Intelligence artificielle et sécurité nationale

Par **Julien BARNU**
Conseiller industrie et numérique du secrétaire général de la défense et de la sécurité nationale (SGDSN)

Dans son rapport « Donner un sens à l'intelligence artificielle » de mars 2018, Cédric Villani désignait le secteur de la défense et de la sécurité nationale comme prioritaire en matière de développement et d'usage de l'intelligence artificielle en France, considérant que ce secteur pouvait « constituer un avantage comparatif de la France ». Dans cet article, je chercherai à mettre en lumière, de façon non exhaustive, les particularités de ce domaine pour l'usage de l'intelligence artificielle, en développant quelques exemples sur le fondement des missions du secrétariat général de la défense et de la sécurité nationale, en particulier dans les domaines de la cyberdéfense et du renseignement technique.

Apports et limites de l'IA : l'exemple de la cyberdéfense

L'IA est susceptible d'être une technologie clé dans le domaine de la cyberdéfense, mais ce domaine illustre également plusieurs de ses limites inhérentes. Ainsi, si l'on considère ce que peut apporter l'IA dans les différentes phases d'une attaque informatique :

- Dans la phase amont de protection des systèmes d'information, l'IA est susceptible de démultiplier les capacités de l'État et des entreprises à effectuer des diagnostics techniques, voire de mettre en place des mesures correctives, à bas coût, en automatisant la recherche de vulnérabilités et de défauts de configuration dans un réseau. De la même manière, l'IA permettra d'automatiser le travail d'évaluation de la sécurité des produits numériques. L'impact de ce progrès technologique sur le niveau global de cybersécurité est cependant discutable. D'une part, parce que ces mêmes outils d'IA, utilisés par des attaquants, leur permettront également d'identifier de façon massive et automatisée des failles dans les réseaux et dans les produits numériques, et, d'autre part, car, au-delà des failles techniques, ce sont les failles humaines qui constituent et continueront à constituer la principale porte d'entrée des attaquants dans les réseaux (introduction de clés USB piégées, ouverture de pièces jointes malveillantes, etc.). Par ailleurs, la généralisation de l'usage d'outils d'IA permettra vraisemblablement aux attaquants de forger du contenu piégé de plus en plus crédible, de façon automatisée, par exemple en parcourant les réseaux sociaux ou le contenu de messageries électroniques. Un tel usage de l'IA, permettant aux attaquants de mieux exploiter les failles humaines, pourrait même compenser son apport en matière de sécurisation technique. Le développement de l'IA dans le domaine cyber pourrait ainsi avoir un effet globalement nocif sur l'état général de cybersécurité du pays.
- C'est probablement dans la phase de détection des attaques que les technologies d'IA sont les plus prometteuses. En effet, les technologies de détection actuelles s'appuient majoritairement sur des « signatures », c'est-à-dire sur des éléments caractéristiques d'un attaquant (adresse IP d'un serveur d'attaque, code malveillant, etc.) connus et intégrés dans un dispositif de détection, comme une sonde ou un antivirus. Ces dispositifs ne peuvent donc pas, par nature, détecter des tentatives d'attaque nouvelles aux caractéristiques inconnues. L'intégration de l'IA, et en particulier des techniques d'apprentissage, dans les systèmes de détection, est en mesure de résoudre cette problématique, en permettant de détecter des anomalies dans un réseau, et donc des modes opératoires d'attaque jamais observés jusqu'alors. Cette application de l'IA à la détection des cyberattaques fait l'objet d'importants travaux, en France comme à l'étranger.

