

L'IA dans les élections municipales 2026 : Des usages qui appellent un regard nuancé



Antoine MARIE

Chercheur postdoctoral

antoinevictor.marie@sciencespo.fr www.antoinemariesci.com

The municipal elections of March 15 and 22, 2026 constitute the first French local election in which generative artificial intelligence has massively entered campaign practices. This research brief offers a balanced analysis of documented uses, distinguishing legitimate applications (poster generation, programmatic chatbots, content synthesis) from confirmed abuses (defamatory deepfakes, misleading images, astroturfing). While these new technological capabilities raise legitimate concerns for the integrity of democratic debate, we argue that their impact must be put in perspective in light of the literature on the limited effects of mass political communication and the actual determinants of disinformation consumption. Rather than a catastrophism that risks fueling generalized distrust, we advocate for an approach grounded in media literacy and the strengthening of professional journalism.

Au cours des élections municipales de 2026, l'intelligence artificielle générative s'est imposée comme un sujet de préoccupation majeur dans le débat public français. Cette mutation tient à l'explosion de la présence des algorithmes et du digital dans la vie et la participation politiques des Français depuis une vingtaine d'années (Cardon, 2015 ; Neihouser et al., 2023). Les reportages se multiplient sur les candidats utilisant *ChatGPT* pour rédiger leurs professions de foi, *Midjourney* pour créer leurs affiches, ou des *chatbots* pour interagir avec les électeurs ou, côté citoyens, pour connaître les programmes. Parallèlement, des cas de *deepfakes* malveillants et d'images trompeuses ont défrayé la chronique, alimentant les craintes d'une manipulation de masse du débat démocratique. Ces développements inquiètent les Français, qui sont 57% à déclarer avoir « assez » ou « beaucoup » peur des développements actuels de l'intelligence artificielle (Muxel & Perrineau, 2026).

Cette situation appelle une analyse¹ qui évite deux écueils symétriques. D'un côté, la minimisation des risques réels que ces technologies font peser sur l'intégrité informationnelle, en particulier la création facilitée de fake news et de *deepfake* audios et vidéos ultra-réalistes (Chavalarias, 2022 ; Gaumont et al., 2018).

[1] L'auteur remercie les chercheurs et professionnels de la communication politique suivants de l'avoir conseillé dans la rédaction de cette note de recherche : Dan Williams, philosophe à l'université de Sussex ; Philippe Moreau Chevrolet, senior advisor chez No Com, enseignant à Sciences Po ; Sacha Altay, chercheur en mésinformation ; Peter Barrett, enseignant en dépoliarisation, ESSEC. Il remercie aussi Océane Herrero, journaliste à *Politico* sur les thèmes liés à l'intelligence artificielle, et Chine Labbe, de *NewsGuard*, d'avoir partagé leur point de vue et leur expertise.

De l'autre, un catastrophisme excessif qui non seulement serait analytiquement incorrect, mais risquerait d'alimenter la défiance généralisée envers toute information et de décourager la participation citoyenne. En effet, comme l'ont montré les travaux sur les effets de la désinformation, les messages politiques déployés par les candidats et leurs équipes de campagne n'ont en moyenne que d'assez faibles effets persuasifs. Ensuite, il convient de se rappeler que lorsqu'il analyse les évolutions technologiques et leurs conséquences, l'esprit humain est sujet à deux biais cognitifs majeurs. Le premier est le *biais de négativité* (Baumeister et al., 2001) : la tendance de notre attention à se trouver capturée par les risques et ce qui peut dysfonctionner, mais à s'habituer immédiatement aux progrès, parfois considérables, qu'ont permis d'accomplir les innovations technologiques. Le second biais est celui de la croyance en l'influencabilité des autres (*third person effect*) : la tendance à considérer que l'on est soi-même bien armé contre les tentatives de manipulations politiques, mais que les *autres* y sont beaucoup plus vulnérables - quitte à surestimer cette manipulabilité des tiers (Golan & Lim, 2016). Combinés ensemble, ces biais de perception peuvent nourrir des « paniques technologiques », qui émergent à chaque vague d'innovation majeure (Orben, 2020, 2020).

Cet arrière-plan cognitif fondamental étant rappelé, cette note propose une cartographie des usages documentés de l'IA générative dans la campagne des municipales 2026, avant de les mettre en perspective au regard des connaissances scientifiques sur la persuasion politique et la consommation de désinformation. Nous couvrirons ensuite les risques spécifiques liés aux ingérences étrangères et à l'érosion de la confiance, avant de souligner les facteurs éducatifs qui peuvent permettre de préserver le débat démocratique des risques technologiques apportés par l'IA générative.

1. Les usages documentés de l'IA générative dans la campagne des municipales 2026 par les candidats

1.1. Des applications légitimes au service de l'efficacité communicationnelle

Les premiers usages de l'IA générative dans cette campagne relèvent d'une logique d'optimisation des ressources communicationnelles, particulièrement précieuse pour les candidats des petites listes disposant de moyens limités. Ces applications s'inscrivent dans une tendance de fond à la démocratisation des outils de communication politique, observable depuis l'avènement des réseaux socio-numériques (Cardon, 2015 ; Neihouser et al., 2023).

Les LLMs comme source d'inspiration programmatique. Le premier usage qui a été fait de l'IA générative par les candidats est la consultation des Grands modèles de langage (LLMs) comme ChatGPT (Open AI), Claude AI (Anthropic), Gemini (Google) ou Le chat (Mistral) pour chercher des idées pour alimenter leurs programmes politiques. La démarche est la même que celle des citoyens ordinaires utilisant les LLMs pour leurs recherches personnelles ou professionnelles : chercher de l'inspiration, et tirer parti des capacités de synthèse d'informations des LLMs. Toutefois, comme le signale Martin Pavanello, co-fondateur et CEO de Mister IA, une entreprise de formation à l'IA, tant les données d'entraînement que les modèles sont la propriété d'entreprises américaines (Barbaza, 2026). La pratique peut donc potentiellement produire des recommandations pour les programmes ne cadrant pas avec les réalités de la politique française (en termes de coûts financiers des projets, par exemple).

La génération d'affiches et de visuels constitue l'usage le plus répandu. Mohamed Tlamsi, candidat à Périgny-sur-Yerres (Val-de-Marne), décrit un « gain de temps énorme » permis par la délégation de la création graphique à l'IA : « *On peut s'occuper d'autres choses. On n'est pas forcément obligé de s'attarder sur tout ce qui est communication* » (Bouvier et al., 2026). À Grenoble, Hervé Gerbi (Horizons) a fait générer son affiche entièrement par intelligence artificielle, avec la mention explicite de son origine synthétique, pour exprimer son intérêt pour « l'innovation » (Figure 1) (Bouvier et al., 2026). À Reims, Stéphane Lang assume utiliser l'IA générative pour « *concevoir des images moins chères et de meilleure qualité* », notamment une affiche de campagne animée, notant que « *le prospectus, on lit, on ne lit pas, et après ça part à la poubelle* » (Ksiazienicki, 2026 ; M.D, 2026). Ces usages s'inscrivent dans une continuité avec les pratiques antérieures de retouche photographique, tout en abaissant considérablement les barrières financières à l'entrée pour les petites listes.

