche collaborative appropriée — intégrant la science de l'IA à la théorie politique — ces limita-

²⁹ Le peu d'attention accordé jusqu'à présent aux utilisations de l'IA pour soutenir les processus démocratiques en dehors des instances délibératives officielles a conduit certains, par exemple, à appeler à un débat sur la manière dont l'IA peut contribuer à la formation et au maintien de sphères publiques démocratiques (Farrell & Han, 2025).

Ces difficultés sont très probablement surmontables, ce qui se traduirait par de réels avantages démocratiques. Nous pensons qu'une collaboration minutieuse de ce type pourrait lever certaines des objections que nous avons formulées ci-dessus concernant l'utilisation démocratique des grands modèles de langage (LLM). En particulier, grâce à une réflexion approfondie et à une définition plus précise des critères permettant de résumer de manière appropriée les réponses aux appels à la participation publique, nous estimons que les LLM *pourraient* sans aucun doute permettre une extraction bien plus efficace des opinions du public sur des questions d'État essentielles.

5 Conclusions

Les progrès récents en matière d'IA sont sans précédent : quelles que soient les lacunes des grands modèles linguistiques, ceux-ci constituent un élargissement radical de l'horizon technologique. Le biais de saillance suffirait à lui seul à justifier l'exploration de l'utilité potentielle des grands modèles linguistiques pour la délibération et la prise de décision démocratiques ; mais le potentiel va clairement bien au-delà. La démocratie est une pratique fondamentalement communicative, et elle est donc aussi une pratique linguistique. De nouveaux outils d'analyse et de production de contenu linguistique seront certainement utiles à un moment ou à un autre du processus démocratique. Cependant, la communication ne se réduit pas, en fin de compte, au seul langage : la démocratie est aussi fondamentalement sociale. C'est une manière de faire émerger et de prendre en compte des intérêts, de parvenir à des compromis et de s'engager dans une sorte d'expérience transformatrice (Paul, 2014). S'il n'y a guère de mal à utiliser les LLM pour soutenir des approches innovantes de la conception participative des chatbots, ces expériences ne peuvent guère inspirer une innovation démocratique plus large. Les LLM sont (du moins à l'heure actuelle) trop opaques pour fournir une procédure décisionnelle démocratique légitime. Et nous ne pouvons pas externaliser les aspects sociaux de la communication démocratique vers des modèles linguistiques : nous avons des raisons participatives qui justifient que nous nous engagions nous-mêmes. On pourrait peut-être en dire davantage sur l'idée de faire en sorte que des agents IA agissent en tant que nos représentants politiques (Burgess, 2022 ; Lechterman, 2024), mais à première vue, cela semble être une mauvaise idée.

Les grands modèles linguistiques (LLM) ne vont pas réparer la démocratie. Ils peuvent toutefois contribuer à favoriser des démocraties plus saines, non pas en introduisant de nouvelles méthodes d'agrégation des jugements ou de représentation algorithmique, mais en nous aidant à mieux naviguer dans la surabondance d'informations à laquelle nous sommes confrontés en ligne. Une sphère publique informelle saine est une condition préalable indispensable à une démocratie saine. La contribution des LLM à cette sphère publique a jusqu'à présent été pour le moins ambiguë ; mais non seulement il y a matière à amélioration, mais il existe un réel potentiel pour y parvenir. Nous encourageons les chercheurs en IA désireux d'utiliser les LLM pour renforcer la démocratie — et les théoriciens politiques computationnels de manière plus générale — à cesser de se concentrer sur les tentatives de mise en place de procédures décisionnelles alternatives visant à résoudre des conflits d'intérêts et de valeurs, ainsi que sur les efforts connexes visant à compléter les formes formelles de délibération démocratique, pour se concentrer plutôt sur l'utilisation des capacités de recherche et de présentation d'informations des LLM afin de donner aux citoyens démocratiques les moyens de naviguer et d'améliorer notre sphère publique informelle.

Remerciements Nous remercions les membres du laboratoire MINT, ainsi que Ricardo Fabrino et John Dryzek pour leurs commentaires sur les versions préliminaires. Nous remercions également les organisateurs et les participants de l'atelier sur l'IA et la cybersécurité de l'Institute for Humane Studies pour leurs commentaires et pour nous avoir donné l'occasion de présenter ces travaux.

Financement Financement en libre accès rendu possible et organisé par le CAUL et ses institutions membres. Le financement a été fourni par l'Australian Research Council (subvention n° FT210100724), Seth Lazar (subvention n° LP210200818), Seth Lazar, l'Institute for Humane Studies et l'université George Mason.

Accès libre Cet article est sous licence Creative Commons Attribution 4.0 International, qui autorise l'utilisation, le partage, l'adaptation, la distribution et la reproduction sur tout support ou format, à condition de citer correctement le ou les auteurs originaux et la source, de fournir un lien vers la licence Creative Commons et d'indiquer si des modifications ont été apportées. Les images ou autres éléments provenant de tiers figurant dans cet article sont inclus dans la licence Creative Commons de l'article, sauf indication contraire dans la mention de crédit relative à ces éléments. Si le contenu n'est pas couvert par la licence Creative Commons de l'article et que l'utilisation que vous envisagez n'est pas autorisée par la réglementation en vigueur ou dépasse l'utilisation autorisée, vous devrez obtenir l'autorisation directement auprès du titulaire des droits d'auteur. Pour consulter une copie de cette licence, rendez-vous [sur http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