- En matière de réponse à une attaque informatique, l'apport de l'IA doit être relativisé pour plusieurs raisons. Tout d'abord parce qu'il est impossible de se reposer sur une analyse technique, fut-elle conduite par une IA, pour attribuer une attaque, c'est-à-dire pour en identifier l'auteur. Certes, l'IA permettra d'associer plus rapidement une attaque donnée à un « mode opératoire », c'est-à-dire à un ensemble d'outils connus et caractéristiques d'un groupe d'attaquants, mais un tel rapprochement ne suffit pas à attribuer une attaque. En effet, les outils d'attaque, dont beaucoup sont en vente sur Internet, peuvent aisément être réutilisés, et certains attaquants de haut niveau sont même connus pour introduire dans leurs outils de fausses preuves techniques afin de conduire à des erreurs d'attribution (en orientant vers un pays donné, par exemple). Là encore, l'IA renforcera vraisemblablement la capacité des attaquants à brouiller les pistes, en automatisant l'introduction de fausses preuves dans leurs outils d'attaque et en renforçant la crédibilité de ces dernières. Le volet humain du travail d'attribution, et en particulier le renseignement d'origine humaine, restera donc sans nul doute tout aussi crucial qu'aujourd'hui pour attribuer une attaque. En matière de réponse, il est encore plus difficile d'imaginer que l'IA puisse provoquer un effet de rupture, d'une part, parce que le choix de répondre à une attaque, qui suppose de l'avoir préalablement attribuée, demeurera un choix politique, et, d'autre part, parce que la nature de la réponse à une attaque restera évaluée au cas par cas, en fonction de nombreux facteurs, notamment géopolitiques, et ne se situe pas forcément dans le champ cyber : une réponse à une cyberattaque peut être diplomatique, politique, économique ou même militaire.

Un rôle de l'IA parfois mal compris : l'exemple du traitement de données hétérogènes

Un des défis majeurs des États dans le champ du renseignement est d'exploiter de façon optimale les importantes quantités de données collectées (données de connexion, images, captations audio, renseignement satellitaire, etc.).

Dans la poursuite de cet objectif, des briques d'intelligence artificielle peuvent naturellement compléter les outils d'analyse de données mis en œuvre, par exemple en dégageant des tendances ou en identifiant des singularités difficilement perceptibles par l'œil humain. Mais dans ce domaine, le rôle de l'IA, au moins à court terme, est souvent surestimé. La principale difficulté rencontrée aujourd'hui est en effet de disposer, avant même de songer à déployer des outils d'IA, d'une capacité souveraine à structurer les données collectées et à fournir aux analystes des outils numériques leur permettant de les exploiter efficacement.

Contrairement à une idée reçue, le succès des solutions logicielles, telles que celles fournies par le *leader* américain Palantir, ne repose pas sur l'IA, qui n'y est en réalité intégrée que très marginalement, mais sur la capacité de cette entreprise à agréger rapidement des données d'origines et de formats variés et à les présenter à l'analyste d'une façon qui lui permette de les visualiser, de les croiser et de les analyser selon son approche métier propre. La force de telles sociétés ne repose pas sur une quelconque avance technologique en matière d'IA, mais sur les développements informatiques qu'elles ont conduits en associant étroitement des clients d'horizons divers (services de renseignement, banques, industriels, etc.). Cette approche leur a permis de disposer d'une palette de plus en plus large de fonctionnalités rapidement déployables, qu'elles sont en outre en mesure d'adapter en permanence aux besoins de leurs nouveaux clients, dans une forme de cercle vertueux.

Le retard de l'Europe et de la France sur ce segment ne résulte donc pas de lacunes dans le domaine de l'IA. La difficulté des industriels français à proposer des produits compétitifs s'explique davantage par leur manque d'agilité et de culture commerciale, qui se manifeste par une incapacité

à s'approprier les besoins de leurs clients et à leur proposer, en travaillant étroitement avec eux, des offres adaptées à leur métier. Certes, des fonctionnalités fondées sur des technologies d'IA complèteront naturellement de tels outils de traitement de données, mais à court terme c'est moins de prouesses technologiques dans le domaine de l'IA que d'une meilleure approche client dont nous avons besoin pour voir émerger une offre nationale crédible.

