LA SÉCURITÉ COGNITIVE : CADRES, ENJEUX ET DÉFIS

M. GUIDERE. Professeur des Universités (Paris 8). Directeur de recherches à l'INSERM
Institut National de la Santé et de la Recherche Médicale, France.
mathieu.guidere@inserm.fr

Résumé : Cet article examine le concept de sécurité cognitive au sein des environnements numériques actuels, en se concentrant sur la protection de l'intégrité psychique, des processus de délibération et de l'économie attentionnelle. En mobilisant des cadres d'analyse issus de la santé mentale, de la psychologie comportementale et de la cybersécurité, il met en évidence les vulnérabilités structurelles induites par l'ingénierie sociale de précision, les interfaces manipulatoires (*Dark Patterns*) et la surcharge informationnelle. L'analyse souligne les menaces associées à l'exploitation systémique de ces failles cognitives, notamment la fatigue décisionnelle, la paralysie analytique, l'émergence de la menace interne et l'instrumentalisation stratégique de l'anxiété géopolitique.

L'article explore également les défis posés par la dérégulation des marchés de l'information et la banalisation de la fraude synthétique, en montrant comment ces phénomènes fragilisent la sécurité épistémique et l'intégrité du débat démocratique. Enfin, il propose un cadre stratégique de résilience, articulant l'adoption d'un paradigme de conception protectrice (*Secure-by-Design*), la systématisation de l'hygiène cognitive et la mise en œuvre d'audits de maturité cognitive. L'ensemble vise à contribuer au passage d'une posture défensive réactive à une culture proactive d'écosystèmes numériques salutogènes, indispensables à l'acceptabilité sociale et politique de la révolution cognitive associée à l'intelligence artificielle.

Mots clés : Sécurité cognitive, économie de l'attention, ingénierie sociale, hygiène cognitive, vulnérabilité cognitive, résilience organisationnelle, audit cognitif, neurodroits.

Abstract: This article examines the concept of cognitive security within current digital environments, focusing on the protection of mental integrity, deliberation processes, and the attention economy. Drawing on analytical frameworks from mental health, behavioral psychology, and cybersecurity, it highlights the structural vulnerabilities induced by precision social engineering, manipulative interfaces (Dark Patterns), and information overload. The analysis underlines the threats associated with the systemic exploitation of these cognitive vulnerabilities, notably decision fatigue, analytical paralysis, the emergence of the insider threat, and the strategic instrumentalization of geopolitical anxiety.

The article also explores the challenges posed by the deregulation of information markets and the normalization of synthetic fraud, demonstrating how these phenomena weaken epistemic security and the integrity of democratic debate. Finally, it proposes a strategic framework for resilience, articulating the adoption of a protective design paradigm (Secure-by-Design), the systematization of cognitive hygiene, and the implementation of cognitive maturity audits. The overall aim is to contribute to the transition from a reactive defensive posture to a proactive culture of salutogenic digital ecosystems, which are essential to the social and political acceptability of the cognitive revolution associated with artificial intelligence.

Keywords: Cognitive security, attention economy, social engineering, cognitive hygiene, cognitive vulnerability, organizational resilience, cognitive audit, neurorights.

Introduction

La structuration des environnements numériques contemporains impose une redéfinition rigoureuse des paradigmes de protection. Jusqu'à récemment, l'effort sécuritaire se concentrait de manière prépondérante sur la cybersécurité, entendue comme la préservation de l'intégrité des infrastructures de communication, des réseaux et des bases de données. À cette dimension matérielle et logicielle s'est progressivement greffée la notion de *souveraineté cognitive*, qui désigne l'indépendance technologique et la capacité d'une collectivité à maintenir la maîtrise de ses outils de production symbolique face à des acteurs hégémoniques exogènes.

