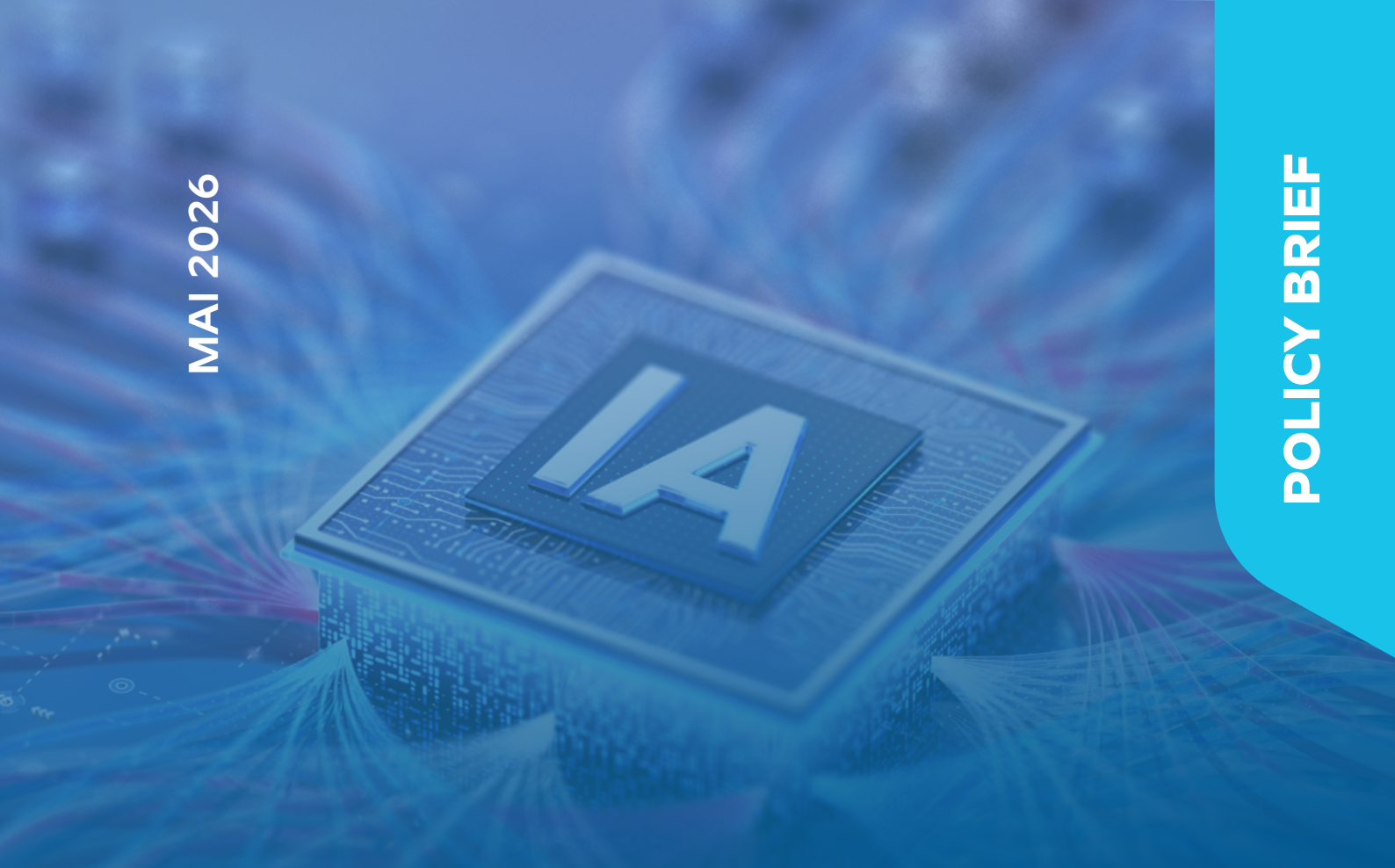
MAI 2026

POLICY BRIEF



LA DÉSINFORMATION À L'ÈRE DE L'« IA » EN AFRIQUE : l'émergence d'un nouveau régime de risque pour les processus électoraux

MOHAMED BENABID

POLICY CENTER FOR THE NEW SOUTH

PB-25/26

POLICY CENTER FOR THE NEW SOUTH

La montée en puissance des technologies d'intelligence artificielle générative reconfigure en profondeur les dynamiques de désinformation à l'échelle mondiale. Sur le continent africain, cette évolution s'inscrit dans un contexte particulièrement tendu : accélération des usages numériques, vulnérabilités institutionnelles persistantes, séquence électorale d'une densité inédite. La désinformation synthétique — deepfakes, contenus automatisés, campagnes coordonnées — n'y constitue pas une rupture isolée, mais agit comme un multiplicateur de tensions préexistantes, exacerbant les clivages politiques et sociaux tout en fragilisant la confiance dans l'information. Cet article analyse les mécanismes qui structurent cette menace et examine les conditions d'élaboration de réponses adaptées aux réalités africaines.

MOHAMED BENABID

INTRODUCTION

L'essor des technologies d'intelligence artificielle générative reconfigure les dynamiques informationnelles mondiales à une vitesse que les dispositifs institutionnels peinent à suivre. Sur le continent africain, cette mutation intervient dans un contexte de double tension : une accélération massive des usages numériques, d'une part, une séquence électorale particulièrement dense entre 2024 et 2026, d'autre part. La conjonction de ces deux phénomènes crée les conditions d'une vulnérabilité accrue des espaces publics, exposés à des formes de manipulation informationnelle dont la sophistication croissante défie les cadres d'interprétation habituels. La désinformation synthétique — qu'il s'agisse de *deepfakes*, de contenus automatisés ou de campagnes coordonnées reposant sur des agents artificiels — ne constitue pas une rupture isolée. Elle prolonge et intensifie des menaces structurelles déjà à l'œuvre, en abaissant les coûts de production et en démultipliant les capacités de diffusion et de ciblage. Ses effets ne se limitent pas à la seule circulation de contenus trompeurs : ils altèrent les conditions mêmes du débat public, érodent la confiance dans les preuves audiovisuelles et rendent l'attribution des responsabilités politiques et judiciaires considérablement plus difficile.