Figure 1 - Affiche (statistique) de campagne générée par intelligence artificielle d'Hervé Gerbi (Horizons) pour la mairie de Grenoble



Figure 2 - Affiche (statistique) de campagne d'Emmanuel Grégoire le posant en duel contre Rachida Dati (Les Républicains), dans un style visuel inspiré de la série Star Wars de G. Lucas



Les chatbots programmatiques reposant sur des LLMs représentent une innovation plus substantielle. À Les Mureaux, Damien Vignier propose « IA-smine », un agent conversationnel personnifié par une abeille humanoïde, répondant aux questions des électeurs sur son programme (Baudoin, 2026). Le candidat justifie ce choix : « *Les gens jettent souvent les tracts à la poubelle, et c'est parfois difficile d'intéresser les jeunes à la politique. L'idée est d'amener un peu de démocratie participative en touchant un public plus large* » (Baudoin, 2026). L'équipe de Gerbi à Grenoble a développé un système similaire (Belhassen & AFP, 2025 ; Bouvier et al., 2026), avec un LLM entraîné sur du « *contenu programmatique, des interviews, des discours* » selon son directeur de campagne Raphaël Meyer, précisant que « *l'IA ne fait que les restituer de façon résumée* ». Ces outils offrent une disponibilité permanente et multilingue, potentiellement utile pour impliquer des populations habituellement éloignées du débat politique, comme les jeunes.

Les podcasts automatisés illustrent une troisième catégorie d'usages. À Chantilly, Florence Woerth (DVD) utilise l'IA générative pour transformer ses réunions publiques en émissions de radio présentées par des « journalistes » synthétiques (de Florival et al., 2026). Sa directrice de campagne, Morgane Geffroy, explique : « *C'est simplement un brief pour donner la ligne de conduite à avoir pour générer ce podcast* ». La candidate souligne qu'« *on précise que c'est fait par l'intelligence artificielle et donc on ne trompe pas sur le sujet* », même si le résultat comporte « *quelques problèmes de prononciation* ».

Ces applications partagent plusieurs caractéristiques : elles visent l'efficacité plutôt que la tromperie, elles sont généralement signalées comme telles. Elles ne créent pas de fausses informations mais reformulent ou présentent différemment du contenu qui, conceptuellement, en termes d'idées politiques, est authentique. Elles participent d'une démocratisation des outils de communication politique et d'une tentative d'élargissement de leur impact (*reach*), réduisant l'avantage dont disposaient traditionnellement les candidats les mieux dotés financièrement. Ces techniques sont également nettement plus faciles à utiliser pour des candidats et des équipes de campagne plus jeunes que pour leurs concurrents seniors.

La régulation française actuelle

Sur le plan légal, il est utile de rappeler que le code électoral français reste muet sur l'intelligence artificielle. Comme le souligne l'avocat Jean-Christophe Ménard, « *pour l'instant, rien n'interdit à un candidat de recourir à l'IA pour créer des affiches, générer un discours ou retoucher des photographies* » (Belhassen & AFP, 2025). Le Conseil d'État a confirmé en octobre 2025 que l'usage de l'IA pour produire des professions de foi était sans effet sur la régularité électorale (*Conseil d'État, 2025*). Toutefois, d'autres textes encadrent ces pratiques. La loi SREN du 21 mai 2024 a renforcé les sanctions pénales contre les *deepfakes* non consentis (deux ans d'emprisonnement et 45 000 euros d'amende, circonstance aggravante en cas de diffusion sur les réseaux sociaux). La loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information permet d'agir en référé dans les trois mois précédant une élection pour faire cesser la diffusion d'informations manifestement fausses (LOI N° 2018-1202 du 22 décembre 2018 Relative à La Lutte Contre La Manipulation de l'information (1), 2018). Quant à l'échelle européenne, le règlement sur l'IA (AI Act), applicable en août 2026, imposera le marquage obligatoire des contenus générés par IA.

1.2. Les dérivés par les candidats : *deepfakes* et images trompeuses

À côté des usages plutôt anodins mentionnés plus haut, plusieurs cas plus contestables de *deepfakes* (ou « hypertrucages ») ont été documentés durant cette campagne.

Les images synthétiques diffamatoires anonymes. L'un des cas les plus clairs de dérive dans le cadre des municipales s'est produit à Guéret (Creuse), où plusieurs candidats ont été ciblés par des images générées par IA et diffusées anonymement sur les réseaux sociaux. La maire sortante Marie-Françoise Fournier a porté plainte, certaines images suggérant faussement un enrichissement personnel (France Info & Ici Creuse, 2026 ; Ny Ando, 2026). Elle estime que la ville « *mérite une campagne municipale propre, avec un débat d'idées* » et dénonce ces pratiques « *ni respectueuses, ni correctes* ». Le candidat socialiste Didier Hoeltgen a été tourné en dérision par un montage le montrant « main dans la main » avec son opposante, une « image qui n'existe pas » selon lui et qu'il considère comme une « calomnie » (Corcos, 2026).

Ces attaques anonymes exploitent le réalisme croissant de l'IA générative pour produire des images diffamatoires plus crédibles que les photomontages traditionnels. Elles illustrent une constante de la mésinformation politique de ton hostile : son exploitation d'intérêts et de motivations psychologiques que les audiences possèdent déjà - attrait pour l'humoristique, le sensationnel - dans le but stratégique de discréditer ses ennemis, de mobiliser et de renforcer les oppositions contre eux, etc. (Marie & Petersen, 2022, 2024, 2025 ; Osmundsen et al., 2021 ; Petersen, 2020).

Les images et vidéos synthétiques à visée rhétorique. Une seconde catégorie de dérives concerne l'utilisation d'images synthétiques pour illustrer des arguments politiques. Virginie Joron (RN), tête de liste à Strasbourg, a diffusé sur TikTok de fausses images générées par IA pour dénoncer la supposée saleté de la ville, provoquant un tollé (Belhassen & AFP, 2025). Elle a ensuite indiqué n'avoir eu « aucune intention malveillante », parlant d'« images améliorées ». À Paris, Sarah Knafo (Reconquête) a publié une vidéo montrant une scène d'agression fictive générée par IA pour illustrer son discours sur l'insécurité, accompagnée de la mention paradoxale : « Ce n'est pas une fiction, c'est la réalité » (Baudoin, 2026) (Figure 3). Knafo a également utilisé l'IA générative pour illustrer les conséquences de son projet de déployer des lampadaires et des forces de police municipale additionnelles permettant de réduire le sentiment d'insécurité des femmes dans les rues de Paris (Figure 4).

