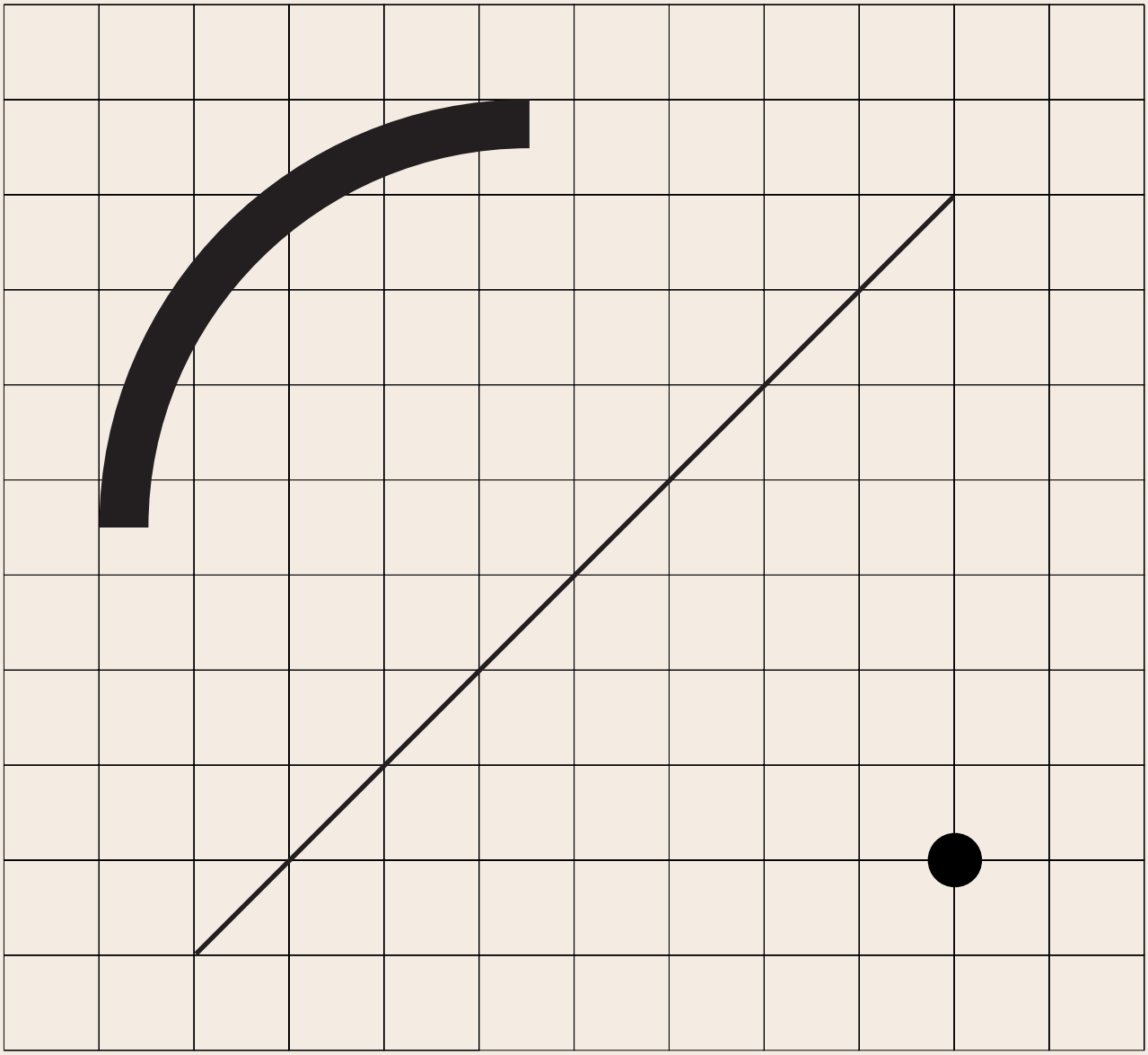
Un Cloud global au service de la recherche sur la désinformation

NOTE POUR L'ACTION - OCTOBRE 2021



groupe d'études geopolitiques

groupe d'études geopolitiques

Un Cloud global au service de la recherche sur la désinformation

AUTEURS
SAIPH SAVAGE, ISHAN SHARMA, CRISTINA MARTINEZ PINTO,VICTOR STORCHAN

AVANT-PROPOS
PATRICE GEOFFRON

45, RUE D'ULM 75005 PARIS
LEGRANDCONTINENT.EU
GEG@ENS.FR

POUR CITER L'ÉTUDE
SAIPH SAVAGE, ISHAN SHARMA, CRISTINA MARTINEZ PINTO,VICTOR STORCHAN, UN CLOUD GLOBAL AU SERVICE DE LA RECHERCHE SUR LA DÉSINFORMATION, NOTE POUR L'ACTION, GROUPE D'ÉTUDES GÉOPOLITIQUES, OCTOBRE 2021.

