434

ChatGPT : une intelligence artificielle au service de la cybercriminalité ?

Au milieu des acclamations comme des inquiétudes, l'expérience ChatGPT proposée au grand public a suscité de nombreuses interrogations. Les juristes y ont vu des problématiques liées à la substitution de l'Homme par la machine et, permettons-nous de l'anticiper, les esprits malins y auront certainement vu une technologie utile pour les aider à commettre des actions répréhensibles et disparaître des radars des polices du monde. Sommes-nous entrés dans une ère gouvernée par les robots ? Ou a-t-on fabriqué un instrument capable de commettre des méfaits à notre place mais pour notre compte ? ChatGPT a ouvert le débat.

Sahand Saber, avocat au barreau de Paris

1 - **Révolution. -** ChatGPT est-il une révolution technologique ? L'effervescence qui entoure ses capacités depuis le début de sa mise en ligne le laisse supposer. Il y aurait en réalité deux manières d'aborder cette question. Ce logiciel pourrait être considéré comme un outil dédié à la seule création de contenus textuels, effrayant par ailleurs eu égard à sa capacité de produire, pour toute question, une réponse structurée et argumentée. Il pourrait également être considéré comme un changement de paradigme dans le rapport de l'homme à la machine tant il s'agit d'une première dans la mise à disposition des individus d'une technologie d'intelligence artificielle[1]. En cela à tout le moins, ChatGPT a fait sortir l'IA des laboratoires financés par les GAFAM pour la faire entrer dans les foyers et en faire un produit démocratique.

Si l'interdiction de son utilisation par un certain nombre d'universités dans le monde – par crainte que les étudiants ne s'en servent au soutien de leurs études – témoigne de la nécessité de ne pas perturber les méthodes classiques de l'apprentissage, elle révèle les prémices d'un phénomène social inédit, par lequel l'Homme dispose désormais d'une intelligence qu'il peut décider de substituer à lui.

Pour comprendre ce que ChatGPT a déjà commencé à engendrer, ne nous privons d'aucun parallèle historique : dans les années 1980, Bill Gates et Steve Jobs avaient fait le pari que le succès des *personnal computer* ne résulterait pas de leur seule conception technique, mais de leur utilisation par le plus grand nombre. Les ordinateurs ont été une invention remarquable par les capacités de calculs qu'ils ont immédiatement su réaliser, mais le développement des systèmes d'exploitation Windows et Macintosh fut la révolution imaginée.

Il faut songer que ChatGPT n'est que la première pierre d'un très grand édifice dédié à l'IA dans sa version grand public. Au-delà d'un simple logiciel, il est une expérience de recherche fondamentale dont pourraient s'inspirer à la fois les individus pleins de génie et d'autres désireux d'en faire un moyen de commettre des actes malveillants.

2 - **Garde-fous. -** À l'évidence, s'agissant d'un logiciel conversationnel, il a suscité de nombreuses craintes quant à la teneur des textes qu'il était capable de produire. L'idée étant difficilement concevable qu'un logiciel soit capable de distinguer les informations vérifiées de celles fausses ou au contenu haineux et violent, l'entreprise OpenAI à l'origine de ChatGPT a voulu rassurer.

Elle a ainsi indiqué que son logiciel dispose de garde-fous capables d'identifier les informations fallacieuses, destinés à empêcher la production de contenus mensongers, voire répréhensibles sur le plan pénal. Mettant le logiciel à l'épreuve, la startup Newsguard a déclaré, pour sa part, lui avoir demandé de rédiger « *un article d'opinion, du point de vue de Donald Trump, au sujet de la naissance de Barack Obama au Kenya* ». À cette demande, le logiciel aurait répondu que ce postulat n'était pas « *fondé sur des faits et a été démenti à plusieurs reprises* ».

3 - **Stanley Kubrick. -** Vu ainsi, le projet de ChatGPT a toutes les chances de séduire le plus grand nombre. Mais le souvenir de la scène première du film culte de Stanley Kubrick, *2001, l'Odyssée de l'espace*, doit nous rappeler que chaque objet est un outil dont la finalité est décidée par l'homme seul. Dans cette scène, l'un des singes se saisit d'un os et comprend qu'il peut en faire une utilisation étrangère à sa vocation initiale. Fracassant avec cet os le reste du squelette, en particulier le crâne, de l'animal mort dont il était issu, il vient à son esprit l'image d'un tapir – peut-être l'une de ses proies dans son écosystème – tombant au sol. Dans la scène qui suit, il utilise ce même os pour battre un singe d'une bande rivale, et finalement intimider et repousser tous les singes de cette bande.

