L'IA et ses conséquences pour les métiers de la sécurité

Gérald Vernez, fondateur et directeur de la fondation digiVolution[1]
Prof. Dr. Diego Kuonen, Université de Genève, fondateur et directeur de la société Statoo Consulting, membre du Comité consultatif de digiVolution

---

Le phénomène ChatGPT

Le 30 novembre 2022, OpenAI rendait publique ChatGPT, le premier agent conversationnel issu de l'IA générative[2], gratuit et accessible à quiconque disposant d'une connexion Internet. Un véritable séisme qui a même surpris de nombreux observateurs avisés et ses propres inventeurs. ChatGPT a popularisé l'IA et déclenché une lame de fond et la prolifération inédite d'instruments et de services faisant usage de briques technologiques basées sur l'IA. Il y a clairement un avant et un après ChatGPT.

ChatGPT et les développements similaires se composent tout d'abord de carburant, c'est-à-dire de données. La seconde composante clé, c'est le cerveau ou les algorithmes de ces LLM, ou Large Language Models[3]. Les LLM sont entraînés sur de vastes quantités de données à partir desquelles ils établissent un modèle statistique leur permettant de proposer, avec une forte probabilité, quel mot vient après quel autre dans une réponse à une question posée ou à générer de nouvelles idées à partir des informations qu'ils possèdent.

Ces éléments clés – données et algorithmes – expliquent pourquoi l'IA est encore largement perfectible. Il suffit que l'un d'eux comporte une erreur pour que le résultat soit directement affecté. «*Garbage in ! Garbage out* !». Et cela est vrai dès les premiers stades de développement et d'entraînement des modèles. Vous empoisonnez l'entrée, vous n'aurez rien de mieux à la sortie. La qualité des données explique la raison pour laquelle les systèmes d'IA produisent de nombreux biais et se rendent parfois même coupables de racisme ou de discrimination, allant même jusqu'à inventer des faits. On parle alors «d'hallucination». Si la question (qui est de la donnée) est mal posée en entrée, l'outil produira une réponse (de la donnée) inadéquate en sortie. Et si le destinataire ne sait pas interpréter le résultat fourni ou le vérifier, des faits erronés seront pris pour des vérités.

Usages de l'IA

Considérer ces défauts de jeunesse comme définitifs et condamner l'IA *ad aeternam* ou limiter l'IA à la seule IA générative et à des agents conversationnels comme ChatGPT, Copilot ou Gemini serait une erreur monumentale. L'IA est en effet déjà établie dans de nombreuses activités et produits courants du quotidien exposés dans le tableau suivant.

Domaines d'emplois	Exemples
Apprentissage automatique (Machine Learning)	- Reconnaissance d'images, diagnostics médicaux. - Prédiction de prix immobiliers, prévisions météorologiques. - Analyse de données génomiques.

[1] https://digivolution.swiss/
[2] L'«IA générative» est une notion large qui se réfère aux systèmes d'IA entraînés sur de grands volumes de données du monde réel et virtuel afin de générer eux-mêmes des données (textes, images, enregistrements sonores, vidéos, simulations, codes, etc.). Ils sont souvent multimodaux, avec p.ex. des entrées et/ou des sorties dans une ou plusieurs modalités (texte, image, vidéo, etc.).
[3] L'intégration de modalités supplémentaires aux LLM permet de créer des LMM (Large Multimodal Models) ; voir https://huyenchip.com/2023/10/10/multimodal.html

Domaines d'emplois	Exemples
	- Systèmes de recommandation. - Création artistique (peintures, musique, mode, architecture).
Traitement automatique du langage naturel (NLP)	- Traduction automatique, génération de texte. - Surveillance des médias sociaux, services clients. - Assistants virtuels, transcription automatique (dictée). - Création de contenu, résumés automatiques.
Vision	- Surveillance vidéo, véhicules autonomes. - Reconnaissance faciale, analyse d'images médicales. - Inspection industrielle, diagnostic médical par imagerie. - Surveillance de la sécurité. - Analyse sportive.
Systèmes de recommandation	- Plateformes de streaming, e-commerce, publicité. - Analyse prédictive (p.ex. dans l'entretien des machines)
Robotique	- Robots de livraison et de construction, véhicules autonomes. - Automatisation industrielle, prothèses robotiques. - Sauvetage et aide en cas de catastrophe.
Systèmes experts	- Assistance dans la prise de décisions cliniques. - Prévision des marchés boursiers, détection de fraudes. - Agriculture (prévention des maladies, suivi du développement des récoltes, amélioration et sélection des cultures, détection et traitement des parasites, etc.).
Cybersécurité	- Détection d'intrusions, surveillance des réseaux. - Analyse des menaces, identification des vulnérabilités, réponses automatisées aux incidents.