Toutefois, l'intensification des interactions entre la cognition humaine et les systèmes algorithmiques requiert aujourd'hui l'intégration d'un troisième concept, fondamentalement distinct mais complémentaire : la *sécurité cognitive*. Celle-ci ne s'attache ni à la robustesse du contenant technologique ni à la territorialité de sa conception, mais vise directement la protection de l'intégrité psychique, des processus de délibération autonome et de l'économie attentionnelle humaine contre les altérations systémiques induites par la diffusion ciblée de l'information.

Ce glissement conceptuel s'inscrit dans un contexte où la nature même de la menace a profondément muté, opérant une transition d'une vulnérabilité strictement technologique vers une vulnérabilité d'ordre psycho-social. Dans cet écosystème, l'individu n'est plus envisagé comme un simple utilisateur opérant au sein d'un espace informationnel neutre ; il constitue désormais la surface d'attaque principale des nouvelles ingénieries de l'influence.

Le cerveau humain, appréhendé dans ce paradigme comme un composant biologique directement interfaçable avec les flux de données, souvent désigné sous le terme de *wetware*, devient l'ultime champ de bataille. L'exploitation systématique des biais heuristiques, l'hyper-sollicitation de l'attention et l'analyse granulaire des comportements transforment ainsi l'architecture cognitive des usagers en un vecteur de compromission, rendant de fait inopérantes les approches sécuritaires classiques centrées exclusivement sur le périmètre des machines.

Face à cette industrialisation de la capture attentionnelle et à l'érosion progressive des capacités d'analyse critique, l'élaboration d'une réponse systémique devient un impératif de premier ordre. La réflexion doit dépasser le simple diagnostic comportemental pour interroger les modalités de gouvernance de ces environnements immersifs.

La problématique fondamentale réside alors dans l'articulation entre protection et développement : comment structurer des cadres de protection technique qui soient véritablement capables de préserver l'intégrité mentale et la sécurité épistémique des personnes, tout en veillant à ne pas entraver de manière disproportionnée la liberté d'innovation numérique ? C'est à l'aune de cette tension complexe que doivent être analysés les mécanismes, les enjeux et les défis inhérents à la sécurité cognitive.

Vulnérabilités et menaces cognitives

L'analyse de la sécurité cognitive ne saurait se circonscrire aux seules confrontations interétatiques ou aux campagnes de désinformation massives orchestrées à l'échelle macro-politique. Elle exige d'examiner avec une granularité plus fine les vecteurs d'insécurité qui structurent notre quotidien numérique, en mettant en lumière la manière dont les environnements socio-techniques contemporains exploitent de façon systémique les limites neurobiologiques des utilisateurs. Dans cette perspective, la menace se déplace des infrastructures matérielles vers l'architecture psychique de l'individu, transformant les processus d'attention et de délibération en véritables surfaces d'attaque.

Cette dynamique s'observe d'abord dans l'évolution de l'ingénierie sociale, qui a opéré une transition d'une logique de diffusion indifférenciée vers une ingénierie de précision. L'industrialisation des cyberattaques s'appuie désormais sur un profilage algorithmique capable de cartographier les vulnérabilités émotionnelles et cognitives d'une cible en temps réel. En exploitant des traces numériques comportementales, ces dispositifs identifient avec précision les leviers psychologiques tels que le sentiment d'urgence, la peur de l'exclusion ou l'empathie, qui sont les plus susceptibles de court-circuiter les mécanismes de l'évaluation rationnelle. L'objectif n'est plus seulement de tromper par un faux message, mais d'activer intentionnellement des heuristiques de jugement, ces raccourcis mentaux qui, bien qu'utiles sur le plan évolutif, s'avèrent dramatiquement inadaptés face à des artefacts numériques conçus pour manipuler.

Cette altération de l'autonomie décisionnelle s'exerce également de manière structurelle par le biais des interfaces elles-mêmes. L'intégration prolifique de « *Dark Patterns* » ou « interfaces trompeuses » dans le design des plateformes constitue une forme documentée d'extorsion cognitive.