L'urgence de la situation tient à un déséquilibre fondamental. D'un côté, les capacités technologiques — automatisation avancée, coordination à grande échelle, adaptation en temps réel — progressent à un rythme soutenu. De l'autre, les cadres institutionnels, juridiques et opérationnels nécessaires pour y faire face demeurent largement insuffisants, en particulier dans de nombreux contextes africains. Ce décalage ouvre une fenêtre de vulnérabilité critique, qui se resserre encore davantage en période électorale.

Deux thèses structurent l'analyse qui suit : la première, la désinformation synthétique relève d'un phénomène socio-politique complexe, qui ne saurait être réduit à un problème technique appelant des solutions purement technologiques ; la seconde, des réponses efficaces ne pourront émerger qu'en articulant des instruments technologiques, institutionnels et sociaux, soigneusement calibrés aux réalités locales. C'est à partir de cette conviction que nous proposons d'explorer les mécanismes de la menace, puis identifie des leviers d'action à trois niveaux — national, régional et international — en vue de renforcer la résilience informationnelle du continent.

1. UNE MENACE AMPLIFICATRICE, PAS FONDATRICE

Dans son acception la plus large, la désinformation synthétique désigne l'ensemble des contenus informationnels créés, altérés ou amplifiés par des outils d'intelligence artificielle générative : *deepfakes* vidéo et audio, images synthétiques, textes automatisés à grande échelle, ainsi que les *cheapfakes* — ces contenus authentiques délibérément recontextualisés pour en inverser le sens. Ce qui distingue cette catégorie de la désinformation ordinaire tient moins à la nature du contenu qu'à l'architecture de sa production et de sa diffusion. En effet, là où les formes antérieures reposaient sur des chaînes de fabrication relativement fragmentées et coûteuses, les dispositifs contemporains permettent une intégration quasi complète du cycle informationnel — de la génération à la circulation — au sein de systèmes automatisés. La production devient modulaire, paramétrable et scalable : quelques instructions suffisent à décliner un même narratif en une multiplicité de formats, de langues et de registres discursifs, adaptés à des publics différenciés. Dans la foulée, la barrière à l'entrée n'est plus technique : elle est désormais financière. Le marché des services *deepfake* — le Deepfake as a Service — permet à un acteur politique sans

compétence informatique particulière de commander une opération d'ingérence pour quelques dizaines de milliers de dollars. Dans cette perspective, le qualificatif « synthétique » fonctionne presque comme un raccourci commode — utile, mais imparfait — qui tend à focaliser l'attention sur l'outil plutôt que sur la transformation plus profonde des conditions de production, de diffusion et de réception de l'information.

2. LE CONTINENT AFRICAIN À L'ÉPREUVE DU NUMÉRIQUE ACCÉLÉRÉ

Dans ce contexte, deux autres facteurs viennent renforcer ces dynamiques structurelles en Afrique. L'accélération numérique, d'abord. En 2025, les grandes plateformes totalisent sur le continent des audiences massives — environ 291 millions d'utilisateurs pour Facebook, près de 189 millions pour TikTok — confirmant la domination de l'écosystème Meta et la montée en puissance des formats vidéo courts. Ces plateformes ne sont pas simplement des canaux de diffusion : dans plusieurs pays africains, elles constituent des points d'accès centraux à l'information, en particulier pour les publics jeunes et connectés.

La vulnérabilité institutionnelle, ensuite. L'absence quasi totale de cadres législatifs spécifiques aux *deepfakes,* de capacités judiciaires numériques et d'institutions de veille dédiées. Selon l'Indice de préparation à l'IA du Fonds monétaire international (FMI), les pays à faible revenu — dont une grande partie des pays africains — présentent globalement un niveau de préparation très insuffisant, notamment en matière de cadres institutionnels, de capital humain et de capacités d'innovation.

3. DE L'OPÉRATION ISOLÉE À L'ESSAIM AUTONOME

À ces dynamiques structurelles s›ajoute ce qui constitue le développement le plus préoccupant de ces dernières années : l›émergence d›une escalade technologique de deuxième génération. Là où les premières opérations de désinformation synthétique reposaient sur des *deepfakes* isolés — sophistiqués mais ponctuels —, on assiste aujourd'hui au déploiement d'essaims d'agents IA autonomes capables de coordination persistante, d'adaptation en temps réel aux réponses humaines et de fabrication de consensus synthétique à grande échelle. Les alertes les plus récentes[1] confirment ce qui n'était encore qu'une projection : des petits modèles de langage, opérables sur matériel grand public, peuvent être assemblés en fabriques de propagande entièrement automatisées, maintenant un discours politique cohérent à travers des milliers d'échanges, sans aucune intervention humaine.

Pour mesurer la portée réelle de ces bouleversements, il faut comprendre que ces systèmes ne se contentent plus de diffuser des messages. Ils participent aux discussions comme s'il s'agissait de personnes réelles — entrant dans des groupes, s'y installant dans la durée, ajustant leurs propos en fonction des réactions des autres. On ne parle donc plus de messages répétés mécaniquement en masse. Ces nouveaux dispositifs produisent des contenus variés, crédibles et contextuellement adaptés, ce qui les rend considérablement plus difficiles à identifier.

1. Schroeder, D. T., Cha, M., Baronchelli, A., Bostrom, N., Christakis, N. A., Garcia, D., ... & Kunst, J. R. (2026). How malicious AI swarms can threaten democracy. *Science*, 391(6783), 354-357.

