skema THINK TANK | PUBLIKA

NOTE
INCERTITUDES

DES CARTES SANS DESTINATIONS ?

Juin 2026

De la description des flux à la mesure des effets dans l'influence numérique

Jean LANGLOIS-BERTHELOT & Paul JANIN

SKEMA PUBLIKA

SKEMA Publika est le think tank international de SKEMA Business School, il analyse les grandes mutations sociales, économiques, technologiques et géopolitiques pour éclairer la décision publique et privée. Sur la base des travaux de l'école et de contributions extérieures validés scientifiquement, le think tank alimente le débat public et émet des recommandations pour les décideurs nationaux et internationaux.

SKEMA Publika mobilise une approche interdisciplinaire et internationale, nourrie par le réseau mondial de campus de SKEMA et une communauté d'experts issus du monde académique et professionnel. Cette dimension internationale ne constitue pas seulement un réseau, mais un mode de pensée. Ainsi, Publika articule les dynamiques locales avec les transformations globales pour proposer un regard décentré et multipolaire sur les grands enjeux contemporains.

Couverture : Un très grand écran de nombreux types de téléviseurs. Eamonn Wang. 2023. Unplash.com

SKEMA Publika
SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Tél. : +33.1.71.13.39.32
Courriel : publika@skema.edu
Site Internet : www.publika.skema.edu

COLLECTION

« Explorer les zones grises de la géopolitique, du numérique et des risques globaux ».

Dans la « collection incertitude » les travaux analysent les tensions, risques émergents et les dynamiques d'instabilité à l'échelle mondiale.

COMITE DE LECTURE

Frédérique Vidal est Directrice du Développement de SKEMA Publika et Directrice de la Stratégie et de l'Impact Scientifique de SKEMA Business School. Professeur des universités en biologie, elle a été Présidente de l'université Nice-Sophia-Antipolis entre 2012 et 2017, puis ministre de l'Enseignement Supérieur, de la Recherche et de l'Innovation, dans les Gouvernements Philippe et Castex de 2017 à 2022. Elle a été Conseillère spéciale auprès du Président de l'EFMD et est aussi actuellement Représentante permanente de la Principauté de Monaco auprès du Programme des Nations unies pour l'environnement et de la Commission baleinière internationale.

Sean Scull, Chargé de projets think tank, est Doctorant en science de l'information et de la communication à l'université Paul Valéry-Montpellier III. Il est diplômé en sciences politiques avec une spécialisation en relations internationales de l'université de Göteborg et d'un Master en politiques internationales avec une spécialisation en politique anglophone de l'université de Toulon. Sean a vécu et travaillé en Suède et aux États-Unis d'Amérique.

AUTEURS

Jean Langlois-Berthelot est responsable chez Kannon Labs, où il travaille dans le domaine de la simulation, de l'intelligence décisionnelle et de l'évaluation technologique appliquée aux environnements stratégiques. Il est actuellement Professeur invité au Center for Advanced Studies de l'Université de Palerme et travaille sur des expérimentations dans le cadre de l'exercice militaire ORION26.

Paul Janin est officier supérieur de l'armée de Terre, spécialisé dans les enjeux de sciences cognitives, d'influence et de guerre cognitive. Désigné par l'Enseignement militaire supérieur scientifique et technique pour conduire une analyse de l'écosystème scientifique, académique et institutionnel travaillant sur ces questions, il a mené un travail d'immersion, d'analyse et d'échange en interface avec les principaux établissements de référence du domaine. Il a notamment été en immersion au CEA Paris-Saclay et en liaison avec les scientifiques travaillant sur les programmes de financement AID/DGA consacrés à la guerre cognitive et à l'influence. Il a publié dans plusieurs revues, dont la Revue Défense Nationale, Polytechnique Insights et ISTE OpenScience.

RESUME

L'influence numérique occupe désormais une place centrale dans le champ cyber. Elle apparaît dans les doctrines de lutte informatique d'influence, dans les politiques publiques contre les manipulations de l'information, dans les travaux sur les plateformes, dans les débats sur l'intelligence artificielle générative et dans les réflexions sur la résilience démocratique. Cette centralité n'est pas artificielle : les opérations contemporaines ne séparent plus clairement l'attaque technique, la manœuvre informationnelle et la désorganisation sociale.

Ce constat a amené les auteurs a à poser la problématique suivante : quelles sont les limites des outils de veille informationnelle lorsqu'il s'agit de mesurer l'efficacité réelle d'une opération d'influence ? Autrement dit, comment passer de la description des flux à la mesure des effets ? L'hypothèse défendue ici est simple : l'influence numérique ne peut être évaluée sérieusement qu'à partir d'un déplacement comportemental rapporté à un scénario contrefactuel. Les récits, les réseaux, les biais cognitifs, les relais sociaux et les métriques de plateforme constituent des variables utiles, mais intermédiaires. Ils ne deviennent décisifs que lorsqu'ils permettent d'estimer ce qui change effectivement par rapport à ce qui se serait probablement produit sans intervention.

TABLE DES MATIERES

INTRODUCTION

L'influence numérique occupe désormais une place centrale dans le champ cyber. Elle apparaît dans les doctrines de lutte informatique d'influence, dans les politiques publiques contre les manipulations de l'information, dans les travaux sur les plateformes, dans les débats sur l'intelligence artificielle générative et dans les réflexions sur la résilience démocratique. Cette centralité n'est pas artificielle : les opérations contemporaines ne séparent plus clairement l'attaque technique, la manœuvre informationnelle et la désorganisation sociale.

Une fuite de données peut être exploitée pour produire un effet réputationnel, une campagne de dénigrement peut accompagner une cyberattaque, une opération de sabotage symbolique peut être amplifiée par les réseaux sociaux, un contenu généré par IA peut accélérer la saturation informationnelle avant même que sa véracité soit vérifiée. Le cyber n'est donc plus seulement un espace d'intrusion, de compromission ou de protection technique. Il est aussi devenu un espace de production d'effets cognitifs, sociaux et politiques.

Cette évolution explique la multiplication des outils de veille, de détection et d'analyse. Les organisations savent aujourd'hui observer des volumes de messages, détecter des coordinations, cartographier des réseaux, suivre des hashtags, comparer des récits, identifier des relais, mesurer des interactions et produire des tableaux de bord. La France a elle-même formalisé cette préoccupation dans la Stratégie nationale de lutte contre les manipulations de l'information 2026-2030, publiée le 11 février 2026 sous l'égide du SGDSN.[1] Mais cette montée en puissance des instruments ne résout pas le problème scientifique central. Au contraire, elle peut même le masquer.

Cela nous amène à poser la problématique suivante : quelles sont les limites des outils de veille informationnelle lorsqu'il s'agit de mesurer l'efficacité réelle d'une opération d'influence ? Autrement dit, comment passer de la description des flux à la mesure des effets ? Les indicateurs disponibles (impressions, engagement, récurrence narrative, volume de comptes, vitesse de propagation, visibilité d'un récit) décrivent un environnement

[1] Secrétariat général de la défense et de la sécurité nationale (SGDSN), Stratégie nationale de lutte contre les manipulations de l'information 2026-2030, 11 février 2026.

informationnel mais ils ne mesurent pas directement l'effet d'une opération. Ils disent qu'un contenu a été vu, repris, commenté ou amplifié mais ils ne disent pas si une population a modifié un comportement, une intention robuste, une décision, une coopération, une mobilisation, une abstention, un achat, une défiance institutionnelle ou une pratique de sécurité.

L'hypothèse défendue ici est la suivante : l'influence numérique ne peut être évaluée sérieusement qu'à partir d'un déplacement comportemental rapporté à un scénario contrefactuel. Les récits, les réseaux, les biais cognitifs, les relais sociaux et les métriques de plateforme constituent des variables utiles, mais intermédiaires. Ils ne deviennent décisifs que lorsqu'ils permettent d'estimer ce qui change effectivement par rapport à ce qui se serait probablement produit sans intervention.