Références

- Abdul Samad, S., Sushma, R., Bharathi Mohan, G., Samuji, P., Repakula, S., & Kothamasu, S. R. (2024). Faire progresser la synthèse abstraite : évaluation des modèles GPT-2, bart, T5-small et pegasus à l'aide des métriques de référence rouge et bleu. *Innovations in Cybersecurity and Data Science* (pp. 119–31). https://doi.org/10.1007/978-981-97-5791-6_10
- Allen, D. (2023). *La justice par la démocratie*. University of Chicago Press.
- Allen, D., & Weyl, E. G. (2024). Les véritables dangers de l'IA générative. *Journal of Democracy*, 35(1), 147–162. <https://muse.jhu.edu/pub/1/article/915355>
- Alnemr, N. (2020). L'émancipation ne peut être programmée : les angles morts de la facilitation algorithmique dans la délibération en ligne. *Contemporary Politics*, 26(5), 531–552. <https://doi.org/10.1080/13569775.2020.1791306>
- Anderson, C. (2005). *Le consentement des perdants : élections et légitimité démocratique*. Oxford University Press.
- Anderson, E. (2009). Démocratie : valeur instrumentale ou non instrumentale. Dans T. Christiano & J. Christman (dir.), *Contemporary debates in political philosophy* (pp. 213–227). Wiley. <https://doi.org/10.1002/9781444310399>
- Premier rapport : Contributions démocratiques à l'IA. (2023). *Dembrane*. <https://www.dembrane.com/blog/report-opena-i-october-2023>
- Anthropic. (2025). *Claude 4.5 Opus Soul Document*. <https://gist.github.com/Richard-Weiss/efe157692991535403bd7e7fb20b6695>
- Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., Sorensen, T., & Wingate, D. (2023). Tirer parti de l'IA pour un discours démocratique : les interventions par chat peuvent améliorer à grande échelle les conversations politiques en ligne sur l' . *Actes de l'Académie nationale des sciences*, 120(41), e2311627120.
- Arterton, F. C. (1987). *Télé démocratie : la technologie peut-elle protéger la démocratie ?* Sage.
- Aytac, U. (2024). Domination numérique : réseaux sociaux et démocratie contestataire. *Political Studies*, 72(1), 6–25.
- Bakker, M., Chadwick, M., Sheahan, H., Tessler, M., Campbell-Gillingham, L., Balaguer, J., Mcaleese, N., Glaese, A., Aslanides, J., Botvinick, M., & Summerfield, C. (2022). Affiner les modèles linguistiques pour trouver un consensus entre des personnes aux préférences diverses. *Advances in Neural Information Processing Systems*, 35, 38176–38189.
- Beguś, G., Dąbkowski, M., & Rhodes, R. (2023). Grands modèles linguistiques : analyse des capacités d' s linguistiques théoriques des LLM. *Prépublication arXiv arXiv:2305.00948*
- Benadé, G., Nath, S., Procaccia, A. D., & Shah, N. (2021). Recueil des préférences pour la budgétisation participative. *Management Science*, 67(5), 2813–2827. <https://doi.org/10.1287/mnsc.2020.3666>
- Bennett, S. & Cutler, N. (2019). Lab long read : Consultations politiques — partie 2 : Un rôle pour la science des données ?, *Policy Lab* (GOV.UK). <https://openpolicy.blog.gov.uk/2019/10/28/lab-long-read-policy-consultation-part-2-role-of-data-science/>
- Bernholz, L. (2016). Créer une société civile numérique : la bibliothèque publique numérique d'Amérique. Dans R. Reich, C. Cordelli et L. Bernholz (dir.), *Philanthropie dans les sociétés démocratiques : histoire, institutions, valeurs* (pp. 178–202). The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226335780003.0008>

