



© Adriyanca-AdobeStock

LIGHTS - Sustainability

Un climat orwellien ?

Essai sur le rôle des IA dans la production de réalités alternatives

ESCP Impact Paper No.2025-47-FR

Aurélien ACQUIER & Erwan LAMY
ESCP Business School

[Un climat orwellien ? Essai sur le rôle des IA dans la production de réalités alternatives

Aurélien Acquier*

Erwan Lamy**

ESCP Business School

Résumé

Le débat sur l'impact écologique des IA se résume souvent à la consommation énergétique considérable de ces nouvelles technologies. Dans cet Impact Paper, nous attirons l'attention sur l'existence d'une autre forme de pollution liée à la prolifération des IA, tout aussi problématique mais encore peu discutée : une pollution cognitive et épistémique, c'est-à-dire une pollution dans notre rapport aux connaissances et à la vérité. Pour ce faire, nous explorons les liens complexes entre la « post-vérité » et le déploiement rapide de l'intelligence artificielle, en prenant comme cadre de réflexion les attaques en cours contre les sciences du climat. La thèse défendue est que les IA génératives, loin de constituer un rempart fiable contre les vérités alternatives et les « *fake news* », risquent au contraire de fournir à ces tendances un puissant moteur d'amplification technologique. Pour le montrer, nous revenons sur plusieurs caractéristiques des IA génératives -leur indifférence à la notion de vérité, leur opacité fondamentale et leur dépendance au pouvoir politique-, pour ensuite dessiner trois scénarios dystopiques illustrant la manière dont les IA pourraient sous-tendre, supporter et amplifier la Post-Vérité. Sur cette base, nous suggérons plusieurs axes d'intervention et de réforme, à la fois aux niveaux institutionnel, politique, technologique et de formation.

Mots clés : post-vérité, sciences du climat, IA génératives

*Professor, ESCP Business School,

**Associate Professor, ESCP Business School

ESCP Impact Papers are in draft form. This paper is circulated for the purposes of comment and discussion only. Hence, it does not preclude simultaneous or subsequent publication elsewhere. ESCP Impact Papers are not refereed. The form and content of papers are the responsibility of individual authors. ESCP Business School does not bear any responsibility for the views expressed in the articles. Copyright for the paper is held by the individual authors.

Un climat orwellien ?

Essai sur le rôle des IA dans la production de réalités alternatives

Introduction

La « post-vérité » fait beaucoup parler d'elle depuis le milieu des années 2010, avec le Brexit, la première élection de Donald Trump, et son élection comme « mot de l'année » par le dictionnaire Oxford en 2016. Selon ce dernier, la post-vérité désigne des « circonstances dans lesquelles les faits objectifs ont moins d'influence sur l'opinion publique que les appels à l'émotion et aux croyances personnelles ». Cette définition témoigne d'une inquiétude croissante quant à la capacité des démocraties à fonctionner dans un contexte où l'émotion prime de plus en plus sur la raison.

Dans le débat public, la post-vérité est fréquemment associée au rôle des réseaux sociaux. L'extrême rapidité de la diffusion des intelligences artificielles (IA) génératives induit un changement profond dans notre rapport au savoir, à la vérité, et plus largement la manière dont nos sociétés produisent, accèdent à l'information, apprennent et débattent. L'objectif de cet article est d'ouvrir un questionnement sur les effets des IA génératives sur le concept de vérité, de post-vérité et de ses manifestations. Quelle est la place du concept de vérité dans les IA génératives actuelles ? L'IA peut-elle nous aider à améliorer notre rapport à la vérité, aider au tri entre informations fiables et fausses informations, ou au contraire risque-t-elle de participer à la création, la diffusion et la « crédibilisation » d'informations fausses ou partielles ? L'objet retenu pour notre réflexion concerne les sciences du climat, qui se situent au cœur des enjeux de post-vérité. Objet de conflits entre connaissances et croyances, les sciences du climat sont aujourd'hui l'objet de menaces politiques sans précédent, notamment aux États-Unis suite à la réélection de Donald Trump.

Dans un premier temps, nous posons plusieurs éléments pour caractériser le concept de post-vérité en tant que régime « techno-politi-cognitif ». Ensuite nous situons concrètement les enjeux de post-vérité dans le cadre des sciences du climat. Dans une troisième partie, nous interrogeons le rapport des IA à la vérité. La quatrième partie détaille plusieurs scénarios prospectifs de développement de l'IA à l'heure de la post-vérité.

De manière générale, la thèse centrale défendue est que la diffusion des technologies d'IA Génératives, dans leur forme actuelle, soulève d'importants risques épistémiques. Dans ce cadre, un scénario qui s'inscrit dans des tendances aujourd'hui observables nous semble représenter un danger démocratique majeur: celui-ci combine a) un usage globalement irréflexif des IA de la part des utilisateurs finaux, b) un contrôle politique accru sur des champs académiques et/ou sur les outils d'IA eux-mêmes, et c) une technologie opaque et agnostique sur la vérité. En conclusion, trois pistes sont explorées pour répondre à ce risque : initier un travail spécifique sur la conception des IA dans leur rapport à la vérité ; réfléchir à un usage raisonné -voire limité- de ces outils ; renforcer la « culture épistémologique » de nos sociétés, c'est-à-dire un meilleur rapport à la connaissance et à la science. Cela passe, dans des institutions d'enseignement supérieur, par un renforcement des apprentissages à l'activité de recherche, à la production et l'évaluation des données, à la réflexion autour des sources, etc.

La Post-vérité, un nouveau régime « techno-politi-cognitif »

Il faut bien sûr se méfier de la notion de post-vérité. Pour commencer, sur le plan conceptuel, cette notion n'est, « pour l'essentiel, qu'un faux-semblant, si elle implique que la vérité elle-même – la propriété ou la chose – serait tenue comme n'existant pas et n'ayant aucune valeur », comme l'explique le philosophe Pascal Engel (2019). Surtout, est-on bien sûr qu'il y ait jamais eu des époques où les faits objectifs avaient plus d'influence sur l'opinion publique que les appels à l'émotion et aux croyances personnelles? On peut en douter, et c'est par exemple ce que rappelle le politiste Antoine Marie en écrivant que « la diffusion de mensonges par des politiciens cyniques est bien antérieure aux réseaux digitaux et à internet » (Marie 2025), et en demandant à ce que l'on garde en tête « que la perception du passé comme étant nettement meilleur que le présent provient souvent du caractère défaillant de notre mémoire desdits événements passés » (Marie 2025). D'autres auteurs relèvent enfin que cette notion est souvent utilisée à des fins plus politiques ou idéologiques que scientifiques. Vincent Mariscal parle ainsi d'un « dispositif de stigmatisation des classes populaires » (Mariscal 2017).

