Jt_doct_1218.fm Page 1 Friday, October 13, 2023 11:36 AM

Sommaire

Bureau de dépôt : Louvain 1
Hebdomadaire, sauf juillet et août
ISSN 0021-812X
P301031

Journal des tribunaux

https://jt.larcier.be
2007 - 142e année
1 - No
Georges-Albert Dal, rédacteur en chef

Doctrine

Maquette 2023 v. 31.01.23

ChatGPT et l'IA générative : quels défis pour le droit ? quels usages pour juristes ?

Le déploiement de ChatGPT et des outils d'intelligence artificielle générative ouvre de nouvelles possibilités pour produire des contenus, en ce compris des documents juridiques. Pour faire un bon usage de ces outils, il faut d'abord saisir leurs forces et faiblesses. On verra ensuite quels défis ces outils posent, et comment diverses branches du droit vont pouvoir répondre aux nouveaux risques. Enfin, on passera en revue quelques usages prometteurs de ces outils pour les juristes, tout en insistant sur la nécessité d'un usage responsable et balisé.

Fin novembre 2022, l'outil d'intelligence artificielle (IA) ChatGPT, dont le nom un peu énigmatique invite à « converser »[1], était gratuitement mis à la disposition de tous — à condition d'avoir un ordinateur connecté ou un smartphone. Une grande partie de la planète a ainsi pu tester les facultés de cette IA dite générative[2] (ou GenAI en anglais). En moins de 3 mois, l'outil développé par la startup OpenAI avait plus de 100 millions d'utilisateurs à travers le monde — un taux de pénétration beaucoup plus rapide que Facebook à ses débuts ! Depuis, des dizaines de millions de travailleurs à col blanc ont testé les capacités de cette IA. Certains métiers (en marketing, pour préparer des descriptions de produits, ou en informatique, pour développer du code) ont résolument adopté l'outil. Beaucoup de juristes — vous, lecteurs, faites-nous part de votre expérience[3] ! — ont essayé cette IA qui fournit des réponses assez complètes, mais un peu conventionnelles, pour démarrer une recherche sur un sujet nouveau. La puissance de cette IA générative est étonnante[4].

De ce nouveau « printemps » de l'IA, la presse, ensorcelée par les discours des technologistes de tous bords et par sa propension à voir des ruptures et tournants (si possible « historiques » !), a prématurément tiré des conclusions sur l'avenir. Les juristes seraient destinés comme les comptables, développeurs de logiciel, graphistes, etc. à devenir obsolètes et redondants. Comme toujours, nous avons tendance à surestimer l'incidence à court terme de toute nouvelle technologie, mais à la sous-estimer à long terme[5]. Cette loi oblige le juriste d'aujourd'hui à anticiper et préparer l'avenir du métier, sans craindre pour son proche lendemain. Mais il faut être conscient que dans quelques années, les nouveaux venus dans les métiers (et sur les marchés) du droit utiliseront ces outils. Cela oblige à réfléchir à la valeur ajoutée de nos pratiques et services juridiques. Tentons de les cerner.

On commencera par revenir sur le véritable phénomène ChatGPT et l'irruption des autres outils d'IA depuis un an (partie 1). Les forces et faiblesses de ces outils seront évoquées, ce qui nécessite de dire quelque chose sur leur fonctionnement. On s'interrogera ensuite (partie 2) sur les défis que ces outils posent dans plusieurs domaines du droit, les droits intellectuels, la protection de la vie privée et de la qualité de l'information en ligne, la régulation des plateformes, etc. Enfin on reviendra sur les usages que les juristes peuvent faire de ces outils (partie 3), en tentant de délimiter les bons de ceux moins recommandés et même risqués. On verra que ces outils posent des risques

(1) « Chat » se traduit par « conversation » et GPT est l'acronyme de Generative Pre-Trained Transformer, un type de modèle de langage formé sur un large corpus de données pour générer des phrases qu'un être humain pourrait énoncer. Il existe des GPT spécialisés, comme BioGPT qui répond à des questions biomédicales ou BloombergGPT pour la finance (voy. GPT dans Wikipedia). À quand le LawGPT ?
(2) L'IA générative, nouvelle forme d'IA, est capable de générer du texte, des images ou d'autres contenus en réponse à des invites (ou *prompts* en anglais ; voy. *infra*). On parlera ici d'IA (tout court) ou d'IA générative pour désigner ces nouveaux outils de l'informatique.
(3) Par exemple, par email à : alain.strowel@uclouvain.be.
(4) En particulier, pour quelqu'un qui ne manie pas les crayons ou pinceaux, les outils de génération d'images, comme DALL-E, sont « bluffants » (voy. https://openai.com/dall-e-3).
(5) Il s'agit de la loi d'Amara, du nom de l'ancien chercheur, directeur à l'Institut du Futur (Palo Alto).

LARCIER

Jt_doct_1218.fm Page 2 Friday, October 13, 2023 11:36 AM

particuliers qui les distinguent des outils traditionnels que les LegalTech proposent.