Figure 3 - Capture d'écran du post vidéo de Sarah Knafo (Reconquête)



Figure 4 - Captures d'écran du post de Sarah Knafo (Reconquête)



De manière plus innocente, des candidats comme Pierre-Yves Bournazel (Horizons) et Ian Brossat (Parti communiste) ont également eu recours à la génération d'images synthétiques afin de représenter leurs propositions d'urbanisation pour les quartiers : verdissement et reboisement des avenues et des quais de Seine à Paris, etc. (Bournazel, 2025 ; Brossat, 2025 ; Herrero, 2026).

Les cas de représentations synthétiques de l'insécurité, en particulier vidéos, soulèvent une question éthique et épistémologique intéressante. Philippe Moreau Chevolet, professeur de communication politique à Sciences Po, note que « *malgré ce que Sarah Knafo affirme, on est dans la fiction totale. Il y a une prise de distance avec la réalité, et c'est assez nouveau pour un clip de campagne* » (Baudoin, 2026). Des prises de distance croissantes avec le réel sont indéniablement à craindre si ces pratiques se généralisent dans les campagnes. Pourtant, il importe de remarquer que la communication politique verbale, orale, a toujours recouru à l'invocation d'archétypes et de généralisations abstraites. Parler dans un meeting politique ou un débat télévisé de « l'insécurité », de « la délinquance » ou du « racisme » sans se référer à des cas particuliers situés précisément dans le temps et l'espace est une pratique discursive inévitable (Heine & Narrog, 2015 ; Lippmann, 1929). L'innovation par l'usage de l'IA génératrice de vidéos réside dans l'extension de cette logique au domaine de la communication *visuelle*, traditionnellement soumis à des exigences de véricité plus strictes. Pour les journalistes professionnels (CDJM, 1918), l'image devait représenter *exactement* son objet, et être datée du même jour que la scène représentée. Le recours à l'intelligence artificielle générative pour créer des représentations visuelles *archétypiques* de phénomènes comme l'insécurité n'est pas fondamentalement différent, mais l'IA rend cette pratique plus facile, bousculant la norme journalistique, et cela tend à choquer.

Convenons que la production de représentations visuelles archétypiques augmente le risque de perceptions distordues d'enjeux sociaux importants - bien documentées dans le domaine de l'immigration, du soutien populaire pour la violence, ou encore de l'hostilité perçue entre démocrates et républicains aux États-Unis (Ahler & Sood, 2018 ; Gorodzeisky & Semyonov, 2020 ; Lees & Cikara, 2019 ; Mernyk et al., 2022 ; Sijlmassi et al., 2025).

1.3. L'*astroturfing* et les bots : la menace de l'amplification artificielle

Au-delà des *deepfakes* visuels, une préoccupation peut-être plus sérieuse concerne l'utilisation de l'IA générative pour amplifier artificiellement certains messages ou créer l'illusion de mouvements d'opinion spontanés, souvent pilotés par la Russie (Keller et al., 2020). Cette pratique, connue sous le nom d'« *astroturfing* » (par opposition au « *grassroots* », mouvement populaire authentique), préexistait aux LLMs. Mais elle se trouve considérablement facilitée par leur capacité à générer du contenu varié et crédible à grande échelle. Pendant la campagne des législatives de 2024, un bot se faisant passer pour une jeune femme avait déjà été identifié, après qu'un utilisateur ait réussi à lui faire générer une recette de tarte aux fraises. D'autres bots avaient été repérés comme étant déployés par le Kremlin pour produire un discours anti-occidental sur la guerre en Ukraine (Bernard, 2024 ; Desaunay, 2024). Lors des municipales de 2026, c'est notamment François Piquemal, candidat d'union de la gauche à Toulouse, et Sébastien Delogu, candidat LFI à Marseille, qui ont été victimes de telles opérations étrangères de dénigrement par des posts de blogs et des pages Facebook et X sur lesquelles l'engagement était piloté par des bots (Le Monde & AFP, 2026 ; Reynaud et al., 2026).

Philippe Moreau Chevrolet souligne que « *l'IA est utilisée depuis un moment pour alimenter des faux comptes sur les réseaux sociaux qui publient des messages en masse, afin de créer des mouvements artificiels. C'est de la désinformation planifiée* » (Baudoin, 2026). Cette pratique vise à persuader directement les citoyens, par exemple qu'un candidat comme Sébastien Delogu a commis des violences conjugales. L'*astroturfing* vise aussi à modifier la perception de ce que pensent « les Français » (ou « les Marseillais ») autour de soi. L'objectif est d'exploiter les mécanismes de conformité sociale et d'influence des normes perçues en faisant croire qu'une opinion est plus majoritaire qu'elle ne l'est réellement. Dans la mesure où les coûts réputationnels associés au fait de défendre une opinion décroissent quand celle-ci est perçue comme répandue (Mercier, 2017, 2020), des messages vus des milliers de fois via l'*astroturfing* peuvent ainsi à influencer les citoyens indirectement (Keller et al., 2020 ; Pinker, 2025).

2. Mise en perspective : ce que la recherche nous apprend sur les conditions du succès de la communication politique de masse

Pour évaluer l'impact réel des nouveaux outils d'IA générative dans les élections municipales comme présidentielles, il est essentiel de les situer dans le contexte des connaissances sur l'efficacité de la communication politique de masse que l'on pourrait qualifier de « statique », i.e., non-interactive (Acerbi, 2019a, 2019b ; Altay et al., 2023 ; Kalla & Broockman, 2020 ; Lewandowsky et al., 2017 ; Mercier, 2017, 2020 ; Tappin, 2026 ; Van Der Linden et al., 2018).

2.1. Les effets limités de la communication politique statique

Le pouvoir persuasif des techniques de communication politiques de masse statiques est plus faible que ne le suggère l'intuition commune ou les craintes médiatiques, et ce succès est fortement conditionné par la possession de croyances favorables au message en question chez les audiences (Adena et al., 2015).

Les études quantitatives d'impact montrent que les interventions de persuasion politique statique - publicités, messages de campagne - produisent des effets moyens de l'ordre de quelques points de pourcentage maximum, et généralement éphémères - se dissipant en quelques jours ou semaines. Ces effets faibles apparaissent contrastés par rapport au pouvoir persuasif plus élevé des interactions pouvant s'adapter aux objections spécifiques des interlocuteurs, comme permettent de le faire les LLMs (Costello et al., 2024), ou activant les capacités au *perspective taking* (de décentrement et d'empathie) des personnes visées lors de porte-à-porte (Broockman & Kalla, 2016). Ce faible pouvoir persuasif de la communication statique de masse s'explique par plusieurs mécanismes bien documentés : l'exposition sélective qui conduit les individus à éviter les messages dissonants par rapport à leurs croyances et identités politiques préexistantes (Hart et al., 2009 ; Stroud, 2010) ; une exposition à un message écrit ou en image qui ne survient qu'une ou quelques fois maximum par personne sans pouvoir s'adapter aux doutes des individus ; la résistance de l'esprit face aux tentatives de persuasion perçues comme telles (Mercier, 2017) ; le fait que le message ne provient pas d'une source humaine incarnée avec laquelle un lien de confiance existe ; et l'ancrage des préférences politiques et de vote dans des identités relativement stables, formées au cours de la socialisation des individus (Bourdieu, 1991 ; Van Bavel et al., 2024).