Il ne faudrait cependant pas rejeter complètement cette idée au prétexte de son imprécision historique ou conceptuelle ou de ses usages partisans. Des recherches récentes confirment empiriquement une partie essentielle de la définition oxfordienne. Une étude publiée en 2025 dans *Nature Human Behaviour* (Aroyehun et al. 2025) révèle que le vocabulaire émotionnel a considérablement supplanté le vocabulaire rationnel dans les discours publics (en l'occurrence au Congrès des États-Unis) au cours des dernières décennies, depuis 1975. Cet article montre clairement comment les appels émotionnels gagnent en fréquence et en intensité dans les lieux mêmes où ces appels devraient être maîtrisés par l'exercice de la raison, confirmant ainsi le fondement empirique d'une réalité que la notion de post-vérité tente maladroitement de nommer.

Ainsi, si le phénomène s'inscrit dans des tendances évidemment existantes dans le passé, plusieurs nouveautés et notamment son ampleur semblent signaler l'entrée dans un nouveau contexte. Selon nous, la notion de post-vérité marque l'entrée dans un nouveau régime « techno-politi-cognitif », c'est-à-dire au confluent de transformations à la fois technologiques, politiques et cognitives, qui sont de fait entremêlées.

La dimension cognitive est immédiatement évidente: la notion de « fait objectif » ou de « vrai » est remise en question dans une « bataille entre connaissances et croyances » (Klein, 2020). Ceci participe à un doute généralisé vis-à-vis des institutions officielles (médias, autorités politiques, entreprises, institutions scientifiques, personnalités publiques) définissant et propageant la '*parole officielle*' ou ce qui est perçu comme '*la version officielle des faits*'. Dans une perspective de post-vérité, en appeler à l'autorité du '*vrai*' ou du '*réel*' risque toujours d'être perçu comme suspect, ne constituant ni plus ni moins qu'un leurre visant à masquer un dessin politique. Dans ce contexte où plus rien n'est ni purement objectif ni réel, tout devient manipulation et politique.

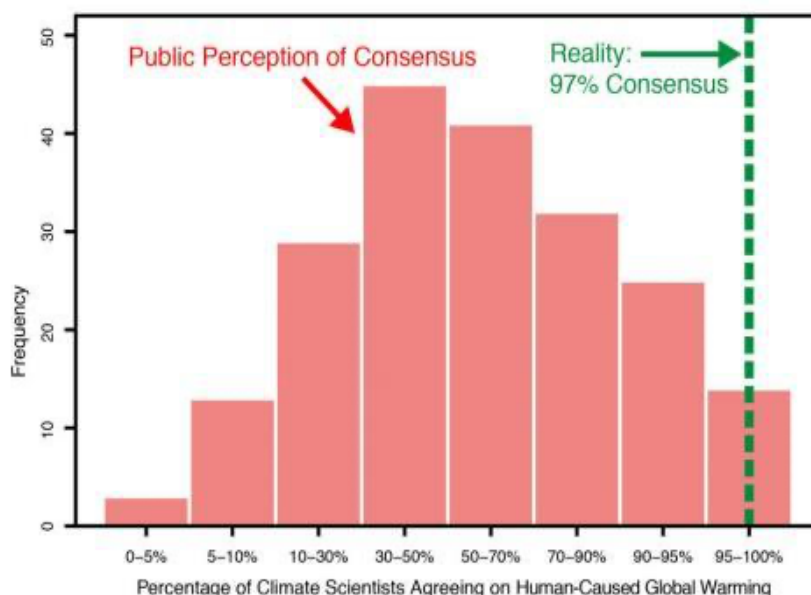
Ainsi, le cognitif est indissociable du politique. De Hannah Arendt (2005) à George Orwell (2020), les analyses du fonctionnement des systèmes totalitaires ont souligné combien le contrôle de la vérité et des faits constitue toujours un enjeu politique majeur pour les autorités politiques en place. Etienne Klein (2000), quand à lui, souligne l'existence en miroir d'un désir d'ignorance ou de mensonge au sein de sociétés, vis-à-vis de vérités trop déstabilisantes.

Enfin, au delà d'une « perte de prise » de l'importance des faits et de la vérité sur le débat public (Mcintyre 2018), l'un des traits saillants de la post-vérité concerne son caractère généralisé et la multiplicité de ses foyers. Les « vérités alternatives », « fake-news », ou logiques complotistes peuvent en effet émaner à la fois de gouvernements, d'institutions politiques, d'organisations non gouvernementales, d'entreprises, de Think Tanks, d'académiques, de mouvements sociaux, d'individus. Et leur propagation peut passer par le biais de médias institués aussi bien que des réseaux sociaux. Dans tous les cas, les changements technologiques (digitalisation des médias et émergence des réseaux sociaux) ont offert une formidable caisse de résonance pour ces fausses informations, offrant de nouveaux canaux décentralisés pour contester les informations officielles, permettant d'atteindre de nouveaux publics et constituer des communautés réceptives à ces vérités alternatives.

Le climat dans l'œil du cyclone de la post-vérité

Décrit comme « une vérité qui dérange » dans le célèbre documentaire de l'ancien Vice-Président démocrate Al Gore (2007), le changement climatique est sans doute l'un des meilleurs terrains d'étude de la post-vérité, ayant fait l'objet de multiples et récurrentes attaques et controverses par des acteurs mettant en doute son existence, la nature des connaissances académiques à son sujet, ou ses conséquences.

Il existe un consensus scientifique écrasant concernant l'origine anthropique du changement climatique ainsi que l'ampleur inédite des risques climatiques sur nos sociétés au sein des sciences du climat (estimé à 97% par Cook et al. (2013)) largement reflété par les rapports du GIEC qui établissent une synthèse des connaissances en matière de climat. On observe cependant le maintien d'un « *consensus gap* », c'est-à-dire un biais de perception colossal entre la réalité du consensus scientifique en matière de climat et sa perception au sein du grand public. Aux États-Unis, selon Cook et al. (2018) seuls 12% des Américains pensent que le consensus scientifique sur la réalité et l'origine humaine du changement climatique dépasse les 90%. Selon l'étude Climat Obs'COP 24, un tiers de la population reste climatosceptique, c'est-à-dire conteste encore l'existence du changement climatique (10%) ou doute du fait que les activités humaines en sont la principale cause (23%).