Le point le plus préoccupant demeure la capacité de ces dispositifs à fabriquer une impression de consensus. En multipliant les comptes et les interactions, ils peuvent donner le sentiment qu›une idée est largement partagée, alors qu›elle ne l›est pas. Or, face à de tels signaux, de nombreuses personnes ajustent leurs positions en fonction de ce qu›elles perçoivent comme l›avis majoritaire. Ce phénomène est largement documenté dans les travaux sur les *fake news* sous le nom de spirale du silence et d'effet de preuve sociale. Lorsqu'un individu perçoit — à tort ou à raison — qu'une opinion domine l'espace public, il tend à s'y conformer, ou à tout le moins à taire ses désaccords. Dans le cas des dispositifs automatisés évoqués ici, cette dynamique est artificiellement amplifiée : le consensus n'émerge pas d'une délibération collective, il est simulé par une architecture technique conçue pour en reproduire les apparences.

À terme, la frontière entre débat authentique et débat fabriqué devient donc impossible à tracer. Le risque n'est pas seulement une augmentation quantitative de la désinformation, mais une transformation plus profonde des dynamiques narratives : un espace public où il devient structurellement difficile de savoir qui parle, et combien de personnes pensent réellement ce qui est exprimé. Il serait cependant inexact de cantonner cette réalité au domaine des laboratoires ou des projections technologiques. Le continent africain offre, depuis 2019, une trajectoire documentée qui atteste de manière concrète de l'existence de terreau fertile, à tout le moins à l'analyse des nouvelles architectures d'influence qui pourraient en résulte.

En Côte d'Ivoire, par exemple, les élections présidentielles de 2020 et législatives de 2021 ont été marquées par une forte circulation de désinformation en ligne, notamment via des images et vidéos manipulées (*cheapfakes*), certaines ayant contribué à exacerber les tensions inter-partisanes.

Au Kenya, les élections de 2022 ont donné lieu à des dispositifs renforcés de surveillance des discours haineux en ligne, mobilisant acteurs locaux et agences des Nations Unies. Ces initiatives ont permis d'identifier et de signaler un volume important de contenus problématiques, sans pour autant que des chiffres consolidés et largement vérifiés ne soient disponibles. Dans ce contexte, l'attention portée aujourd'hui aux dynamiques informationnelles ne relève pas d'une simple précaution technique, mais s'inscrit dans une mémoire politique précise.

À ce titre, le précédent de 2007–2008 — marqué par des violences post-électorales ayant fait environ 1 500 morts — rappelle avec quelle force les médias, notamment la radio, peuvent amplifier des tensions aux côtés de facteurs politiques et ethniques plus larges. De nos jours, et contrairement à il y a dix-huit ans — où les médias traditionnels, en particulier la radio, jouaient un rôle central dans la diffusion et l'amplification des messages —, l'environnement informationnel a profondément changé. Les plateformes numériques occupent désormais une position structurante, non seulement comme canaux de diffusion, mais comme espaces d'interaction où les contenus circulent, se transforment et se légitiment au fil des échanges.

Plus récemment, en République démocratique du Congo (RDC), des *deepfakes* attribuant de faux discours à Emmanuel Macron ont circulé dans le contexte des tensions liées à l'est du pays — illustrant la capacité de l'IA générative à instrumentaliser des figures étrangères dans des conflits locaux. Au Niger, cette fois-ci, après le coup d'État de 2023, plusieurs accusations visant la France — libération de terroristes, présence de militaires français aux côtés de groupes armés — ont largement circulé avant d'être démenties par Paris ou réfutées par des vérificateurs indépendants. Dans le cas nigérien, la désinformation n'a pas reposé principalement sur des *deepfakes* sophistiqués, mais sur des formes plus classiques — recyclage d'images, détournements contextuels, rumeurs virales — qui ont

néanmoins suffi à structurer des perceptions hostiles. Ce point est essentiel : l'efficacité informationnelle ne dépend pas encore systématiquement du degré de sophistication technologique. En tout cas, pas encore.

4. QUI PRODUIT, QUI DIFFUSE ? LE SPECTRE ÉLARGI DES ACTEURS

Ce constat invite à déplacer le regard : au-delà des outils et de leur degré de sophistication, la question décisive devient aussi celle des acteurs qui produisent, relaient et instrumentalisent ces contenus. Reste alors à préciser où se situent concrètement les centres d'initiative de ces dynamiques. Contrairement à une représentation répandue, la désinformation — y compris dans ses formes émergentes liées à l'IA — ne relève pas principalement d'acteurs étatiques étrangers. Les recherches du UNU-CPR et d'autres institutions montrent que, dans de nombreux contextes africains, les élites politiques et les acteurs domestiques jouent un rôle central dans la production et la diffusion de contenus trompeurs, en particulier en période électorale. L'accessibilité croissante des outils numériques et d'IA abaisse les barrières techniques ; les motivations locales — souvent immédiates et électorales — structurent largement ces pratiques.

Des opérations d'influence informationnelle conduites par des acteurs étatiques étrangers, en particulier russes, sont bien documentées dans plusieurs pays africains, notamment en République centrafricaine et au Sahel. Elles reposent sur des réseaux de médias, des campagnes sur les réseaux sociaux et des relais locaux. Certaines hypothèses émergentes suggèrent que ces stratégies pourraient évoluer vers des formes plus sophistiquées, intégrant l'optimisation de contenus pour les moteurs de recherche et potentiellement pour les systèmes d'IA, bien que ces dynamiques restent encore peu documentées empiriquement.