- Bernstein, M., Christin, A., Hancock, J., Hashimoto, T., Jia, C., Lam, M., Meister, N., Persily, N., Piccardi, T., Saveski, M., Tsai, J. L., Ugander, J., & Xu, C. (2023). Intégrer les valeurs sociétales dans les algorithmes d' des réseaux sociaux. *Journal of Online Trust and Safety*, 2(1), 1–13. <https://doi.org/10.54501/jots.v2i1.148>
- Besley, T. (2006). *Des agents guidés par des principes ? L'économie politique de la bonne gouvernance*. Oxford University Press.
- Bharathi Mohan, G., Prasanna Kumar, R., Parathasarathy, S., Aravind, S., Hanish, K. B., & Pavithria, G. (2023). Résumé de textes pour l'analyse des mégadonnées : une revue exhaustive des approches GPT 2 et BERT. Dans R. Sharma, G. Jeon, & Y. Zhang (Eds.), *Data Analytics for Internet of Things Infrastructure* (pp. 247–264). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33808-3_14
- Bjerknes, G., Ehn, P., & Kyng, M. (1987). *Les ordinateurs et la démocratie : un défi scandinave*. Gower Publishing.
- Bommasani, R., Creel, K. A., Kumar, A., Jurafsky, D., & Liang, P. S. (2022). S'en prendre toujours à la même personne : la monoculture algorithmique conduit-elle à une homogénéisation des résultats ? *Advances in Neural Information Processing Systems*, 35, 3663–3678. https://proceedings.neurips.cc/paper_files/paper/2022/hash/17a234c91f746d9625a75cf8a8731ee2-Abstract-Conference.html
- Bonchek, M. S. (1995). Les mouvements populaires dans le cyberspace : le recrutement de membres sur Internet. Dans *53e réunion annuelle de l' de la Midwest Political Science Association, Chicago* (pp. 6–8).
- Boulos, M., Van Opheusden, B., Fogelman, B., Albrecht, J., Huang, V., & Richardson, A. (2023). The need for trust: Analyzing the individual comments to the Ntía's AI RFC. *Imbue*: <https://imbue.com/perspectives/ntia-rfc-analysis-individual/>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Les émotions influencent la diffusion de contenus à caractère moral sur les réseaux sociaux. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618923114>
- Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (2016). *Handbook of computational social choice*. Cambridge University Press.
- Brill, M., et al. (2019). Démocratie interactive : nouveaux défis pour la théorie du choix social. Dans J.-F. Laslier, H. Moulin, M. R. Sanver et W. S. Zwicker (dir.), **The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers** (p. 59–66). Springer International Publishing. https://doi.org/10.1007/978-3-030-18050-8_10
- Buchstein, H. (1997). Bytes that bite: The Internet and deliberative democracy. *Constellations*, 4(2), 248–263.
- Bullock, J. G., & Lenz, G. S. (2019). Partisan bias in surveys. *Annual Review of Political Science*. <https://doi.org/10.1146/annurev-polisci-051117-050904>
- Burgess, P. (2022). L'augmentation algorithmique de la démocratie : examiner si la technologie peut renforcer les concepts de démocratie et d'État de droit à travers quatre scénarios hypothétiques. *AI & Society*, 37(1), 97–112. <https://doi.org/10.1007/s00146-021-01170-8>
- Burrell, J. (2016). Comment la machine « pense » : comprendre l'opacité des algorithmes d'apprentissage automatique. *Big Data & Society*, 3(1), 2053951715622512. <https://doi.org/10.1177/2053951715622512>
- Chan, C.M., Chen, W., Su, Y., Yu, J., Xue, W., Zhang, S., Fu, J., & Liu, Z. (2023). ChatEval : Vers de meilleurs évaluateurs basés sur les LLM grâce à des débats multi-agents. *Prépublication arXiv* : <https://arxiv.org/abs/2308.07201>
- Cheng-Ping, H., Sun, S., Krizan, S., Acharya, S., Rekesh, D., Jia, F., Zhang, Y., & Ginsburg, B. (2024). Ruler : Quelle est la taille réelle du contexte de vos modèles linguistiques à long contexte ? *Prépublication arXiv* : <https://arxiv.org/abs/2404.06654>
- Christiano, T. (2004). L'autorité de la démocratie. *Journal of Political Philosophy*, 12(3), 266–290. <https://doi.org/10.1111/j.1467-9760.2004.00200.x>
- Coeckelbergh, M. (2024). *Pourquoi l'IA sape la démocratie et comment y remédier*. John Wiley & Sons.
- Coeckelbergh, M. (2025). LLM, vérité et démocratie : aperçu des risques. *Science and Engineering Ethics*, 31(1), 4. <https://doi.org/10.1007/s11948-025-00529-0>
- Cohen, J. (1986). Une conception épistémique de la démocratie. *Ethics*, 97(1), 26–38. <https://doi.org/10.1086/292815>
- Cohen, J., & Fung, A. (2021). La démocratie et la sphère publique numérique. Dans L. Bernholz, H. Landemore, & R. Reich (dir.), *Technologie numérique et théorie démocratique* (pp. 23–61). The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226748603.003.0002>
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Réduire durablement les croyances conspirationnistes grâce à des dialogues avec l'IA. *Psychology & Psychiatry Journal*, 385(6714), eadq1814. <https://doi.org/10.1126/science.adq1814>
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Dagger, R. (1997). *Civic virtues: Rights, citizenship, and republican liberalism*. Oxford University Press.