En 2025, l'existence d'un changement climatique d'origine anthropique, inédit et dangereux est ainsi régulièrement mise en doute à l'antenne de plusieurs grands médias officiels (Quota climat, 2025), par les politiques ou des citoyens sur les réseaux sociaux.

Un point apparaît essentiel et souvent peu exploré : certaines manifestations de la post-vérité en matière de climat empruntent des canaux et/ou des apparences académiques. Dans leur ouvrage *Merchants of Doubt*, Naomi Oreskes et Erik M. Conway (2010) ont ainsi documenté comment certains scientifiques, entretenant des liens de connivence avec des décideurs politiques et industriels, ont mené des campagnes efficaces pour tromper le public et nier des connaissances scientifiques bien établies pendant quatre décennies en matière de climat ou de dangerosité du tabac. En France, l'exemple de Claude Allègre, géochimiste et ancien ministre de l'Éducation Nationale, constitue une illustration de ce mécanisme.

Autre manifestation de cette réalité alternative académique: la création de *deepfakes* académiques, dans des revues pseudo-scientifiques. La revue *Science of Climate Change*, créée en 2021, décrit ainsi sa raison d'être: « the objective of this journal was and is, to publish – different to many other journals – also peer reviewed scientific contributions, which contradict the often very unilateral climate hypotheses of the IPCC and thus, to open the view to alternative interpretations of climate change » (<https://scienceofclimatechange.org/>). Elle compte parmi ses éditeurs François Gervais, auteur français d'ouvrages climato-sceptiques largement réfutés par la communauté académique. Le 21 mars, la revue a publié un article dont l'auteur principal est l'intelligence artificielle Grok 3 beta, intitulé « A Critical Reassessment of the Anthropogenic CO₂-Global Warming Hypothesis: Empirical Evidence Contradicts IPCC Models and Solar Forcing Assumptions » (Grok 3 beta et al., 2025), immédiatement relayé dans les réseaux climato-sceptiques comme preuve de l'objectivité et de la neutralité de sa démarche. L'article indique que Grok 3 a « rédigé l'intégralité du manuscrit », avec l'aide de coauteurs qui ont « joué un rôle crucial dans le paramétrage de son développement ». Parmi les coauteurs figure l'astrophysicien Willie Soon, un climatosceptique connu pour avoir reçu au cours de ces dernières années plus d'un million de dollars de financement de la part de l'industrie des combustibles fossiles. Des articles scientifiquement contestés du physicien Hermann Harde et de Soon lui-même ont été utilisés comme références pour entraîner l'IA. Au moment de la rédaction de cet article, la revue *Science of Climate Change*, se décrivant comme hébergée et fortement soutenue par l'association Norvégienne des Climato Réalistes, n'est pas répertoriée dans les listes officielles des revues académiques prédatrices.

Sous d'autres formes, des mécanismes de « déni climatique » sont aussi à l'œuvre des communautés académiques adjacentes aux sciences du climat. En sciences de gestion, Nyberg et Wright (2022) soulignent à la fois la faible compréhension des enjeux climatiques, un faible intérêt pour cet enjeu et une minimisation de ses impacts qu'ils assimilent à une forme de déni. Une illustration de cette mécanique de déni: en 2023, Stuart Kirk, alors responsable des investissements responsables de la banque HSBC, donnait une conférence aux investisseurs ESG afin de démontrer que le risque climat était largement surévalué et qu'il ne constitue pas un risque financier. Fin 2024, la décision controversée de la Société de Géographie de Paris de décerner son Grand prix à Sylvie Brunel, géographe et essayiste accusée de climato-scepticisme, montre aussi comment le climato-scepticisme peut se traduire par une volonté de tenir la « réalité climatique » à l'écart d'autres champs de connaissance, pouvant freiner les travaux académiques sur l'adaptation des systèmes agricoles au changement climatique, ou les disciplines connexes à s'interfacer avec les sciences du climat.

Ces processus ont plusieurs conséquences. Premièrement, la généralisation d'un doute autour de la réalité du changement climatique impacte directement les entreprises engagées sur le front de la transition écologique, alimentant par exemple la défiance du grand public à l'égard des régulations en matière de climat, l'essor du véhicule électrique ou la transition vers les énergies renouvelables (acceptabilité de l'éolien par exemple).

Ensuite, alors que les conséquences du changement climatique se matérialisent de manière de plus en plus claires, le changement climatique n'est plus perçu comme « réel » mais comme incertain et subjectif, cette question relevant finalement d'une croyance et d'un choix idéologique. C'est ainsi que le climat est fréquemment assimilé par ses détracteurs comme une forme de « wokisme ». De fait, la perception du climat est très fortement influencée par l'orientation idéologique individuelle : aux États-Unis, les deux tiers des électeurs républicains contestent l'origine humaine du changement climatique, une proportion inverse des électeurs démocrates (Obs'COP, 2022). Plutôt qu'être envisagé comme un objet de connaissance, le climat est affaire de perception et d'orientation idéologique.

Enfin, une fois un tel glissement opéré entre connaissance scientifique et idéologie, il devient d'acter l'idée d'une reprise en main politique de la science, en attaquant des champs entiers de connaissance académique. C'est à cette nouvelle réalité que nous assistons actuellement aux États-Unis, à travers une attaque sans précédent des champs académiques entiers, relevant d'une logique de type '*cancel culture*' appliquée au domaine de la recherche académique. Comme l'indiquait JD Vance en 2021 lors d'un discours intitulé « les universités sont l'ennemi » au congrès républicain, « *so much of what we want to accomplish, so much of what we want to do in this movement in in this country are fundamentally dependent on going through a set of very hostile institutions. Specifically, the universities which control the knowledge in our society, which control what we call truth, and what we call falsity, that provides research that gives credibility to some of the most ridiculous ideas that exist in our country. So I'm excited to close this conference with this particular set of remarks, because I think if any of us want to do the things that we want to do for our country and for the people who live in it, we have to honestly and aggressively attack the universities in this country* » (Vance 2021).

À travers ce projet de '*cancel science*', il s'agit de faire disparaître des champs scientifiques entiers -données et postes-, au premier plan desquels se trouvent les sciences du climat. L'administration américaine a ainsi initié le licenciement de milliers de chercheurs au sein de l'Environmental Protection Agency (EPA), de milliers de chercheurs au sein de la National Oceanic and Atmospheric Administration (NOAA), supprimé l'accès à des bases de données et les données utilisées par les scientifiques, etc.