Avant-propos

PATRICE GEOFFRON •
PROFESSEUR D'ÉCONOMIE ET DIRECTEUR DU CENTRE DE GÉOPOLITIQUE DE L'ÉNERGIE ET DES MATIÈRES PREMIÈRES DE L'UNIVERSITÉ PARIS-DAUPHINE

Ne cachons pas que le travail ici proposé suscite tout d'abord la sidération : il fait malheureusement peu de doutes que les phénomènes énumérés dans le préambule du document sont bien advenus, notamment pour ce qui relève de l'intégrité des processus électoraux.

Le dénombrement de 639 campagnes distinctes de désinformation relatives aux élections américaines de 2020 (par l'Electoral Integrity Project) nous projette directement dans une dystopie… Et ce ne sont pas les presque deux années de crise sanitaire qui auront estompé cette sensation, avec parmi les néologismes de cette période le « infodémic ».

Comment on est arrivé là ? Peut-on repérer une montée des opérations de désinformation et un « tipping point » à partir duquel un effet d'emballement exponentiel est observable ?

Retracer le cheminement des menaces protéiformes de la désinformation est une tâche complexe. Ce qui explique que le volume de recherches consacrées à la désinformation croisse fortement. La base de données bibliographique ScienceDirect (consultée le 9 octobre 2021) indique que 126 articles comportent le mot-clé « Disinformation » depuis 1987 qui en marque la première apparition (article alors consacré au SIDA en Afrique). Si, pendant plusieurs décennies, cette base n'a recensé qu'un article par an (et très souvent aucun), 21 ont été publiés en 2020 et, d'ores et déjà, 54 en 2021 (année en cours…). Sur les publications de 2021, 33 mentionnent à Twitter, 28 Facebook, 23 la Chine, 14 la Russie, ce qui donne une vue panoramique quelques nœuds du problème. Quelques titres parmi les papiers les plus récents donnent un éclairage sur les questions soulevées : « Examining characteristics of prebunking strategies to overcome public relations disinformation attacks », « Misinformation, disinformation, and fake news : Cyber risks to business », « Relating foreign disinformation through social media, domestic online media fractionalization », « CoVerifi: A COVID-19 news verification system »,… Notons également que, dès 2013, les World Economic Forum's Global Risk Report avait mis en évidence le risque de désinformation numérique, en relation directes avec les cyberattaques.

Nous avons, indubitablement, besoin de travaux de recherche qui analysent les ressorts des désinformations et leur impact. Dans cette perspective, le premier mérite du papier de S. Savage, I. Sharma, C.M. Pinto et V. Storchan est d'analyser les « coûts » de l'analyse de la désinformation : la combinaison d'une puissance de calcul massive, l'accès à des données pertinentes, la disposition de modèles d'IA sophistiqués, … Et ces conditions, les grandes plateformes technologiques sont mieux à même de les réunir, alors même qu'elles sont parmi les canaux via lesquels circulent les désinformations, de sorte qu'émergence une menace de « pervasive bias in the large-scale disinformation research that is conducted » (selon les termes de auteurs).

Au-delà de ces observations, et dès lors que la robustesse de démocraties très matures a été éprouvée par des opérations de désinformation, les auteurs appellent à réduire les asymétries de moyens observées pour conduire des recherches pertinentes et bâtir un espace de coopération entre universités, ONG, institutions publiques et entreprises. La force de leur proposition, au-delà de la « pétition de principe », est de définir très finement les arcanes du dispositif, à la fois en termes techniques (un cloud global et décentralisé, alimentation progressive d'une base de données pertinente) et de gouvernance (de façon à éviter une « malédiction des communs » et permettre les indispensables interactions pluridisclinaires). Leur proposition n'est sans doute que le début d'un long processus, mais est consonante avec les initiatives de clouds de recherche et d'infrastructures de calcul publiques avec lesquelles elle a vocation à former un ecosystème.