Ces contraintes expliquent pourquoi certains spécialistes de la persuasion politique de masse considèrent - en ce qui concerne les messages écrits notamment - la maximisation du degré d'*exposition* plutôt que l'optimisation du caractère *persuasif* du message comme le facteur le plus déterminant (Tappin, 2026). Un calcul simple l'illustre : toucher 1 000 000 de personnes avec un message politique qui persuade 1% d'entre elles produit 10 000 « conversions » (en faveur d'un vote, par exemple). Toucher 1 000 personnes avec un message 10x plus persuasif (10%) n'en produit que 100. L'IA générative de tweets ou de posts écrits, en abaissant les coûts de production de contenu, s'inscrit surtout dans cette stratégie de maximisation de la portée plutôt que dans une hypothétique « super-persuasion » algorithmique. Il faut s'attendre à ce que chaque message politique écrit, pris isolément, reste relativement peu persuasif.

2.2. Le plus grand pouvoir persuasif et désorientant des *deepfakes* vidéos ?

Ces conclusions sur les effets limités de la communication écrite ou visuelle statique méritent toutefois d'être nuancées s'agissant des *deepfakes* vidéos hyperréalistes. Comparée aux messages statiques, la vidéo constitue un médium sensoriellement riche, capable d'activer des réactions émotionnelles et morales plus fortes par le biais d'indices non verbaux (expressions faciales, ton de la voix, réalisme des scènes). Or, le photoréalisme et donc le pouvoir persuasif de ces *deepfakes* vidéos risque de ne faire que croître dans les prochaines années. Dans les cas d'opacité sur l'identité et les intentions de leur source (Mercier, 2017), ce photoréalisme pourrait accroître leur pouvoir d'influence dans les cas où les *deepfakes* vidéos semblent émaner de figures perçues comme fiables.

Cela étant dit, les études empiriques disponibles invitent à la nuance. Vaccari & Chadwick, (2020), dans une étude employant des images synthétiques mettant en scène l'ex-Président américain Obama, trouvent que les *deepfakes* « trompent » environ 14% des répondants, instillant plutôt un état d'incertitude chez 35% des participants (et un rejet chez le reste). Dans une étude plus récente, Barari et al., (2025) trouvent que les *deepfakes* vidéos ne sont ni plus ni moins persuasifs que des messages présentant le même contenu en format texte ou image. Il convient de noter que les deux études furent conduites sur des participants états-uniens, pas des Français. Le « saut persuasif » attribué à cette technologie reste donc à démontrer expérimentalement, même s'il faut s'attendre à des progrès techniques rapides dans le photo-réalisme des *deepfakes* vidéos politiques et à leur utilisation probable lors des élections présidentielles de 2027 en France.

2.3. La désinformation : un problème d'offre mais aussi de demande

Un argument fréquemment avancé est que l'IA générative va « inonder » l'espace informationnel de fausses nouvelles, submergeant les capacités de discernement des citoyens. Cette vision repose sur une conception implicite de la désinformation comme problème d'offre : plus il y a de faux contenus disponibles, plus les gens y seront exposés, et influencés.

S'il est clair que l'IA générative permet une massification de la production et de l'exposition, il importe de se rappeler que la quantité de désinformation effectivement consommée (vue, lue) dépend principalement de l'intensité de la demande. Les études sur le partage de *fake news* et de théories complotistes sur les réseaux sociaux montrent que l'essentiel des partages et des lectures est concentré sur un petit nombre de « *superspreaders* ». Ceux-ci sont souvent de plus gros consommateurs d'information politique, plus polarisés et davantage à la recherche de visibilité que la moyenne des citoyens, ou caractérisés par une faible confiance dans les institutions (Arceneaux et al., 2021 ; A. Bor et al., 2026 ; M. B. P. Bor Mathias Osmundsen, Alexander, 2021 ; A. Guess et al., 2019 ; Osmundsen et al., 2021 ; Petersen et al., 2023).

La majorité des citoyens continue de s'informer principalement via des sources établies et fiables, en nombre relativement restreint (Altay et al., 2022 ; A. M. Guess, 2021). Ainsi, augmenter l'offre de désinformation n'augmente pas mécaniquement la demande : il faut « aller chercher » les audiences, ce qui requiert des compétences de ciblage et de distribution que l'IA ne fournit pas automatiquement. Cela étant dit, l'accès croissant aux informations via les timelines des réseaux sociaux (plutôt que via les visites directes de sites type lemonde.fr) crée une vulnérabilité accrue en exposant davantage les utilisateurs à des contenus potentiellement non vérifiés circulant dans leurs réseaux.

2.4. Les LLMs comme outils d'accès à l'information experte pour les citoyens

Un aspect généralement négligé du débat public sur les progrès de l'intelligence artificielle générative concerne le potentiel des LLM pour *améliorer*, plutôt que dégrader, la qualité de l'information politique accessible aux citoyens. Selon des sondages Ipsos BVA, la proportion de Français ayant recours à l'IA générative pour effectuer des recherches est en croissance très rapide : environ 20% en 2024, 39% en 2025 et 68% en janvier 2026, avec en tête ChatGPT, utilisé par un Français sur deux (Zulfikarpasic, 2026). Près d'un Français sur deux (48%) a déjà utilisé ou envisage d'utiliser l'IA générative pour se renseigner sur la politique, cherchant principalement à connaître les mesures d'un programme (39%) ou les prises de position d'une personnalité (31%).

Or, les capacités de synthèse de ces outils sont remarquables et pourraient contribuer à réduire certains effets délétères du fractionnement des sources médiatique et de la polarisation des croyances qui en résulte, en facilitant l'établissement de faits communs (*ground truths*) (Matthews, 2026 ; Williams, 2025). C'est en tout cas l'avis donné par un nombre croissant de chercheurs en intelligence artificielle et en sciences humaines interrogés dans le cadre de cette note de recherche², ou s'étant exprimés sur le site de blogging *Substack* (Matthews, 2026 ; Williams, 2025). Un exemple frappant nous vient des États-Unis, où des tests ont montré que les LLM produisent des résumés remarquablement équilibrés et factuels d'événements politiques controversés comme la mort d'Alex Pretti assassiné par des agents fédéraux des Border Patrol, en janvier 2026. L'administration Trump et *Fox News* ont relayé la thèse fautive selon laquelle Pretti avait menacé les agents avec son arme, alimentant des « réalités parallèles » distinctes chez les supporters du camp Trump et les Démocrates. Les synthèses générées par les LLMs, en revanche - y compris *Grok* de l'entreprise X, dont Elon Musk, proche de D. Trump, est propriétaire - offraient une synthèse beaucoup plus factuelle, distinguant ce qui relevait de l'établi (la mort de Pretti par balles) de ce qui demeurerait contesté (l'accusation que Pretti constituait une menace). Cette observation d'une séparation, par *Grok*, des faits établis sur la mort de Pretti du récit de l'administration Trump a été répliquée par l'auteur de cette note le 15 mars 2026.