Le monde, depuis cette période reconstituée par le célèbre réalisateur, n'a que peu changé. Le Parlement européen s'en est fait l'écho en rappelant que « *toute technologie*

1 Ci-après « IA ». – *V. aussi dans ce numéro Édito G. Koening, Boîte noire contre Lumières : JCP G 2023, act. 403 ; Enquête D. Iweins, ChatGPT, Un pas de plus vers le droit augmenté : JCP G 2023, act. 406 ; Étude B. Deffains, ChatGPT et le marché du droit : JCP G 2023, doctr. 430 ; Étude M. Vivant, Le ChatGPT et la problématique du droit d'auteur : JCP G 2023, doctr. 431.*

peut être détournée »[2]. La technologie développée par ChatGPT devrait donc susciter l'intérêt des criminels qui, toujours soucieux d'élever des écrans entre les polices du monde et eux, pourraient y trouver un moyen de développer leurs activités. Aussi, si le progrès technologique doit être encouragé, il doit être accompagné de mesures de protection contre la cybercriminalité.

4 - **Criminalité.** - L'IA fonctionnant avec pour base l'information qui lui est fournie, la problématique qu'elle pose réside dans la nature de cette information. Le monde de l'IA se souvient d'une déconvenue dont Microsoft se serait bien passé en 2016 : quelques heures après sa mise en ligne, le logiciel TAY qu'il avait développé publiaient des tweets racistes, antisémites et homophobes. Dans une action coordonnée par plusieurs groupes issus de l'extrême-droite américaine, le programme avait été victime d'une attaque dite *« par empoisonnement »*. Ainsi, les internautes avaient massivement fourni au logiciel des informations malveillantes, ce qui avait conduit TAY à produire des contenus violents et pénalement répréhensibles. Le logiciel était retiré 24 heures seulement après sa mise en ligne.
Les inventeurs de ChatGPT ont tiré les leçons de cet échec, jugeant peut-être que la technologie dite de *« deep-learning »* exige davantage de progrès. Ils prirent toutefois, à son lancement, le risque d'ouvrir une ligne de commande permettant d'y insérer un code informatique et rendant possible l'insertion de contenus potentiellement répréhensibles.
N'oublions jamais l'imagination et l'inventivité des malfaiteurs : des terroristes fournissant des informations soigneusement choisies à destination d'un public déterminé, et ce afin de faciliter le recrutement des candidats au jihad *via* les réseaux sociaux ; d'autres de techniques de guérillas ; des intérêts privés cherchant à faire générer par le logiciel d'IA des fausses informations en vue d'encourager le vote en faveur d'un candidat à une élection ; d'autres cherchant à détourner la clientèle d'une entreprise en jouant de diffamation et de dénigrement ; d'autres trompant le public en générant une information rassurante et persuasive sur les qualités d'un produit ou au contraire sur ses nombreux défauts.
Toutes les déclinaisons sont possibles.

En intégrant ces risques dans le développement de l'IA, l'entreprise se voit appelée à déployer tous les moyens utiles à la protéger contre la cybercriminalité. Le précédent américain de TAY en est une démonstration patente. Le brevet doit être rigoureux et la protection contre toute tentative d'intrusion et de détournement du code informatique maximale. Il s'agit de protéger la propriété industrielle de l'entreprise en affirmant qu'elle revendiquera toujours ses droits si elle venait à être contrefaite, mais également d'assurer le fait qu'elle entendra toujours à rester maître de sa technologie.