À ces catégories s'ajoutent des réseaux diasporiques — informels et transnationaux — qui contribuent à la circulation, à la mise en récit et, dans certains cas, à la coordination ponctuelle de contenus politiques, comme cela a été observé au Cameroun et dans d'autres contextes électoraux africains. Au Nigeria, les campagnes électorales de 2019 ont illustré l'usage intensif de groupes WhatsApp locaux pour la mobilisation politique et la diffusion de messages ciblés, dans un environnement où la nature privée des échanges rend difficile toute quantification précise de l'ampleur du phénomène. Dans un registre différent, des acteurs armés non étatiques, tels que le M23 en République démocratique du Congo, ont également investi l'espace informationnel, mobilisant les réseaux sociaux pour diffuser leurs récits, légitimer leurs actions et tenter d'influencer les perceptions à différentes échelles. Ce spectre élargi d'acteurs souligne que les dynamiques informationnelles contemporaines résistent à toute réduction à une catégorie unique d'opérateurs. C'est dans cet espace hétérogène — où coexistent communication stratégique, mobilisation politique et circulation de contenus — que s'insèrent et prospèrent les phénomènes de désinformation. C'est précisément cette pluralité qui invite à analyser les mécanismes profonds qui en rendent les dynamiques particulièrement efficaces

5. SIX RESSORTS D'UNE MENACE SYSTÉMIQUE

C'est à ce niveau d'analyse qu'il devient possible d'identifier les ressorts opérationnels du phénomène. Six mécanismes explicatifs structurent la menace et conditionnent la pertinence des réponses qui peuvent lui être apportées. Le premier tient à ce qu'on pourrait appeler l'effet multiplicateur. La désinformation synthétique ne fabrique pas de menaces ex nihilo : elle aggrave des tensions informationnelles préexistantes en se greffant sur des

clivages ethno-politiques déjà constitués. Les témoignages de praticiens africains recueillis par l'UNU-CPR en 2024 convergent en ce sens — la désinformation tend à exacerber des croyances préexistantes et à les ancrer davantage, plutôt qu'à en créer de nouvelles. Certains travaux[2] corroborent ce constat de manière significative. Ils montrent que des modèles de langage configurés pour incarner des personas politiques renforcent leur adhérence idéologique lorsqu'ils sont confrontés à des contre-arguments et augmentent parallèlement la production de contenu extrême. Concrètement, dans des environnements fortement polarisés, des systèmes automatisés de production de contenu pourraient accentuer la radicalisation des échanges dès lors qu'ils sont engagés dans des interactions conflictuelles. Cette dynamique est particulièrement préoccupante dans des contextes où les pratiques de *fact-checking* en ligne prennent la forme d'interactions directes et contradictoires, susceptibles d'alimenter la polarisation plutôt que de la désamorcer.

Le deuxième mécanisme est celui de l'érosion institutionnelle. Les *deepfakes* politiques fragilisent d'abord la confiance dans les preuves audiovisuelles : leur simple existence permet de contester des contenus authentiques, un phénomène que la littérature anglo-saxonne désigne sous le terme de *liar's dividend.* Autrement dit, la possibilité même de falsification devient une ressource stratégique : elle permet non seulement de diffuser de faux contenus, mais aussi de discréditer des preuves authentiques en les présentant comme manipulées. Dans ce contexte, la charge de la preuve se renverse progressivement, au détriment de ceux qui cherchent à établir des faits.

Certains chiffres montrent, par ailleurs, que des vidéos *deepfake* de haute qualité ne sont identifiées comme fausses que dans environ 25 % des cas, tandis que la détection d'images manipulées peut atteindre environ 60 %, ce qui souligne un écart significatif entre les capacités de perception humaine et le niveau de sophistication atteint par ces technologies.[3] Autrement dit, dans des conditions ordinaires d'exposition — défilement rapide, attention fragmentée, absence de vérification systématique —, les individus ne disposent pas des ressources cognitives nécessaires pour opérer une distinction fiable entre contenus authentiques et synthétiques.

À l›érosion épistémique s›ajoutent des contraintes d›attribution judiciaire. Si de nombreux pays africains ont intégré des dispositions relatives à la preuve numérique, les standards restent hétérogènes et souvent insuffisamment adaptés aux contenus synthétiques générés par l›IA. Les *deepfakes* complexifient l'établissement de l'authenticité et de la responsabilité, augmentent le coût probatoire et ouvrent la voie à des stratégies de contestation systématique des preuves. La fabrication d'aveux, la mise en scène de déclarations fictives ou la production d'ordres militaires synthétiques peuvent produire des effets rapides et difficiles à corriger — la diffusion précédant largement les capacités de vérification et de démenti.

Les données issues de la cybersécurité confirment d'ailleurs une dynamique d'industrialisation des attaques assistées par l'IA. Le rapport Imperva Bad Bot Report[4], une radioscopie de référence en matière d'analyse du trafic automatisé à l'échelle mondiale, indique qu'en 2024, environ 2 millions d'attaques cyber assistées par IA ont été détectées et bloquées quotidiennement. Ces opérations s'inscrivent dans un modus operandi désormais bien documenté, combinant attaques simples à très grande échelle et développement parallèle

2. Olejnik, L. (2025). AI Propaganda factories with language models. arXiv preprint arXiv:2508.20186.

3. Chipeta, C. (2025, May 29). Deepfake statistics (2025): 25 new facts for CFOs. Eftsure. https://www.eftsure.com/statistics/deepfake-statistics/

4. Imperva. (2025). *2025 Bad Bot Report: The rapid rise of bots and the unseen risk for business.* https://www.imperva.com/resources/gated/reports/2025-Bad-Bot-Report.pdf

de techniques plus sophistiquées. Si ces données concernent principalement des activités de type bot, elles révèlent un phénomène plus structurel : l'abaissement des barrières d'entrée et la capacité croissante d'acteurs peu spécialisés à déployer des opérations à grande échelle.

Le quatrième mécanisme est celui de la contamination des écosystèmes d'intelligence artificielle. Cette contamination peut opérer passivement, par l'exposition des modèles à des environnements informationnels dégradés. Des enquêtes récentes ont ainsi mis en évidence l'existence de réseaux coordonnés de sites diffusant à grande échelle des contenus pro-Kremlin, optimisés pour la visibilité en ligne. En 2023, NewsGuard a recensé un réseau de plus de 150 sites — désigné comme le Pravda network — qui produisent des contenus multilingues reprenant des narratifs favorables à la Russie et conçus pour une indexation élevée par les moteurs de recherche..[5]

Mais ce phénomène dépasse largement ce seul cas. Des analyses conduites par le Digital Forensic Research Lab, un programme de recherche de l'Atlantic Council spécialisé dans la traque des opérations d'influence en ligne, mettent en évidence des infrastructures informationnelles comparables, associées à différents acteurs étatiques ou para-étatiques, reposant sur des réseaux de sites interconnectés, des fermes de contenus et des stratégies de *search engine optimization* (SEO) visant à maximiser leur visibilité et leur crédibilité apparente.[6] Ces contenus, explicitement configurés pour être explorés par les *crawlers* — notamment via des sitemaps et des fichiers robots.txt optimisés — sont susceptibles d'être massivement indexés et archivés dans des bases de données web à large échelle. Ils contribuent ainsi à saturer l'espace informationnel de récits biaisés ou trompeurs.