[2] Dan Williams, philosophe à l'université de Sussex ; Philippe Moreau Chevolet, senior advisor chez No Com, enseignant à Sciences Po ; Sacha Altay, chercheur en mésinformation ; Peter Barrett, enseignant en dépoliarisation, ESSEC ; Benjamin Icard, chercheur IA sur la désinformation ; David Mas, chief AI officer, [Make.org](https://www.make.org) ; Paul Égré, chercheur au CNRS ; Dimitri Courant, chercheur en science politique, Sciences Po ; Chiara Marcocchia, philosophe, ENS ; Thomas Souverain, chercheur en éthique de l'IA, CEA Paris-Saclay ; Lionel Page, économiste à l'université de Queensland, Australie.

L'observation tend à se généraliser actuellement (mars 2026) à la plupart des thèmes controversés comme le changement climatique, l'efficacité des vaccins ou les effets de l'immigration sur l'emploi. De même pour la recherche d'un programme de candidat aux élections municipales. Un LLM fournit en un clic un résumé structuré de l'état des connaissances factuelles ou scientifiques sur le sujet en question. Il ne cherche pas à vous enrôler dans un camp. Il s'en tient aux faits et isole les interprétations idéologiques qui peuvent en être faites, comme pour nous avertir. Les grands LLMs actuels sont paramétrés pour donner l'éventail des positions sur les questions controversées. Ce qui peut aider les citoyens à prendre conscience de la complexité des enjeux, et les encourage à l'humilité intellectuelle. Ainsi, là où les réseaux sociaux ont pu servir à la démocratisation du débat public, les LLM actuels peuvent donc constituer une force d'amélioration de l'information (Matthews, 2026 ; Williams, 2025). L'une des raisons de ce phénomène, c'est que les entreprises qui développent les LLMs sont en concurrence entre elles pour fournir de l'information fiable dans une myriade de domaines - du bricolage aux compétitions sportives en passant par la géopolitique et la finance - afin d'attirer la confiance d'un large public, et de multiplier les abonnements.

Cette analyse optimiste doit certes être tempérée par les risques d'hallucinations des LLM (génération de fausses informations présentées comme des faits). Ces hallucinations sont toutefois en chute rapide depuis 2022, à mesure que les entreprises qui les créent sont en compétition intense pour en éliminer l'occurrence. Cette analyse optimiste de l'influence des LLM ne doit pas non plus mettre de côté les questions de souveraineté sur les algorithmes comme sur les données que pose la domination d'outils américains dans ce marché. Mais elle invite à ne pas laisser le discours public sur l'impact de l'IA sur l'information politique se focaliser sur ses seuls effets négatifs, et à résister au biais de négativité (Altay et al., 2023 ; Baumeister et al., 2001).

3. Les risques réels : ingérence étrangère et érosion de la confiance

3.1. La menace documentée des ingérences étrangères (faux sites, *astroturfing*, etc.)

Les ingérences dans les élections françaises émanant de la Russie et de ses relais constituent une menace plus sérieuse, disposant de ressources et de motivations que les acteurs domestiques n'ont généralement pas. Ces opérations ne visent pas nécessairement à faire élire un candidat local particulier, mais plutôt à fragmenter l'opinion publique française, éroder la confiance dans les institutions démocratiques, et nuire à la réputation de candidats pro-européens et soutenant activement l'implication militaire de la France en Ukraine. Avec en ligne de mire, au-delà des élections municipales, l'élection présidentielle de 2027.

Dans le cadre des municipales, les services de l'État ont alerté sur des stratégies de manipulation sophistiquées : faux sites imitant des médias régionaux, mais non maintenus par des journalistes professionnels, et campagnes coordonnées sur les réseaux sociaux notamment (Singh, 2026). En ce qui concerne les faux sites d'information régionale, « *Soit ce sont des sites créés de toutes pièces, soit ils ont vocation à imiter des médias déjà existants de manière à peser sur le débat public en s'inspirant du crédit et de la confiance qu'ont ces marques en termes d'information* », commente Nicolas Arpagian, directeur de la stratégie Jizô AI et expert en cybersécurité (Singh, 2026).

Figure 5 - Faux sites d'information régionale générés par des services russes



Le candidat Renaissance-Horizons, Pierre-Yves Bournazel, a été la cible d'une opération d'ingérence numérique attribuée au réseau russe *Storm-1516*, documentée par *Viginum* dans un rapport publié début mars 2026 (France Info, 2026 ; Libération & AFP, 2026). L'attaque a pris la forme d'un site internet usurpant l'identité visuelle de sa campagne officielle, diffusant une fausse information selon laquelle le candidat prévoyait de transformer le Centre Pompidou en centre d'accueil pour migrants. Le but était de discréditer des candidats perçus comme pro-européens ou atlantistes, et d'attiser les tensions sociales autour de l'immigration. L'opération fut immédiatement dénoncée par l'équipe de campagne et plusieurs médias (Delorme, 2026).

Le gouvernement a surveillé plusieurs dizaines de villes, dont Le Havre où se représente l'ancien Premier ministre Édouard Philippe, qui s'est déclaré à la course à l'Élysée (Singh, 2026). Les autorités considèrent ces municipales comme un « test » avant l'élection présidentielle de 2027, période où les tentatives d'ingérence seront vraisemblablement intensifiées.

3.2. Le “dividende du menteur” : quand le soupçon généralisé devient le problème

Un effet pervers de la médiatisation des capacités de l'IA générative dans les campagnes politiques mérite attention : la généralisation du soupçon envers tout contenu numérique. Lorsque chaque image, chaque vidéo, chaque déclaration peut être suspectée d'avoir été fabriquée, c'est la notion même de *preuve visuelle* qui se trouve mise en doute. L'image et la vidéo, jadis considérées comme preuves irréfutables, ne suffisent plus à convaincre.

Ce phénomène, parfois appelé « dividende du menteur » (*liar's dividend*), permet à des acteurs comme des politiciens de mauvaise foi de rejeter toute preuve compromettante comme potentiellement falsifiée (Altay et al., 2020a ; Schiff et al., 2025 ; Vaccari & Chadwick, 2020). Il alimente un scepticisme informationnel qui peut conduire au désengagement civique : si l'on ne peut plus rien croire, à quoi bon s'informer³ ? Cette dynamique est potentiellement favorable aux théories complotistes, émanant d'acteurs qui peuvent désormais invoquer la possibilité technique du trucage pour rejeter tout élément contredisant leurs narratifs⁴. Alors que les électeurs se sentent déjà étrangers à la politique (Cevipof, 2026), il est possible que le fake aggrave la défiance.

Conclusion : Vers une analyse vigilante mais sans catastrophisme de la diffusion récente de l'IA générative dans la communication politique

Au terme de cette analyse, plusieurs pistes se dégagent pour analyser les défis posés par l'IA générative aux élections sans verser dans un catastrophisme excessif.