I. MESURER L'IMPACT DES OPERATIONS D'INFLUENCE : SORTIR DE LA CONFUSION ENTRE VISIBILITE ET EFFET

Ce point avait déjà été posé comme difficulté opérationnelle dans une restitution scientifique de l'Agence de l'innovation de défense au laboratoire Cognition Humaine et Artificielle de l'École Pratique des Hautes Etudes en 2018 : les systèmes d'analyse de l'influence reposaient majoritairement sur des variables accessibles parce qu'observables, mais insuffisantes pour établir un impact comportemental. La formulation actuelle du problème ne fait donc pas apparaître une nouveauté radicale. Elle confirme plutôt un diagnostic ancien, étant celui d'un champ qui sait de mieux en mieux décrire les signaux, mais il peine encore à estimer les effets.

La littérature scientifique permet de comprendre pourquoi cette question est difficile. Depuis les années 1940, les recherches sur l'influence ont progressivement réglé plusieurs dimensions du problème, sans pour autant résoudre la mesure de l'effet causal. Le premier moment important est celui de la communication de masse. En 1948, Harold Lasswell formule[2] la question devenue classique : qui dit quoi, par quel canal, à qui, avec quel effet ? Cette formulation a le mérite de placer l'effet au cœur du problème, mais elle ne fournit pas encore une méthode robuste pour l'estimer. Elle ouvre le champ plutôt qu'elle ne le ferme.

Le deuxième moment se situe entre 1944 et 1955 avec les travaux de Paul Lazarsfeld, Bernard Berelson, Hazel Gaudet puis Elihu Katz.[3] *The People's Choice*, publié en 1944, puis *Personal Influence*, publié en 1955, déplacent l'analyse. L'influence ne passe pas directement des médias vers les individus. Elle transite par des relais, des leaders d'opinion, des groupes primaires et des médiations sociales. C'est ce que l'on surnomme le modèle du *two-step*

[2] Harold D. Lasswell, « The Structure and Function of Communication in Society », in Lyman Bryson (dir.), The Communication of Ideas, New York, Harper & Brothers, 1948, p. 37-51.

[3] Paul F. Lazarsfeld, Bernard Berelson et Hazel Gaudet, The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign, New York, Columbia University Press, 1944 ; Elihu Katz et Paul F. Lazarsfeld, Personal Influence: The Part Played by People in the Flow of Mass Communications, Glencoe, Free Press, 1955.

flow : l'effet d'un message dépend de la structure sociale dans laquelle il s'insère. Pour le cyber, ce résultat reste fondamental dans la mesure où il explique pourquoi une information techniquement visible peut rester socialement faible, tandis qu'un contenu moins diffusé peut produire des effets forts s'il passe par des relais crédibles ou des communautés denses.

Ce modèle relationnel est renforcé par les recherches contemporaines sur les réseaux. Les travaux de Duncan Watts, notamment au début des années 2000, montrent[4] que la diffusion dépend de la topologie du réseau, des seuils d'adoption et des propriétés locales des connexions. Manuel Castells formalise, dans les années 1990 et 2000, la société en réseaux comme structure centrale des sociétés contemporaines. Sinan Aral analyse ensuite, dans les années 2010, la propagation de l'information en ligne et les conditions dans lesquelles les contenus se diffusent. Le cyber hérite directement de cette tradition, celle d'un environnement numérique qui n'est pas une simple addition d'individus exposés, mais un système de relations, de canaux, de communautés, de plateformes et d'intermédiaires.

L'étude de Bond et al., publiée dans Nature en 2012, constitue ici un repère empirique majeur.[5] Elle porte sur 61 millions d'utilisateurs de Facebook lors des élections américaines de 2010 et montre que des messages sociaux peuvent influencer l'expression politique, la recherche d'information et le vote réel. Mais l'apport principal n'est pas seulement la taille de l'expérience. Il tient au rôle des relations sociales : l'effet ne provient pas d'une exposition abstraite, mais d'une exposition socialement située, notamment via les amis et les amis d'amis. Pour le champ cyber, cela signifie qu'un indicateur de portée brute est insuffisant. Ce qui compte est la combinaison entre exposition, crédibilité du relais, position dans le réseau et proximité comportementale de la cible.

Le troisième moment scientifique majeur est celui de la psychologie cognitive et de la décision. Dans les années 1970 et 1980, Daniel Kahneman

[4] Duncan J. Watts, Six Degrees: The Science of a Connected Age, New York, W. W. Norton, 2003 ; Manuel Castells, The Rise of the Network Society, Oxford, Blackwell, 1996 ; Sinan Aral, The Hype Machine, New York, Currency, 2020.

[5] Robert M. Bond et al., « A 61-million-person experiment in social influence and political mobilization », Nature, vol. 489, 2012, p. 295-298.

et Amos Tversky montrent[6] que les individus ne traitent pas l'information comme des calculateurs rationnels. Ils utilisent des heuristiques, subissent des biais d'ancrage, de disponibilité, de cadrage et d'aversion à la perte. Le message ne vaut donc pas seulement par son contenu, mais par la manière dont il active un raccourci cognitif. En 1986, Richard Petty et John Cacioppo proposent l'Elaboration Likelihood Model, distinguant la voie centrale de persuasion, qui suppose un traitement approfondi, et la voie périphérique, qui repose sur des indices contextuels. En 1984, Robert Cialdini formalise dans Influence plusieurs principes robustes (preuve sociale, autorité, rareté, engagement, réciprocité) qui structurent encore la communication persuasive.

Ces travaux règlent une question essentielle : l'influence n'est pas l'exposition. Une personne peut voir un message sans le traiter, le traiter sans le croire, le croire sans agir et les environnements cyber amplifient cette dissociation. Les utilisateurs sont exposés à des flux massifs, souvent contradictoires, dans des conditions d'attention fragmentée. La surcharge informationnelle, les formats courts, la vitesse de circulation et les signaux de popularité modifient les conditions de traitement. Le nombre de vues devient alors un indicateur très pauvre. Il mesure l'accès potentiel au stimulus, pas sa transformation en décision.

Le quatrième moment concerne les cadres discursifs et les récits. Erving Goffman publie *Frame Analysis* en 1974.[7] George Lakoff développe ensuite, dans les années 1980 et 1990, une théorie des métaphores conceptuelles et du framing politique. Teun van Dijk consolide l'analyse critique du discours en reliant structures linguistiques, cognition sociale et rapports de pouvoir. Ces travaux ont depuis longtemps montré que le langage n'est pas une couche superficielle. Il organise les perceptions, active des cadres d'interprétation, désigne des responsables, hiérarchise les menaces, stabilise des oppositions morales. Les approches contemporaines centrées

[6] Daniel Kahneman et Amos Tversky, « Judgment under Uncertainty: Heuristics and Biases », Science, vol. 185, no 4157, 1974, p. 1124-1131 ; Daniel Kahneman et Amos Tversky, « Prospect Theory: An Analysis of Decision under Risk », Econometrica, vol. 47, no 2, 1979, p. 263-291 ; Richard E. Petty et John T. Cacioppo, Communication and Persuasion: Central and Peripheral Routes to Attitude Change, New York, Springer, 1986 ; Robert B. Cialdini, Influence: The Psychology of Persuasion, New York, William Morrow, 1984.

[7] Erving Goffman, Frame Analysis: An Essay on the Organization of Experience, Cambridge, Harvard University Press, 1974 ; George Lakoff et Mark Johnson, Metaphors We Live By, Chicago, University of Chicago Press, 1980 ; Teun A. van Dijk, Discourse and Power, Basingstoke, Palgrave Macmillan, 2008.

sur les récits ne découvrent donc pas un continent vierge. Elles réactualisent, dans les environnements numériques, un acquis ancien : les messages n'opèrent pas comme des unités isolées, mais comme des configurations narratives et interprétatives.