En orchestrant une asymétrie informationnelle, en masquant délibérément les options alternatives ou en induisant une friction ergonomique artificielle, ces architectures de choix contraignent l'utilisateur à concéder des données ou à valider des actions à l'encontre de ses intérêts. Cette manipulation de l'environnement visuel instrumentalise la fatigue décisionnelle, réduisant le consentement à un simple réflexe d'évitement et privant le sujet de sa capacité d'arbitrage réfléchi.

Au-delà de la manipulation directe, c'est l'économie même des plateformes, fondée sur la captation continue de l'attention, qui instaure une *vulnérabilité systémique* par la surcharge informationnelle. L'exposition ininterrompue à un flux massif, fragmenté et hyper-stimulant induit un phénomène d'infobésité dont les répercussions neuro-physiologiques sont délétères.

Ce stress cognitif chronique se traduit par une saturation rapide de la mémoire de travail et une sous-activation du cortex préfrontal, région cérébrale impliquée dans le contrôle exécutif et la pensée critique.

À l'échelle des environnements professionnels et institutionnels, cette surabondance de signaux engendre une véritable paralysie analytique. Loin d'éclairer la prise de décision, l'excès de stimuli entrave la capacité d'intégration sémantique, poussant les décideurs vers l'inaction ou, à l'inverse, vers des conclusions précipitées dictées par des biais de confirmation.

Enfin, ces vulnérabilités psychologiques bouleversent les paradigmes traditionnels de la sécurité organisationnelle. L'employé hyper-connecté, soumis à une injonction paradoxale de réactivité permanente et de vigilance absolue, devient mécaniquement le maillon faible de la chaîne défensive.

La fatigue mentale altère sa capacité à détecter les signaux faibles d'une compromission, transformant l'épuisement cognitif en faille de sécurité exploitable.

Ce risque interne est aujourd'hui démultiplié par l'émergence de la *fraude synthétique*, portée par la démocratisation des *deepfakes* audio et vidéo. Lorsqu'un algorithme génératif permet d'usurper avec une fidélité biométrique l'identité d'un dirigeant ou d'un collègue, c'est le fondement même de la confiance interpersonnelle qui s'effondre. L'incapacité croissante à distinguer une instruction humaine d'une simulation automatisée instaure un climat de *doute épistémique permanent*, menaçant à la fois la fluidité opérationnelle et la cohésion sociale des organisations.

L'anxiété géopolitique comme vulnérabilité systémique

L'émergence conceptuelle de l'anxiété géopolitique constitue une illustration paradigmatique de la manière dont les failles psychologiques sont aujourd'hui instrumentalisées dans le cadre des rivalités interétatiques. Née à l'intersection de la santé mentale et des relations internationales, ce phénomène se définit comme un état d'inquiétude ou de peur provoqué par l'instabilité du système mondial et la perception de menaces globales échappant à toute maîtrise individuelle.

Contrairement à une peur ponctuelle, l'anxiété géopolitique s'installe dans la durée sous la forme d'un stress chronique et incontrôlable.

Au sein des environnements numériques, cette anxiété se manifeste selon trois axes structurants : un sentiment d'impuissance face aux décisions des puissants, une perte de repères liée à l'imprévisibilité systémique des crises, et l'intériorisation d'une menace perçue comme existentielle.

L'exposition continue aux conflits de haute intensité via les médias d'information continue et les plateformes sociales engendre une saturation cognitive, amplifiée par le phénomène de « défilement morbide » (*doomscrolling*), qui affaiblit durablement la capacité de résilience émotionnelle des populations.

Cette fragilité neuro-psychologique de masse est désormais militarisée par des puissances rivales, marquant un basculement stratégique de la propagande traditionnelle vers une véritable « capture cognitive ».

Plutôt que de diffuser massivement un message uniforme, les algorithmes adverses analysent les traces numériques pour cartographier les vulnérabilités collectives telles que les croyances négatives, les peurs incapacitantes ou les griefs sociaux, permettant ainsi une micro-segmentation informationnelle d'une précision redoutable.