Des bénéfices pour les médias professionnels affichant un non-usage de l'IA générative ?

Une conséquence potentiellement bénéfique de la banalisation des sources soupçonnées d'être trompeuses pourrait être de pousser les citoyens à se recentrer sur des sources d'information dont ils savent qu'elles sont tenues par des journalistes professionnels suivant des règles de déontologie. Le phénomène a déjà été observé dans une étude couvrant plusieurs pays européens et les États-Unis pendant la pandémie de Covid-19, au cours de laquelle une majorité de citoyens se tournèrent vers les médias traditionnels afin de se documenter sur les développements de la pandémie et les gestes à adopter (Altay et al., 2022).

[3] Ce phénomène de scepticisme généralisé est fréquent dans les contextes autocratiques où, suspectant tout message institutionnel ou politique de relever de la propagande du régime, certains citoyens non-favorables au gouvernement en place cessent de croire à tout média officiel dans leur pays.

[4] Il convient toutefois de souligner que les acteurs susceptibles de bénéficier du dividende du menteur risquent d'être ceux ne bénéficiant déjà *pas* d'une réputation épistémique de source fiable - le président américain Donald Trump en constituant l'illustration parfaite. Une représentante politique dont *on attend* qu'elle soit honnête et dise des choses vraies risque d'endommager gravement sa réputation épistémique si elle vient à contredire de front le témoignage convergent de multiples médias ou acteurs jugés bien informés. Une étude de Altay et al., (2020b) suggère qu'une réputation d'informateur fiable se perd vite et se reconstitue difficilement, de façon asymétrique. Tenter d'exploiter le dividende du menteur peut donc s'avérer risqué.

Pour que ce recours aux médias professionnels et fiables fonctionne, deux conditions au moins doivent être remplies. Premièrement, il est crucial d'enseigner les règles sur lesquelles repose le journalisme professionnel auprès de la population afin que les citoyens comprennent l'origine de cette fiabilité : factualité (vs. interprétations partisans), recoupement des sources, interrogation des experts, distinction entre information et opinion, correction des erreurs, évitement du sensationnalisme. La diffusion d'une connaissance critique des sources et des mécanismes de production de l'information par les journalistes professionnels était déjà essentielle avant l'essor de l'IA générative, et elle le devient davantage encore (Hubé & Kervella, 2024 ; Scheufele & Krause, 2019).

Deuxièmement, pour que le soupçon envers les contenus synthétiques ne glisse pas vers une défiance généralisée, il est essentiel que les médias professionnels signalent clairement (i.e., fasse connaître) qu'ils ne font pas d'usage non supervisé de l'IA dans la production de leurs textes et images. La crédibilité du journalisme professionnel est plus que jamais un bien public à protéger activement, y compris par des mécanismes de financement qui préservent son indépendance et sa capacité à produire une information professionnelle de qualité (Cagé & Huet, 2021).

Exiger la transparence plutôt que l'interdiction générale

Les usages légitimes de l'IA générative en campagne (génération d'affiches, *chatbots* programmatiques, synthèses), qu'elles soient municipales ou présidentielles, ne posent pas de problème démocratique majeur s'ils sont signalés comme tels. L'enjeu est plutôt d'imposer une transparence sur leur utilisation. Le *marquage obligatoire* des contenus générés par IA, s'il peut être techniquement appliqué et contrôlé, constitue une piste régulatoire prometteuse qui permettrait aux citoyens de calibrer leur confiance en connaissance de cause.

Concentrer les efforts sur l'*astroturfing* et les ingérences étrangères

Les *deepfakes* visuels, comme les images synthétiques d'agression de rue de Sarah Knafo, captent l'attention médiatique. Toutefois, la menace la plus sérieuse réside peut-être dans l'amplification artificielle de certains messages textes via des réseaux de bots et de faux comptes, ainsi que dans les opérations d'ingérence étrangère. L'amplification de messages par des puissances étrangères est souvent pernicieuse car les messages cherchent à « surfer » sur des croyances et des récits politiques déjà bien implantés auprès de la population (Acerbi, 2019a ; Marie et al., 2023 ; Mercier, 2017). Les efforts de régulation et de surveillance, notamment ceux de *Viginum* (Service de vigilance et de protection contre les ingérences numériques étrangères), devraient prioritairement cibler ces pratiques d'*astroturfing* et ces opérations coordonnées.

Les efforts pour produire du contre-discours sont eux aussi importants, mais selon des modalités qui permettent à ce contre-discours de persuader. En évitant, par exemple, de rendre manifeste que la source est gouvernementale (avec un logo Marianne), car de telles opérations n'ont guère de chances de persuader les publics déjà bas en confiance institutionnelle et tentés par le complotisme (Cordonier & Cafiero, 2024 ; Oliver & Wood, 2014 ; Petersen et al., 2023).

Éviter le catastrophisme technologique qui alimente la défiance

Enfin, la couverture médiatique et le discours politique sur le développement de l'IA générative dans la communication politique doivent éviter d'exagérer les capacités manipulatoires de l'IA, et résister au biais de négativité qui nous piège tous quand nous raisonnons sur les questions de progrès technologique et social (Baumeister et al., 2001). Un catastrophisme excessif dans le discours sur les usages de l'IA est non-justifié, et risque paradoxalement de renforcer l'effet redouté : une défiance généralisée envers toute information, favorable au désengagement civique. Et, dans le pire des cas, favorable aux manipulateurs qui pourront rejeter toute preuve comme potentiellement falsifiée (si les citoyens sont déjà acquis à l'idée qu'aucune information n'est fiable). La communication sur ces sujets doit s'efforcer de garder un équilibre entre *informer* sur les risques réels, indéniables, sans *alimenter* une panique disproportionnée par les effets performatifs de constats trop pessimistes.

Références

- Acerbi, A. (2019a). Cognitive attraction and online misinformation. *Palgrave Communications*, 5(1), 1–7.
- Acerbi, A. (2019b). *Cultural evolution in the digital age*. Oxford University Press.
- Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of The Nazis in Prewar Germany *. *The Quarterly Journal of Economics*, 130(4), 1885–1939.
- Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *Journal of Politics*, 80(3), 964–981.
- Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media and Society*, 9(1).
- Altay, S., Hacquin, A. S., & Mercier, H. (2020a). Why do so few people share fake news? It hurts their reputation. *New Media and Society*.
- Altay, S., Hacquin, A.-S., & Mercier, H. (2020b). Why do so few people share fake news? It hurts their reputation. *New Media & Society*, 1461444820969893.
- Altay, S., Nielsen, R. K., & Fletcher, R. (2022). Quantifying the “infodemic”: People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media*, 2.

Arceneaux, K., Gravelle, T. B., Osmundsen, M., Petersen, M. B., Reifler, J., & Scotto, T. J. (2021). Some people just want to watch the world burn: The prevalence, psychology and politics of the “Need for Chaos.” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1822).

Barari, S., Lucas, C., & Munger, K. (2025). Political deepfakes are as credible as other fake media and (sometimes) real media. *The Journal of Politics*, 87(2), 510–526.