Un Cloud global au service de la recherche sur la désinformation

SAIPH SAVAGE • DIRECTEUR DU NORTHEASTERN CIVIC A.I. LAB ET PROFESSEUR ASSOCIÉ À L'UNIVERSITÉ DE NORTHEASTERN
ISHAN SHARMA • MEMBRE DE LA FÉDÉRATION DES SCIENTIFIQUES AMÉRICAINS ET ANALYSTE POLITIQUE AU DAY ONE PROJECT
CRISTINA MARTINEZ PINTO • FONDATRICE DE PIT POLICY LAB ET CHERCHEUR SENIOR AU LABORATOIRE DES INTERACTIONS HOMME-MACHINE DE LA WEST VIRGINIA UNIVERSITY
VICTOR STORCHAN • ANCIEN ÉLÈVE DE STANFORD, INGÉNIEUR EN IA DANS L'INDUSTRIE FINANCIÈRE

La désinformation génère de plus en plus une défiance généralisée des citoyens affectant la sécurité nationale, l'intégrité des élections, la santé publique et d'autres secteurs critiques aux États-Unis [8,21]. Les 7,2 millions de mentions de rumeurs de violence antifa et les 6,2 millions de mentions relatives à la théorie du complot de QAnon sur les réseaux sociaux auraient ainsi contribué à l'attaque du 6 janvier contre le Capitole des États-Unis[1]. Le partenariat pour l'intégrité des élections (EIP) a identifié 639 campagnes distinctes de désinformation liées aux élections américaines de 2020 sur 15 plateformes de réseaux sociaux, dont 72% visaient explicitement à délégitimer l'élection[2].

Notons que les campagnes de désinformation ne constituent pas des incidents isolés qui ne toucheraient que les États-Unis. Après l'ingérence de la Russie dans la campagne présidentielle américaine de 2016, la désinformation est devenue un problème mondial. Les démocraties européennes ont été menacées par des campagnes de désinformation [26,27]. Plusieurs de ces campagnes ont eu pour objectif de semer la méfiance à l'égard des élections européennes, et de promouvoir les divisions sociales afin de créer le chaos à travers l'Europe dans le but par exemple de renforcer l'influence du Kremlin [3, 16].

Combler l'asymétrie de moyens entre l'industrie, la science ouverte multilatérale et les initiatives universitaires.

Les capacités technologiques les plus récentes, comme les grands modèles de langue en IA, sont des outils puissants qui peuvent accélérer le déploiement des opérations d'influence hostiles et en accroître l'ampleur [6]. Alors que de nombreuses démocraties tentent de se protéger contre la désinformation par le biais de leurs organes législatifs [1], elles doivent s'appuyer sur des coalitions flexibles, pluridisciplinaires et plurilatérales pour apporter l'expertise sur d'experts une technologie qui évolue rapidement et en permanence.

Au-delà de la technologie elle-même, une compréhension plus approfondie des mécanismes, de la portée et de l'impact des campagnes et des dynamiques de désinformation est nécessaire pour répondre à ces questions urgentes [7,31]. Cependant, pour mener des recherches à grande échelle sur la désinformation afin de comprendre comment ces réseaux et relations complexes se déroulent dans la nature, il est nécessaire de disposer d'une puissance de calcul massive et coûteuse et d'avoir accès aux données pertinentes. À l'heure actuelle, les seules entités capables de mener de telles recherches à l'échelle sont les grandes plateformes technologiques (par exemple, Google, Facebook, Twitter, etc.), qui disposent de ressources informatiques considérables et du hardware (composants électroniques) le plus récent pour entraîner des modèles à milliards de paramètres (GPT-3 d'OpenAI ou Switch-C de Google). Étant les seules à disposer des données générées par les utilisateurs sur les plateformes, elles deviennent de facto des fondés de pouvoir pour la recherche sur la désinformation [2, 11, 24].

Il en résulte un biais généralisé dans les recherches à grande échelle sur la désinformation qui sont menées. En particulier, elles ont eu tendance à se concentrer sur les études qui correspondent aux intérêts et à l'agenda des grandes plateformes technologiques [2, 37] (d'autant plus que ce sont elles qui fournissent les données et les calculs pour mener les études). Les équipes de recherche traitant des sujets qui ne sont pas alignés avec leurs agendas se retrouvent fréquemment à devoir composer avec

1 — Misinformation 2020: What the Data Tells Us About Election-Related Falsehoods, Defense One, November 2020.