Dès lors, des essaims d'agents conversationnels pilotés par l'intelligence artificielle infiltrent les réseaux sociaux et les communautés fermées. Leur finalité n'est plus le simple parasitage informationnel, mais la radicalisation incrémentale des opinions sur des sujets clivants, visant à fragmenter le corps social et à rendre la nation organiquement ingouvernable.

Le cas de la France offre à cet égard un terrain d'observation particulièrement révélateur des dynamiques de cette guerre psychocognitive. Affichant un « Indice d'Anxiété Géopolitique » (IAG) de 87 %, la plaçant au premier rang mondial, la population française se caractérise par un « découplage cognitif » profond : une majorité de la population déclare un bien-être personnel satisfaisant (55 %) tout en nourrissant une vision apocalyptique et décliniste du destin national.

Ce pessimisme endémique, perçu comme une faiblesse psychologique, est directement converti en arme stratégique par des dispositifs hostiles (*weaponized pessimism*). Un algorithme adverse n'a en effet nul besoin d'inventer des menaces *ex nihilo* ; il lui suffit de s'adosser à ce sentiment de déclassement inéluctable pour transformer une déprime nationale en un levier de déstabilisation. Le message subliminal infusé en continu par ces agents intelligents, postulant que le pays est condamné et l'action politique vaine, vise à opérer un glissement du pessimisme totalisant vers un défaitisme incapacitant. L'axiome stratégique sous-jacent est implacable : si une population est intimement convaincue qu'elle a d'ores et déjà perdu, elle renoncera à se mobiliser.

L'objectif ultime de cette *militarisation de l'anxiété* via la saturation de l'espace numérique n'est donc plus de substituer un mensonge à la vérité, mais de rendre le réel purement et simplement indiscernable. En injectant massivement des contenus génératifs déstabilisants (tels que des *deepfakes* ou de fausses alertes), la guerre cognitive cherche à paralyser le cycle de décision stratégique, ou boucle OODA (Observer-Orienter-Décider-Agir), de l'adversaire.

L'exploitation systémique de l'anxiété géopolitique par l'ingénierie algorithmique s'impose ainsi comme l'un des défis majeurs de la sécurité cognitive, frappant les nations non plus sur leurs infrastructures matérielles, mais au cœur même de leur volonté de puissance et de leur cohésion mentale.

Défis et stratégies de résilience

Face à l'industrialisation des agressions ciblant l'architecture psychique, la préservation de la souveraineté et de l'intégrité mentale impose l'élaboration d'une véritable défense cognitive. Cette démarche de résilience ne saurait se limiter à un constat de vulnérabilité ou à une posture strictement réactive ; elle exige le déploiement de solutions structurelles et de stratégies de mitigation capables de sanctuariser les processus de délibération.

L'enjeu est de bâtir un environnement numérique où la sécurité n'est plus une option périphérique, mais le fondement même de la conception technologique et de la formation citoyenne.

Ainsi, l'atténuation des vulnérabilités systémiques requiert un changement de paradigme dans l'ingénierie des systèmes d'information : l'adoption d'une approche « *Secure-by-Design* » qui intègre, dès la phase de conception algorithmique et visuelle, les limites neurobiologiques de la cognition humaine. Alors que l'économie de l'attention exploite la friction manipulatoire pour extorquer le temps de cerveau disponible, l'objectif consiste à inverser cette asymétrie par l'implémentation d'architectures de choix fondées sur l'introduction de frictions protectrices.

Concrètement, cette approche vise à ralentir artificiellement certaines interfaces numériques afin de restaurer le temps nécessaire à l'évaluation rationnelle, contrant ainsi l'immédiateté qui favorise les heuristiques fallacieuses. Ce ralentissement stratégique s'incarne notamment dans l'instauration de protocoles de validation cognitive. Face à des stimuli induisant un sentiment d'urgence extrême ou lors de la transmission d'instructions critiques, il est impératif d'imposer une vérification par des canaux totalement déconnectés (rencontre physique, radio sécurisée ou validation matérielle) pour court-circuiter l'influence manipulatoire d'une potentielle intelligence artificielle adverse opérant dans l'environnement numérique.