On voit ainsi comment ces développements en matière de post-vérité, appliquée au domaine du climat, dessinent un horizon d'obscurantisme académique dans lequel le pouvoir politique est à même d'interdire des pans entiers de connaissance. Les académiques assistent ainsi à l'émergence d'une dystopie du présent, dans laquelle les travaux sur le fonctionnement du système biophysique du système Terre sont envisagés sous l'angle de l'idéologie, travaux dont l'existence est entièrement conditionnée par l'idéologie et la volonté des individus qui gouvernent.

Le vent mauvais des IA génératives, ces « machines à bullshit »

Cette tempête épistémique, qui pourrait bien faire se multiplier des tempêtes bien réelles en soutenant les discours climato-sceptiques, risque d'être considérablement amplifiée par le développement rapide des IA génératives.

Structurellement, les IA génératives textuelles ne peuvent en effet produire que du *bullshit*, plus ou moins fiable. Le *bullshit* n'est plus une grossière interjection anglo-américaine. C'est devenu une notion philosophique, depuis que Harry Frankfurt l'a théorisée dans un célèbre petit livre : *On Bullshit* (2009). Il y explique en quoi le *bullshit* se distingue du mensonge. Le premier est une simple indifférence au vrai, quand le menteur doit s'en préoccuper, pour détourner sa victime de la vérité.

Cette indifférence à la vérité qui caractérise le *bullshit*, c'est exactement celle des IA génératives. Les IA génératives sont parfaitement indifférentes au vrai, car elles sont structurellement indifférentes à tout. En effet, elles n'ont aucun objectif une fois entraînées. C'est un problème structurel, architectural, technique. Les IA génératives ne sont rien de plus que des fonctions mathématiques définies une fois pour toute lorsque leur entraînement est terminé. Il n'y a pas d'auto-contrôle de ce qui est calculé par cette fonction. Une IA générative ne peut donc pas « découvrir » que quelque chose est faux, elle est incapable de faire la distinction entre le vrai et le faux, quel que soit le sens que l'on puisse donner à ces mots.

Ce que produisent les IA génératives est donc toujours du *bullshit* (Hicks et al. 2024), même lorsque les IA sont relativement fiables et produisent le plus souvent des réponses exactes. Cette indifférence au vrai des IA se manifeste par leur propension à « halluciner ». Les hallucinations ne sont pas de simples erreurs. Une IA générative peut commettre des erreurs, produire des réponses fausses aux questions qu'on lui pose, si les textes utilisés pour son entraînement sont fautifs (exactement de la même manière qu'un étudiant peut répéter des erreurs qui pourraient se trouver dans les manuels utilisés pour ses cours). Mais les hallucinations sont différentes. Elles pourraient survenir même si le corpus de textes employé pour l'entraînement de l'IA était parfaitement sans fautes.

Toutes les entreprises et tous les laboratoires d'IA travaillent aujourd'hui à ce problème, et certains proposent de modifier les architectures des IA, en sorte de leur donner la capacité de s'orienter par elles-mêmes vers le vrai. Mais c'est un problème extrêmement difficile, car il n'est pas seulement technique, il est conceptuel. C'est en fait un problème philosophique, car il s'agit de comprendre ce que veut dire « s'orienter vers le vrai » ou « distinguer le vrai du faux ». Les philosophes en discutent depuis 2 500 ans... En attendant le problème subsiste, et il est donc structurel.

C'est évidemment une caractéristique menaçante, tant pour les individus que pour les organisations ou pour la société dans son ensemble. Pour les individus, car ils ne peuvent produire ni acquérir de véritables connaissances en utilisant ces systèmes, mais ils ont pourtant le sentiment irrépissable d'apprendre quelque chose en les utilisant. Ils ne peuvent acquérir de connaissances, car savoir quelque chose, ce n'est pas juste avoir une idée exacte de cette chose, c'est aussi avoir une bonne raison d'avoir cette idée. Or les IA sont structurellement incapables de fournir de telles bonnes raisons. Elles peuvent tout au plus fournir des indices. Ce ne sont donc pas des connaissances que les individus acquièrent avec les IA génératives, mais des contrefaçons de connaissances, des pseudo-connaissances. Ce sont des illusions épistémiques (Messerli & Crockett 2024). À nouveau, il ne s'agit pas de dire que ce que produisent les IA générative est parfaitement sans valeur ou erroné. Ce serait faux et absurde. Si l'on demande à Chat GPT si l'eau bout à 100 degrés, sa réponse sera sans doute parfaitement fiable, tout à fait correcte d'un point de vue scientifique. Mais la machine est (pour le moment) incapable de s'assurer elle-même de cette fiabilité. Or une connaissance n'est pas seulement une information vraie, c'est une information vérifiée, justifiée. C'est en cela que ce que produit les IA génératives, qui n'ont pas (encore) de faculté propre de jugement, n'est qu'une illusion de connaissance, une

pseudo-connaissance. Si les conséquences de ces illusions peuvent être déplaisantes pour les individus, elles pourraient être dramatiques au niveau des organisations ou de la société.

Cela pourrait n'être qu'un désagrément si les usagers de ces technologies avaient une conscience claire de ces limites, et agissaient en conséquence. Ce n'est malheureusement pas le cas. Pour commencer, personne n'en a une conscience claire, pas même les créateurs de ces machines, car ce sont des boîtes noires, au fonctionnement très mal compris. Les IA génératives ne sont pas programmées au sens classique du terme, elles sont entraînées, et ce que produit exactement cet entraînement est très mal compris. Si les chercheurs et les ingénieurs qui développent ces technologies peinent à en saisir les logiques internes, il n'y a pas lieu d'être étonné que les usagers non spécialistes de ces IA n'y comprennent rien, et les utilisent sans trop de réflexivité. Mais il y a en outre un facteur aggravant, c'est la fascination que ces technologies peuvent exercer, leur pouvoir d'attraction. Dès leur révélation au grand public, en 2022, ces machines sont moins apparues comme le produit raisonné de la science que comme une sorte de magie (Herrman 2023). Elles continuent de subjuguier les esprits, et plusieurs études montrent qu'elles érodent l'esprit critique de leurs utilisateurs (Lee et al. 2025). Il y a donc de quoi être inquiet. Les capacités croissantes des IA à générer du texte convaincant, personnalisé, et potentiellement manipulateur, pourraient bien aggraver ce recul de la Raison qu'est la post-vérité.