- Dahl, R. A. (1989). *La démocratie et ses détracteurs*. Yale University Press.
- Darapaneni, N., Prajeesh, R., Dutta, P., Pillai, V. K., Karak, A., & Paduri, A. R. (2023). Résumé de texte à l'aide des modèles BERT et GPT-2, dans la Conférence internationale 2023 sur le traitement du signal, le calcul, l'électronique, l'énergie et les télécommunications (IConSCEPT) (IEEE). <https://doi.org/10.1109/IConSCEPT57958.2023.10170093>
- Davoodijam, E., & Mohsen. (2024). Indicateurs d'évaluation de la synthèse de texte : étude exhaustive. *Knowledge and Information Systems*, 66(12), 7717–7738. <https://doi.org/10.1007/s10115-024-02217-0>
- Devine, F., Krasodonski-Jones, A., Miller, C., Lin, S. Y., Cui, J.W., Marnette, B., & Wilkinson, R. (2023). Recursive public : expérimentation d'un engagement démocratique connecté dans le cadre de la gouvernance de l'IA. https://vtaiwan-o-penai-2023.vercel.app/Report_%20Recursive%20Public.pdf
- Diamond, L. (2020). Sortir de la crise démocratique. *Journal of Democracy*, 31(1), 36–50.
- Edelman, J. (2023). Democratic fine-tuning, *Lesswrong*. <https://www.lesswrong.com/posts/ncb2ycEB3ymNqzs93/democratic-fine-tuning>
- Edelman, J., & Klingefjord, O. (2023). OpenAI X Dft : le premier graphe moral. Meaning Alignment Institute. <https://meaningalignment.substack.com/p/the-first-moral-graph>
- Farrell, H., & Han, H. (1er août 2025). L'IA et les publics démocratiques. *Knight First Amendment Institute*. <http://knightcolumbia.org/content/ai-and-democratic-publics>
- Fischli, R. (2024). La démocratie fondée sur la propriété des données : l'autonomisation des citoyens par la propriété des données. *European Journal of Political Theory*, 23(2), 204–223. <https://doi.org/10.1177/14748851221110316>
- Fish, S., Gözl, P., Parkes, D. C., Procaccia, A. D., Rusak, G., Shapira, I., & Wüthrich, M. (2025). Choix social génératif. *Prépublication arXiv* : <https://arxiv.org/abs/2309.01291>
- Flanigan, B., Gözl, P., Gupta, A., Hennig, B., & Procaccia, A. D. (2021). Fair algorithms for selecting citizens' assemblies. *Nature*, 596(7873), 548–552. <https://doi.org/10.1038/s41586-021-03788-6>
- Ford, B. (2021). Technolisation de la démocratie ou démocratisation de la technologie ? Une perspective d'architecture en couches sur les potentiels et les défis. Dans R. Reich, L. Bernholz et H. Landmore (dir.), *Technologie numérique et théorie démocratique* (p. 274–321). The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226748603.003.0011>
- Fukuyama, F. (2020). La fin de l'histoire ? Dans S. Seidman & J. C. Alexander (dir.), *The New Social Theory Reader* (2e éd., pp. 298–304). Routledge.
- Gastil, J., & Davies, T. (2020). Démocratie numérique : Épisode IV — Un nouvel espoir* : Comment une société de logiciels publics pourrait transformer l'engagement numérique pour le gouvernement et la société civile. *Digital Government: Research and Practice*, 1(1), 1–15. <https://doi.org/10.1145/3342194>
- Goldberg, B., Acosta-Navas, D., Bakker, M., Beacock, I., Botvinick, M., Buch, P., DiResta, R., Donthi, N., Fast, N., Iyer, R., Zaria, J., Konya, A., Kwak Danciu, G., Landmore, H., Marwick, A., Miller, C., Ovadya, A., Saltz, E., Schirch, L., Shalom, D., Siddharth, D., Sieker, F., Small, C., Stray, J., Tang, A., Tessler, M., & Zhang, A. (2024). L'IA et l'avenir des espaces publics numériques. *Prépublication arXiv* : <https://arxiv.org/abs/2412.09988>
- Goodin, R. E., & Spiekermann, K. (2018). *Une théorie épistémique de la démocratie*. Oxford University Press.
- Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Pas ce à quoi vous vous attendiez : compromettre des applications réelles intégrant des modèles de langage de grande envergure (LLM) par injection indirecte de prompts. *Prépublication arXiv* : <https://arxiv.org/abs/2302.12173>
- Groper, R. (1996). Le courrier électronique et la revitalisation de la démocratie américaine. *Social Science Computer Review*, 14(2), 157–168. <https://doi.org/10.1177/089443939601400203>
- Grossi, D., Hahn, U., Mäs, M., Nitsche, A., Behrens, J., Boehmer, N., Brill, M., Endriss, U., Grandi, U., Haret, A., Heitzig, J., Janssens, N., Jonker, C. M., Keijzer, M. A., Kistner, A., Lackner, M., Lieben, A., Mikhaylovskaya, A., Murukannaiah, P. K., Proietti, C., Revel, M., Rouméas, É., Shapiro, E., Sreedurga, G., Swierczek, B., Talmon, N., Turrini, P., Terzopoulou, Z., & Van De Putte, F. (2024). Favoriser la renaissance démocratique numérique : un programme de recherche pour la démocratie numérique. *Prépublication arXiv* : <https://arxiv.org/abs/2401.16863>
- Grumbach, J. M. (2023). Les laboratoires du recul démocratique. *American Political Science Review*, 117(3), 967–984. <https://doi.org/10.1017/S0003055422000934>
- Guan, M. Y., Joglekar, M., Wallace, E., Jain, S., Barak, B., Helyar, A., Dias, R., Vallone, A., Ren, H., Wei, J., Chung, H. W., Toyer, S., Heidecke, J., Beutel, A., & Glaese, A. (2025). *Alignement délibératif : le raisonnement « » permet de créer des modèles linguistiques plus sûrs*. *arXiv* <https://doi.org/10.48550/arXiv.2412.16339>
- Gudiño-Rosero, J., Grandi, U., & Hidalgo, C. A. (2024). Les grands modèles linguistiques (LLM) en tant qu'agents d'une démocratie augmentée. *Prépublication arXiv* : <https://arxiv.org/abs/2405.03452>