Parallèlement, le déploiement technique d'algorithmes intelligents d'audit permet de repérer les biais cognitifs dans les interfaces et de neutraliser préventivement les mécanismes de manipulation subliminale ciblant les utilisateurs.

L'hygiène cognitive

Si l'aménagement ergonomique des infrastructures technologiques constitue un premier rempart indispensable, la résilience à long terme repose in fine sur l'aguerrissement psychologique des individus face aux menaces hybrides. Cette démarche d'hygiène cognitive a pour vocation de doter les personnes de compétences métacognitives robustes, leur permettant d'identifier consciemment leurs propres biais (tels que l'effet d'ancrage ou le biais de négativité), de réguler leur consommation attentionnelle et de cultiver une forme d'agnosticisme numérique face aux flux d'informations non certifiés.

Dans la sphère professionnelle, institutionnelle et militaire, cet aguerrissement se traduit par des exercices d'entraînement sous stress cognitif : les décideurs sont plongés dans des simulations saturées de fausses informations ultra-réalistes et de

deepfakes, dans le but explicite de forger une « mémoire immunitaire » résiliente face à la manipulation. Cette posture de défense implique également une gestion experte et systémique des biais cognitifs au sein des instances dirigeantes.

À l'échelle de la société civile, cette nécessité impose de repenser l'éducation pour l'ériger en un véritable système d'arme défensif. Il s'agit de structurer des programmes d'« auto-défense intellectuelle » et de psychoéducation aux médias sociaux. L'objectif est d'enseigner de manière analytique et explicite comment les algorithmes de recommandation sont délibérément conçus pour capter l'attention et polariser les opinions, protégeant ainsi le corps social contre les risques de subversion idéologique et de fragmentation.

La vigilance cognitive

L'évolution des environnements numériques et la porosité croissante des frontières de l'entreprise imposent une redéfinition de la sûreté de l'information, plaçant la vigilance cognitive au cœur des dispositifs de protection. Cette approche se paramètre spécifiquement pour la détection précoce des menaces internes, communément qualifiées d'« Insider Threat ».

Loin de se limiter à une surveillance purement technique des flux de données ou des accès réseaux, la vigilance cognitive déploie une analyse comportementale et psychocognitive fine. Elle s'appuie notamment sur des outils de Traitement Automatique du Langage (TAL) pour opérer des analyses de rupture, permettant ainsi d'identifier l'émergence de clusters sémantiques caractéristiques de la déloyauté ou d'un processus profond de désaffiliation organisationnelle.

Sur le plan de l'analyse comportementale, cette méthodologie mobilise des cadres théoriques tels que le modèle MICE (Money, Ideology, Coercion, Ego) afin d'évaluer avec précision les leviers classiques qui sous-tendent et précipitent le passage à l'acte malveillant.

L'enjeu central est de capter les signaux faibles et les altérations psychiques, particulièrement les distorsions cognitives d'ordre paranoïaque ou messianique, qui précèdent fréquemment les actes de sabotage ou de compromission systémique.

La typologie de ces menaces s'avère hautement polymorphe : elle s'étend du sabotage informatique perpétré par un collaborateur en rupture émotionnelle consécutive à une frustration professionnelle, à l'employé radicalisé se réclamant de mouvances souverainistes pour rejeter toute autorité contractuelle ou institutionnelle, jusqu'à l'ingénierie sociale impliquant la manipulation de prestataires externes à des fins d'exfiltration de données de recherche et développement sensibles.

Face à l'identification de telles vulnérabilités, la doctrine de riposte préconise l'application stricte d'audits restrictifs des journaux d'accès selon le principe du « Zero Trust », couplée à une intervention immédiate des ressources humaines sous la supervision étroite de la direction de la sûreté.

Il est impératif de souligner que l'ensemble de ces démarches analytiques et opérationnelles s'inscrit rigoureusement dans un cadre légal défini, garantissant le respect des normes telles que le RGPD et le secret des correspondances, en limitant l'examen algorithmique aux seules données expressément autorisées par les chartes d'entreprise.