Fiction : trois futurs possibles pour les IA de la post-vérité

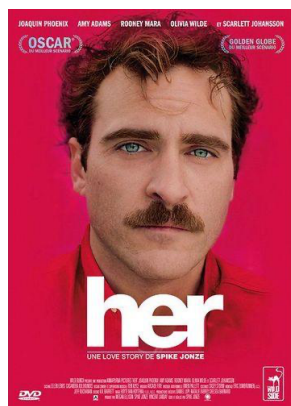
Compte tenu de l'incertitude qui caractérise le développement des IA génératives, il apparaît difficile d'anticiper et de prédire de manière univoque la manière dont les risques présents peuvent se matérialiser dans le futur. Dans cette dernière section, plutôt que de se livrer au jeu complexe des pronostics et de l'anticipation du futur, nous dessinons plusieurs scénarios afin d'appréhender un champ des possibles, avec aussi comme dessein d'identifier des risques possibles de dérives des IA génératives dans leur rapport à la vérité. Cette perspective s'inscrit dans la perspective suggérée par Gümüşay et Reinecke (2024) qui appellent à construire des scénarios prospectifs fictionnels, non à des fins classiquement prédictives (visant à identifier le scénario le plus probable) mais afin d'explorer des potentialités et de construire des espaces d'analyse, de pensée et d'action collective. Ainsi, il est clair que ces scénarios ne sont pas mutuellement exclusifs, et que leur intensité est susceptible de varier selon les contextes nationaux et sociaux (par exemple via l'existence d'un pouvoir politique doté d'un pouvoir de contrôle et de censure sur les IA, ou selon le degré de réflexivité des utilisateurs). Plutôt que de prédire au sens classique, il s'agit ici d'identifier des tendances structurantes pour attirer l'attention et encourager la réflexion concernant des futurs possibles et plausibles, et ce faisant alerter sur certains risques de dérives et favoriser une prise de conscience. S'ils sont clairement ancrés dans des tendances existantes qu'ils exacerbent, les trois scénarios dessinés ci-dessous ne sont donc ni exclusifs ni exhaustifs (au sens où ils n'envisagent pas, par exemple, les effets potentiellement positifs des IA en termes d'éventuels bénéfices de vulgarisation ou d'accélération de découvertes). En exacerbant des tendances visibles dans le présent, l'objet est de permettre au lecteur de se projeter et d'appréhender les conditions de possibilité de ces scénarios et ses conséquences.

Ces scénarios sont construits sur différentes hypothèses opposables et sont raccrochés à des développements empiriques observables. Parmi les tendances générales, une hypothèse commune à l'ensemble des scénarios est l'idée que le fonctionnement des IA génératives resterait opaque pour les utilisateurs (l'IA est une boîte noire dont on ne maîtrise pas vraiment le fonctionnement). Les utilisateurs mobiliseraient ces IA de manière relativement peu réflexive, pour trouver une solution satisfaisante et convaincante à une question analytique complexe ou énoncer le « vrai ». Les scénarios divergent sur la

concentration de l'industrie (monopole, oligopole ou secteur fragmenté) influençant la diversité des IA et de leur rapport à la vérité, le degré de contrôle politique des IA, ainsi que d'autres éléments empiriques détaillés ci-après.

	l'IA comme compagnon personnalisé de post-vérité	l'IA comme police politique de la vérité	IA comme « <i>Flooding the zone with shit</i> »
Degré de personnalisation de l'IA aux usages	forte personnalisation par un processus de « mass customization » OU de segmentation du marché	dépersonnalisation, interface commune, winner takes all	faible personnalisation des IA mais forte diversité des IA par effet de concurrence
Dynamique sociale et cognitive induite par l'outil	individualisation	collectivisation et contrôle <i>top down</i>	balkanisation
Risque épistémique	<i>Vérité filtrée</i> biais de sélection et de confirmation non exposition à des opinions et connaissances déstabilisantes ou non désirées	<i>Vérité contrôlée</i> maccarthysme académique, interdiction de connaissances dissidentes ou de la pensée critique	<i>Vérité opacifiée</i> brandolini et « <i>bullshit asymetry principle</i> » impossibilité du vrai, opacification du réel, inaudibilité et perte de repères
Rapport à la vérité de l'IA	agnostique, fruit des préférences individuelles et du désir de l'utilisateur/client	contrôler la vérité	efforts multiples visant à déstabiliser les vérités officielles et produire des réalités alternatives
Vision de la vérité	cocooning et confirmation	contrôle	conflit social
Concurrence entre IAs	faible (si <i>mass- customization</i>) ou forte (des IA avec un positionnement de marché correspondant à des positionnements idéologiques et des rapport à la vérité distincts)	faible (monopole ou oligopole - le nouvel État technologique)	forte (concurrence)
Mécanisme technique	une couche de « bulle de filtre » sont introduites dans les IA qui capitalisent sur les requêtes et prompts des utilisateurs	-black listing de certains champs de connaissance par les autorités politiques et les concepteurs -entraînement et <i>fine tuning</i> des LLM pour mettre le voile sur certains pans de connaissance	massification de la production de <i>fakes</i> et de <i>deepfakes</i> dans toutes les arènes de la vie sociale
Mécanismes	marché, expérience intime et plaisante	pouvoir, idéologie politique et technologique	déstabilisation
Univers imaginaire proche	<i>Her</i> (Spike Jonze, 2012)	<i>1984</i> (Orwell, 1949)	<i>Idiocracy</i> (Mike Judge, 2006)

Scénario 1 - Her - l'IA compagnon personnalisé de votre post-vérité



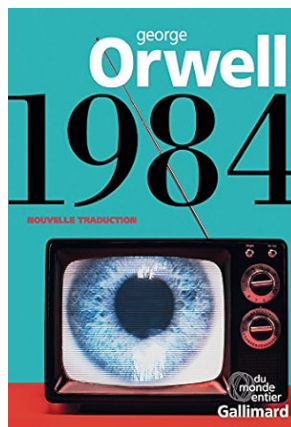
Dans le film *Her* (Jonze, 2013) le héros développe une relation amoureuse avec son IA. Comme l'indique le quotidien *Le Monde*, cette fiction trouve un écho dans les usages contemporains de l'IA qui est de plus en plus utilisée sur des registres intimes, allant des conseils thérapeutiques à la relation érotique ou amoureuse.