Barbaza, Q. (2026, March 12). Comment l’IA a bouleversé la campagne aux municipales. *Sud Radio*. <https://www.sudradio.fr/politique/slogans-programmes-affiches-strategies-comment-lia-a-bouleverse-la-campagne-des-municipales>.

Baudoin, P. de. (2026, June 2). *Chatbot, images générées artificiellement... Comment “l’IA a contaminé les campagnes électorales” pour les municipales ?* France 3 Paris Île-de-France. <https://france3-regions.franceinfo.fr/paris-ile-de-france/yvelines/chatbot-images-geneeres-artificiellement-comment-l-ia-a-contamine-les-campagnes-electorales-pour-les-municipales-3294018.html>.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.

Belhassen, A., & AFP. (2025, November 18). Élections municipales 2026 : “L’IA peut être un formidable outil”, quand la technologie s’incruste dans la campagne électorale. France 3 Auvergne-Rhône-Alpes. <https://france3-regions.franceinfo.fr/auvergne-rhone-alpes/isere/grenoble/elections-municipales-2026-l-ia-peut-etre-un-formidable-outil-quand-la-technologie-s-incruste-dans-la-campagne-electorale-3251104.html>.

Bernard, H. (2024, June 24). Ce bot d’IA sévit sur X (Twitter) pour parasiter les élections législatives. *Rotek*. <https://rotek.fr/bot-ia-sevit-x-twitter-parasiter-elections-legislatives/>.

Bor, A., Marie, A., Pradella, L., & Petersen, M. B. (2026). Social media users experience more political hostility in less economically equal and less democratic societies. *Nature Human Behaviour*.

Bor, M. B. P., Mathias Osmundsen, Alexander. (2021). Beyond Populism: The Psychology of Status-Seeking and Extreme Political Discontent*. *In The Psychology of Populism*. Routledge.

Bourdieu, P. (1991). *Language and Symbolic Power*. Harvard University Press.

Bournazel, P.-Y. (2025, December 6). *Post sur X [Tweet]*. Twitter. <https://x.com/pybournazel/status/1997375662727082384>.

Bouvier, M., Salem, N., & Murviedro, N. (2026, January 12). *Municipales 2026 : Ces candidats qui se servent de l’intelligence artificielle pour faire campagne*. Franceinfo. https://www.franceinfo.fr/replay-jt/france-2/13-heures/municipales-2026-ces-candidats-qui-se-servent-de-l-intelligence-artificielle-pour-faire-campagne_7736668.html.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 3352(6282), 81–87.

Brossat, I. (2025, May 21). *Post X* [Tweet]. Twitter. <https://x.com/lanBrossat/status/1925076635365085208>.

Cagé, J., & Huet, B. (2021). *L'information est un bien public. Refonder la propriété des médias*. Le Seuil.

Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*. Le Seuil.

CDJM. (1918). Les chartes. *CDJM*. <https://cdjm.org/les-chartes/>.

Cevipof. (2026). *Baromètre de la confiance politique CEVIPOF 2026 : La confiance s'effondre en politique, la proximité fait figure de refuge*. Cevipof. <https://www.sciencespo.fr/cevipof/fr/actualites/barometre-de-la-confiance-politique-cevipof-2026-la-confiance-s-effondre-en-politique-la-proximite-fait-figure-de-refuge/>.

Chavalarias, D. (2022). *Toxic Data*. Flammarion.

Conseil d'État, Inédit au recueil Lebon ____ (Tribunal 2025). https://www.legifrance.gouv.fr/ceta/id/CETATEXT000052398197?init=true&page=1&query=505407&searchField=ALL&tab_selection=all.

Corcos, L. (2026, January 6). *La maire de Guéret porte plainte après la diffusion d'images trompeuses générées par intelligence artificielle - ICI*. ICI, le média de la vie locale. <https://www.francebleu.fr/nouvelle-aquitaine/creuse-23/gueret/la-maire-de-gueret-porte-plainte-apres-la-diffusion-d-images-trompeuses-generées-par-l-intelligence-artificielle-5477174>.

Cordonier, L., & Cafiero, F. (2024). Public sector corruption is fertile ground for conspiracy beliefs: A comparison between 26 Western and non-Western countries. *Social Science Quarterly*.

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814.

de Florival, G., Bouvier, M., & et al. (2026, January 29). *Municipales : Quand l'intelligence artificielle s'invite dans la campagne*. Franceinfo. https://www.franceinfo.fr/elections/municipales/municipales-quand-l-intelligence-artificielle-s-invite-dans-la-campagne_7773023.html.

Delorme, T. (2026, March 13). *Un candidat à la mairie de Paris victime d'une ingérence étrangère ?* Sopra Steria FR. <https://www.soprasteria.fr/perspectives/details/un-candidat-a-la-mairie-de-paris-victime-d-une-ingerence-etrangere>.

Desaunay, D. (2024, June 20). *Surprise, les trolls russes utilisent ChatGPT pour leur propagande*. RFI. <https://www.rfi.fr/fr/podcasts/un-monde-de-tech/20240620-surprise-les-trolls-russes-utilisent-chatgpt-pour-leur-propagande>.

France Info. (2026, March 6). *Municipales 2026 : Pierre-Yves Bournazel, candidat à la mairie de Paris, a été la cible d'une attaque informationnelle russe*. Franceinfo. https://www.franceinfo.fr/elections/municipales-2026-pierre-yves-bournazel-candidat-a-la-mairie-de-paris-a-ete-la-cible-d-une-attaque-informationnelle-russe_7849616.html.

France Info, & Ici Creuse. (2026, January 6). *La maire de Guéret compte porter plainte après la diffusion d'images trompeuses générées par IA*. Franceinfo. https://www.franceinfo.fr/france/nouvelle-aquitaine/creuse/la-maire-de-gueret-compte-porter-plainte-apres-la-diffusion-d-images-trompeuses-generees-par-ia_7724302.html.

Gaumont, N., Panahi, M., & Chavalarias, D. (2018). Reconstruction of the socio-semantic dynamics of political activist Twitter networks - Method and application to the 2017 French presidential election. *PLOS ONE*, 13(9), e0201879.

Golan, G. J., & Lim, J. S. (2016). Third-Person Effect of ISIS's Recruitment Propaganda: Online Political Self-Efficacy and Social Media Activism. *International Journal of Communication*, 10(0), Article 0.

Gorodzeisky, A., & Semyonov, M. (2020). Perceptions and misperceptions: Actual size, perceived size and opposition to immigration in European societies. *Journal of Ethnic and Migration Studies*, 46(3), 612–630.

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*, 65(4), 1007–1022.

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Asian-Australasian Journal of Animal Sciences*, 32(2), 1–9.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555.

Heine, B., & Narrog, H. (2015). *The Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Herrero, O. (2026, January 16). *Municipales : Comment l'IA générative fait irruption dans la campagne*. POLITICO. <https://www.politico.eu/article/municipales-comment-lia-generative-fait-irruption-dans-la-campagne/>.