L'audit de maturité cognitive

Pour transcender la simple réaction face à l'incident et instaurer une véritable culture de la sécurité mentale et opérationnelle, les organisations doivent se soumettre à une démarche structurée d'audit de maturité cognitive. Cet instrument de diagnostic systémique vise à évaluer la robustesse de l'entité à travers cinq piliers fondamentaux de résilience, permettant d'objectiver la capacité de la structure à faire face aux manipulations algorithmiques et humaines.

Le premier de ces piliers, consacré à la formation et aux compétences, constitue le socle indépassable de l'autonomie épistémique des collaborateurs. Il requiert non seulement que le personnel soit expressément formé aux limites inhérentes aux intelligences artificielles génératives, mais également qu'il développe l'acuité analytique nécessaire pour comprendre et identifier les biais algorithmiques véhiculés par ces systèmes.

Dans une perspective de prévention de l'atrophie intellectuelle et de la dette cognitive, ce pilier exige l'entretien continu des compétences fondamentales d'écriture et de raisonnement autonome, soutenu par la pratique régulière d'exercices opérationnels réalisés sans aucune assistance d'outils d'intelligence artificielle.

Cette architecture évaluative se complète par l'analyse approfondie de quatre autres dimensions critiques de l'entreprise : la Gouvernance et la Décision, le Management et l'Organisation, la Sécurité informationnelle, ainsi que la Gestion de crise. L'aboutissement de cet audit exhaustif permet aux directions d'établir un diagnostic précis et de calculer un « Indice de dette cognitif ».

Cette métrique exécutive offre aux instances dirigeantes une cartographie claire de leurs vulnérabilités structurelles. Sur la base de ce diagnostic, l'organisation est alors en mesure de déployer des simulations de crise tactiques, ou *wargaming*, afin de tester l'escalade des scénarios d'urgence (des signaux faibles jusqu'au passage à l'acte) et d'éprouver la solidité de ses réflexes décisionnels face à une menace interne évolutive.

De la protection des données à la protection des cerveaux

Face aux mutations de l'écosystème numérique, le paradigme centré exclusivement sur la protection des données personnelles (à l'instar du RGPD) s'avère désormais insuffisant. L'enjeu ne réside plus seulement dans la sécurisation de la trace numérique de l'utilisateur, mais dans la préservation de son intégrité psychique face à des dispositifs conçus pour en altérer le fonctionnement.

L'accélération des neuro-technologies et la sophistication de l'économie de l'attention imposent une révision profonde de notre corpus juridique. Les droits humains fondamentaux, conçus à une époque où le sanctuaire de l'esprit semblait inviolable, doivent être adaptés pour faire face à la perméabilité nouvelle de la cognition humaine. C'est dans ce contexte qu'émerge le concept de « neurodroits », qui vise à étendre la protection juridique à la sphère neurobiologique.

Cette nouvelle architecture normative repose sur trois piliers fondamentaux. Le premier est le droit à la « vie privée mentale » (Mental Privacy), qui sanctuarise les données neurales et interdit l'inférence non consentie des états émotionnels ou psychologiques par des systèmes algorithmiques.

Le deuxième est le droit à la « continuité psychocognitive », qui protège l'utilisateur contre les altérations invisibles de sa personnalité induites par un profilage et un ciblage continus.

Enfin, le droit à la « liberté cognitive » garantit la capacité du sujet à prendre des décisions sans subir de manipulations algorithmiques visant à contourner son rationalisme ou à forcer son consentement par l'épuisement de ses ressources délibératives.

La reconnaissance de ces neurodroits implique une redéfinition du statut de *l'attention humaine*. Longtemps considérée comme une variable d'ajustement marketing, l'attention doit désormais être juridiquement qualifiée de ressource critique et finie, au même titre que certaines ressources environnementales.

L'économie numérique contemporaine repose sur un modèle extractiviste où les plateformes maximisent le « temps d'écran » par la conception d'interfaces addictives et de flux ininterrompus.