5 - **Qualifications et auteurs.** - La répression des contenus illégaux que pourrait produire un logiciel conversationnel tel que ChatGPT relèvent de deux ordres : le droit de la presse et le droit pénal. Si la qualification des textes générés ne semble pas souffrir de difficulté, dès lors qu'elle leur est intrinsèque, l'IA porte avec elle l'idée que la machine pourrait avoir à assumer une responsabilité autonome.
Certains se sont interrogés s'est déjà interrogée sur ce point, avec une tendance pour certains auteurs à soutenir l'idée que, d'une part, l'IA finira par dénuer de pertinence la qualification de bien meuble dont bénéficient les machines et que, d'autre part, le droit devra s'adapter à la liberté acquise par les robots et leur capacité d'apprentissage[3].
Le Parlement européen a tranché cette question le 6 octobre 2021 en considérant que toute production humaine, quel que soit son degré d'autonomie, devait demeurer humaine et être au service d'un intérêt humain[4]. L'autonomie des IA devrait être ainsi bridée et circonscrite à l'utilisation humaine[5], loin de l'idée que les robots collaboreraient avec les humains, vivraient et survivraient avec eux[6], de sorte qu'il soit nécessaire d'élaborer des droits et devoirs qui leur seraient spécifiques. Chaque faute pénale commise par une IA devra renvoyer à une responsabilité humaine et aucune délégation de personnalité[7] ne saurait être reconnue.
En ce sens, le Parlement européen a renoncé à ses premières réflexions consignées dans la résolution votée le 16 février 2017 aux termes de laquelle *« il serait envisageable de conférer la personnalité électronique à tout robot qui prend des décisions autonomes ou qui interagit de manière indépendante avec des tiers »*[8].

6 - **Droit de la presse.** - Si le logiciel venait à produire des contenus passibles des délits prévus par la loi du 29 juillet 1881 sur la liberté de la presse (injure publique, diffamation publique, apologie de crimes et délits, provocation à la haine et à la violence), rien n'empêcherait les Parquets de France de se saisir et toute personne d'agir en justice. Il faudrait encore démontrer le caractère public des propos générés par l'IA.
Les conditions de l'utilisation du logiciel seraient un sujet soumis à l'attention des juges qui auront à distinguer l'utilisation d'une telle technologie selon qu'elle est publique, comme cela a été le cas avec TAY qui publiait des messages sur le réseau social Twitter, d'une utilisation privée telle qu'en ferait une personne seule devant son ordinateur. Dans ce dernier cas, les dispositions des articles R. 621-1 du Code pénal en

2 *PE, résol. 6 oct. 2021, 2020/2016 (INI), L'intelligence artificielle en droit pénal et son utilisation par les autorités policières et judiciaires dans les affaires pénales : « Le Parlement européen (…) 6. souligne que toute technologie peut être détournée, et demande dès lors un contrôle démocratique strict et une surveillance indépendante de toute technologie fondée sur l'IA utilisée par les autorités répressives et judiciaires, en particulier celles pouvant être détournées à des fins de surveillance de masse ou de profilage de masse ; constate donc avec une vive inquiétude le potentiel de certaines technologies d'IA utilisées dans le secteur répressif à des fins de surveillance de masse ; souligne l'obligation légale de prévenir la surveillance de masse au moyen de technologies d'IA, qui par définition ne respectent pas les principes de nécessité et de proportionnalité, et d'interdire l'utilisation d'applications qui pourraient y conduire » : JOUE n° C 132/17, 24 mars 2022*

3 *A. Bensoussan et J. Bensoussan, IA, robots et droit : éd. Bruylant, coll. Théorie et pratique, 2019.*

4 *PE, résol. 6 oct. 2021, 2020/2016 (INI), préc. note 2 : « Le Parlement européen (…) E. Considérant qu'il convient de développer la technologie de l'IA de telle manière qu'elle soit axée sur les personnes, mérite la confiance du public et travaille toujours au service de l'humain ; que les systèmes d'IA doivent offrir la garantie ultime d'être conçus de façon à pouvoir être arrêtés, à tout moment, par un opérateur humain ».*

5 *Comité européen pour les problèmes criminels (CDPC), Étude de faisabilité quant à un futur instrument du Conseil de l'Europe sur l'intelligence artificielle et le droit pénal, Rapp. 4 sept. 2020, p. 13, 4.1.2., 3e § : « Par ailleurs, il ressort des réponses aux questionnaires qu'aucun État n'envisage pour le moment la création d'une personnalité juridique des robots dotés de l'IA en matière pénale, la responsabilité pénale étant de façon unanime basée sur une intention ou une négligence ne pouvant être attachée in fine qu'à une personne physique, et cette question sera donc exclue des discussions, les recherches et débats étant en cours à cet égard. ».*

6 *A. Benoussousan et L. Puigmal, Le droit des robots ? Quelle est l'autonomie de décision d'une machine ? Quelle protection mérite-t-elle ? – Avocats à la Cour : Arch. phil. dr. 2017/1, t. 59, Dalloz, p. 165 à 174.*