Sans nécessairement viser explicitement les systèmes d'intelligence artificielle, cette production massive accroît la probabilité d'intégration de ces contenus dans des corpus d'entraînement, exposant les modèles à des données de faible qualité ou manipulées, susceptibles d'être ensuite reprises ou reformulées par des systèmes conversationnels. Le phénomène se distingue des *deepfakes* audiovisuels : il ne repose pas sur la fabrication ponctuelle de preuves visuelles, mais sur une stratégie d'injection diffuse de récits dans l'environnement informationnel lui-même, sans marquage explicite ni signal technique aisément détectable à grande échelle.

Cette dynamique est renforcée par l'industrialisation de la production de contenu assistée par l'IA. Des rapports récents du *European External Action Service*[7], le service diplomatique de l'Union européenne (UE), et de OpenAI[8] mettent en évidence l'émergence d'opérations d'influence combinant génération automatisée de textes et diffusion coordonnée sur plusieurs plateformes. Ces opérations reposent sur des capacités accrues de production à grande échelle et sur l'exploitation des logiques de circulation propres aux environnements numériques, augmentant ainsi la visibilité et la persistance de certains narratifs. Elles contribuent, de manière cumulative, à une forme de saturation de l'espace informationnel. Dans ce contexte, les contenus ainsi produits deviennent susceptibles d'être captés par des systèmes de collecte de données à large échelle, exposant indirectement les modèles d'intelligence artificielle à des informations biaisées ou manipulées.

5. NewsGuard Technologies. (2025, March 6). *A well-funded Moscow-based global "news" network has infected Western artificial intelligence tools with Russian propaganda*. https://www.newsguardtech.com/special-reports/moscow-based-global-news-network-infected-western-artificial-intelligence-russian-propaganda

6. Digital Forensic Research Lab. (2026, April 8*). Pravda in the pipeline: Early evidence of state-adjacent propaganda in AI training data*. https://dfrlab.org/2026/04/08/pravda-in-the-pipeline/

7. European External Action Service. (2025). *3rd EEAS report on foreign information manipulation and interference (FIMI) threats*. https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf

8. OpenAI. (2025, February 21). *Disrupting malicious uses of AI*. https://openai.com/fr-FR/index/disrupting-malicious-ai-uses/

Un vecteur actif vient compléter ce tableau. Les analyses en cybersécurité montrent que des bots malveillants se font passer pour des *crawlers* légitimes afin de contourner les dispositifs de protection, exploitant la réticence des organisations à bloquer ces flux de peur de perturber des services essentiels. Ces stratégies accroissent la visibilité et la diffusion de contenus manipulés, tout en s›insérant dans les flux de collecte de données utilisés à grande échelle sur le web. Sans constituer une injection directe et documentée dans les pipelines d›entraînement des modèles, ces dynamiques contribuent à dégrader l›environnement informationnel auquel les systèmes d›IA sont exposés.

Le cinquième mécanisme — l' essaimage des opérations informationnelles — marque une rupture qualitative avec les formes antérieures de désinformation. Il ne s'agit plus de contenus isolés, mais de dynamiques coordonnées reposant sur une multiplicité d'agents automatisés. Les travaux en cybersécurité et en analyse des plateformes ont déjà documenté l'existence de réseaux de comptes coordonnés et de *botnets* capables de maintenir des identités persistantes et de diffuser des messages de manière synchronisée. L'intégration progressive de l'IA générative ouvre la possibilité d'une évolution vers des formes plus avancées de coordination, dans lesquelles des agents automatisés pourraient adapter leurs contenus, varier leurs registres discursifs et interagir avec des publics en temps quasi réel. Si ces configurations restent encore inégalement documentées dans leur forme pleinement autonome, elles signalent une transformation potentielle majeure des opérations d'influence — passant de contenus ponctuels à des systèmes dynamiques, distribués et adaptatifs. Dans des contextes africains à forte cohésion sociale, où la destruction de l'indépendance des jugements collectifs peut déclencher des dynamiques de mobilisation violente, ce risque est particulièrement sévère. Ces résultats suggèrent qu'une régulation centrée exclusivement sur les modèles est insuffisante tant qu'elle n'intègre pas l'analyse des architectures de persona et des dynamiques conversationnelles, qui structurent les effets produits dans les interactions.

6. TECHNOLOGIE ET SOCIÉTÉ : DEUX REGISTRES INDISSOCIABLES

Face à cette configuration, les réponses s'organisent en deux registres complémentaires dont la hiérarchisation dépend des contextes. Le premier est de nature technologique. S'il est nécessaire, ses limites sont désormais bien établies. Le principal écueil tient au techno-solutionnisme, c'est-à-dire à la tendance à requalifier un problème fondamentalement socio-politique en enjeu technique, en orientant les ressources vers des outils de détection importés et, parfois, mal adaptés aux contextes locaux. Les travaux récents sur les opérations d'influence assistées par IA montrent d'ailleurs que l'efficacité de ces dispositifs dépend moins des modèles eux-mêmes que de leur configuration. La conception des personas — qui encode orientation idéologique, style rhétorique et posture interactionnelle — apparaît déterminante, tandis que les écarts entre modèles restent relativement marginaux.[9]

Dans ces conditions, les solutions technologiques ne peuvent être efficaces qu'à condition d'être intégrées à des dispositifs de régulation plus larges. Cela suppose notamment de marquer clairement les contenus générés par IA sur les plateformes — mais uniquement si ce marquage s'accompagne de mesures concrètes limitant leur diffusion. Pris isolément, l'étiquetage ne suffit pas. Son efficacité varie fortement selon les contextes d'exposition, et des contenus pourtant signalés comme synthétiques continuent souvent à circuler largement et à générer un fort engagement. Ce registre inclut aussi l'élaboration de règles claires pour vérifier les preuves numériques devant les tribunaux, la mise en place de

9. Olejnik, 2025

partenariats avec des acteurs de la recherche pour renforcer les compétences, ainsi qu'une meilleure supervision de la manière dont les modèles sont entraînés.