Ce modèle d'accumulation rend indispensable une régulation institutionnelle stricte de la captation algorithmique. Il s'agit d'encadrer juridiquement les techniques d'ingénierie comportementale qui exploitent les vulnérabilités du cerveau (récompenses aléatoires, défilement infini), en allant bien au-delà de la seule régulation des grands modèles de langage (LLMs). La loi doit imposer des limites à l'intensité et à la fréquence des sollicitations numériques, afin de prévenir l'épuisement cognitif de masse et de restaurer la souveraineté attentionnelle des citoyens.

L'insécurité cognitive

L'insécurité cognitive individuelle se traduit inévitablement par une vulnérabilité démocratique à l'échelle collective. Le bon fonctionnement des institutions repose sur une *sécurité épistémique partagée*, c'est-à-dire un socle commun de réalités factuelles et de consensus scientifiques. Or, la dérégulation des marchés de l'information, dominés par des *logiques de viralité et de polarisation*, menace directement ces fondements.

Pour endiguer cette fragmentation, le cadre juridique doit opérer un glissement de la responsabilité : il ne s'agit plus seulement de traquer le créateur d'un contenu malveillant, mais de responsabiliser les architectures de diffusion elles-mêmes. Les algorithmes de recommandation et les systèmes créant des « bulles de filtres » doivent être tenus juridiquement responsables des externalités négatives qu'ils génèrent. En privilégiant systématiquement l'engagement émotionnel au détriment de la pertinence factuelle, ces dispositifs aggravent *l'insécurité cognitive* et nécessitent une régulation contraignante pour garantir l'intégrité et la pluralité du débat public.

La gouvernance cognitive

La nature transnationale des flux informationnels et la déterritorialisation des infrastructures algorithmiques rendent obsolètes les réponses purement étatiques. La sécurité cognitive appelle l'émergence d'une gouvernance globale, seule capable d'imposer un cadre normatif aux conglomérats technologiques et d'endiguer la militarisation de l'espace mental.

Cette exigence soulève le défi inédit de la coopération internationale à un moment de reconfiguration

géopolitique majeur. À l'instar des accords régulant les armes de destruction massive, la communauté internationale doit envisager la conception de traités de non-prolifération spécifiquement dédiés aux *armes de manipulation cognitive*. Il s'agit d'interdire l'usage étatique et para-étatique d'intelligences artificielles génératives à des fins de subversion psychocognitive de masse ou de polarisation artificielle des sociétés civiles.

Parallèlement à cette diplomatie normative, la gouvernance globale doit s'appuyer sur des mécanismes de contrôle techniques et institutionnels stricts. Le rôle d'audits algorithmiques indépendants devient central. De la même manière que des agences certifient la sûreté des installations industrielles ou pharmaceutiques, de nouvelles autorités de régulation devront être mandatées pour délivrer une « certification de sûreté cognitive » aux plateformes numériques et aux grands modèles de langage.

Ces audits devront évaluer de manière transparente les poids synaptiques des modèles, les biais d'optimisation de l'attention et les risques de capture comportementale, conditionnant ainsi l'accès au marché de ces systèmes à leur stricte conformité avec les impératifs de la santé mentale publique.

Conclusion

Au terme de cette analyse, il apparaît que la sécurité cognitive ne saurait être réduite à un simple ajustement technique ou à une contrainte juridique périphérique. Face à l'industrialisation des menaces psychocognitives, à la saturation attentionnelle et à l'exploitation algorithmique de nos vulnérabilités heuristiques, la préservation de l'intégrité psychique et des capacités délibératives s'impose comme le défi majeur des sociétés hautement connectées.

Le déploiement d'une véritable hygiène cognitive à l'échelle organisationnelle dessine les contours d'une nouvelle souveraineté de l'esprit. Loin de constituer un frein au progrès technologique ou une entrave à l'innovation, l'édification de cette architecture de sécurité cognitive en représente, au contraire, la condition *sine qua non* de son acceptabilité sociale et de sa légitimité politique.