Ce rappel est important pour le débat français. Les approches récentes qui insistent sur les récits, la sérialité des discours ou la structuration narrative apportent une contribution utile à l'analyse des environnements informationnels. Elles permettent de mieux suivre les continuités, les reprises, les motifs, les personnages collectifs, les schémas d'accusation, les séquences de victimisation ou de délégitimation. Elles sont pertinentes pour comprendre ce qui circule mais elles ne règlent pas la mesure de l'effet. Un récit cohérent n'est pas nécessairement influent et un récit visible n'est pas nécessairement comportementalement actif. Puis, un récit marginal peut produire un effet s'il atteint des acteurs proches d'un seuil d'action. À l'inverse, un récit massif peut rester une simple pollution de fond.

Le cinquième moment est celui de la mesure empirique des effets, et il est beaucoup plus sévère qu'on ne le croit souvent. La méta-analyse de Joshua Kalla et David Broockman, publiée en 2018 dans *l'American Political Science Review*,[8] examine les effets du contact de campagne et de la publicité politique sur les choix électoraux. Leur conclusion est nette : dans les élections générales américaines, la meilleure estimation moyenne des effets persuasifs sur le choix du candidat est proche de zéro ; leur synthèse inclut notamment une méta-analyse systématique de 49 expériences de terrain. Ce résultat ne signifie pas que toute influence est impossible. Il signifie que l'effet ne peut pas être présumé à partir de l'exposition. Dans des environnements saturés, polarisés, identitaires ou fortement préstructurés, un message supplémentaire peut ne rien déplacer.
Ce résultat est capital pour l'influence numérique et le cyber dans la mesure où il remet en cause l'idée selon laquelle plus de messages, plus de visibilité ou plus de rapidité produiraient mécaniquement plus d'effet. L'augmentation du volume peut renforcer une campagne ; elle peut aussi saturer, banaliser, provoquer un rejet ou produire une amplification adverse. La logique cyber pousse souvent à réagir vite, parce que l'environnement technique valorise la détection et la réponse. Mais l'influence n'est pas un malware : son

[8] Joshua L. Kalla et David E. Broockman, « The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments », American Political Science Review, vol. 112, no 1, 2018, p. 148-166.

traitement ne se réduit pas à l'identification d'une menace et à son éradication. Une réponse informationnelle peut être efficace, inutile ou contre-productive selon le public, le canal, le moment, la crédibilité de la source et la proximité des individus avec un seuil de comportement.

C'est ici que le problème de mesure devient central. Le champ dispose désormais d'une masse considérable d'acquis : les réseaux comptent, les relais sociaux comptent, les biais cognitifs comptent, les cadres discursifs comptent, les signaux de plateforme comptent. Mais ces éléments ne suffisent pas à dire qu'une opération a produit une influence.

Dans une perspective rigoureuse, l'effet d'influence doit être compris comme un déplacement comportemental par rapport à une situation sans intervention. C'est précisément le cœur du modèle défendu ici : les variables observables (production, visibilité, circulation, engagement, coordination, récurrence narrative) décrivent l'environnement informationnel, mais ne mesurent pas en elles-mêmes l'influence. L'objet de mesure doit être le déplacement comportemental contrefactuel, c'est-à-dire la différence entre ce qu'une population fait sous intervention et ce qu'elle aurait probablement fait sans intervention.

Il ne s'agit pas de dire que les approches narratives, les méthodes d'ingénierie ou les programmes de résilience cognitive sont inutiles. Il faut dire plus précisément qu'ils occupent aujourd'hui le débat parce qu'ils traitent des objets visibles, institutionnellement commodes et scientifiquement manipulables. Les récits se décrivent, les réseaux se cartographient et les dispositifs se conçoivent. La résilience se teste par questionnaires ou expériences. Mais la question centrale, celle du comportement qui aurait eu lieu sans intervention, reste souvent absente, car elle est la plus difficile.

Cette difficulté est encore plus importante dans le domaine cyber. Les opérations sont hybrides, les publics fragmentés, les données incomplètes, les plateformes instables, les effets différés, les adversaires adaptatifs. Une campagne peut viser à faire partager, mais aussi à faire douter, à décourager, à désorganiser, à détourner l'attention, à produire une fatigue, à fragiliser la confiance dans une procédure, à rendre coûteuse une décision ou à rendre acceptable une action future. Beaucoup de ces effets ne se lisent pas directement dans les métriques de plateforme. Ils apparaissent dans des comportements faibles, distribués, parfois retardés.

La conséquence scientifique est claire : la prochaine frontière ne consiste pas à produire une typologie supplémentaire des récits ni un tableau de bord plus riche. Elle consiste à relier les indicateurs disponibles à une théorie de l'effet. Cette théorie doit distinguer l'exposition, la susceptibilité, la cognition, la norme sociale, le coût d'action, la faisabilité et le comportement. Elle doit aussi comparer les scénarios : réponse publique, réponse discrète, inoculation, correction, silence, changement de canal, mobilisation de relais locaux. Sans comparaison, l'évaluation devient une narration après coup.

La première conclusion est donc la suivante : le champ de l'influence numérique dans le cyber n'est pas en manque de données, ni même en manque de concepts. Il est en manque d'une architecture de mesure qui mette chaque concept à sa place. Les récits décrivent les structures de sens, les réseaux décrivent la circulation, les métriques d'engagement décrivent l'interaction et les approches cognitives décrivent la réception. Mais l'influence, au sens strict, ne commence que lorsqu'un comportement est déplacé par rapport à ce qui aurait eu lieu sans intervention.

II. LES LIMITES DE L'ANALYSE NARRATIVE, DE L'INGENIERIE COGNITIVE ET DE LA RESILIENCE DU RECEPTEUR POUR MESURER L'INFLUENCE

Le diagnostic posé en 2018 par le rapport de l'Agence de l'innovation de défense permet de lire avec plus de netteté les orientations qui occupent aujourd'hui le débat français sur l'influence numérique. Depuis cette date, le champ ne s'est pas appauvri, il s'est au contraire densifié. La France dispose désormais d'une stratégie nationale de lutte contre les manipulations de l'information 2026-2030, publiée le 11 février 2026, qui articule quatre priorités : résilience de la Nation, encadrement des plateformes et de l'IA générative, consolidation des capacités de détection-attribution-réponse, et coopération européenne et internationale. Ce texte confirme que l'influence numérique est désormais traitée comme un enjeu de sécurité nationale, à l'intersection du cyber, de la protection du débat public, de l'OSINT, de la régulation des plateformes et de la résilience collective. Il acte aussi un fait important : la lutte contre les manipulations de l'information ne relève plus seulement de la communication, mais d'un dispositif interministériel de protection de l'espace informationnel.

Cette montée en puissance institutionnelle donne un cadre à des recherches et à des pratiques jusque-là dispersées mais elle ne règle pas le problème de mesure, elle le déplace. Trois approches occupent aujourd'hui une place disproportionnée dans le débat : l'analyse narrative, l'ingénierie cognitive des dispositifs, et la résilience cognitive du récepteur. Ces trois approches sont utiles ; aucune ne doit être rejetée en bloc. Leur faiblesse n'est pas d'être fausse, mais de se situer en amont ou à côté de l'objet qu'il faut mesurer. Elles décrivent des structures, conçoivent des interventions ou renforcent des capacités individuelles, mais elles ne permettent pas encore, à elles seules, d'estimer l'effet comportemental d'une intervention d'influence dans un environnement cyber.