- Habermas, J. (1991). *La transformation structurelle de la sphère publique : enquête sur une catégorie de la société bourgeoise d'...* The MIT Press.
- Habermas, J. (1998). *Entre faits et normes : contributions à une théorie discursive du droit et de la démocratie*. The MIT Press.
- Habermas, J. R. (1984). *La théorie de l'action communicative*. Beacon Press.
- Hartz-Karp, J., & Sullivan, B. (2014). La promesse non tenue de la délibération en ligne. *Journal of Public Deliberation*, 10, 16. <https://doi.org/10.16997/jdd.191>.
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT, c'est n'importe quoi. *Ethics and Information Technology*, 26(2), 38. <https://doi.org/10.1007/s10676-024-09775-5>
- Hill, K. A., & Hughes, J. E. (1998). *Cyberpolitics: Citizen activism in the age of the internet*. Rowman & Littlefield Publishers Inc.
- Himmelreich, J. (2022). Devrions-nous automatiser la démocratie ? Dans C. Veliz (dir.), *Oxford handbook of digital ethics* (pp. 655–684). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.33>
- Himmelreich, J. (2023). Against « Democratizing AI ». *AI & Society*, 38(4), 1333–1346. <https://doi.org/10.1007/s00146-021-01357-z>
- Huang, S., Siddarth, D., Lovitt, L., Liao, T. I., Durmus, E., Tamkin, A., & Ganguli, D., (2024). IA constitutionnelle collective : aligner un modèle linguistique sur les contributions du public. Dans *ACM Conference on Fairness, Accountability, and Transparency* (pp. 1395–1417). <https://doi.org/10.1145/3630106.3658979>
- Huq, A., & Ginsburg, T. (2018). Comment perdre une démocratie constitutionnelle. *UCLA L. Rev.*, 65, 78. <https://www.uclalawreview.org/lose-constitutional-democracy/>
- Jarrett, D., Pislari, M., Bakker, M. A., Tessler, M. H., Koster, R., Balaguer, J., Elie, R., Summerfield, C., & Tacchetti, A. (2025). Les agents linguistiques en tant que représentants numériques dans la prise de décision collective. *Prépublication arXiv* : <https://arxiv.org/abs/2502.09369>
- Ji, Z., Yu, T., Xu, Y., Lee, N., Ishii, E., & Fung, P. (2023). Vers l'atténuation des hallucinations des modèles de langage à grande échelle (LLM) par l'autoréflexion. Dans H. Bouamor, J. Pino, & K. Bali (Eds.), *Résultats de l'Association for Computational Linguistics : EMNLP 2023* (pp. 1827–1843). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-emnlp.123>
- Jia, C., Lam, M. S., Mai, M. C., Hancock, J., & Bernstein, M. S., (2023). Intégrer les valeurs démocratiques dans les IA des réseaux sociaux d'... via des fonctions objectives sociétales. *Prépublication arXiv* : <https://arxiv.org/abs/2307.13912>
- Jungherr, A., & Rauchfleisch, A. (2025). L'intelligence artificielle dans la délibération : la pénalité de l'IA et l'émergence d'un nouveau fossé délibératif. *Government Information Quarterly*, 42(4), 102079. <https://doi.org/10.1016/j.giq.2025.102079>
- Kahng, A., Mackenzie, S., & Procaccia, A. D. (2021). Démocratie liquide : une perspective algorithmique. *Journal of Artificial Intelligence Research*, 70, 1223–1252. <https://doi.org/10.1613/jair.1.12261>
- Kenthapadi, K., Sameki, M., & Taly, A. (2024). Ancrage et évaluation des grands modèles linguistiques : Prati-
Défis actuels et enseignements tirés (enquête). Dans *les Actes de la 30e conférence ACM SIGKDD sur la découverte de connaissances et l'exploration de données* (pp. 6523–6533). <https://doi.org/10.1145/3637528.3671467>
- Khan, L. M. (2016). Le paradoxe antitrust d'Amazon. *Yale Law Journal*, 126, 710. <https://yalelawjournal.org/note/amazons-antitrust-paradox>
- Kim, Y., Chang, Y., Karpinska, M., Garimella, A., Manjunatha, V., Lo, K., Goyal, T., & Iyyer, M. (2024). FABLES : Évaluation de la fidélité et de la sélection de contenu dans la synthèse de livres. *Pré- ation arXiv* : <https://arxiv.org/abs/2404.01261>
- Kingzett, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). Comment la polarisation affective sape le soutien aux normes démocratiques. *Public Opinion Quarterly*, 85(2), 663–677. <https://doi.org/10.1093/poq/nfab029>
- Kolodny, N. (2023). *La hiérarchie sociale : un problème philosophique*. Harvard University Press.
- Konya, A., Schirch, L., Irwin, C., & Ovadya, A. (2023). Élaboration de politiques démocratiques à l'aide de dialogues d' collective et de l'IA. *arXiv* : <https://doi.org/10.48550/arXiv.2311.02242>
- Konya, A., Thorburn, L., Almasri, W., Leshem, O. A., Procaccia, A. D., Schirch, L., & Bakker, M. A. (2025). Utilisation de dialogues collectifs et de l'IA pour trouver un terrain d'entente entre les artisans de la paix s israéliens et palestiniens. *Prépublication arXiv* : <https://arxiv.org/abs/2503.01769>
- Kreps, S., & Kriner, D. (2023). Comment l'IA menace la démocratie. *Journal of Democracy*, 34(4), 122–131. <https://muse.jhu.edu/pub/1/article/907693>
- Landemore, H. (2012a). *La raison démocratique : politique, intelligence collective et la règle de la majorité*. Princeton University Press.