Sans cette garantie d'intégrité mentale, c'est le contrat social lui-même qui se trouve menacé par l'érosion de la confiance épistémique et l'aliénation de l'autonomie citoyenne.

Ce constat appelle, en conséquence, un dépassement des logiques défensives traditionnelles. L'enjeu réside dans la transition d'une approche strictement réactive, consistant à corriger les vulnérabilités exploitées, à colmater les brèches psychologiques et à atténuer les biais *a posteriori*, vers une approche foncièrement proactive englobant toutes les dimensions de la cognition. Il s'agit d'impulser la conception d'écosystèmes et d'environnements informationnels pensés dès leur genèse pour générer de la santé mentale et favoriser l'épanouissement intellectuel.

L'horizon ultime de la sécurité cognitive ne se limite donc pas à la seule sanctuarisation du cerveau contre les ingénieries de manipulation ; il réside dans l'invention d'infrastructures de la connaissance capables de soutenir l'attention, d'élever la pensée critique et de renforcer durablement les capacités neurocognitives humaines au lieu de les épuiser.

Références

Danet, D. (2025). Des opérations d'influence à la sécurité cognitive: Faire face aux menaces dans le champ informationnel. In *La guerre hybride* (No. 4, pp. 135-142). Presses universitaires de Rennes.

Jost, C. (2026). Ingérence russe: la guerre cognitive 2.0. *Dossiers Cairn*, (14).

Valette, M. (2024). Guerre cognitive, culture et récit national. *Ingénierie cognitique*, *7*(1), 6-12.

Bazalgette, D., Boisset, M. O., & Langlois-Berthelot, J. (2023). Opérations cognitives. *DSI (Défense et Sécurité Internationale)*, (163), 92-97.

Claverie, B. (2024). "Cognitive Warfare"–Une guerre invisible qui s' attaque à notre pensée. *Faut-il s' Inquiéter?*, 89-115.

Claverie, B. (2025). La guerre cognitive: le nouveau champ de bataille qui exploite nos cerveaux. *Polytechnique Insights*.

Fassler, T., & Hardy, M. (2024). La guerre cognitive vue par le Japon: un concept et une posture récents et situés. *Ingénierie cognitique*, *7*(1), pp-23.

Achmirowicz, J., & Langlois-Berthelot, J. (2023). Jeux vidéo et métaverse: nouveaux terrains de la guerre cognitive des groupes terroristes?. *Revue Défense Nationale*, *865*(10), 52-56.

Darses, F. (2025). Charge et surcharge cognitive en situation opérationnelle: quelques repères théoriques. *Revue Défense Nationale*, (HS16), 85-97.

de Swielande, T. S., Orinx, K., & Peiffer, S. (2024). La guerre cognitive au cœur de la stratégie chinoise de socialisation.

Futorjanski, J., & Ahmadi, R. (2024). Formation cognitive personnalisée: France et Allemagne. In *Manifeste pour une Europe à impact* (pp. 64-70). Vuibert.

Allard, C., & Delorme, T. (2026). De la massification à l'individualisation: comment l'IA promet de transformer la guerre cognitive. Le cas GoLaxy. *Revue Défense Nationale*, *886*(1), 57-64.

Valle-Cruz, D., García-Contreras, R., & Gil-Garcia, J. R.

(2024). Analyse des effets négatifs de l'intelligence artificielle dans l'administration: la face cachée des algorithmes intelligents et des machines cognitives. *Revue Internationale des Sciences Administratives*, *90*(2), 281-297.

Lassalle, B., & De Gliame, B. (2022). Sensibiliser et armer les citoyens face à la guerre cognitive. *Revue Défense Nationale*, *848*(3), 88-93.

Dugoin-Clément, C., & Henot, S. (2023). L'influence du sentiment d'insécurité dans le comportement et la prise de décision: étude de cas. *Question (s) de management*, *46*(5), 83-90.

Guidère M. (2025). *L'anxiété géopolitique: Comprendre et agir dans un monde incertain*. Paris : Strategikia.