La première orientation est celle du récit. Elle consiste à considérer les manipulations informationnelles non comme une addition de fausses

propositions, mais comme des structures narratives répétées, stabilisées et cumulatives. L'approche est intellectuellement solide, puisqu'elle prolonge des travaux anciens sur le cadrage, la narrativité et les structures discursives. Dans le débat français récent, cette orientation est explicitement formulée à travers l'idée de lire la désinformation comme un « récit sériel », c'est-à-dire comme une construction discursive qui progresse par épisodes, personnages, conflits récurrents et schémas interprétatifs continus. Paul Charon défend cette approche dans un article publié en novembre 2024 dans Le Rubicon, présenté comme une approche littéraire des manipulations de l'information.[9]

L'apport est évident : cette approche corrige la naïveté qui consisterait à traiter chaque contenu faux ou manipulatoire comme un objet isolé. Une opération d'influence fonctionne rarement par un seul message. Elle fonctionne par répétition, par familiarisation, par stabilisation d'un monde interprétatif. Dans le champ cyber, cette intuition est utile : après une fuite de données, une attaque par rançongiciel, une compromission institutionnelle ou une opération de sabotage numérique, le récit qui accompagne l'événement peut produire un cadrage durable. Il peut transformer une vulnérabilité technique en preuve d'incompétence politique, une attaque opportuniste en humiliation stratégique, ou une perturbation locale en signe d'effondrement systémique. L'analyse narrative aide donc à comprendre comment une séquence cyber devient une séquence informationnelle.

Mais cette approche ne mesure pas l'influence : elle décrit la structure du discours et ne dit pas si ce discours a déplacé un comportement. En d'autres termes, elle permet de savoir qu'un récit est cohérent, qu'il se répète, qu'il s'articule à des épisodes antérieurs, qu'il construit un univers moral mais elle ne permet pas de déterminer si une population a changé d'attitude pratique : coopérer ou non avec une institution, relayer ou non une information, appliquer ou non une consigne de cybersécurité, migrer ou non vers un canal, participer ou non à une mobilisation, faire confiance ou non à une autorité. Dans un environnement cyber, cette distinction est décisive. Un récit hostile peut être très structuré et rester confiné à des publics déjà acquis. Il peut être spectaculaire et sans conséquence comportementale. À

[9] Paul Charon, « Lire la désinformation comme un récit sériel : pour une approche littéraire des manipulations de l'information », Le Rubicon, 13 novembre 2024.

l'inverse, un signal narratif faible peut produire un effet important s'il atteint un segment déjà proche d'un seuil d'action.

Le problème n'est donc pas l'analyse narrative comme telle. Le problème est sa prétention implicite à occuper le centre de l'évaluation. Elle enrichit l'analyse de l'environnement informationnel, mais elle n'est pas une mesure de l'effet. Elle travaille sur une variable d'entrée : le récit. L'objet scientifique à mesurer est une variable de sortie : le comportement déplacé. Entre les deux, il manque la chaîne qui relie exposition, crédibilité, normes perçues, coût, faisabilité, compatibilité identitaire et action. C'est précisément ce que les tableaux de bord narratifs ne fournissent pas. Ils peuvent dire qu'un récit monte. Ils ne peuvent pas dire, sans modèle supplémentaire, si cette montée modifie ce que les individus font.

La deuxième orientation est celle de l'ingénierie cognitive et du design des dispositifs. Elle est représentée par des travaux qui proposent de passer d'une logique d'observation à une logique de conception : comprendre finement les publics, prototyper des réponses, expérimenter, ajuster, intégrer les retours. L'article d'Axel Ducourneau publié en février 2024, « Un design lab. pour la sécurité cognitive »,[10] plaide pour un dispositif pérenne fondé sur six principes : centrage de l'analyse sur l'acteur, attitude prospective, cohérence horizontale et verticale, agilité expérimentale, rapidité d'exécution et intégration itérative des résultats dans une logique de prototypage. Le texte insiste aussi sur une compréhension « émique » des populations ciblées, c'est-à-dire fondée sur leurs propres concepts et systèmes de pensée.

Cette approche corrige une faiblesse fréquente des dispositifs d'influence : leur tendance à produire des messages depuis le centre, avec une compréhension insuffisante des publics visés. Dans le cyber, ce défaut est fréquent. Une administration, une armée, une entreprise ou une plateforme peut répondre à une crise informationnelle avec des catégories qui lui sont propres, mais qui ne correspondent pas aux perceptions des publics. Une alerte de cybersécurité peut être techniquement exacte et socialement inaudible. Une réponse institutionnelle peut être cohérente pour l'émetteur et contre-productive pour le récepteur. Une correction peut stabiliser une

[10] Axel Ducourneau, « Un design lab. pour la sécurité cognitive », Ingénierie cognitique, vol. 7, no 1, 2024, p. 88-93, DOI : 10.21494/ISTE.OP.2024.1094.

communauté déjà confiante et aggraver la défiance d'une communauté sceptique. Le centrage sur l'acteur est donc indispensable.

Mais l'ingénierie cognitive rencontre une limite symétrique de l'approche narrative. Elle améliore la conception des interventions, mais elle ne fournit pas nécessairement le critère permettant d'établir qu'une intervention a produit un effet. Tester un dispositif dans un contexte donné n'indique pas automatiquement ce qui se serait produit sans lui. Sans *baseline*, sans comparaison entre options, sans estimation de l'effet net et sans prise en compte des risques d'amplification, l'expérimentation peut devenir une exploration bien organisée plutôt qu'une évaluation.

Cette limite est particulièrement forte dans les environnements cyber. Une opération informationnelle liée à un incident cyber se déploie dans un contexte instable : annonces officielles, fuites secondaires, commentaires d'experts, interprétations journalistiques, réactions adverses, rumeurs techniques, peur des utilisateurs, contraintes juridiques, temporalité de la remédiation. Si une réponse est mise en place et que le bruit informationnel diminue, il est tentant d'attribuer cette baisse à l'intervention. Mais elle peut aussi provenir de l'épuisement naturel du cycle médiatique, de l'émergence d'un autre sujet, de la correction d'un problème technique, de la fermeture de comptes, d'une décision de plateforme ou d'un repositionnement adverse. Sans estimation contrefactuelle, l'efficacité reste interprétée plutôt que mesurée.

L'ingénierie cognitive doit donc être replacée dans la chaîne d'évaluation. Elle est utile au stade de la conception des actions ; elle n'est pas suffisante au stade de la mesure de l'effet. Elle produit des hypothèses d'action, des prototypes et des boucles d'amélioration mais elle ne peut pas, sans architecture causale, trancher entre plusieurs options : réponse publique, réponse discrète, non-réponse, recours à un relais local, inoculation préalable, modification du canal, ou action technique sur l'infrastructure de diffusion. Le risque, sinon, est de confondre l'activité avec l'efficacité : parce qu'un dispositif est agile, rapide, contextualisé et itératif, il serait supposé pertinent. Or, dans l'influence numérique, une action bien conçue peut être inutile, une action utile localement peut être nuisible globalement ; une action silencieuse peut être supérieure à une réponse visible.

La troisième orientation consiste à déplacer l'analyse vers le récepteur, en mettant l'accent sur ses capacités cognitives : discernement, résistance aux

biais, aptitude à identifier des techniques de manipulation. Cette approche s'inscrit dans une tradition bien plus ancienne qu'elle ne le laisse parfois entendre.

Dès les années 1960, les travaux de William J. McGuire introduisent[11] le concept d'inoculation psychologique : exposer les individus à des versions affaiblies d'arguments manipulatoires permettrait de renforcer leur résistance future. Cette idée a été largement reprise, testée et discutée dans les décennies suivantes, notamment dans la littérature en psychologie sociale et en communication persuasive.

Les travaux récents permettent d'être plus concret. Des interventions de type *Bad News* ou *Go Viral!* exposent les participants à des versions affaiblies de techniques de manipulation (usurpation, polarisation, appel émotionnel, conspirationnisme, discrédit de la source) puis mesurent leur capacité à reconnaître ces procédés dans des contenus expérimentaux.[12] Le discernement peut donc progresser mais le résultat reste situé. Il porte d'abord sur une capacité de jugement mesurée en contexte contrôlé, non sur la modification durable de pratiques comme le partage effectif d'un contenu, l'application d'une consigne de sécurité, le refus de cliquer, la migration vers un canal fiable ou la confiance accordée à une institution en situation de crise.[13]

Cependant, cette orientation n'est ni nouvelle, ni exempte de limites, et elle a fait l'objet de critiques récurrentes.