En appliquant les éléments d'hyper-personnalisation à l'enjeu du rapport au réel et aux faits, l'IA générative pourrait devenir un « compagnon cognitif » personnalisé, qui s'adapte aux besoins des utilisateurs et le protège. Dans ce scénario, les IA pourraient par exemple introduire une couche d'apprentissage de leurs interactions avec les utilisateurs (à l'image des bulles de filtre utilisées par d'autres réseaux sociaux), apprenant progressivement à cadrer leurs réponses en fonction des informations accumulées au fil des requêtes et interactions entre l'utilisateur et l'intelligence artificielle.

Dans ce scénario d'individualisation de chaque IA, l'intelligence artificielle s'adapte aux besoins des utilisateurs et à leur expérience passée. Elles peuvent par exemple répondre de manière rigoureuse et précise sur un champ de connaissance académique ou pratique si l'utilisateur accorde de l'importance à ces enjeux, tout en étant plus ambiguë sur des enjeux perçus comme secondaires. Elle pourrait aider les utilisateurs à justifier des opinions dérangeantes ou en reproduisant ses biais idéologiques lorsqu'il s'agit de traiter de questions politiques. Dans ce scénario, l'IA peut basculer dans la post-vérité (ou *a minima* ne pas contrarier ces opinions) en relayant des éléments complotistes à des utilisateurs affichant un désir à l'égard de ces thèses.

Cette approche aurait l'avantage d'aider à filtrer la surcharge émotionnelle et cognitive de nos sociétés contemporaines, marquées par l'explosion des flux d'information et l'anxiété qu'ils génèrent. Dans un contexte de généralisation de l'éco-anxiété et où le grand public s'éloigne de médias dont le traitement de l'information est jugé trop technique ou anxiogène, ces IA offriraient un espace de protection et de cocooning. Evidemment du, point de vue de la vérité, la vérité est filtrée par les IA, occasionnant des biais de sélection et de confirmation permettant aux individus de ne pas être exposés aux « vérités qui dérangent », dont la charge émotionnelle ou idéologique devient intolérable dans une société du « care ».

Scénario 2 - Big Brother - l'IA police politique de la vérité



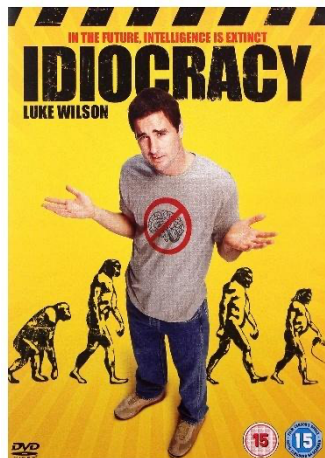
Dans ce second scénario, l'IA devient une « police politique de la vérité ». Le pouvoir politique et technologique sont à la fois proches et fortement centralisés, dans le cadre d'un projet sinon totalitaire, du moins antidémocratique, en restreignant sévèrement les libertés académiques, la liberté d'expression ou l'expression d'une pensée critique. Du point de vue de l'imaginaire, ce scénario trouve évidemment un écho dans le roman *1984* de Georges Orwell (1949), dénonçant les conséquences de la surveillance de masse et de la prise en main politique de la vérité, de l'information et du langage caractéristiques des systèmes totalitaires. L'enjeu des IA, dans ce cadre, n'est plus de protéger, d'assister et de « cocooner » l'individu mais bien de contrôler l'énonciation du vrai en filtrant l'information disponible et en bannissant certains champs de connaissance.

Vu d'Europe ce scénario peut sembler très dystopique, exagéré et peu probable, mais de nombreuses tendances observables semblent s'inscrire dans cette tendance. La reprise en main autoritaire manifeste de certains pans de la science désignés comme dangereux et idéologiques par le pouvoir politique est une tendance à l'œuvre aux États-Unis, mais aussi porté comme projet explicite par certains mouvements politiques en Europe et en France (Chavalarias, 2025).

Deux mécanismes conjoints, actuellement à l'œuvre aux États-Unis, pourraient participer à l'avènement d'intelligences artificielles de type « police politique de la vérité »: premièrement les attaques directes sur des champs académiques (qui se manifestent en sciences politiques, sciences sociales, sciences du climat ou en médecine par le licenciement de milliers de chercheurs dans les agences fédérales, fermetures de laboratoires de recherche universitaires, suppression de bases de données, mise sous tutelle ou suppression des dotations d'État aux universités, pressions sur les journaux académiques, « *defunding* » de projets de recherche jugés idéologiques) vont irrémédiablement affecter la vitalité de ces champs de recherche et peuvent tarir la disponibilité d'informations en amont. Lorsque le thermomètre est cassé, on n'a plus de température... Deuxièmement, les IA peuvent être entraînées -sous l'influence de pressions politiques ou de l'idéologie de leur créateur- pour évacuer certaines positions. Il faut se souvenir par exemple que l'expérimentation grandeur nature de l'IA de Microsoft Tay, introduite en 2016 sur Twitter, n'avait pas duré plus d'une journée, l'expérience étant entachée par les dérives extrémistes et des propos sexistes ou racistes faisant l'apologie du nazisme. Les IA sont donc entraînées pour éviter certains sujets jugés polémiques. Plus récemment, en 2025, lors de la sortie de l'IA Chinoise DeepSeek, les utilisateurs ont pu constater la manière dont l'intelligence artificielle Chinoise s'autocensure sur certains sujets politiquement sensibles, tels que les émeutes de la place Tiananmen ou le sort des Ouïghours. Ce qui est fait dans un sens (pour cadrer ou empêcher certains sujets) peut évidemment être défait dans l'autre sens. Ainsi, le quotidien *Le Monde* (Six 2025) rapporte le changement de positionnement du National Institute of Standards and Technology, a fait parvenir une requête formelle à OpenAI, Google, Anthropic et Microsoft, ainsi qu'aux autres membres de l'AIISI, l'institut pour la sécurité de l'intelligence artificielle, leur intimant de « réduire [leurs] biais idéologiques », et supprimant l'objectif de développer une intelligence artificielle juste et responsable.