- Landemore, H. (2012b). Pourquoi la majorité est plus intelligente que la minorité et pourquoi cela importe. *Journal of Deliberative Democracy*, 8(1), 7. <https://doi.org/10.16997/jdd.129>
- Landemore, H. (2021a). Démocratie ouverte et technologies numériques. Dans L. Bernholz, H. Landemore et R. Reich (dir.), *Technologies numériques et théorie démocratique* (p. 62–89). Presses universitaires de l' . <https://doi.org/10.7208/chicago/9780226748603.003.0003>
- Landemore, H. (2021b). Un argument épistémique en faveur de la démocratie. Dans M. Hannon et J. de Ridder (dir.), *The Routledge Handbook of Political Epistemology* (pp. 363–373). Routledge.
- Landemore, H. (2023). L'intelligence artificielle peut-elle mettre la délibération à la portée du grand public ? Dans R. Chang & A. Srinivasan (dir.), *Conversations in philosophy, law, and politics* (pp. 39–69). Oxford University Press. [h https://doi.org/10.1093/oso/9780198864523.003.0003](https://doi.org/10.1093/oso/9780198864523.003.0003)
- Lazar, S. (2023). Justice communicative et répartition de l'attention. *Knight First Amendment Institute* : <http://knightcolumbia.tierradev.com/content/communicative-justice-and-the-distribution-of-attention>
- Lazar, S. (2024a). Légitimité, autorité et devoirs démocratiques d'explication. *Oxford Studies in Political Philosophy*, 10, 28–56. <https://doi.org/10.1093/oso/9780198909460.003.0002>
- Lazar, S. (2024b). Autorités automatiques : pouvoir et IA. Dans A. Sethumadhavan & M. Lane (dir.), *Intelligence collaborative : comment les humains et l'IA transforment notre monde* (pp. 37–60). MIT Press. <https://doi.org/10.7551/mitpress/14962.003.0004>
- Lazar, S. (2025). Governing the Algorithmic City. *Philosophy & Public Affairs*. <https://doi.org/10.1111/papa.12279>
- Lazar, S., Thornburn, L., Jin, T., & Belli, L. (2024). The moral case for using language model agents for recommendation. *Prepublication arXiv* : <https://arxiv.org/abs/2410.12123>
- Lechterman, T. M. (2024). Le politicien parfait. Dans D. Edmonds (dir.), *AI morality* (pp. 53–63). Oxford University Press. <https://doi.org/10.1093/oso/9780198876434.003.0006>
- Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., See, D., Noothigattu, R., Lee, S., Psomas, A., & Procaccia, A. D. (2019). WeBuildAI : cadre participatif pour la gouvernance algorithmique. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–35. <https://doi.org/10.1145/3359283>
- Lehdonvirta, V. (2022). *Cloud empires : Comment les plateformes numériques prennent le pas sur l'État et comment nous pouvons reprendre le contrôle*. MIT Press.
- Lijphart, A. (1977). *La démocratie dans les sociétés plurielles : une étude comparative*. Yale University Press. Lin, J., Dai, X., Xi, Y., Liu, W., Chen, B., Zhang, H., Liu, Y., Wu, C., Li, X., Zhu, C., Guo, H., Yu, Y., Tang, R., & Zhang, W. (2023). Comment les systèmes de recommandation peuvent-ils tirer parti des grands modèles linguistiques : une étude sur l' . *Prepublication arXiv* : <https://arxiv.org/abs/2306.05817>
- Lin, S., Hilton, J., & Evans, O. (2022). *TruthfulQA : mesurer la manière dont les modèles imitent les mensonges humains* . <https://doi.org/10.48550/arXiv.2109.07958>
- Lin, T.-A. (2024). « Démocratiser l'IA » et la question de l'injustice algorithmique. *Philosophy & Technology*, 37(3), 103. <https://doi.org/10.1007/s13347-024-00792-2>
- Liu, R., Jia, C., Wei, J., Xu, G., & Vosoughi, S. (2022). Quantification et atténuation des biais politiques dans les modèles linguistiques de l' . *Artificial Intelligence*, 304, 103654. <https://doi.org/10.1016/j.artint.2021.103654>
- Lovett, A. (2024). *Les échecs de la démocratie et l'éthique de la démocratie*. University of Pennsylvania Press. Lovett, A., & Zuehl, J. (2022). La possibilité d'une autonomie démocratique. *Philosophy & Public Affairs*, 50(4), 467–498. <https://doi.org/10.1111/papa.12223>
- Macedo, S. (1990). *Les vertus libérales : citoyenneté, vertu et communauté dans le constitutionnalisme libéral*. Clarendon Press.
- Malkin, C. et N. Alnemr. (2024). *Projets délibératifs portés par les géants de la tech* (Document technique n° 5). Global Citizens' Assembly Network . <https://glocan.org/wp-content/uploads/2024/05/Technical-Paper-5-20-24-Malkin-and-Alnemr.pdf>
- Mansbridge, J. J. (1983). *Beyond adversary democracy*. University of Chicago Press.
- Margolis, M., Resnick, D., & Tu, C.-C. (1997). Campagne électorale sur Internet : les partis et les candidats sur le World Wide Web lors de la saison des primaires de 1996. *Harvard International Journal of Press/politics*, 2(1), 59–78. <https://doi.org/10.1177/1081180X97002001006>
- McDonald, D., Papadopoulos, R., & Benningfield, L. (2024). Réduire les hallucinations des modèles de langage à grande échelle (LLM) grâce à la distillation des connaissances : une étude de cas avec Mistral Large et le benchmark MMLU. Prépublications. <https://doi.org/10.36227/techrxiv.171665607.76504195/v1>
- McLuhan, M. (1964). *Les médias : les prolongements de l'homme*. McGraw Hill.
- Mendoza, G. B. (2023) Peut-on utiliser l'IA pour enrichir les consultations démocratiques ? *Rappler*. <https://www.rappler.com/technology/features/generative-ai-use-enriching-democratic-consultations/>