Premièrement, la littérature souligne depuis longtemps que les effets de l'inoculation restent contextuels et dépendants des conditions d'exposition. Les travaux de Sander van der Linden et d'autres auteurs montrent que ces

[11] William J. McGuire, « Inducing Resistance to Persuasion: Some Contemporary Approaches », in Leonard Berkowitz (dir.), Advances in Experimental Social Psychology, vol. 1, New York, Academic Press, 1964, p. 191-229.

[12] John A. Banas et Stephen A. Rains, « A Meta-Analysis of Research on Inoculation Theory », Communication Monographs, vol. 77, no 3, 2010, p. 281-311 ; Cecilie S. Traberg, Jon Roozenbeek et Sander van der Linden, « Psychological Inoculation against Misinformation: Current Evidence and Future Directions », The ANNALS of the American Academy of Political and Social Science, vol. 700, no 1, 2022, p. 136-151.

[13] Jon Roozenbeek et Sander van der Linden, « Fake News Game Confers Psychological Resistance against Online Misinformation », Palgrave Communications, vol. 5, art. 65, 2019 ; Melisa Basol, Jon Roozenbeek et Sander van der Linden, « Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity against Fake News », Journal of Cognition, vol. 3, no 1, 2020.

effets peuvent décroître dans le temps, varier selon les publics et dépendre fortement du format et du contexte de présentation. L'inoculation n'est pas un mécanisme universel, mais une intervention située.

Deuxièmement, plusieurs recherches en psychologie et en science politique ont mis en évidence une dissociation persistante entre amélioration du jugement et modification du comportement. Les individus peuvent améliorer leur capacité de détection sans modifier leurs pratiques effectives. Cette limite est cohérente avec des résultats plus généraux, notamment ceux de Joshua Kalla et David Broockman, qui montrent que les effets persuasifs sont souvent faibles dans des environnements saturés.

Troisièmement, cette approche tend à isoler la cognition de son environnement social. Or, comme l'ont montré Cristina Bicchieri ou encore les travaux sur les normes sociales et les comportements collectifs,[14] la décision dépend autant de ce que les individus croient que de ce qu'ils pensent que les autres font ou attendent. La capacité à identifier une manipulation ne suffit pas à agir différemment si les normes perçues, les coûts ou les contraintes restent inchangés.

Dans les environnements numériques et cyber, cette limite est particulièrement visible. Les utilisateurs peuvent être formés, sensibilisés et capables de reconnaître certaines techniques (phishing, deepfake, usurpation) sans pour autant modifier leurs comportements. La pression temporelle, les contraintes pratiques, la surcharge informationnelle et les dynamiques sociales continuent de peser sur la décision. En ce sens, la résilience cognitive constitue une condition favorable, mais non une mesure de l'influence. Elle agit sur une variable intermédiaire, celle de la capacité de jugement mais sans garantir le passage à l'action. L'amélioration du discernement est mesurable. L'effet sur l'action l'est beaucoup moins.

Les trois approches qui occupent le débat français doivent ainsi être hiérarchisées. L'analyse narrative permet de décrire les structures de sens. L'ingénierie cognitive permet de concevoir des interventions plus adaptées. La résilience cognitive permet de renforcer certaines capacités du récepteur. Mais aucune de ces approches ne constitue, seule, une mesure de

[14] Cristina Bicchieri, Norms in the Wild: How to Diagnose, Measure, and Change Social Norms, Oxford, Oxford University Press, 2016.

l'influence. Elles traitent respectivement le discours, le dispositif et le récepteur. L'objet manquant demeure le comportement déplacé.

Cette hiérarchisation n'est pas une critique externe, elle découle du problème même que ces approches cherchent à traiter. Si l'influence numérique est un effet, alors il faut définir cet effet. Si l'effet est comportemental, alors il faut identifier un comportement cible. Si ce comportement peut être influencé par plusieurs facteurs, alors il faut estimer ce qui aurait eu lieu sans intervention. Et si cette situation sans intervention n'est pas observable, alors il faut construire une approximation contrefactuelle. Le modèle présenté dans l'article de référence formule précisément ce point : les variables observables décrivent l'environnement, mais l'influence doit être estimée comme déplacement comportemental par rapport à un scénario sans intervention.

Le débat français tend aujourd'hui à s'organiser autour d'objets pertinents mais incomplets. Le récit attire parce qu'il est lisible, interprétable, intellectuellement riche. Le *design lab* attire parce qu'il donne une impression d'action, d'agilité et de modernité opérationnelle. La souveraineté cognitive attire parce qu'elle propose une réponse humaine, démocratique et défensive à la saturation informationnelle. Ces trois objets sont nécessaires à une politique d'influence numérique ; aucun ne suffit à l'évaluation scientifique de l'influence.

L'enjeu est de reformuler le problème dans des termes opérationnels. En cybersécurité, on distingue habituellement les indicateurs de compromission, les indicateurs d'attaque, les impacts métiers, la remédiation et le risque résiduel. Une logique analogue doit être appliquée à l'influence. Les signaux informationnels sont des indicateurs d'activité, pas nécessairement des indicateurs d'impact. La cohérence narrative est un indicateur de structuration, pas nécessairement un indicateur d'effet. L'engagement est un indicateur d'interaction, pas nécessairement un indicateur de transformation. La résilience cognitive est un indicateur de capacité, pas nécessairement un indicateur de comportement.

Cette analogie cyber permet de clarifier le cœur du problème. Personne ne confondrait sérieusement le nombre de paquets réseau avec le dommage métier d'une intrusion. Pourtant, dans l'influence numérique, on confond encore souvent le volume de messages ou le niveau d'engagement avec l'effet réel. C'est cette erreur de niveau qu'il faut corriger. Le champ a besoin

d'une doctrine de mesure qui sépare l'activité, l'exposition, la réception, la décision et le comportement. Tant que ces niveaux restent confondus, les débats continueront à produire des concepts utiles mais non décisifs.

La deuxième conclusion provisoire est donc la suivante : les approches aujourd'hui dominantes ne sont pas à abandonner ; elles doivent être subordonnées à la mesure de l'effet. Le récit devient une variable d'entrée et l'ingénierie devient une méthode de production d'options. La résilience devient un facteur de modification des états cognitifs mais la question centrale demeure : quelle option déplace quel comportement, dans quel segment de population, avec quel risque d'amplification, par rapport à l'absence d'action ? C'est cette question qui ouvre la troisième et dernière partie.

III. PASSER DE L'OBSERVATION CYBER A L'ESTIMATION DE L'EFFET

Les deux premières parties conduisent à un constat précis : le champ de l'influence numérique ne manque ni de données, ni d'outils, ni de traditions théoriques. Il sait observer la circulation des contenus, analyser les récits, cartographier les réseaux, concevoir des dispositifs et renforcer certaines capacités cognitives du récepteur. La difficulté n'est donc pas de savoir si l'influence numérique existe, ni même de savoir par quels canaux elle peut circuler. La difficulté est de savoir comment l'évaluer comme un effet. Plus exactement : comment déterminer qu'une intervention informationnelle, dans un environnement cyber, a modifié un comportement par rapport à ce qui se serait produit sans elle.

Ce problème est familier en cybersécurité, même s'il est rarement formulé ainsi dans le champ de l'influence. Lorsqu'une organisation évalue une cyberattaque, elle ne se contente pas de mesurer l'activité technique. Elle distingue les indicateurs d'exposition, les indicateurs de compromission, les impacts opérationnels, les coûts de remédiation et les risques résiduels. Un pic de trafic réseau n'est pas automatiquement un dommage métier. Une tentative d'intrusion n'est pas automatiquement une compromission. Une compromission n'est pas automatiquement une interruption d'activité. La

chaîne d'évaluation suppose de relier plusieurs niveaux : activité hostile, vulnérabilité exploitée, effet technique, effet métier, coût et décision.