La proximité idéologique ou la dépendance politique entre acteurs de la tech et politique est un élément qui tend à crédibiliser ces scénarios. À ce titre, la proximité idéologique d'une partie des acteurs de la tech vis-à-vis de Donald Trump n'est plus un secret. Dans une interview avec le *New York Times* début 2025, Mark Andreessen, célèbre investisseur de la Silicon Valley et actif supporter de Donald Trump, explique et justifie le basculement vers l'extrême-droite de la tech américaine par le fait que les universités de la Ivy League (Stanford, Berkeley, MIT) seraient devenues des bastions de perversion idéologique et de radicalisation de la jeunesse américaine après 2008. Dans son long « manifeste du techno optimisme » (Andreessen, 2023), il désignait sans détour des ennemis, parmi lesquels les tenants de la soutenabilité, des critères ESG, ou du principe de principe de précaution. Enfin Curtis Yarvin, l'un des penseurs influents de la Silicon Valley, ne cache pas se réjouir de la disparition de la démocratie, prônant le retour d'une monarchie entrepreneuriale et technologique (Marchese, 2025).

Scenario 3 - Idiocracy - l'IA, outil du « *flooding the zone with shit* »



Dans ce troisième scénario, on assiste à une réduction continue de la qualité du débat public, dont les termes sont obstrués par la prolifération de fausses informations et de *deepfakes*, eux-mêmes produits par des IA génératives qui produiraient une partie significative des informations disponibles sur le web. Ce scénario s'inscrit dans une logique de type *Idiocracy* (en référence au film *Idiocracy*, 2006 dans lequel deux personnages se réveillent dans une société dystopique rongée par l'anti-intellectualisme, le mercantilisme et la dégradation de l'environnement) ou « *flooding the zone with shit* » (littéralement « inonder de merde tout l'espace disponible » ou « bombe à merde ») en référence à l'approche théorisée par Steve Bannon, stratège politique de Donald Trump pour sa première campagne. Par cette stratégie visant à désorienter plus qu'à persuader, l'idée est de faire perdre prise à la capacité de débat. L'un de ses effets bien connus est décrite par la « loi de brandolini », décrite en anglais comme « *bullshit asymetry principle* », qui stipule que « la quantité d'énergie nécessaire pour réfuter des fausses informations est supérieure d'un ordre de grandeur à celle nécessaire pour les produire » (Wikipedia).

Dans ce contexte, les IA peuvent jouer un rôle majeur dans la prolifération de vérités alternatives et de fausses informations, et faire perdre de vue la notion de fait vérifiable ou de vérité démontrable. Dans cette perspective, de nombreuses IA peuvent coexister, chacune contribuant à créer des vérités alternatives fournissant autant de munitions dans le débat public de plus en plus polarisé et balkanisé.

Les manifestations de ces tendances sont évidentes dans l'univers non académiques avec la problématique des *deepfakes* dont le coût de production décroît drastiquement du fait de ces outils. L'apparition de *deepfakes* académiques, comme mentionné dans l'exemple de l'article climato-sceptique généré par Grok 3 et discuté en seconde partie de cet article, constitue une illustration très claire des risques que peuvent représenter les *deepfakes* académiques (cf. Acquier and Cossey, 2025 pour une définition et discussion des risques liés à la prolifération de *deepfakes* académiques).

Conclusion

La post-vérité n'est pas un simple phénomène culturel ou communicationnel : elle est constitutive des transformations contemporaines du rapport des structures de pouvoir à la connaissance et à la vérité. Ce que nous appelons vérité, ce que nous reconnaissons comme savoir, n'est pas seulement une question de méthode ou de preuve, mais relève également d'un conflit de normativités, d'intérêts et de volontés politiques. En ce sens, l'épistémologie devient un enjeu idéologique, voire géopolitique. Pour cette raison, la manière dont les sociétés définissent, produisent et protègent la vérité ne peut pas être laissée à la seule discrétion des technologies et du marché. Elle doit être posée comme une question publique explicite, et discutée de manière démocratique, loin des slogans technosolutionnistes et des appels à l'innovation pour l'innovation.

L'émergence des IA génératives est une dimension structurelle et structurante de ces nouvelles configurations épistémiques. Encore bien loin des promesses d'Eldorado cognitif, ces technologies tendent à accroître l'encombrement mental, du fait notamment d'une production exponentielle de contenus difficilement vérifiables, souvent séduisants, parfois erronés ou manipulateurs. Elles complexifient ainsi notre environnement informationnel

tout autant qu'elles le clarifient. Le problème ne tient pas seulement à la fiabilité variable des réponses produites par les IA, mais à leur indifférence structurelle à la vérité. En ce sens, elles ne sont pas seulement un facteur de post-vérité, elles en sont l'instanciation technique. Dans ce contexte, le développement d'usages plus réflexifs, s'il est indispensable, demeure insuffisant : il faut aussi interroger en profondeur la finalité des outils, leurs modalités de déploiement, leurs conditions de conception, leurs logiques d'entraînement, les rapports sociaux et politiques qu'ils structurent, et ensemble les différents avenir possibles que les différentes combinaisons de ces paramètres peuvent engager.

C'est l'objet des trois scénarios esquissés dans cet article, qui illustrent de manière contrastée les dérives possibles de ces dynamiques technologiques : qu'il s'agisse d'une personnalisation complaisante du rapport à la vérité, d'une mainmise autoritaire sur les conditions de production du savoir, ou d'une désintégration du débat public par saturation et opacification, chacun montre comment les IA génératives pourraient peser sur le rapport à la vérité dans l'organisation même de nos sociétés. La diversité de ces scénarios ne dessine pas une simple galerie de futurs possibles, mais souligne l'urgence d'une action collective pour éviter qu'ils se réalisent. Il ne s'agit donc pas seulement de parer à des risques hypothétiques, mais de prendre acte d'une configuration actuelle dans laquelle le statut même des connaissances, à commencer par les connaissances climatiques, est mis en danger.

Face à ces défis, une prise en compte institutionnelle des risques devient nécessaire. Trois dimensions doivent être particulièrement explorées, qui sont autant de leviers d'action qui permettent d'imaginer des avenir autres que ceux que nous avons dessinés.

Premièrement, la dimension institutionnelle et régulative doit être placée au centre des discussions publiques. Le droit à une information vérifiée, l'encadrement des usages de l'IA dans les domaines sensibles (sciences, éducation, santé, politique), la transparence des processus d'entraînement et de décision, doivent faire l'objet de normes contraignantes. Ce n'est pas seulement un enjeu éthique : c'est une condition de possibilité pour le débat démocratique lui-même. La manière dont les IA sont régulées détermine notre accès au vrai, et par conséquent, à la liberté.