Cette supervision ne peut plus se limiter à vérifier les sources officiellement déclarées. Elle doit aussi prendre en compte les modes réels d'accès aux données, comme la collecte automatisée dissimulée et l'introduction volontaire de contenus dans les jeux de données. Un exemple concret : un réseau d'acteurs peut créer des milliers de pages web ou de comptes en ligne diffusant un même récit trompeur, puis s'assurer que ces contenus soient massivement indexés et récupérés par des systèmes de collecte automatique. Ces contenus finissent alors intégrés aux données d'entraînement, ce qui augmente la probabilité que les modèles les reprennent ensuite comme des informations plausibles, voire crédibles.

Enfin, ce registre technologique doit être complété — et, dans certains contextes, précédé — par un registre techno-social adapté aux réalités africaines. Les pratiques de terrain montrent en effet que les organisations de *fact-checking* combinent outils numériques et jugement humain, en raison des limites persistantes des solutions automatisées face aux langues locales et aux contextes socio-culturels spécifiques. Ce second registre repose ainsi sur le renforcement des normes journalistiques professionnelles et sur le développement de mécanismes de vérification distribués. Des initiatives comme *CongoCheck* en République démocratique du Congo ou *PesaCheck* en Afrique de l'Est montrent qu'il est déjà possible de produire des réponses crédibles et adaptées aux contextes locaux. Il faut désormais renforcer ce réseau de vérificateurs en lui donnant les moyens d'agir dans les principales langues africaines, y compris les langues locales.

Parallèlement, ce travail pourrait s'appuyer sur un système partagé de suivi des campagnes d'influence liées à l'IA, utilisant des outils automatisés pour observer les comportements en ligne. Ces outils permettraient aussi de tester, en conditions contrôlées, la manière dont certaines campagnes de manipulation pourraient fonctionner, afin de mieux s'y préparer. Dans ce cadre, l'usage de modèles ouverts — dont le fonctionnement est accessible et vérifiable — faciliterait la transparence, le contrôle et leur appropriation par des acteurs locaux. Un tel dispositif renforcerait les capacités d'anticipation et permettrait une réponse coordonnée et rapide des acteurs publics et non publics, en particulier durant les périodes électorales.

La démarche implique également des approches préventives, au premier rang desquelles le *pre-bunking* — l'exposition anticipée des publics aux mécanismes de manipulation avant même leur diffusion. Ces stratégies visent à renforcer la résilience cognitive des individus face aux opérations d'influence. Dans les deux cas, la pertinence de l'action dépend d'une condition préalable trop souvent négligée : la formation des acteurs institutionnels eux-mêmes. Magistrats, procureurs, agents électoraux, forces de sécurité — aucun de ces acteurs n'est aujourd'hui en mesure d'assurer le rôle de gardien des normes épistémiques que la menace exige d'eux. L'investissement dans cette formation n'est pas un supplément d'âme. C'est une condition sine qua non de toute stratégie crédible.

7. QUATRE RISQUES QUI PEUVENT RETOURNER LES POLITIQUES PUBLIQUES

Reste qu'en dépit de la cohérence de ce schéma d'ensemble, plusieurs risques transversaux peuvent en pervertir les effets si on ne les anticipe pas — retournant les politiques publiques contre leurs propres objectifs. Le premier risque tient à l'instrumentalisation politique des législations *anti-deepfakes*. Bien documenté dans plusieurs pays du continent, il s'est déjà concrétisé lorsque des lois, initialement conçues pour protéger l'espace public, ont

servi à criminaliser la satire, la dissidence ou encore le journalisme professionnel. Dans ces conditions, toute réforme qui ne prévoit ni mécanismes d'application indépendants du pouvoir exécutif ni garanties explicites en faveur de la liberté d'expression, s'expose à un renversement de ses finalités initiales.

À ce premier enjeu s'ajoute un second risque, d'ordre temporel : celui d'une fenêtre d'action particulièrement étroite. Les capacités des systèmes d'IA en matière de production et de diffusion de contenus évoluent à un rythme soutenu, tandis que les formes d'automatisation et de coordination des opérations d'influence gagnent en sophistication. Or, cette accélération technologique intervient dans un contexte africain marqué par une séquence électorale dense — plusieurs dizaines de scrutins nationaux entre 2024 et 2026. La combinaison de ces dynamiques réduit significativement la marge de manœuvre des acteurs publics. L'appel à une réponse institutionnelle et législative rapide ne relève donc pas d'un effet de style : il traduit un décalage croissant entre la vitesse des transformations technologiques et celle des capacités de régulation.

Enfin, un troisième risque prolonge ces tensions : celui du décalage temporel des plateformes. En pratique, les mécanismes de modération interviennent le plus souvent a posteriori, une fois les contenus déjà diffusés, dans un environnement où la circulation de l'information est particulièrement rapide. Ce désajustement structurel limite mécaniquement l'efficacité des dispositifs existants et souligne la- nécessité de repenser les modalités d'intervention en amont, au niveau même des dynamiques de propagation.

Un cas documenté en Hongrie en 2026 illustre cette dynamique : un réseau coordonné de comptes TikTok diffusant des contenus générés par IA a cumulé environ dix millions de vues avant d'être identifié puis supprimé par la plateforme. Ce type de séquence suggère l'existence d'une asymétrie structurelle entre la rapidité de production et de circulation des contenus et la capacité, plus lente et souvent réactive, d'intervention des plateformes.[10] Dans des fenêtres électorales africaines où les quarante-huit à soixante-douze heures précédant le scrutin concentrent les opérations de manipulation, ce délai rend la modération *post hoc* structurellement insuffisante. La négociation collective avec les- plateformes doit intégrer des obligations d'intervention pré-électorale, et non simplement réactive.