L'influence numérique devrait être traitée avec la même rigueur. Un volume de messages, une visibilité élevée et une cohérence narrative ne sont pas un effet. Puis, une amélioration du discernement n'est pas encore un effet comportemental. Ce sont des signaux, des conditions, des variables intermédiaires ou des facteurs de risque. L'effet se situe plus loin : dans une modification observable, ou raisonnablement inférable, d'un comportement cible.

Cette distinction est essentielle pour le champ cyber, parce que les opérations contemporaines combinent fréquemment action technique et action informationnelle. Une fuite de données peut viser autant la compromission technique que la désorganisation sociale. Une attaque par rançongiciel peut être accompagnée d'une campagne de pression publique contre la victime. Une opération de hack-and-leak peut chercher moins à révéler des informations qu'à détruire la confiance dans une institution. Une campagne de *phishing* peut être renforcée par un récit de crise, d'urgence ou d'autorité. Une manipulation informationnelle peut préparer une compromission technique en abaissant la vigilance des utilisateurs. Dans tous ces cas, l'effet pertinent n'est pas seulement informationnel. Il est comportemental : cliquer, relayer, céder, payer, se taire, accuser, désobéir, se désengager, contourner une procédure, perdre confiance ou modifier une pratique de sécurité.

La question scientifique doit donc être formulée de manière stricte : comment estimer le déplacement comportemental produit par une intervention d'influence numérique, lorsque les données sont fragmentaires, les publics hétérogènes, les plateformes instables et les groupes de contrôle rarement disponibles ? Cette formulation impose quatre clarifications.

La première clarification porte sur la variable dépendante. Il n'y a pas de mesure d'influence sans comportement cible. Dans beaucoup d'analyses, l'objet reste flou : on parle d'influence, de perception, d'adhésion, d'exposition, de manipulation ou de mobilisation comme s'il s'agissait d'un même phénomène. Or ce sont des niveaux distincts. Une influence peut viser une action numérique, comme partager un contenu, cliquer sur un lien, rejoindre un canal, signaler un compte, quitter une plateforme, modifier un

paramètre de sécurité. Elle peut viser une action sociale, comme participer à une manifestation, refuser une consigne, soutenir une position, boycotter une marque, diffuser une rumeur dans un groupe privé. Elle peut viser une action institutionnelle, comme dégrader la confiance dans une administration, faire pression sur une décision publique, ralentir une procédure ou rendre politiquement coûteuse une réponse. Tant que le comportement visé n'est pas défini, la mesure reste flottante.

La deuxième clarification porte sur la chaîne causale. Entre un contenu et une action, plusieurs médiations interviennent. L'exposition signifie qu'un individu ou un groupe a pu rencontrer un contenu. L'attention signifie qu'il l'a effectivement traité. La crédibilité signifie qu'il l'a jugé plausible ou légitime. La compatibilité identitaire signifie que le contenu n'entre pas en conflit avec ses appartenances ou ses valeurs. La norme perçue signifie qu'il croit que d'autres personnes pertinentes pensent ou agissent de la même manière. La faisabilité signifie qu'il peut concrètement agir. Le coût signifie qu'il accepte le risque associé à l'action. L'action n'apparaît qu'au terme de cette chaîne, et parfois seulement lorsqu'un seuil est franchi. Un utilisateur peut être exposé à un faux message d'alerte cyber sans cliquer ; il peut le croire sans le relayer ; il peut le relayer sans modifier ses pratiques ; il peut modifier ses pratiques sans que cette modification soit durable. Mesurer l'exposition en lieu et place de l'action revient donc à interrompre la chaîne au mauvais endroit.

La troisième clarification porte sur l'hétérogénéité des publics. Dans un environnement cyber, la population n'est jamais homogène. Les segments diffèrent par leur niveau de confiance institutionnelle, leur littératie numérique, leur proximité idéologique avec un récit, leur exposition aux plateformes, leur sensibilité aux preuves techniques, leur appartenance professionnelle, leur niveau de vulnérabilité ou leur position dans un réseau. Une même intervention peut produire des effets opposés sur plusieurs segments. Une réponse publique peut rassurer un public déjà favorable, irriter un public sceptique, et donner de la visibilité à un contenu hostile dans un public indécis. Une correction technique peut être comprise par des experts et produire de la confusion chez des non-spécialistes. Une campagne de sensibilisation peut renforcer la vigilance d'une partie du public et produire de la fatigue attentionnelle chez une autre. L'effet moyen est alors peu informatif si l'on ne connaît pas la distribution des publics proches d'un seuil d'action.

La quatrième clarification porte sur le contrefactuel. Une évolution observée après une intervention ne suffit pas à établir que l'intervention a causé cette évolution. En environnement cyber, cette difficulté est constante. Une campagne hostile peut décliner naturellement parce que le cycle médiatique s'épuise. Une réponse publique peut sembler efficace alors que la plateforme a supprimé des comptes en parallèle. Une rumeur peut progresser malgré une bonne réponse parce qu'un événement exogène l'alimente. Une baisse d'engagement peut indiquer une perte d'intérêt, mais aussi une migration vers des canaux fermés. Sans estimation de ce qui se serait produit sans intervention, l'attribution reste fragile.

Ces quatre clarifications déplacent le champ d'une logique de *monitoring* vers une logique d'évaluation. Le *monitoring* répond à la question : que se passe-t-il dans l'environnement informationnel ? L'évaluation répond à une question plus exigeante : quelle différence l'intervention produit-elle ? Les deux démarches sont complémentaires, mais elles ne sont pas équivalentes. Une organisation peut disposer d'un excellent dispositif de monitoring et d'une faible capacité d'évaluation. C'est même la situation la plus courante : beaucoup de signaux, peu de causalité.

Pour progresser, il faut traiter l'influence numérique comme un problème d'estimation sous contrainte. Dans un monde idéal, l'effet d'une intervention serait mesuré par expérimentation contrôlée : un groupe exposé, un groupe non exposé, des comportements observés, des conditions comparables. Dans le monde opérationnel cyber, ce dispositif est rarement disponible. Les populations ne sont pas assignées au hasard, les campagnes adverses se déroulent en temps réel, les plateformes évoluent, les données sont incomplètes, les comportements importants se déplacent parfois vers des espaces fermés, et la décision doit être prise avant que toutes les informations soient stabilisées. Il faut donc recourir à des estimations imparfaites, mais explicites.

C'est ici que la notion de population synthétique ou de contrefactuel synthétique prend son importance.[15] Il ne s'agit pas de prétendre reproduire parfaitement une population réelle. Il s'agit de construire une approximation disciplinée de ce qui aurait probablement eu lieu sans intervention, puis de

[15] Alberto Abadie, Alexis Diamond et Jens Hainmueller, « Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program », Journal of the American Statistical Association, vol. 105, no 490, 2010, p. 493-505.

comparer cette trajectoire avec différents scénarios d'action. Cette approche est déjà familière dans d'autres domaines : prévision épidémiologique, simulation de crise, planification opérationnelle, modélisation économique, évaluation de politiques publiques. Elle devient nécessaire dans l'influence numérique dès lors que l'objet mesuré est un effet causal et que le contrôle expérimental est absent. Il faut un modèle qui remplisse cette exigence. En effet, lorsque les contrôles expérimentaux ou quasi expérimentaux ne sont pas disponibles, une population synthétique devient l'architecture minimale rationnelle pour estimer les effets comportementaux sous rareté de données. Elle ne prédit pas les esprits ; elle spécifie une ligne de base, compare des scénarios et produit une décision bornée par l'incertitude.