L'IA est désormais au centre de la mutation numérique. Comme en biologie, «mutation» signifie qu'il s'agit d'un processus irréversible. Que nous le voulions ou non, avec l'IA la société va vers toujours plus d'automatisation[4] et « d'intelligence augmentée ».

Les bénéfices apportés par l'IA sont indéniables, mais toute pièce de monnaie comporte une seconde face et la liste des risques et des usages malveillants est aussi longue que celle des avantages. C'est ainsi que l'IA a fait une entrée remarquée dans le champ politique avec l'élection présidentielle américaine en 2016 et elle est omniprésente dans les leviers qui animent les opérations d'influence, la radicalisation et les théories du complot notamment. Il ne se passe plus un jour sans que l'IA ne soit associée à des dérapages, aux conflits en cours ou à des cyberattaques. On observe par ailleurs que les criminels et les fraudeurs sont souvent plus rapides que bon nombre d'entreprises et de services publics dans l'adoption de technologies tel que l'IA pour renforcer leurs activités. L'IA est également omniprésente dans les armements et les débats sur les robots tueurs sont légion, toutefois sans résultats vraiment tangibles jusqu'ici. Et l'on ne parle pas de tous les problèmes juridiques, de propriété intellectuelle ou encore de clonage vocal et visuel (deepfakes) liés. La liste est sans fin. Et parmi les prochains coups d'accélérateur vient l'informatique quantique.

[4] L'automatisation et la rationalisation des tâches répétitives impliquant de grandes quantités de données devraient intervenir dans une situation stable, dont les règles s'appliquent aujourd'hui et demain, dont l'avenir ressemble au passé et dont personne ne peut enfreindre les règles. De nombreux domaines de la politique ou d'activités se caractérisent en effet par des situations où règne l'incertitude et où les règles ne peuvent pas être appliquées telles quelles. C'est le cas par exemple en cuisine, lorsqu'il s'agit de répondre aux souhaits individuels des convives ou de réagir correctement dans des situations d'urgence inattendues pendant la préparation du repas, telle l'absence d'ingrédients ou d'ustensiles indispensables. Dans son livre *Klick*, publié en 2021, le psychologue allemand Gerd Gigerenzer décrit cette incertitude en évoquant à quel point il serait difficile de jouer aux échecs « si le roi pouvait violer les règles sur un coup de tête et si la dame pouvait quitter l'échiquier en protestant après avoir mis le feu aux tours ». (Source: chronique « *Gesucht: Kochroboter/in – Gefunden: Schachroboter/in*» [Diego Kuonen, Walliser Bote du 18 novembre 2021]). L'IA n'est prometteuse que dans des situations stables. En cas d'incertitude, elle peut néanmoins fournir une aide aux décideurs humains.

Futur de l'IA et superintelligence

Une politique de développement responsable comme celle promue par la Confédération[5] met l'humain et son contrôle de la technologie au centre des processus, mais le futur verra inexorablement augmenter la part des systèmes d'IA, y compris dans les domaines complexes ou des tâches à haute valeur ajoutée, comme la planification stratégique ou la prise de décision. Si le public a toujours plus l'impression de converser et d'interagir avec une réelle intelligence, en réalité il échange avec des systèmes d'IA dont l'intelligence réside encore dans des statistiques. Les robots quadrupèdes ressemblent à de sympathiques toutous et les bipèdes se voient affublés de caractéristiques humanoïdes. Pourtant ce ne sont encore que des systèmes d'IA frustes.

Divers indices annoncent toutefois très prochainement d'importants progrès. Il se pourrait que nous approchions plus rapidement qu'anticipé de« l'intelligence artificielle générale » et de ses déclinaisons. Le terme clé est celui de la «singularité technologique», ce moment où la technologie disposerait d'une propre conscience et autonomie lui permettant de s'autoaméliorer et de concevoir de nouvelles générations technologiques de plus en plus intelligentes et toujours plus rapidement. Ce serait alors une «explosion d'intelligence» dépassant de loin l'intelligence humaine. Ray Kurzweil, inventeur du concept, prédit ce passage pour 2045. Cette évolution et cette date font cependant l'objet d'âpres débats et de spéculations. Certains estimant par exemple que les récents développements technologiques la rendraient possible dès 2028 déjà. Les conséquences sociales et politiques seront majeures, mais il n'y a aucun consensus[6].

L'IA fascine autant qu'elle effraie et on observe une intense activité législative pour tenter d'en maîtriser le développement. Mais le problème est kafkaïen. D'un côté il y a le risque de ne pas imposer des règles assez tôt et assez puissantes et donc sans effet, conduisant à une perte de contrôle sur l'IA et ses usages. À l'opposé il y a le risque de trop légiférer et d'étouffer les développements chez soi pendant que d'autres progressent sans entraves et en tirent d'importants avantages stratégiques, militaires et financiers. Peur du monstre ou peur pour les affaires? Comment réguler quelque chose que l'on ne comprend pas et pour lequel les experts eux-mêmes sont en désaccord? La réponse s'appelle « principe de précaution » et vigilance (donc renseignement) illustré ci-contre par un symbole de ce domaine (produit par une IA), à la condition toutefois que tout le monde respecte les mêmes règles.