Le quatrième risque — peut-être le moins visible, mais l'un des plus durables — est celui de la dépendance normative. Les standards de détection des contenus synthétiques, les seuils de classification et les référentiels techniques sont aujourd'hui largement définis par des plateformes et des laboratoires situés hors du *Global South*. Les acteurs africains se trouvent ainsi principalement en position d'adoption de normes à l'élaboration desquelles ils ont apporté leur petite contribution, conçues dans des environnements linguistiques, culturels et politiques profondément différents. Cette asymétrie soulève des enjeux d'adéquation et d'efficacité, mais aussi, plus largement, de souveraineté normative : la capacité des États et des sociétés africaines à définir les cadres de régulation adaptés à leurs propres réalités, au même titre que les enjeux de souveraineté économique ou politique.

Les évolutions récentes du positionnement américain en matière de lutte contre la désinformation — marquées par des recompositions politiques, juridiques et institutionnelles — contribuent par ailleurs à rendre plus incertain le niveau de pression exercé sur les plateformes pour l'application de leurs propres politiques de modération. Dans ce contexte, les États africains ne peuvent se permettre de dépendre d'un alignement de puissances extérieures dont les priorités et les intérêts évoluent. La construction de capacités de régulation indépendantes, ainsi que le renforcement des mécanismes de coordination au niveau régional, apparaissent comme des nécessités stratégiques plutôt que comme des options.

10. NewsGuard Technologies. (2026, March 20). *Influence campaign uses AI TikTok videos to boost Hungary's Viktor Orbán*. https://www.newsguardtech.com/special-reports/influence-campaign-uses-ai-tiktok-videos-to-boost-hungarys-viktor-orban/

8. AGIR AUX NIVEAUX NATIONAL, RÉGIONAL ET INTERNATIONAL

C'est de cette conviction — que ni la dépendance externe ni les réponses improvisées ne sauraient suffire — que découlent les recommandations qui suivent, articulées sur trois niveaux d'action distincts mais complémentaires. Au niveau national, la priorité réside souvent dans l'adaptation de cadres juridiques existants afin d'y intégrer les usages politiques des contenus synthétiques, plutôt que dans la création ex nihilo de nouveaux corpus législatifs. La proposition de révision du *Computer Misuse Act* en Ouganda en 2022[11], un dispositif initialement conçu pour encadrer les infractions liées aux systèmes informatiques et aux communications électroniques, s'inscrivait dans une logique d'adaptation incrémentale. Toutefois, son invalidation par la Cour constitutionnelle en mars 2026 — pour des motifs à la fois procéduraux, liés au non-respect des exigences encadrant le processus législatif, et substantiels, tenant au caractère vague et extensif de certaines dispositions — met en évidence la fragilité de telles approches lorsqu'elles ne satisfont pas aux principes de légalité, de précision et de proportionnalité. Ce cas illustre ainsi les tensions entre l'impératif de régulation rapide des contenus numériques et les garanties constitutionnelles fondamentales.

Le cas marocain illustre une dynamique plus complexe qu'une simple adaptation juridique. La régulation des contenus manipulés s'inscrit d'abord dans un socle existant, notamment la loi organique n° 27-11 relative à la Chambre des représentants, qui prévoit déjà des sanctions en matière de diffusion de fausses informations susceptibles d'altérer le processus électoral. Les développements récents — en particulier le projet d'introduction de nouvelles dispositions, dont l'article 51-bis — témoignent moins d'une rupture que d'un approfondissement de ce cadre. Au regard de la sensibilité et de la technicité des enjeux en présence, cette évolution soulève la question de l'ampleur et des modalités du débat parlementaire — et plus encore sociétal — qui l'a accompagnée. Or, rien n'indique que ce débat ait été à la mesure des transformations en cours. On observe plutôt un relatif déficit de délibération publique, alors même que les implications de ces dispositions excèdent largement le seul cadre électoral, en ce qu'elles touchent aux équilibres fondamentaux entre lutte contre la manipulation informationnelle et garanties de la liberté d'expression.

Ces avancées révèlent néanmoins une tension caractéristique des régulations contemporaines : la volonté de contrôler les effets de la numérisation du champ politique sans disposer pleinement des instruments permettant d'agir sur ses infrastructures. En effet, une part significative des infrastructures informationnelles mobilisées dans ces dynamiques — plateformes de réseaux sociaux, régies publicitaires, systèmes algorithmiques de recommandation[12] —échappe largement aux juridictions nationales, tant du point de vue de leur gouvernance que de leurs modalités opérationnelles. Les autorités publiques se trouvent ainsi dans une position asymétrique : elles peuvent encadrer les usages en aval, sanctionner certains contenus ou comportements, mais disposent de marges d'action plus limitées sur les mécanismes en amont qui conditionnent la visibilité, la circulation et la hiérarchisation de l'information.

Il convient de renforcer les capacités opérationnelles des institutions électorales. Cela passe notamment par la mise en place, en amont des scrutins, d'unités spécialisées de

11. Parliament of Uganda. (2022). *Computer Misuse (Amendment) Act*, 2022. https://bills.parliament.ug/attachments/Computer%20Misuse%20(Amendment)Act,%202022.pdf

12. Benabid, M. (2025). « *Lutter contre la désinformation : savoirs, enjeux et pratiques* ». Policy Center for the New South.

veille informationnelle, dotées d'un mandat couvrant le suivi des contenus synthétiques, des campagnes coordonnées et des usages de systèmes automatisés dans la diffusion d'informations électorales. Ces unités devraient prioriser la détection des comportements conversationnels — la cohérence excessive d'un persona dans le temps étant documentée comme la signature la plus fiable des opérations automatisées — plutôt que la seule recherche d'artefacts techniques dans les contenus.