Dans un environnement cyber, cette architecture permettrait de comparer plusieurs options. Face à une campagne hostile après une fuite de données, faut-il répondre publiquement, laisser décliner le récit, produire une réponse technique, mobiliser des relais indépendants, cibler les publics exposés, agir auprès des plateformes, ou préparer une inoculation préventive ? Chaque option modifie des variables différentes. Une réponse publique augmente la visibilité, mais peut réduire l'incertitude. Une réponse discrète limite l'amplification, mais peut ne pas atteindre les publics les plus vulnérables. Un relais tiers augmente la crédibilité, mais diminue le contrôle du message. Une action plateforme réduit la circulation, mais peut nourrir un récit de censure. Une inoculation préventive améliore la résistance future, mais n'éteint pas nécessairement une crise immédiate. Une non-réponse peut être la meilleure option si le récit est faible et si l'intervention risque de le rendre visible.

Ce type de comparaison est impossible avec de simples métriques d'engagement. Il exige une structure d'évaluation qui distingue l'activité, l'exposition, la réception et le comportement. Il exige aussi que les risques soient intégrés dans le score de décision. En influence numérique, l'effet utile n'est pas seulement ce que l'action produit positivement ; c'est ce qu'elle produit après soustraction des risques d'amplification, de *backlash*, de perte de crédibilité, de saturation et de déplacement vers des espaces plus difficiles à observer.

L'amplification est probablement le risque le plus sous-estimé. Dans le champ cyber, la réponse officielle à une rumeur technique peut donner une légitimité nouvelle à cette rumeur. La réfutation d'un contenu marginal peut l'introduire dans le débat principal. La publication d'un démenti peut fournir

à l'adversaire des éléments de langage, des captures, des preuves apparentes d'inquiétude ou un prétexte à relance. Une réponse mal calibrée peut transformer une attaque informationnelle faible en événement médiatique. Les systèmes de mesure qui valorisent la rapidité et la visibilité de la réponse encouragent ce risque. Ils récompensent l'activité visible plutôt que l'effet net.

Le *backlash* constitue un second risque. Une intervention peut produire une réaction inverse chez certains publics, en particulier lorsque l'émetteur est perçu comme illégitime. Une institution peut dire vrai et perdre en crédibilité si la source n'est pas acceptée par le public visé. Dans un contexte de défiance, la correction peut confirmer le soupçon. Dans un contexte cyber, cette dynamique est fréquente : un acteur institutionnel qui dément une accusation de faille ou de compromission peut être perçu comme juge et partie. Un relais indépendant, local ou technique peut alors être plus efficace, même s'il est moins visible.

La saturation constitue un troisième risque. Les environnements numériques sont déjà surchargés. Ajouter un message supplémentaire ne produit pas nécessairement de la clarté. Cela peut produire du bruit. Les publics peuvent se désengager, ignorer les alertes, banaliser les risques ou développer une fatigue informationnelle. Dans la cybersécurité, ce phénomène est bien connu avec la fatigue aux alertes. Il existe aussi dans l'influence : trop d'alertes aux manipulations peuvent finir par affaiblir le discernement au lieu de le renforcer.

La perte de crédibilité constitue un quatrième risque. Une réponse imprécise, prématurée ou trop générale peut affaiblir durablement la source. Dans un environnement cyber, où les faits techniques peuvent évoluer, la prudence est nécessaire. Une attribution hâtive, une minimisation excessive ou une communication incomplète peuvent être exploitées par l'adversaire. La mesure de l'effet doit donc intégrer non seulement le comportement immédiat, mais aussi l'état futur de crédibilité de la source.

Ces risques montrent pourquoi la question scientifique ne peut pas être réduite à la détection. Détecter une campagne est nécessaire, mais insuffisant. Il faut savoir si elle est comportementalement dangereuse, pour quels segments, à quel horizon, et quelle réponse présente le meilleur rapport effet-risque. Cette logique est proche de la gestion cyber : une vulnérabilité détectée n'a pas la même priorité selon son exploitabilité, son

exposition, son impact métier et la disponibilité d'un correctif. De même, un récit hostile détecté n'a pas la même priorité selon sa crédibilité, sa diffusion, son public, sa proximité avec un seuil d'action et le risque de l'amplifier.

L'évaluation de l'influence numérique doit donc intégrer une logique de seuil. Dans beaucoup de situations, une opération ne transforme pas une population entière ; elle déplace un segment déjà proche d'un comportement. Ce segment peut être petit mais décisif. Dans une crise cyber, il peut s'agir des clients sur le point de quitter un service, des salariés susceptibles de contourner une consigne, des journalistes hésitant entre plusieurs cadres d'interprétation, des décideurs publics confrontés à une pression, ou des utilisateurs susceptibles de relayer une fausse alerte. L'exposition moyenne de la population n'est alors pas l'indicateur principal. Ce qui compte est la masse de publics proches du seuil et la capacité d'une intervention à les déplacer.

Cette logique explique pourquoi les résultats empiriques sur la persuasion sont souvent faibles en moyenne mais non négligeables dans certains contextes. Les campagnes générales touchent beaucoup de personnes, mais souvent éloignées du seuil. Les interventions sociales ou normatives peuvent toucher moins de personnes, mais modifier davantage celles qui sont prêtes à agir. L'étude de Bond et al. montre l'importance des relations sociales dans la mobilisation ; Kalla et Broockman montrent la faiblesse des effets persuasifs moyens dans des environnements politiques saturés. Ces deux résultats ne se contredisent pas. Ils indiquent que l'effet dépend de la position des individus dans une chaîne sociale et comportementale.
La recherche doit alors passer d'une question générale portant sur le degré d'influence d'une campagne à une question structurée : « quel segment, exposé par quelle source, sous quelle norme perçue, avec quel coût d'action, est déplacé vers quel comportement, par rapport à quelle ligne de base ? » Cette formulation paraît plus lourde, mais elle est scientifiquement plus correcte. Elle empêche de traiter l'influence comme une propriété globale du message. Elle l'inscrit dans une relation entre un contenu, une source, un public, un contexte et un comportement.

Elle permet aussi de réintégrer proprement les trois approches discutées plus haut. L'analyse narrative renseigne le contenu et les cadres interprétatifs. L'ingénierie cognitive renseigne la conception et l'adaptation des interventions. La résilience cognitive renseigne la capacité du récepteur à traiter ou neutraliser certains stimuli. Mais la mesure de l'effet exige de

relier ces trois dimensions au comportement. Ce n'est pas une opposition. C'est une architecture.

Pour le champ cyber, cette architecture pourrait être organisée autour de cinq niveaux. Le premier niveau est le contexte : incident technique, crise politique, événement médiatique, vulnérabilité exploitée, état de confiance préalable. Le deuxième niveau est l'environnement informationnel : récits dominants, sources, plateformes, relais, signaux adverses, coordination. Le troisième niveau est la population : segments, niveaux de confiance, exposition, compétences numériques, proximité avec un comportement cible. Le quatrième niveau est l'intervention : réponse publique, réponse discrète, action plateforme, relais tiers, inoculation, silence. Le cinquième niveau est le comportement : action numérique, action sociale, action institutionnelle, décision de sécurité, modification de confiance, mobilisation ou retrait.

Une telle architecture ne fournit pas une certitude. Elle fournit une discipline. Elle oblige à expliciter les hypothèses, à distinguer les niveaux, à estimer l'incertitude et à comparer les options. C'est déjà un progrès considérable par rapport aux tableaux de bord qui accumulent des signaux sans les relier à une décision. Dans l'influence numérique comme dans le cyber, l'objectif n'est pas l'omniscience ; c'est la réduction de l'erreur décisionnelle.

La validation d'un tel cadre devra être progressive. Elle peut commencer par des analyses rétrospectives : prendre des crises informationnelles associées à des incidents cyber, reconstruire les données disponibles avant la décision, simuler plusieurs options, puis comparer les recommandations avec les trajectoires observées. Elle peut ensuite passer par des exercices prospectifs : produire des estimations avant intervention et vérifier leur calibration. Elle peut enfin comparer le jugement expert seul avec le jugement expert assisté par modèle. Le critère ne sera pas la prédiction parfaite, mais l'amélioration de la cohérence, la réduction des erreurs d'amplification, la meilleure identification des segments à risque et la capacité à justifier une non-réponse lorsque celle-ci est supérieure.