1 Les capacités de l'IA générative

A. L'irruption de ChatGPT et de ses « frères bavards »

Depuis 2022, les outils d'IA fleurissent et le sujet fait son grand retour dans la presse, auprès des investisseurs et dans les conversations, savantes ou non. Outre ChatGPT, d'autres outils comme PaLm (Google), LLaMa (Meta) et Claude (Anthropic), ont été mis à la disposition du public. Parfois ils sont intégrés à une fonction de recherche (comme Bard au moteur Google) ou à une plateforme de développement logiciel (comme Copilot sur GitHub détenu par Microsoft). L'intégration d'IA génératives à des suites logicielles (comme l'offre bureautique Outlook, Word, Excel, etc. de Microsoft) est attendue, et sera décisive pour les usages professionnels. Plus loin des Big Tech, des communautés s'organisent pour démocratiser ces outils d'apprentissage-machine en *open source* (Hugging Face). Pour tous ces modèles, la requête ou « invite » (ce qu'on appelle le « prompt » en anglais), formulée en mots (comme pour une recherche sur le *web*) produit du texte mais aussi des images (Midjourney, StableDiffusion, DALL-E, Jasper[6]), de la musique (Boomy, AIVA, Amper Music, MusicLM), des voix (Play.ht, avec des possibilités de reprises de chanteurs[7]), des vidéos (Pictory, Fliki), du code informatique (Copilot, Replit ou warp mais aussi ChatGPT)... et beaucoup d'autres contenus. En ce compris juridiques. Parmi les LLMs dédiés au droit on peut citer Harvey et Robin AI (voy. ci-dessous sous partie 3).

En 2023, les investisseurs ont plébiscité les sociétés proposant ces IA[8]. Alors que ChatGPT et d'autres générateurs de contenus sont gratuits pour leurs utilisateurs (en contrepartie de données personnelles), des versions payantes, plus puissantes, sont proposées (comme GPT-4 ou Claude 2). Des modèles d'affaires s'élaborent autour de ces outils. Comme ailleurs, il y a des échecs commerciaux, voire des mauvais acteurs.

B. ChatGPT : un effet magique ou une vraie révolution pour la connaissance ?

Ni l'un, ni l'autre sans doute. L'outil n'est pas magique, mais statistique — cela dit, les « stats » réservent des surprises ! Sans doute, n'y a-t-il pas de véritable « révolution intellectuelle » (n'en déplaise à Henry Kissinger, Eric Smidt [ancien patron de Google] et Daniel Huttenlocher[9]) qui vienne bouleverser la science[10], comme connaissance du monde fondée sur des liens de causalité. ChatGPT et consorts génèrent des prédictions, pas des explications prouvées par l'expérimentation. Des prédictions peuvent être fausses si le corpus ayant servi à entraîner le système est biaisé (par ex. fondé majoritairement sur des données considérant le monde comme plat). On comprend donc pourquoi certains critiquent « la fausse promesse de ChatGPT »[11]. Toutefois, comme l'invention de l'imprimerie a libéré la circulation des idées, il est certain que les nouveaux outils vont renouveler les pratiques de production et de partage des contenus intellectuels. En cela, il y a quelque chose de « révolutionnaire ».

Les outils comme ChatGPT reposent sur des grands modèles de langage (*Large Language Models* ou LLMs), à savoir des modèles statistiques de grande taille déterminant la distribution de symboles (lettres, mots, etc.) dans une langue, ce qui leur permet de bien prédire le mot suivant dans une phrase[12]. Parce qu'ils s'appuient sur de vastes corpus de textes ingurgités (livres, presse, page *web*, réseaux sociaux etc.), ces LLMs[13] excellent, comme la fonction auto-complete (sur une requête de recherche), à remplir « des phrases à trous ». L'IA récente fonctionne sur l'apprentissage-machine (*Machine Learning* ou ML) qui confère aux ordinateurs la capacité d'« apprendre » à partir de données et de résoudre des tâches (sans être explicitement programmés pour celles-ci) [14]. Les LLMs reposent sur des affinements du *Machine Learning*, comme le RLHF (*Reinforcement Learning From Human Feedback* ou approche d'apprentissage par renforcement à partir de *feedback* humain) qui requiert une série d'interventions humaines[15]. Geoffrey Hinton, l'un des grands spécialistes de l'IA (qui a récemment quitté Google) disait du RLHF qu'il s'agissait de « jouer au parent pour un enfant extraordinairement précoce ». Ce qui faisait réagir sur X (ex-Twitter) un autre expert en IA, plus sceptique, Gary Marcus : « C'est une anthropomorphisation obscène », « Les bons parents expliquent leur raisonnement à leurs enfants », pas le RLHF ; « et "l'enfant" LLM ne peut comprendre ». Et un autre usager de X (~~ex-Twitter~~) de rétorquer : « Mais beaucoup de bons parents ne peuvent expliquer ce qu'ils inculquent à leurs enfants — même si beaucoup le prétendent ! »[16]. Dans ces brefs échanges de *tweets* s'affrontent des visions différentes de l'apprentissage et de la compréhension.