2 — Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory (2021). The Long Fuse: Misinformation and the 2020 Election. Stanford Digital Repository: Election Integrity Partnership. v1.3.0.

un accès limité aux ressources de calcul et aux données pour mener leurs investigations [2,13,30]. Le cas du récent incident entre Facebook et des chercheurs de l'université de New York (NYU) en est l'illustration : leur accès aux données aurait été révoqué après que leurs recherches aient révélé comment Facebook a profité de publicités politiques manipulatrices[3]. Plus récemment encore, des courriels internes de Facebook ont révélé que dans le cadre d'une initiative de recherche dans laquelle l'entreprise partageait des données avec des universitaires, seule la moitié des données relatives aux interactions des utilisateurs de Facebook était partagée, et non la totalité, comme l'entreprise l'avait initialement affirmé[4]. Cet incident a eu pour conséquence d'invalider les résultats de la recherche réalisée au cours des trois dernières années et qui utilisaient les données de Facebook[5]. L'accès limité aux calculs et aux données pour les enquêtes sur la désinformation ne rend pas seulement difficile le progrès des connaissances, il inhibe également les capacités des gouvernements à créer des stratégies de défense pour combattre le phénomène au niveau national, la course à la maîtrise de cette technologie étant devenue un enjeu géopolitique[6].

Enfin, il est important de souligner que les modèles d'IA utilisés pour lutter contre la désinformation et, plus généralement, les technologies utilisées dans le monde entier, nécessitent un examen plus approfondi. Les initiatives de science ouverte (comme le Big Science Workshop ou le groupe EleutherAI) ainsi que le monde universitaire devraient être mobilisés et habilités à fournir des audits impartiaux et transparents de ces technologies, en particulier lorsqu'elles ont un effet critique sur le débat public et les processus démocratiques [5, 14].

Concevoir un Cloud global et décentralisé au service de la recherche sur la désinformation

Pour répondre à cette urgence, nous proposons la création d'un Cloud global et décentralisé au service de la recherche sur la désinformation. Cet outil doit permettre à un nombre plus important et plus diversifié de parties prenantes (universitaires, ONG, gouvernements, entreprises) de collaborer dans un cadre plurilatéral afin de mener en toute sécurité des recherches à grande échelle sur la désinformation. Une partie du défi consiste inévitablement à identifier les meilleures pratiques pour partager les données de désinformation et les ressources computationnelles avec chacunes des parties prenantes sans compromettre la vie privée, la sécurité ou les intérêts exclusifs (propriété intellectuelle notamment). C'est pourquoi nous proposons un modèle de gouvernance du cloud qui facilite les collaborations entre les secteurs autour de la recherche sur la désinformation. Cette gouvernance comprend la définition de la manière dont les parties prenantes accèdent aux ressources et les partagent entre elles pour mener des recherches sur la désinformation, ainsi que la façon dont elles collaborent dans un cadre multidisciplinaire avec des acteurs variés qui dépassent le strict cadre de leur domaine d'activité. Nous espérons qu'en créant des synergies autour de ces collaborations multidisciplinaires, les acteurs seront en mesure d'élaborer des solutions concrètes plus exploitables pour faire face aux phénomènes complexes de la désinformation dont les répercussions sur la sécurité des nations, la sauvegarde de nos démocraties et la santé publique ne cessent d'être constatées. Des premières esquisses de solutions existent comme le concept de « middleware » , un software de désintermédiation produit par un tiers, point d'entré indépendant et intégré aux plateformes (API de Google, Facebook, Twitter...) pour modérer le contenu, réguler les feeds, filtrer ou classer de l'information[7] mais devraient faire l'objet d'une étude d'impact à grande échelle plus systématique. De même, le projet de collaboration entre NYU et le Poynter Institute for Media Studies visant à former chercheurs et décideurs publics sur l'utilisation de la technologie pour la vérification des faits pourrait ainsi passer à l'échelle et être pérennisé[8].

Notons que la garantie d'une gouvernance multipartite non seulement élimine un lourd fardeau de coordination pour les grandes plateformes numériques (par exemple pour Facebook), mais peut aussi limiter les théories conspirationnistes concernant la censure de certains travaux de recherche, voire de certaines

3 — NYU Researchers Were Studying Disinformation On Facebook. The Company Cut Them Off, NPR, August 2021.

4 — Facebook made big mistake in data it provided to researchers, undermining academic work, The Washington Post, September 2021.