Cette dernière capacité est essentielle. Un système d'évaluation sérieux doit pouvoir recommander de ne pas agir. Si toute détection entraîne une réponse, le système n'est pas un instrument d'évaluation ; il devient un mécanisme de justification de l'activité. Or, dans l'influence numérique,

l'inaction peut être rationnelle. Elle peut éviter l'amplification, préserver la crédibilité, laisser un récit s'épuiser ou réserver la réponse à un moment plus favorable. Le cyber connaît déjà cette logique : toute alerte ne déclenche pas la même réponse ; toute vulnérabilité ne justifie pas la même urgence ; toute anomalie ne devient pas incident majeur. L'influence doit adopter la même maturité.

La question scientifique finale peut donc être formulée ainsi : comment construire, pour l'influence numérique en environnement cyber, un cadre d'estimation qui relie les signaux informationnels observables aux comportements cibles, intègre l'hétérogénéité des publics, compare plusieurs options d'action et estime ce qui se serait produit sans intervention ?

Cette question marque la vraie frontière du champ. Elle ne consiste pas à produire une nouvelle typologie des récits. Elle ne consiste pas à multiplier les prototypes d'intervention. Elle ne consiste pas seulement à entraîner la résilience cognitive. Elle consiste à transformer ces acquis en système d'évaluation. Tant que cette étape n'est pas franchie, le champ restera riche en descriptions et pauvre en mesure. Il saura dire ce qui circule, mais pas ce qui change. Il saura détecter des campagnes, mais pas toujours hiérarchiser leur danger. Il saura produire des réponses, mais pas toujours établir qu'elles étaient préférables au silence.

L'enjeu est d'autant plus important que la cybersécurité entre dans une phase où les frontières entre technique, information et comportement deviennent moins nettes. Les attaques ne ciblent plus seulement des systèmes ; elles ciblent des décisions. Elles exploitent des vulnérabilités techniques, mais aussi des vulnérabilités cognitives, organisationnelles et sociales. Mesurer l'influence numérique revient donc à protéger une chaîne de décision. Ce n'est pas une question secondaire de communication. C'est une question de sécurité.

IV. RECOMMANDATIONS : CONSTRUIRE UNE DOCTRINE DE MESURE DE L'INFLUENCE NUMERIQUE

Premièrement, définir le comportement cible avant de mesurer.
Une opération d'influence ne devrait pas être évaluée à partir d'un indicateur générique de visibilité. Elle doit d'abord être rapportée à un comportement cible : cliquer, relayer, s'abstenir, appliquer une consigne de sécurité, quitter une plateforme, rejoindre un canal, faire confiance à une source, exercer une pression sur une institution ou modifier une pratique professionnelle. Sans comportement cible, les métriques décrivent seulement un bruit informationnel.

Deuxièmement, séparer les indicateurs d'activité, de réception et d'impact.
Les tableaux de bord doivent distinguer trois niveaux. Les indicateurs d'activité décrivent ce qui circule : volumes, comptes, hashtags, temporalité, coordination. Les indicateurs de réception décrivent ce qui est traité : attention, crédibilité, mémorisation, adhésion déclarée. Les indicateurs d'impact décrivent ce qui change : décision, action, non-action, retrait, mobilisation, coopération ou pratique de sécurité. Confondre ces niveaux conduit à surestimer l'efficacité des campagnes visibles.

Troisièmement, subordonner les récits, le design et la résilience à l'évaluation de l'effet.
L'analyse narrative, l'ingénierie cognitive et la résilience du récepteur doivent rester dans la chaîne d'évaluation, mais à leur juste place. Le récit renseigne les cadres de sens. Le design produit des options d'intervention. La résilience cognitive renforce une capacité de jugement. Aucun de ces éléments ne prouve seul une influence. Ils doivent être reliés à un effet comportemental observé ou estimé.

Quatrièmement, construire une ligne de base contrefactuelle.
Toute évaluation doit comporter une estimation de ce qui se serait probablement produit sans intervention. Cette ligne de base peut être imparfaite, mais elle doit être explicite : tendance naturelle de la conversation, cycle médiatique attendu, publics exposés, relais habituels,

niveau préalable de confiance, trajectoire probable du récit adverse. Une réponse ne doit être considérée comme efficace que si elle fait mieux que cette trajectoire de référence.

Cinquièmement, comparer plusieurs options, y compris le silence.
L'évaluation ne doit pas seulement répondre à la question : faut-il agir ? Elle doit comparer plusieurs options : réponse publique, réponse discrète, relais tiers, inoculation préalable, correction technique, action auprès des plateformes, changement de canal ou non-réponse. Dans certaines situations, le silence contrôlé peut produire un meilleur effet net qu'un démenti visible, notamment lorsque le risque d'amplification est supérieur au gain attendu.

Sixièmement, intégrer les risques d'amplification, de backlash et de saturation.
Le score d'efficacité doit intégrer les effets négatifs possibles de l'intervention : donner de la visibilité à un récit marginal, renforcer une défiance préexistante, fournir des éléments de relance à l'adversaire, fatiguer le public ou affaiblir la crédibilité de la source. Une action informationnelle doit donc être évaluée en effet net, non en activité visible.

Septièmement, prioriser les segments proches d'un seuil d'action.
Les efforts doivent porter sur les publics capables de changer de comportement, et non seulement sur les publics les plus exposés. Une campagne peut toucher beaucoup d'individus déjà convaincus sans produire d'effet. À l'inverse, un segment réduit mais proche d'un seuil, salariés tentés de contourner une procédure, usagers hésitant à quitter un service, journalistes cherchant un cadrage, décideurs soumis à une pression, peut être décisif.

Huitièmement, organiser l'évaluation autour de cinq niveaux.
La doctrine de mesure doit articuler cinq niveaux : le contexte de crise, l'environnement informationnel, la population, l'intervention et le comportement. Cette architecture impose de préciser ce qui est observé, ce qui est inféré, ce qui est décidé et ce qui est finalement modifié. Elle permet aussi de justifier une réponse proportionnée et de conserver la trace des hypothèses utilisées.

Neuvièmement, valider progressivement les modèles par retour d'expérience.
Les modèles d'évaluation doivent être testés sur des cas rétrospectifs, puis dans des exercices prospectifs. Il faut comparer les recommandations produites avant décision avec les trajectoires observées après coup : réduction de l'amplification, meilleure identification des publics à risque, précision des hypothèses, capacité à recommander une non-réponse. L'objectif n'est pas la prédiction parfaite, mais la réduction de l'erreur décisionnelle.

CONCLUSION

Le champ de l'influence numérique dispose déjà de nombreuses briques : réseaux, récits, biais, relais, métriques, dispositifs expérimentaux et programmes de résilience. Ce qui manque encore est l'architecture qui permet de les ordonner autour de l'effet. Dans le cyber, cette architecture doit imposer une distinction stricte entre visibilité et impact, activité et efficacité, récit et comportement, action et résultat. Cette exigence conduit à replacer le contrefactuel au centre de l'analyse. Il ne suffit pas de savoir ce qui s'est passé après une intervention. Il faut estimer ce qui se serait probablement passé sans elle, puis comparer les options disponibles en intégrant les risques d'amplification, de *backlash*, de saturation et de perte de crédibilité.

C'est à cette condition que l'influence numérique pourra être traitée comme un objet cyber à part entière : non plus seulement surveillée, commentée ou cartographiée, mais mesurée dans ses effets probables, ses risques et ses alternatives d'action. Le véritable progrès ne consistera pas à produire une carte plus détaillée des flux. Il consistera à savoir si cette carte mène quelque part.

SKEMA Publika
SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Tél. : +33.1.71.13.39.32
Courriel : publika@skema.edu
Site Internet : www.publika.skema.edu