Hubé, N., & Kervella, A. (2024). Introduction du dossier : (Ré)éduquer aux médias. Une action publique contre les « désordres informationnels ». *Politiques de communication*, 23(2), 5–31.

Kalla, J. L., & Broockman, D. E. (2020). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review*, 114(2), 410–425.

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. *Political Communication*, 37(2), 256–280.

Ksiazienicki, S. (2026, February 3). À Reims, Stéphane Lang fait parler son affiche grâce à l'IA. L'Hebdo du Vendredi. <https://www.lhebdoouvendredi.com/article/54883/a-reims-stephane-lang-fait-parler-son-affiche-grace-a-lia>.

Le Monde, & AFP. (2026, March 28). À Toulouse, François Piquemal dépose un recours sur l'élection municipale pour « ingérences numériques étrangères » après sa défaite face à Jean-Luc Moudenc. *Le Monde*. https://www.lemonde.fr/politique/article/2026/03/28/a-toulouse-francois-piquemal-depose-un-recours-sur-l-election-municipale-pour-ingerences-numeriques-etrangees-apres-sa-defaite-face-a-jean-luc-moudenc_6675068_823448.html.

Lees, J., & Cikara, M. (2019). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369.

Libération, & AFP. (2026, March 6). *Municipales 2026 : Pierre-Yves Bournazel, candidat à Paris, visé par une opération d'ingérence liée à un groupe russe*. Libération. https://www.liberation.fr/politique/municipales-2026-pierre-yves-bournazel-candidat-a-paris-ise-par-une-operation-dingerence-liee-a-un-groupe-russe-20260306_Z2K5JP2RSNBXHKWW26SIMIUBA4/.

Lippmann, W. (1929). *Public Opinion: By Walter Lippmann*. Macmillan Company.

LOI N° 2018-1202 du 22 décembre 2018 Relative à La Lutte Contre La Manipulation de l'information (1), 2018-1202 (2018).

Marie, A., Altay, S., & Strickland, B. (2023). Moralization and extremism robustly amplify myside sharing. *PNAS Nexus*, 2(4), pgad078.

Marie, A., & Petersen, M. B. (2022). Political conspiracy theories as tools for mobilization and signaling. *Current Opinion in Psychology*, 48.

Marie, A., & Petersen, M. B. (2024). Moralization of Rationality Can Stimulate Sharing of Hostile and False News on Social Media, but Intellectual Humility Inhibits it. *Political Communication*, 1–28.

Marie, A., & Petersen, M. B. (2025). Motivations to connect with like-minded audiences increase partisan sharing on social media. *PNAS Nexus*, 4(7), pgaf197.

Matthews, D. (2026). *Pro-social media. Substack*. <https://substack.com/home/post/p-184342590?selection=c8fed537-cfa9-44e9-a2b1-73dc38bb3604>.

M.D. (2026, February 1). *L'intelligence artificielle s'invite dans la campagne de Stéphane Lang à Reims*. <https://www.lunion.fr/id783241/article/2026-02-01/lintelligence-artificielle-sinvite-dans-la-campagne-de-stephane-lang-reims>.

Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology*, 21(2), 103–122.

Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proceedings of the National Academy of Sciences*, 119(16), e2116851119.

Muxel, A., & Perrineau, P. (2026). *Inventaire des peurs françaises - Éditions Odile Jacob*. [www.odilejacob.fr. https://www.odilejacob.fr/catalogue/sciences-humaines/questions-de-societe/inventaire-des-peurs-francaises_9782415014421.php](https://www.odilejacob.fr/catalogue/sciences-humaines/questions-de-societe/inventaire-des-peurs-francaises_9782415014421.php).

Neihouser, M., Haute, T., Nostitz, F.-C. von, & Sandri, G. (2023). Les réseaux socionumériques dans la campagne présidentielle de 2022 en France : Diversité des usages, diversité des usager·e·s? *Revue Française de Science Politique*, 72(6), 977.

Ny Ando, A. (2026, January 7). Des images IA embarrassantes d'élus circulent à Guéret, la maire porte plainte. *LEBIGDATA.FR*. <https://www.lebigdata.fr/des-images-ia-embarrassantes-delus-circulent-a-gueret-la-maire-porte-plainte>.

Oliver, J. E., & Wood, T. J. (2014). Conspiracy Theories and the Paranoid Style(s) of Mass Opinion. *American Journal of Political Science*, 58(4), 952–966.

Orben, A. (2020). The Sisyphean Cycle of Technology Panics. *Perspectives on Psychological Science*, 15(5), 1143–1157.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, 115(3), 999–1015.

Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: How disinformation and demagogues coordinate rather than manipulate. *Current Opinion in Psychology*, 35, 71–75.

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2023). The “Need for Chaos” and Motivations to Share Hostile Political Rumors. *American Political Science Review*, 117(4), 1486–1505.

Pinker, S. (2025). *When Everyone Knows That Everyone Knows...: Common Knowledge and the Mysteries of Money, Power, and Everyday Life*, chapter “The Canceling Instinct.” Scribner.

Reynaud, F., Rof, G., & Leloup, D. (2026, March 10). Sébastien Delogu, François Piquemal et LFI ciblés par une campagne de désinformation en pleine campagne des municipales. *Le Monde*. https://www.lemonde.fr/politique/article/2026/03/09/sebastien-delogu-francois-piquemal-et-lfi-cibles-par-une-campagne-de-desinformation-en-pleine-campagne-des-municipales_6670119_823448.html.

- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), 7662–7669.
- Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025). The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? *American Political Science Review*, 119(1), 71–90.
- Sijlmassi, A., Mercier, H., Safra, L., & Chevallier, C. (2025). Information about immigrants' deservingness reduces misperceptions and opposition to immigration. *PNAS Nexus*, 4(12), pgaf349.
- Singh, K. (2026, September 2). *Faux sites et images générées par l'IA : Qui cherche à nous manipuler sur Internet à l'approche des élections municipales?* France 3 Normandie. <https://france3-regions.franceinfo.fr/normandie/seine-maritime/faux-sites-et-images-generées-par-l-ia-qui-cherche-a-nous-manipuler-sur-internet-a-l-approche-des-elections-municipales-3293664.html>.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556–576.
- Tappin, B. (2026). For Digital Mass Persuasion, Exposure Matters More Than Persuasiveness. *PsyArXiv*. https://osf.io/preprints/psyarxiv/5ewbu_v1.
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6(1), 2056305120903408.
- Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*, 56, 101787.
- Van Der Linden, S., Leiserowitz, A., & Maibach, E. (2018). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour*, 2(1), 2–3.
- Williams, D. (2025, March 3). How AI Will Reshape Public Opinion. *Substack*. <https://substack.com/@conspicuouscognition/p-188535644>.
- Zulfikarpasic, A. (2026, May 2). Les LLM's : Nouveaux acteurs des campagnes électorales? *Fondation Jean-Jaurès*. <https://www.jean-jaures.org/publication/les-llms-nouveaux-acteurs-des-campagnes-electorales/>.