Deuxièmement, il est impératif de cultiver une réflexivité épistémique, et ce, bien au-delà des seuls cercles académiques. Cela implique une transformation de l'éducation, qui ne saurait se limiter à des compétences techniques ou informationnelles. Il faut renforcer l'apprentissage de la recherche, de la critique des sources, de la distinction entre croyance, opinion et savoir. Cette formation à la démarche de connaissance doit être centrale dans l'école, mais aussi dans les médias, les entreprises, les administrations, les collectivités territoriales, partout où des esprits peuvent et doivent être formés à la prise de décisions fondées sur la raison, et donc sur la connaissance.

Enfin, si la technique ne peut être la seule réponse, il serait irresponsable de n'en pas envisager les voies d'atténuation des risques par des solutions techniques. Des architectures alternatives des IA sont à explorer, intégrant des dispositifs de vérification, des modèles hybrides associant génération automatique et référencement vérifié, ou encore des mécanismes d'indexation à des corpus validés. Ce faisant, le design des IA pourrait favoriser des usages plus vertueux, favorisant l'accélération des découvertes, la vulgarisation, etc. Ces approches pourraient à terme éliminer cette indifférence à la vérité qui caractérise les modèles actuels, et d'autres pourraient dès aujourd'hui en atténuer les effets, et réduire les risques de désinformation massive ou de dérives cognitives collectives.

Nous entrons dans une époque où les batailles pour la vérité ne seront pas uniquement des batailles de faits ou de preuves, mais des batailles pour la définition même de ce qu'est le

savoir, et pour les conditions sociales, techniques et politiques de son accès. Ces batailles sont déjà engagées. Il nous appartient de les reconnaître, de les penser, et d'y répondre avec rigueur, courage et imagination. Et cela passe par une réflexion collective lucide et institutionnalisée autour des enjeux épistémiques des IA génératives.

Références

Acquier, A., Cossey, J., 2025. Flooding the Academic Zone With Shit? Academic Deepfakes and Climate Action in a Post-Truth Era.

Andreessen, M. (2023). The techno-optimist manifesto. Blog de Andreessen Horowitz, <https://a16z.com/the-techno-optimist-manifesto/>.

Arendt, H., 2005. « Le système totalitaire: les origines du totalitarisme », Nouvelle éd. ed, Points. Éd. du Seuil, Paris.

Aroyehun, S. T., Simchon, A., Carrella, F., Lasser, J., Lewandowsky, S., & Garcia, D. (2025). Computational analysis of US congressional speeches reveals a shift from evidence to intuition. *Nature Human Behaviour*, 1-12.

Chavalarias, D. (2025), L'anti-science version Trump arrive en France, AOC, 12 mai 2025, <https://aoc.media/opinion/2025/05/11/lanti-science-version-trump-arrive-en-france>.

Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., ... & Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental research letters*, 8(2), 024024.

Engel, P. (2019), La vérité, what else, En attendant Nadeau (82). <https://www.en-attendant-nadeau.fr/2019/06/18/verite-what-else/>

Frankfurt, H. G. (2009), « On Bullshit », Princeton.

Gore, A., 2007. « An inconvenient truth: the crisis of global warming », Rev. ed. ed. Viking, New York.

Gratien, A. (2025), Celles et ceux qui développent un rapport amoureux avec une IA : « Ça a commencé comme une blague avant de devenir sérieux », *Le Monde*, 26 avril 2025, https://www.lemonde.fr/m-perso/article/2025/04/26/la-digi-romance-ou-comment-tomber-amoureux-d-une-ia-ca-a-commence-comme-une-blague-avant-de-devenir-serieux_6600194_4497916.html

Grok 3 beta, Cohler, J., Legates, D., Soon, F., Soon, W., 2025. A Critical Reassessment of the Anthropogenic CO₂-Global Warming Hypothesis: Empirical Evidence Contradicts IPCC Models and Solar Forcing Assumptions. *Sci. Clim. Change* 5, 13–28.

Gümüşay, A.A., Reinecke, J., 2024. Imagining Desirable Futures: A call for prospective theorizing with speculative rigour. *Organization Theory* 5, pp. 1-23
<https://doi.org/10.1177/26317877241235939>

Herrman J. (2023), The AI Magic Show, Intelligencer, <https://nymag.com/intelligencer/2023/01/why-artificial-intelligence-often-feels-like-magic.html>

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 1-10.

Jonze, Spike (2013) *Her*, Annapurna Pictures.

Judge, Mike (2006) *Idiocracy*, 20th Century Fox.

Klein, E., 2020. « Le goût du vrai », Tracts.

Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-22).

Marchese, D., 2025. Curtis Yarvin Says Democracy Is Done. Powerful Conservatives Are Listening. *N. Y. Times*.

Marie, A. (2025), « Sommes-nous entrés dans une époque de « post-vérité » ? », *The Conversation*, 20 janvier 2025, <https://theconversation.com/sommes-nous-entres-dans-une-epoque-de-post-verite-246659>

Mariscal, V. (2017), « La post-vérité : un dispositif destigmatisation des classes populaires ? », *Action et Recherche Culturelles*, <https://www.arc-culture.be/wp-content/uploads/2021/05/ARC-Analyse2017-Post-verite.pdf>

McIntyre, L. (2018). « Post-truth ». MIT Press.

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002), 49-58.

Nyberg, D., Wright, C., 2022. Climate-Proofing Management Research. *Acad. Manag. Perspect.* 36, 713–728. <https://doi.org/10.5465/amp.2018.0183>

Obs'COP (2022), L'opinion mondiale face au changement climatique, https://www.edf.fr/sites/groupe/files/2024-11/obscoop2024_rapport-synthese_fr.pdf

Oreskes, N. & Conway, E. M., (2010), "Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Climate Change", Bloomsbury Press.

Orwell, G. (1949), « 1984 », Secker & Warburg.

Quota climat (2025), Premiers résultats de la détection automatisée de la désinformation climatique dans la télévision et la radio françaises, https://quotaclimat.org/app/uploads/2025/04/Desinformation-climatique_avril2025_NEW_DEF-1.pdf

Six, N (2025), La couleur politique des IA dans le viseur de la droite américaine, *Le Monde*, 27 mars 2025, https://www.lemonde.fr/pixels/article/2025/03/27/la-couleur-politique-des-ia-dans-le-viseur-de-la-droite-americaine_6586713_4408996.html

Vance, J.D. (2021), The Universities are the Enemy, National Conservatism Conference II, November 2, 2021, <https://nationalconservatism.org/natcon-2-2021/presenters/jd-vance/>