5 — Facebook admits it messed up again, Insider, September 2021.

6 — Facebook reportedly provided inaccurate data to misinformation researchers, The Verge, September 2021.

7 — Francis Fukuyama, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed, Marietje Schaake, Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy, Standford University.

8 — Institute for Data, Democracy and Politics, The George Washington University.

voix, qui ne seraient pas alignées sur les intérêts des grandes plateformes numériques [38]. Récemment, par exemple, des décideurs publics britanniques travaillant directement au Parlement européen ont alimenté des théories du complot sur la censure des voix conservatrices par Facebook, en invoquant une forte baisse de l'engagement dans leurs propres pages politiques sur le réseau social [15, 18, 39].

L'outil que nous proposons fournit à de multiples parties prenantes les données et la puissance de calcul dont elles ont besoin pour mener des recherches sur la désinformation. Notre proposition, que nous avons déjà présenté conjointement avec la Federation of American Scientists [29], se compose de trois parties principales pour faciliter les collaborations multidisciplinaires et décentralisées autour de la désinformation : (1) un réseau décentralisé de serveurs qui sera facilement accessible aux parties prenantes pour mener des recherches à grande échelle sur la désinformation ; (2) une «bibliothèque de données» qui contient des données pour aider les chercheurs à mener leurs recherches sur la désinformation ; (3) un modèle de gouvernance pour coordonner les parties prenantes décentralisées afin de partager les ressources informatiques et les données, tout en facilitant les collaborations multidisciplinaires sur la désinformation.

Notre projet de Cloud de recherche présente deux avantages essentiels par rapport aux approches traditionnelles pour mener des recherches sur la désinformation : 1) l'échelle, c'est-à-dire la capacité de créer et de coordonner des équipes de recherche décentralisées de manière dynamique en fonction des intérêts de la recherche ; 2) la diversité, c'est-à-dire la facilité avec laquelle les parties prenantes de divers secteurs et disciplines peuvent être réunies pour mener des recherches sur la désinformation tout en ayant accès à la puissance de calcul et aux données dont elles ont besoin. Nous pensons qu'en permettant à un nombre plus important et plus diversifié de parties prenantes d'accéder à la capacité de calcul, aux données et de bénéficier des mécanismes d'une gouvernance décentralisée, nous permettons la démocratisation de l'étude de la désinformation. Cela conduira à des études plus diversifiées autour de la désinformation, nous aidant à mieux comprendre comment la désinformation cible les minorités à travers les plateformes. Notre outil nous aidera également à concevoir des solutions sociotechniques plus efficaces pour résoudre le problème de la désinformation, ainsi qu'à faciliter la recherche ouverte et les collaborations transnationales en matière de désinformation [23-25, 28].

Gouvernance de la donnée

Notons également que pour qu'un tel outil fonctionne, il est nécessaire d'avoir accès aux données pertinentes. À cet égard, notre Cloud international détient une bibliothèque pour stocker les données de désinformation que les différentes parties prenantes peuvent utiliser pour leurs études. Pour obtenir l'accès initial aux données des plateformes numériques, notre outil fonctionne avec un accès API par abonnement. Nous travaillons ici avec les API que Facebook et Twitter [17, 19, 22, 34], proposent déjà aux scientifiques, ainsi qu'avec les API que différentes entreprises, comme Meltwater [32], ont mis à disposition. Meltwater est la première société de veille des médias en ligne au monde et propose une vaste gamme d'API pour la collecte de données provenant de différentes sources. Plus précisément, Meltwater propose des données provenant de 15 canaux différents de plateformes du numérique (y compris l'accès à l'ensemble du contenu de Twitter), de blogs, des commentaires de différents posts en ligne, d'articles d'actualité et de revues de produits. Meltwater peut également fournir des données pour plus de 270 000 sources d'information dans le monde, plus de 25 000 podcasts, et offre plusieurs options différentes pour collecter des données sur les émissions de télévision et de radio. Grâce à cela, nous alimentons notre Cloud de recherche en données initiales.