- Miller, D. (1995). *On Nationality*. Clarendon Press.
- Miller, N. R. (1983). Pluralisme et choix social. *American Political Science Review*, 77(3), 734–747. <https://doi.org/10.2307/1957271>
- Millière, R. (2024). Les modèles linguistiques en tant que modèles de langage. *Prépublication arXiv* : <http://arxiv.org/abs/2408.07144>
- Muldoon, J. (2022). *Le socialisme des plateformes : comment reprendre notre avenir numérique aux géants de la tech*. Pluto Press.
- Mumford, L. (1964). Techniques autoritaires et démocratiques. *Technology and Culture*, 5(1), 1–8. <https://doi.org/10.2307/3101118>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A comprehensive overview of large language models. *Prépublication arXiv* : <https://arxiv.org/abs/2307.06435>
- OpenAI. (12 février 2025). Partage des dernières spécifications du modèle. *OpenAI*. <https://openai.com/index/sharing-the-latest-model-spec/>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Siemens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Entraînement de modèles linguistiques pour suivre des instructions à l'aide de retours d'information humains. *Advances in Neural Information Processing Systems*, 35, 27730–27744. <https://proceedings.neurips.cc/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html>
- OVADYA, A. (2023). Repenser la démocratie pour l'IA. *Journal of Democracy*, 34(4), 162–170. <https://muse.jhu.edu/pub/1/article/907697>
- Papaioannou, K., Pantazi, M., & Van Prooijen, J. W. (2023). La démocratie est-elle menacée ? Pourquoi la croyance dans les théories du complot prédit des attitudes autocratiques. *European Journal of Social Psychology*, 53(5), 846–856. <https://doi.org/10.1002/ejsp.2939>
- Paul, L. A. (2014). *Expérience transformatrice*. Oxford University Press.
- Pettit, P. (2012). *Selon les conditions du peuple : une théorie républicaine et un modèle de démocratie*. Cambridge University Press.
- Quinn, D. P., & Woolley, J. T. (2001). Démocratie et performance économique nationale : la préférence pour la stabilité de l'. *American Journal of Political Science*, 45(3), 634–657. <https://doi.org/10.2307/2669243>
- Rawls, J. (1993). *Le libéralisme politique*. Columbia University Press.
- Reich, R., Sahami, M., & Weinstein, J. M. (2021). *Erreur systémique : où les géants de la tech ont échoué et comment nous, les citoyens, pouvons redémarrer*. HarperCollins Publishers.
- Rettenberger, L., Reischl, M., & Schutera, M. (2025). Évaluation des biais politiques dans les grands modèles linguistiques. *Journal of Computational Social Science*, 8(2), 1–17. <https://doi.org/10.1007/s42001-025-00376-w>
- Ronfeldt, D. (1992). La cyberocratie arrive. *Information Society*, 8(4), 243–296. <https://doi.org/10.1080/01972243.1992.9960123>
- Rosenblum, N. L., & Muirhead, R. (2019). *A lot of people are saying*. Princeton University Press.
- Rousseau, J.J. (1762). *Du Contrat social ; ou Principes du droit politique*. Amsterdam : Chez Marc Michel Rey.
- Sætra, H. S., Borgebund, H., & Coeckelbergh, M. (2022). Éviter de diluer la démocratie par des algorithmes. *Nature Machine Intelligence*, 4(10), 804–806. <https://doi.org/10.1038/s42256-022-00537-w>
- Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). Quelles opinions les modèles linguistiques reflètent-ils ? Dans *Actes de la 40e Conférence internationale sur l'apprentissage automatique* (pp. 29971–30004). <https://proceedings.mlr.press/v202/santurkar23a.html>
- Scheppele, K. L. (2018). Le légalisme autocratique. *University of Chicago Law Review*, 85(2), 545–584. <https://lawreview.uchicago.edu/print-archive/autocratic-legalism>
- Scheppele, K. L. (2022). How Viktor Orbán wins. *Journal of Democracy*, 33(3), 45–61. <https://muse.jhu.edu/pub/1/article/860244>
- Schlossberg, D., Zavestoski, S., & Shulman, S. W. (2008). Démocratie et élaboration des règles en ligne : technologies web, participation et potentiel de délibération. *Journal of Information Technology & Politics*, 4(1), 37–55. https://doi.org/10.1300/J516v04n01_04
- Schmitt-Beck, R., & Grill, C. (2020). Du salon à la salle de réunion ? Les discussions politiques des citoyens dans le système délibératif. *Political Communication*, 37(6), 832–851. <https://doi.org/10.1080/10584609.2020.1760974>
- Schuler, D. (2020). La technologie peut-elle soutenir la démocratie ? *Digital Government: Research and Practice*, 1(1), 1–14. <https://doi.org/10.1145/3352462>