IA et sécurité

Comment les métiers de la sécurité doivent-ils appréhender les défis posés par l'IA? Le monde ne cesse de se complexifier et les dirigeants ne peuvent plus «simplement faire comme avant». Il n'existe pas de réponse simple. Ils doivent impérativement s'adapter et voici cinq recommandations de *digi*Volution que les auteurs de cet article pratiquent et mettent en œuvre dans leurs activités de conseil et de soutien opérationnel.

1. Information et formation – Qu'on le veuille ou non, la technologie avance. Les cadres et les collaborateurs doivent être informés et formés aux usages des technologies et pouvoir agir de façon éclairée face aux risques (voulus ou non) de l'IA et aux opportunités

[5] « Code de bonnes pratiques de la Confédération pour une science des données et IA centrée sur l'être humain et digne de confiance » ; voir https://www.bfs.admin.ch/bfs/fr/home/dscc/dscc.assetdetail.29325685.html
[6] https://arxiv.org/abs/2401.02843

qu'elle ouvre. Interagir avec l'IA est désormais une compétence indispensable. Le carburant de l'IA étant les données, de même que les résultats produits par les outils basés sur de l'IA, il est impératif de promouvoir une littératie des données. C'est une condition essentielle pour une société informée[7]. Elle doit s'imposer également dans tous les métiers de la sécurité. Chaque individu doit acquérir une culture des données et de l'IA (« *Data and AI Literacy for Everyone! Leave No One Behind!* »)[8].

2. La clé de voûte réside dans la gouvernance – Les décideurs doivent à tout prix éviter de verser dans l'angélisme technophile ou dans la paranoïa technophobe. Le suivi des développements technologiques doit être intégré à leur gestion des risques stratégiques afin de prendre à temps toute mesure de maîtrise des développements disruptifs liés à l'IA ou de tirer avantage de ses apports. Les différents services doivent (et cela n'est pas une option) mettre en place une stratégie permettant aux dirigeants de profiter des progrès de l'IA et de maîtriser les conséquences qui en découlent.
3. Politique forte de gestion des données – Le carburant des décisions, ce sont les données. Chaque entreprise ou organisation doit identifier ses données, les gérer, les protéger et les valoriser dans un processus dédié. Il ne s'agit pas uniquement de stocker – de manière protégée lorsque cela est nécessaire – les données, mais, au sein d'une organisation spécifique, de les rendre utilisables pour tous ceux qui pourraient en bénéficier et de les faire travailler pour produire des connaissances supplémentaires.
4. Stratégie d'intégration technologique – Un usage éclairé de l'IA en tant que « co-pilote cognitif » doit permettre aux organisations / entreprises de produire d'importantes plus-values et des économies dans des tâches répétitives et à faible valeur ajoutée. Ces économies doivent permettre aux collaborateurs de se concentrer sur ce qu'ils savent faire le mieux. Le volume, la vitesse et la volatilité de l'information augmentant massivement, le contrôle de son intégrité prend une importance majeure. Il s'agit donc d'investir dans les tâches de vérification de l'authenticité des informations avant de les intégrer dans le processus décisionnel. L'IA détruirait des jobs? Commençons par les transformer. Ce n'est pas l'IA qui va vous prendre votre travail, mais les gens qui travaillent avec l'IA.
5. Rester sûr et résilient à tout prix – Les promesses des technologies sont certes immenses, mais elles sont loin d'être infaillibles. Alors quid des accidents, pannes et vulnérabilités? Soigner son affinité technologique, oui, sans aucun doute, mais en se prémunissant de toute dépendance systémique et en étant prêt à gérer les situations de crise.

Ces recommandations et les faits exposés plus haut illustrent le tournant où se trouve la société. Ignorer cette situation revient juste à prendre du retard sur un développement inéluctable et attendre, c'est risquer de le rendre irrattrapable. Dans un récent interview, Peter Brabeck-Letmathe estimait que « *nous avons perdu le contrôle du temps et du changement* ». Nous invitons nos lecteurs à y réfléchir sérieusement. Chez *digi*Volution nous partageons ce constat mais c'est pourquoi nous nous engageons pour élaborer des solutions qui renforcent la confiance, la résilience et la souveraineté numériques de la Suisse. Et au centre de nos actions se trouvent les principes d'anticipation et de précaution.

[7] https://bigdata-dialog.ch/fr/promouvoir-la-litteratie-des-donnees/ et https://fr.data-literacy.ch/
[8] https://upd-initiative.ch/data-literacy-leave-no-one-behind/