Vie privée et considérations éthiques

Notre cloud sera fermé, ce qui signifie que seules les institutions et les personnes accréditées par l'équipe de gouvernance pourront accéder aux données et aux ressources de calcul. L'équipe de gouvernance elle-même résulterait d'une concertation entre les grands organismes de recherche tels que le Centre de recherche et développement fédéralement financé (FFRDC) aux Etats-Unis ou le Centre Européen de recherche (ERC) en Europe. À l'instar des pratiques en vigueur dans les collaborations et les archives de données pour les données sensibles [33, 35], nous serons attentifs au fait que l'accès aux données ne doit en aucun cas constituer un premier pas vers de potentielles violations ou atteintes à la vie privée des utilisateurs des plateformes. Les données de désinformation qui seront partagées seront agrégées et anonymisées, suivant des règles strictes pour garantir la confidentialité. L'équipe de gouvernance travaillera avec des comités d'examen institutionnels et des éthiciens sur un ensemble de principes de fonctionnement garantissant que toutes les recherches sur la désinformation produites sur le Cloud respectent les droits de l'homme et la législation sur la vie privée.

Un point qu'il est essentiel d'examiner est la manière de permettre le partage des données entre l'Europe et les autres continents. Il est important de noter que ces mouvements de données sont devenus de plus en plus complexes depuis l'arrêt de la Cour de justice de l'Union européenne qui a invalidé le privacy shield, mécanisme issu du RGPD qui permettait un partage simplifié des données entre l'Europe et le continent américain[9]. La Cour européenne a estimé qu'en raison du US Cloud Act et des lois connexes du continent américain, ce dernier n'était pas en mesure de garantir aux citoyens européens que leurs données seraient traitées dans un cadre aussi protecteur que celui du RGPD[10]. Il faut donc étudier de nouveaux mécanismes qui puissent permettre un partage efficace des données entre l'Europe et les États-Unis.

Clouds de recherche européens et infrastructures de calcul publiques aux Etats-Unis et en Amérique latine

Compte tenu des graves menaces que la désinformation fait peser sur le modèle des démocraties libérales, il est crucial que les nations démocratiques coordonnent leurs efforts de recherche et élaborent des stratégies cohérentes. Nous pensons qu'ensemble, l'Europe, les États-Unis et l'Amérique latine devraient établir un partenariat trilatéral solide en matière de technologie afin de placer les valeurs démocratiques communes et l'État de droit en tête des priorités. Notre proposition d'un Cloud pour la recherche sur la désinformation s'appuie sur les Clouds de recherche transnationaux existants, tels que l'Open Science Cloud (EOSC) de l'European Cloud Initiatives et l'initiative franco-allemande GAIA-X [10,12,20,36]. Cette dernière est présentée comme un «Cloud de recherche»[11], un «écosystème fédéré de coopération» permettant de gérer les problèmes publics dans une perspective multipartite et d'utiliser la puissance de calcul pour élaborer de meilleures politiques. Son rapport sur l'architecture technique fournit un cadre pour l'architecture de base, la gouvernance, l'écosystème d'exploitation, les éléments de sécurité et les dispositions relatives à la protection des données qui pourraient servir à notre proposition de Research Cloud. L'initiative franco-allemande GAIA-X fait valoir que la création de ce type d'infrastructure de données publiques à l'échelle européenne permettra à l'Union européenne de devenir plus compétitive et de promouvoir des processus, des produits et des services innovants [12]. Nous pensons que notre Cloud global et décentralisé sur la désinformation pourrait fonctionner dans le cadre de la stratégie numérique de l'Europe, du projet Gaia-X, de VIGINUM[12] ou de l'Observatoire italien des médias numériques[13] les nouvelles agences française et italienne de lutte contre la désinformation. Ensemble, nous créerons un écosystème florissant dans lequel les universitaires, les entreprises, les pouvoirs publics et les ONG travailleront ensemble pour résoudre le problème de la désinformation. Le modèle de gouvernance que nous proposons facilitera la mise en place d'un cadre réglementaire et normatif harmonisé grâce auquel les parties prenantes de différents pays européens pourront accéder à des données ouvertes et publiques sur la désinformation et les utiliser pour comprendre comment ces campagnes fonctionnent dans un large éventail de scénarios différents. Une collaboration étroite avec ces initiatives européennes permettra également d'unir les efforts de recherche fragmentés autour de la désinformation, et aidera à comprendre comment les campagnes de désinformation étrangères peuvent potentiellement monter les pays les uns contre les autres [4,9]. Le lien avec le projet Gaia-X et l'Observatoire italien des médias numériques aidera également les pays européens à atteindre la souveraineté numérique autour de l'étude à grande échelle de la désinformation. Nous pensons également qu'il sera utile d'établir des liens avec la Commission européenne qui a lancé une consultation publique sur son paquet de mesures relatives aux services numériques (DSA ou IA Act)[14]. Travailler avec la Commission européenne nous aidera probablement à mieux identifier les déséquilibres du marché qui existent dans différents secteurs pour étudier la désinformation à grande échelle, ainsi qu'à créer des stratégies globales pour mieux soutenir ces secteurs.