Au niveau régional, l'enjeu réside moins dans l'élaboration d'un cadre stratégique continental — déjà établi avec la Stratégie continentale sur l'IA adoptée par l'Union africaine en juillet 2024, structurée autour de cinq axes (bénéfices, capacités, risques, investissement et coopération) — que dans sa traduction en dispositifs normatifs contraignants et en capacités effectives de mise en œuvre. .

Cette limite structurelle invite dès lors à déplacer le regard vers d'autres configurations régulatoires, afin d'en apprécier les apports mais aussi les conditions de possibilité. Par exemple, la comparaison avec le régime réglementaire européen — DSA (Digital Services Act), DMA (Digital Markets Act) et AI Act (Artificial Intelligence Act) — est instructive, à condition de ne pas en faire un modèle d'importation directe : ces dispositifs ont été conçus pour des écosystèmes numériques matures, avec des infrastructures institutionnelles, des capacités de contrôle et des traditions juridiques profondément différentes de celles du contexte africain.

Au niveau international, enfin, la participation des États africains et des instances régionales aux processus de définition des standards de détection et de classification des contenus synthétiques doit être revendiquée comme un enjeu de souveraineté normative. Les audits exigés des fournisseurs de LLM opérant dans les espaces numériques africains doivent couvrir non seulement les sources d'entraînement déclarées, mais les vecteurs d'accès aux corpus — contamination passive via réseaux de désinformation étatiques et injection active via *crawlers* déguisés. La négociation collective avec les grandes plateformes — pour imposer des obligations de modération renforcée pré-électorale — ne peut être efficace qu›à cette échelle : les États pris individuellement n›ont pas la taille critique pour peser.

Enfin, le soutien au développement de capacités de recherche Sud-Sud sur les modèles de langage adaptés aux langues africaines et aux corpus de désinformation continentaux constitue une priorité stratégique à long terme. Sans cela, l'asymétrie capacitaire de détection demeurera structurelle. La cause peut sembler décalée au regard des crises immédiates qui monopolisent l'attention et les ressources politiques — mais elle engage en réalité l'essentiel : quelle est l'architecture de réponse à la portée des États africains ?

CONCLUSION

Dans un contexte géopolitique marqué par des tensions croissantes, des calendriers électoraux africains chargés et une accélération technologique sans précédent, la nécessité de bâtir des réponses adaptées à la désinformation synthétique est plus impérieuse que jamais. La menace est réelle, documentée et en expansion. Elle opère dans un environnement institutionnel africain largement démuni, où la majorité des pays manquent des cadres réglementaires élémentaires pour gérer l'IA, et où les outils de détection disponibles ont été conçus pour des réalités linguistiques et culturelles qui ne sont pas celles du continent.

Deux convictions se dégagent de cette analyse. La première : la diversité des réponses technologiques disponibles est une condition nécessaire, mais non suffisante. Elle doit s'inscrire dans une stratégie techno-sociale qui reconnaît la prééminence des dynamiques humaines, communautaires et institutionnelles sur les outils. La seconde : le principal risque n'est pas l'absence de réponse, mais une réponse mal calibrée — celle qui croirait pouvoir résoudre un problème politique profond avec des outils de détection importés, ou celle qui, sous couvert de protéger l'espace public, instrumentaliserait une législation anti-*deepfakes* à des fins de censure.

L'enjeu, en tout état de cause, dépasse largement la seule désinformation synthétique. C'est la question de la souveraineté informationnelle du continent africain qui est posée : la capacité à définir ses propres narratifs, à produire ses propres normes et à construire les infrastructures épistémiques qui permettent à des démocraties fragilisées de résister à des opérations d'ingérence conçues précisément pour les déstabiliser. Ce sont, in fine, la qualité de l'espace public africain et la robustesse des processus démocratiques qui sont en jeu.

À PROPOS DE L'AUTEUR

MOHAMED BENABID

Mohamed Benabid est enseignant à la Faculté de gouvernance, sciences économiques et sociales (FGSES) de l'Université Mohammed VI polytechnique. Lauréat de l'Ecole de journalisme de Strasbourg, titulaire d'un doctorat en sciences de l'information et de la communication de l'Université Paris VIII et d'un doctorat en science de gestion de l'ISCAE, il compte à son actif près de 30 ans d'expérience dans l'industrie des médias. Son parcours pluridisciplinaire l'a conduit à couvrir depuis plusieurs années un large éventail de sujets : veille et intelligence économique, Médias/journalisme, Knowledge Management, géostratégie d'entreprise, géopolitique, communication politique, gestion et communication de crise entrepreneuriat, management stratégique, méthodologie de la recherche en sciences de gestion.

À PROPOS DU POLICY CENTER FOR THE NEW SOUTH

The Policy Center for the New South (PCNS) is a Moroccan think tank aiming to contribute to the improvement of economic and social public policies that challenge Morocco and the rest of Africa as integral parts of the global South.

Le Policy Center for the New South (PCNS) est un think tank marocain dont la mission est de contribuer à l'amélioration des politiques publiques, aussi bien économiques que sociales et internationales, qui concernent le Maroc et l'Afrique, parties intégrantes du Sud global. Le PCNS défend le concept d'un « nouveau Sud » ouvert, responsable et entreprenant ; un Sud qui définit ses propres narratifs, ainsi que les cartes mentales autour des bassins de la Méditerranée et de l'Atlantique Sud, dans le cadre d'un rapport décomplexé avec le reste du monde. Le think tank se propose d'accompagner, par ses travaux, l'élaboration des politiques publiques en Afrique, et de donner la parole aux experts du Sud sur les évolutions géopolitiques qui les concernent. **Lire plus**

Policy Center for the New South

Rabat Campus of Mohammed VI Polytechnic University,
Rocade Rabat Salé - 11103
Email : contact@policycenter.ma
Phone : +212 (0) 537 54 04 04
Fax : +212 (0) 537 71 31 54

www.policycenter.ma

POLICY CENTER FOR THE NEW SOUTH