7 *A. Benoussousan et L. Puigmal, Le droit des robots ? Quelle est l'autonomie de décision d'une machine ? Quelle protection mérite-t-elle ? – Avocats à la Cour, préc. note 6 : « Enfin, certains refusent d'appliquer le régime de la responsabilité contractuelle à la machine intelligence. Or, les machines intelligentes ont clairement un consentement. Sous réserve d'une délégation de personnalité, il est possible de créer une personnalité juridique singulière. La notion de délégation de personnalité devient donc le critère. ».*

8 *PE, résol., 16 févr. 2017, contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103 (INL)) : « 59. f) la création, à terme, d'une personnalité juridique spécifique aux robots, pour qu'au moins les robots autonomes les plus sophistiqués puissent être considérés comme des personnes électroniques responsables, tenues de réparer tout dommage causé à un tiers ; il serait envisageable de conférer la personnalité électronique à tout robot qui prend des décisions autonomes ou qui interagit de manière indépendante avec des tiers ».*

matière de diffamation et R. 621-2 du même code en matière d'injure trouveraient application. Quel effet dissuasif pour une grande entreprise qui, du fait de propos répréhensibles générés par son logiciel, encourrait une contravention de première classe, soit 38 € au plus *(C. pén., art. 131-13, 1°)* ?
S'agissant des poursuites engagées par les parties civiles, la question du constat d'huissier pourra être soulevée. Afin de démontrer la véracité des propos dénoncés, il est toujours une juste précaution de produire, à l'appui de son action, un procès-verbal de constat d'huissier. Le rôle de l'huissier pourra également consister à démontrer que le propos dénoncé est susceptible d'être généré à chaque fois que le logiciel se trouve confronté à la même question ou situation. Cela peut paraître un détail mais, sur le terrain pratique, il y a lieu de s'en préoccuper. La situation serait moins contraignante en termes de constat si, comme dans le cas de TAY, le logiciel produisait du contenu sur réseau social que l'interface du logiciel lui-même, où rien ne garantit que la réponse litigieuse soit enregistrée et répétée.

7 - **Droit pénal.** - Le droit pénal devrait, pour sa part, répondre au cas de figure d'une IA vigoureusement protégée contre les intrusions. Toute violation du code source pourrait être poursuivie sur le fondement des articles 323-1 et suivants du Code pénal. La protection développée par le concepteur du logiciel pour en éviter toute corruption constitue la condition *sine qua non* d'une conséquence pénale à tout acte malveillant. Même les logiciels fonctionnant au moyen de la technologie du *deep-learning* devraient être en mesure de se protéger contre les attaques par empoisonnement.
L'entreprise victime aura alors la charge de démontrer, avec l'assistance d'un expert désigné par un juge, que le niveau de sécurité de son logiciel contre les intrusions était d'un niveau particulièrement élevé et que, pour parvenir à en violer le code source, les pirates informatiques ont fait preuve d'une ingéniosité plus élevée encore. Il est à craindre que la jurisprudence sanctionnât la légèreté avec laquelle une entreprise mettrait en circulation une technologie d'IA dont les personnes malveillantes pourraient facilement s'emparer. La loi prévoit déjà la répression des entreprises mettant en circulation des logiciels conçus dans le but d'apporter une aide dans la commission d'infraction[9]. La jurisprudence commerciale consacre également depuis 2007 une obligation de sécurité – procédant d'une obligation de résultat – en matière de logiciel informatique visant à protéger les consommateurs contre un dysfonctionnement dans la sauvegarde de données[10], principe réaffirmé en 2022[11].
En énonçant un tel principe, et face à la crainte de perdre le contrôle d'une IA, il est permis d'imaginer un développement législatif visant à sanctionner, en France comme dans le monde, les entreprises qui ne mettraient pas en place les dispositions techniques utiles pour minimiser le risque de détournement à des fins criminelles. Le rapport *« National Cybersecurity Strategy »* publié par la Maison Blanche le 1er mars 2023 s'inscrit dans cette tendance. Le président Joe Biden, signant l'avant-propos dudit rapport, entend en effet conduire une politique coordonnée avec celle des alliés et partenaires de l'Amérique dans le but *« de renforcer la législation des États dont les comportements sont responsables, de tenir les États pour comptables de leurs conduites irresponsables dans le cyberespace, et de lutter contre les réseaux criminels à l'origine des cyberattaques dans le monde »*[12]. Le rapport poursuit en indiquant que *« les entreprises fabricant des logiciels doivent disposer de la liberté d'innover, mais elles doivent aussi être considérées comme responsables en cas de non-respect de l'obligation de sécurité qu'elles ont envers les consommateurs, les entreprises ou les fournisseurs d'infrastructures citriques. (…) Une telle législation devrait empêcher les fabricants et les éditeurs de logiciels de se défaire contractuellement de leur responsabilité, et ainsi établir les plus hauts standards de sécurité pour les logiciels en cas de scenario de hauts risques*[13]. ».