S'inscrivant aussi dans le contexte américain, notre proposition s'inspire des efforts déployés par le Stanford

9 — The European Court of Justice has ruled that Privacy Shield is invalid, Wired, July 2020.

10 — EU-U.S. Privacy Shield Program Update, Privacy Shield Framework.

11 — GAIA-X: A Franco-German pitch towards a European data infrastructure, Ministerial talk and GAIA-X virtual expert forum 4 June 2020, Livestream.

12 — LE SERVICE DE VIGILANCE ET DE PROTECTION CONTRE LES INGÉRENCES NUMÉRIQUES ÉTRANGÈRES (VIGINUM).

13 — The Digital Media Observatory Against Fake News and Disinformation is Born, Teller Report, September 2021.

14 — The Digital Services Act package, European Commission.

Institute of Human-Centered Artificial Intelligence (HAI) pour créer un Cloud de recherche national pour l'IA, notamment en créant un groupe de travail national sur les ressources de recherche en IA chargé d'étudier la faisabilité et l'opportunité d'une plateforme basée sur le Cloud pour la recherche en IA à grande échelle. Dans le contexte de l'Amérique latine, nous souhaitons nous appuyer sur l'infrastructure publique de calcul distribué que nous avons conçue précédemment avec la présidence mexicaine[15][16].

Si un Cloud de recherche national pour l'IA est créé aux États-Unis ou en Amérique latine, il y aura sans aucun doute des leçons à partager de l'autre côté de l'océan Atlantique. Ces leçons peuvent être codifiées dans un partenariat de recherche unifié pour contrer la désinformation, une force déstabilisante persistante pour une démocratie intrinsèquement fragile. En bref, si nous voulons renforcer la résilience de nos normes démocratiques, nous devons apprendre et mettre en œuvre des solutions qui fonctionnent, ensemble.

15 — BlockChain Hack MX, Gobierno de México.

16 — Modelo de Gobernanza para implementar la RedBlockchain México, Gobierno de México.

RÉFÉRENCES

[1] Law n° 2018-1202 of december 22 2018 on fighting disinformation. 2018.

[2] M. Abdalla and M. Abdalla. The grey hoodie project: Big tobacco, big tech, and the threat on academic integrity. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pages 287-297, 2021.

[3] M. T. Bastos and D. Mercea. The brexit botnet and user-generated hyperpartisan news. Social science computer review, 37(1):38-54, 2019.

[4] M. Baumann. 'propaganda fights' and 'disinformation campaigns': the discourse on information warfare in russia-west relations. Contemporary politics, 26(3):288-307, 2020.

[5] H. P. Beck and P. Charitos. The Economics of Big Science: Essays by Leading Scientists and Policymakers. Springer Nature, 2021.

[6] M. M. Ben Buchanan, Andrew Lohn and K. Sedova. Truth, lies, and automation: How language models could change disinformation. 2021.

[7] N. Bliss, E. Bradley, J. Garland, F. Menczer, S. W. Ruston, K. Starbird, and C. Wiggins. An agenda for disinformation research. arXiv preprint arXiv:2012.08572, 2020.

[8] A. Bovet and H. A. Makse. Influence of fake news in twitter during the 2016 us presidential election. Nature communications, 10(1):1-14, 2019.

[9] S. Bradshaw and P. N. Howard. The global organization of social media disinformation campaigns. Journal of International Affairs, 71(1.5):23-32, 2018.

[10] P. Budroni, J. Claude-Burgelman, and M. Schouppe. Architectures of knowledge: the european open science cloud. ABI Technik, 39(2):130-141, 2019.

[11] A. Carstens. Big tech in finance and new challenges for public policy. speech to FT Banking Summit, 2, 2018.

[12] E. Celeste. Digital sovereignty in the eu: challenges and future perspectives. Data Protection Beyond Borders: Transatlantic Perspectives on Extrater ritoriality and Sovereignty, pages 211-228, 2021.

[13] A. Esthose Suresh. People over profit: saving social media from big tech. LSE Business Review, 2021.