Une spécificité du domaine de la sécurité nationale : la disponibilité de la donnée

Contrairement à d'autres domaines où le développement d'une IA souveraine se trouve freiné par la faible quantité de données disponibles, ces dernières étant largement aux mains de grandes entreprises technologiques étrangères, le domaine de la défense et de la sécurité nationale est particulièrement riche en données nationales : images captées par satellite, communications électroniques et métadonnées captées sur les réseaux télécoms, captations sonores ou vidéos, données informatiques collectées à des fins de cybersécurité, etc.

Mais, davantage que dans les autres domaines, il existe d'importantes contraintes en matière d'utilisation et d'ouverture de ces données : la protection du secret et le principe du « besoin d'en connaître » imposent en effet un cloisonnement strict de ces données. En particulier, l'accès aux données collectées par les services de renseignement sur le territoire national est très strictement encadré par la loi.

Constituer des jeux d'apprentissage largement accessibles à des fins de développement de technologies d'IA pour les besoins de la sécurité nationale constitue donc un défi de taille, qui nécessite une refonte profonde de la gouvernance de la donnée au sein des ministères régaliens, et une politique claire de gestion et de valorisation de la donnée, prévoyant des procédures de mise à disposition des données pour l'entraînement des algorithmes d'IA ou de test des algorithmes sur des données réelles.

Ce défi est d'autant plus grand que le développement de l'IA constitue un changement de paradigme pour la sphère de la défense : historiquement en avance sur l'innovation civile, elle est aujourd'hui contrainte de s'appuyer sur des acteurs du monde numérique qu'elle connaît relativement mal et avec lesquels elle doit imaginer de nouveaux modes de travail : mettre à disposition d'entreprises innovantes des lacs de données opérationnelles, en veillant à un contrôle strict des accès, s'appuyer sur des outils civils disponibles en source ouverte et les adapter, ou choisir d'investir dans la conception de systèmes d'IA spécifiques – quitte à ce que ces derniers soient moins performants, au moins temporairement, que des solutions commerciales étrangères.

La confiance dans l'IA : un enjeu crucial dans le domaine de la sécurité nationale

Les algorithmes d'IA produisent parfois des résultats aberrants pour la perception humaine et les systèmes d'IA peuvent échouer de manière inattendue. Selon les mots d'une ancienne directrice de la DARPA : "*The problem is that when they're wrong, they are wrong in ways that no human would ever be wrong.*" Les techniques d'apprentissage présentent par ailleurs un risque de biais, qui peut être involontaire (données d'apprentissage non représentatives) ou volontaire (données d'apprentissage modifiées par un tiers pour influencer le comportement du système). Ces techniques peuvent enfin présenter des résultats opaques, difficilement explicables – cette question bien connue de l'explicabilité de l'IA faisant l'objet d'importants travaux internationaux.

Ces limites de l'IA, qui posent la question de la confiance qu'on peut avoir dans un système d'IA, sont particulièrement problématiques en ce qui concerne l'usage de l'IA à des fins de défense et de sécurité nationale, et la rendent inadaptée à de nombreuses applications militaires. Il n'est pas possible à ce jour de dire si ces limites sont surmontables ou non.

Dans l'impossibilité de garantir avec certitude la confiance dans un système d'IA, nous serons amenés à fixer, pour chaque application de l'IA dans le champ de la défense et de la sécurité nationale, un « niveau de confiance » exigible. Cette approche nous impose toutefois de développer une capacité souveraine à évaluer un système d'IA, afin d'être en mesure d'estimer ce niveau de confiance, voire de le certifier – sur le modèle des certifications de sécurité délivrées par l'Agence nationale de la sécurité des systèmes d'information pour les solutions numériques traditionnelles. Or, la certification de l'IA, et en particulier des techniques d'apprentissage, est encore un objet de recherche, et le cadre et les modalités d'une telle certification restent à construire. Pour le développement des usages de l'IA dans le secteur de la défense et de la sécurité nationale, ce chantier constitue un défi majeur, et l'un des facteurs les plus limitants. Sans avancée substantielle dans ce domaine, l'IA restera cantonnée à des usages qui seront en mesure de faire gagner du temps à l'opérateur humain, mais non de créer un effet de rupture et de bouleverser les rapports de force.