- Seger, E., Ovadya, A., Siddarth, D., Garfinkel, B., & Dafoe, A. (2023). Démocratiser l'IA : significations, objectifs et méthodes multiples. *Dans Actes de la conférence AAAI/ACM 2023 sur l'IA, l'éthique et la société* (pp. 715–22). <https://doi.org/10.1145/3600211.3604693>
- Selbst, A. D., & Barocas, S. (2018). L'attrait intuitif des machines explicables. *Fordham Law Review*, 87, 1085–1139. <https://ir.lawnet.fordham.edu/flr/vol87/iss3/11>
- Shah, R., Feuillade-Montixi, Q., Pour, S., Tagade, A., Casper, S., & Rando, J. (2023). Méthodes évolutives et transférables de « jailbreak » de type « boîte noire » pour les modèles linguistiques via la modulation de persona. *Prépublication arXiv* : <https://arxiv.org/abs/2311.03348>
- Siddarth, D. (2024). Comment l'IA et la démocratie peuvent se renforcer mutuellement. <https://www.youtube.com/watch?v=H6 MVJm9ElMA>
- Siegel, S. (2022). Principes de saillance pour la démocratie. Dans S. Archer (dir.), *Salience : A philosophical inquiry* (pp. 235–266). Routledge.
- Small, C. T., Vendrov, I., Durmus, E., Hadjar, H., Barry, E., Cornebise, J., Suzman, T., Ganguli, D., & Megill, C. (2023). Opportunités et risques des LLM pour une délibération évolutive avec polis. *Pré-publication sur arXiv* : <https://arxiv.org/abs/2306.11932>
- Sterling, T. D. (1986). La démocratie dans une société de l'information. *Information Society*, 4(1–2), 9–47. <https://doi.org/10.1080/01972243.1986.9960016>
- Straub, V. J., Hashem, Y., Bright, J., Bhagwanani, S., Morgan, D., Francis, J., Esnaashari, S., & Margetts, H. (2024). L'IA au service de la productivité bureaucratique : mesurer le potentiel de l'IA pour automatiser 143 millions de transactions du gouvernement britannique. *arXiv* : <https://doi.org/10.48550/arXiv.2403.14712>
- Sulots, N. D. D. L. (2023). Résultats du premier sondage délibératif mondial annoncés par le laboratoire de démocratie délibérative de Stanford. *Centre de Stanford sur la démocratie, le développement et l'État de droit*. <https://cddrl.fsi.stanford.edu/news/results-first-global-deliberative-pollr-announced-stanfords-deliberative-democ-racy-lab>
- Susskind, J. (2022). *La république numérique : sur la liberté et la démocratie au XXIe siècle*. Bloomsbury Publishing.
- Taubenfeld, A., Dover, Y., Reichart, R., & Goldstein, A. (2024). Systematic biases in LLM simulations of debates. *Prépublication arXiv* : <https://arxiv.org/abs/2402.04049>
- Tessier, M. H., Bakker, M. A., Jarrett, D., Sheahan, H., Chadwick, M. J., Koster, R., Evans, G., Campbell-Gillingham, L., Collins, T., Parkes, D. C., Botvinick, M., & Summerfield, C. (2024). L'IA peut aider les humains à trouver un terrain d'entente dans la délibération démocratique. *Science*. <https://doi.org/10.1126/science.adq2852>
- Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). Un mauvais ouvrier blâme ses tweets : les conséquences de l'utilisation incivile de Twitter par les citoyens lorsqu'ils interagissent avec les candidats des partis. *Journal of Communication*, 66(6), 1007–1031. <https://doi.org/10.1111/jcom.12259>
- Tsebelis, G. (1999). Les acteurs du veto et l'élaboration des lois dans les démocraties parlementaires : une analyse empirique. *American Political Science Review*, 93(3), 591–608. <https://doi.org/10.2307/2585576>
- Vats, A., Jain, V., Raja, R., & Chadha, A. (2024). Exploring the impact of large language models on recommendation systems: An extensive review. *Prépublication arXiv* : <https://arxiv.org/abs/2402.18590>
- Viehoff, D. (2014). Égalité démocratique et autorité politique. *Philosophy & Public Affairs*, 42(4), 337–375. <https://www.jstor.org/stable/26605216>
- Vredenburg, K. (2022). Le droit à une explication. *Journal of Political Philosophy*, 30(2), 209–229. <https://doi.org/10.1111/jopp.12262>
- Westin, A. F. (1971). *Les technologies de l'information dans une démocratie*. Harvard University Press.
- Weyl, E. G., Tang, A., & Community. (2024). *Pluralité numérique : L'avenir de la technologie collaborative et de la démocratie*. Publication indépendante.
- Wiener, N. (1950). *L'utilisation de l'être humain par l'être humain ; Cybernétique et société*. Houghton Mifflin.
- Winner, L. (1984). Mythinformation à l'ère de la haute technologie. *Bulletin of Science, Technology & Society*, 4(6), 582–596.
- Winthrop, D. (1978). Aristote et la démocratie participative. *Polity*, 11(2), 151–171.
- Wright, N. (2018). Comment l'intelligence artificielle va redessiner l'ordre mondial. *Foreign Affairs*, 10. <https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order>
- Wu, F., Black, E., & Chandrasekaran, V. (2024). Monoculture générative dans les grands modèles linguistiques. *Prépublication arXiv* : <https://arxiv.org/abs/2407.02209>
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., Li, B., Jiang, L., Zhang, X., & Wang, C. (2023). Autogen : Vers des applications LLM de nouvelle génération grâce à un cadre de conversation multi-agents. *Prépublication arXiv* : <https://arxiv.org/abs/2308.08155>

- Wu, T. (2018). *The curse of bigness*. Penguin Random House.
- Wu, T. (2019). « Le premier amendement est-il obsolète ? » Dans D. E. Pozen (dir.), *The perilous public square: Structural threats to free expression today* (pp. 15–61). Columbia University Press. <https://doi.org/10.36644/mlr.117.3.first>
- Young, I. M. (2002). *Inclusion and democracy*. Oxford University Press.
- Zaremba, W., Dhar, A., Ahmad, L., Eloundou, T., Santurkar, S., Agarwal, S., & Leung, J. (2023). Contributions démocratiques à l'IA. (OpenAI) : <https://openai.com/blog/democratic-inputs-to-ai>
- Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024). Évaluation comparative de grands modèles linguistiques pour la synthèse d'actualités. *Transactions of the Association for Computational Linguistics*, 12, 39–57. https://doi.org/10.1162/tacl_a_00632

Note de l'éditeur : Springer Nature reste neutre en ce qui concerne les revendications juridictionnelles figurant sur les cartes publiées et les affiliations institutionnelles.