[14] S. Fachard, S. C. Murray, A. R. Knodell, and K. Papangeli. The fortress of eleutherai: New insights from survey, architecture, and epigraphy. Hesperia: The Journal of the American School of Classical Studies at Athens, 89(3):475-549, 2020.

[15] K. Fangen and C. R. Holter. The battle for truth: How online newspaper commenters defend their censored expressions. Poetics, 80:101423, 2020.

[16] Y. Gorodnichenko, T. Pham, and O. Talavera. Social media, sentiment and public opinions: Evidence from# brexit and# uselection. European Economic Review, page 103772, 2021.

[17] D. Jacobson, G. Brail, and D. Woods. APIs: A strategy guide. « O'Reilly Media, Inc.», 2012.

[18] A. Juhász and P. Szicherle. The political effects of migration-related fake news, disinformation and conspiracy theories in europe. Friedrich Ebert Stiftung, Political Capital, Budapest, 2017.

[19] R. Kimmons, J. Rosenberg, and B. Allman. Trends in educational technology: What facebook, twitter, and scopus can tell us about current research and practice. TechTrends, pages 1-12, 2021.

[20] K. Komaitis. Europe's ambition for digital sovereignty must not undermine the internet's values. Computer Fraud & Security, 2021(1):11-13, 2021.

[21] S. Kumar, R. West, and J. Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In Proceedings of the 25th international conference on World Wide Web, pages 591-602, 2016.

[22] S. Lomborg and A. Bechmann. Using apis for data collection on social media. The Information Society, 30(4):256-265, 2014.

[23] I. V. Pasquetto, B. Swire-Thompson, M. A. Amazeen, F. Benevenuto, N. M. Brashier, R. M. Bond, L. C. Bozarth, C. Budak, U. K. Ecker, L. K. Fazio, et al. Tackling misinformation: What researchers could do with social media data. The Harvard Kennedy School Misinformation Review, 2020.

[24] N. Persily and J. A. Tucker. Conclusion: The challenges and opportunities for social media research. Social Media and Democracy: The State of the Field, Prospects for Reform, page 313, 2020.

[25] N. Persily and J. A. Tucker. Social Media and

Democracy: The State of the Field, Prospects for Reform. Cambridge University Press, 2020.

[26] F. Pierri, A. Artoni, and S. Ceri. Investigating italian disinformation spreading on twitter in the context of 2019 european elections. PloS one, 15(1):e0227821, 2020.

[27] O. Pollicino and E. Bietti. Truth and deception across the atlantic: a roadmap of disinformation in the us and europe. Italian J. Pub. L., 11:43, 2019.

[28] M. Saad, A. Ahmad, and A. Mohaisen. Fighting fake news propagation with blockchains. In 2019 IEEE Conference on Communications and Network Security (CNS), pages 1-4. IEEE, 2019.

[29] S. Savage. A national cloud for conducting disinformation research at scale. Policy Brief for the Federation of American Scientists 2021., page 1, 2021.

[30] J. Schoenfeld. Big tech, cloud computing, and accounting. Tuck School of Business Working Paper, (3596065), 2021.

[31] K. Starbird. Online rumors, misinformation and disinformation: The perfect storm of covid-19 and election2020. 2021.

[32] I. Stavrakantonakis, A.-E. Gagiu, H. Kasper, I. Toma, and A. Thalhammer. An approach for evaluation of social media monitoring tools. Common Value Management, 52(1):52-64, 2012.

[33] I. Susha, M. Janssen, and S. Verhulst. Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. 2017.

[34] R. Tromble. Where have all the data gone? a critical reflection on academic digital research in the post-api age. Social Media+ Society, 7(1):2056305121988929, 2021.

[35] S. Verhulst, A. Young, and P. Srinivasan. An introduction to data collaboratives. Creating Public Value by Exchanging Data. The GovLab, UNICEF, Omidyar Network, New York, 2017.

[36] I. I. Veršić and J. Ausserhofer. Social sciences, humanities and their interoperability with the european open science cloud: What is sshoc? Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 72(2):383-391, 2019.

[37] Z. Vorhies and K. Heckenlively. Google Leaks: A Whistleblower's Exposé of Big Tech Censorship. Simon and Schuster, 2021.

[38] R. Willis. Observations online: Finding the ethical boundaries of facebook research. Research Ethics, 15(1):1-17, 2019.

[39] D. Yoldas. Eu-russia information war, human right, and democracy-fake news, fact-checking, conspiracy theories and hate-speech in post-truth and illiberal democracies age.