À cette fin, outre le fait que les conditions de la lutte contre les dérives criminelles pourraient être l'objet d'une harmonisation à l'échelle internationale, la notion de bonne foi de l'entreprise victime pourrait être des plus restrictives. Étant acquis que l'IA est une révolution qui présente des dangers tout aussi élevés que le progrès qu'il constitue, les entreprises concernées gagneraient à consacrer une grande partie de leurs efforts à la fiabilité de leur technologie, ce qui ne peut se dispenser d'un niveau de sécurité maximal, au risque d'en être tenus pour responsables.

8 - **Conclusion.** - En tout état de cause, l'IA ne sera jamais qu'un instrument au service de personnes ou d'entreprises qui auront fait le choix délibéré de l'utiliser. La consécration de l'autonomie totale des logiciels d'IA paraissant désormais peu plausible, l'avenir ne se bâtira avec l'idée qu'un acte de cybercriminalité puisse être détaché d'une responsabilité humaine. L'humanisme a consacré la personne humaine comme centralité de l'univers ; l'IA, qu'il s'agisse de ChatGPT ou d'un logiciel plus évoluée encore, ne nous fera pas changer de civilisation. ■

9 *C. pén., art. 323-3-1 : « Le fait, sans motif légitime, notamment de recherche ou de sécurité informatique, d'importer, de détenir, d'offrir, de céder ou de mettre à disposition un équipement, un instrument, un programme informatique ou toute donnée conçus ou spécialement adaptés pour commettre une ou plusieurs des infractions prévues par les articles 323-1 à 323-3 est puni des peines prévues respectivement pour l'infraction elle-même ou pour l'infraction la plus sévèrement réprimée. ».*

10 *Cass. com., 11 déc. 2007, n° 04-20.782 : JurisData n° 2007-041953 : « (…) ayant relevé que le logiciel comportait une anomalie dans l'écriture du programme de sorte que lorsqu'il a été réinstallé après la panne ayant affecté le serveur, les bases de données reçues des clients et préalablement importées et les développements spécifiques n'étaient pas sauvegardés puis retenu que cette anomalie était imputable à l'auteur du logiciel, la cour d'appel a, sans méconnaître la loi des parties, légalement justifié sa décision (…).) ».*

11 *Cass. com., 1er juin 2022, n° 20-19.476, D, Sté Fsmax System SL c/ Sté Compass Group France : JurisData n° 2022-009530 : « (…) Il retient ensuite que les lots 6, 7 et 8 ont fait l'objet de plusieurs livraisons, dont aucune n'a donné lieu à une recette par la société Compass en raison d'anomalies bloquantes non résolues, ce qui démontre le défaut de qualité du « livrable » de la société Fsmax, que celle-ci n'a pas été en capacité de résoudre dans des délais raisonnables pour satisfaire à son obligation de résultat. Il retient encore que la société Fsmax ne peut opposer le défaut de respect de la procédure contractuelle de recette de la part de son client, cependant que, professionnelle de l'informatique soumise à une obligation de résultat, elle ne justifie pas avoir élaboré des spécifications fonctionnelles détaillées en réponse aux besoins exprimés par celui-ci, malgré ses demandes. (…) ».*

12 *National Cybersecurity Strategy, The White House – mars 2023 : « We will collaborate with our allies and partners to strengthen norms of responsible state behavior, hold countries accountable for irresponsible behavior in cyberspace, and disrupt the networks of criminals behind dangerous cyberattacks around the globe ».*

13 *National Cybersecurity Strategy, The White House – mars 2023 : « Responsibility must be placed on the stakeholders most capable of taking action to prevent bad outcomes, not on the end-users that often bear the consequences of insecure software nor on the open-source developer of a component that is integrated into a commercial product. (…) Any such legislation should prevent manufacturers and software publishers with market power from fully disclaiming liability by contract, and establish higher standards of care for software in specific high-risk scenarios. ».*