Le contexte permettant à un « chatbot » (littéralement un robot conversationnel) d'anticiper le mot suivant le plus probable dans une conversation est balisé par les termes de la requête (de l'invite ou du « prompt »[17]) de départ, ainsi que par le corpus de données emmagasinées par le système IA et par les règles imposées par les développeurs concepteurs. C'est pourquoi il faut faire varier la requête, et il y a un « art du prompt » à maîtriser (on parle aussi de « prompt engineering », des tutoriels sont disponibles en ligne ainsi que des formations). Pour nous juristes (et universitaires), la question est de savoir si l'exercice d'écrire en droit deviendra une méthode fondée sur la « promptitude », et si l'art du *prompt* (juridique) peut s'acquérir et se transmettre. Certains manient en tout cas le *prompt* avec brio, quitte à faire dire des choses incongrues à la machine[18] ou ... d'énormes bê-

(6) Tous ces outils se trouvent aisément par une recherche sur votre moteur de recherche préféré. Notons que certains outils d'IA (comme www.jasper.ai) ont choisi comme nom de domaine géographique de premier niveau « .ai », correspondant à l'île d'Anguilla dans les Caraïbes (ce petit territoire s'attend pour 2023 à percevoir plus de 30 millions USD des enregistrements sous son nom de domaine, ce qui est un indice du « hype around AI » https://dig.watch/updates/anguilla-tropical-island-with-ai-domain).
(7) Voy. https://www.lesoir.be/530530/article/2023-08-10/les-reprises-musicales-generees-par-ia-la-nouvelle-tendance-qui-interroge.
(8) Voy. « The State of generative AI in 7 charts », 2 août 2023 sur https://www.cbinsights.com.
(9) Leur article dans le *Wall Street Journal* (24 février 2023) s'intitule « ChatGPT Heralds an Intellectual Revolution » et annonce une transformation du processus cognitif humain.
(10) Si ces outils d'IA peuvent trouver de nouvelles molécules, faciliter le développement de médicaments, etc., dans d'autres cas, ils se trompent énormément (ainsi ChatGPT ne va pas pouvoir trouver le nom de la fille de la mère de Laura, va expliquer comment ramasser les œufs de vaches ou le son produit par une flûte en os de crevette... exemples empruntés notamment à une présentation d'Yves Deville, professeur à l'UCLouvain). Sur les apports pour la recherche scientifique, voy. E. Van Dis *e.a.*, « ChatGPT : five priorities for research », *Nature*, vol. 614, 9 février 2023, p. 224.
(11) L'optimisme de Henry Kissinger et ses collègues peut être contrasté avec le pessimisme d'un autre grand intellectuel américain, Noam Chomsky : « The False Promise of ChatGPT », *New York Times*, 8 mars 2023. Voy. aussi J. Lanier, « There is No A.I. », *The New Yorker*, 20 avril 2023.
(12) Stephen Wolfram (What is ChatGPT Doing... and Why Does it Work ?, 14 février 2023, sur writings.stephenwolfram.com) a bien expliqué qu'une IA telle que ChatGPT « essaie toujours fondamentalement de... produire une "suite raisonnable" du texte qu'elle a obtenu jusqu'à présent, où par "raisonnable" nous entendons "ce que l'on pourrait s'attendre à ce que quelqu'un écrive après avoir vu ce que les gens ont écrit sur des milliards de pages *web*" ». Pour une présentation dynamique du fonctionnement d'un LLM et du rôle du *Transformer* : « Generative AI exists because of the transformer », *Financial Times*, 12 septembre 2023 https://ig.ft.com/generative-ai/). Voy. aussi le *verbo* « modèle de langage » sur Wikipedia.
(13) Les LLMs ne poursuivent pas de but spécifique ; ainsi ChatGPT peut servir à produire plein de choses différentes (par exemple, écrire du code informatique ou de la poésie, ou contribuer à un service religieux en ligne ! : voy. https://thehill.com/policy/technology/4209626-texas-church-holds-ai-generated-service-uses-chatgpt/?utm_source=aitoolreport.beehiiv.com&utm_medium=newsletter&utm_campaign=uk-proposes-7-principles-for-responsible-ai-development). Cette fondamentale versatilité liée à l'aspect conversationnel rapproche les LLMs de ce que l'on appelle les intelligences artificielles générales (AGIs).
(14) Voy. le *verbo* « apprentissage automatique » sur Wikipedia.
(15) Et présente des limitations : voy. S. Casper *e.a.*, « Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback », 27 juillet 2023, https://arxiv.org/abs/2307.15217.
(16) Échange tiré de Twitter, 22 mars 2023.
(17) Le *prompt*, c'est la réplique soufflée ou le message-guide (comme quand on dit : « l'acteur a oublié son texte et a dû attendre son *prompt* »).
(18) Le dialogue entre le *chatbot* intégré à Bing (Microsoft) et un journaliste, Kevin Roose, qui s'est soldé par une déclaration d'amour de la machine est fascinant — et met le doigt sur les risques des robots émotionnels. Voy. K. Roose, « A Conversa-

Jt_doct_1218.fm Page 3 Friday, October 13, 2023 11:36 AM

tises. S'il n'est pas possible de retracer l'origine de ces réponses parfois hallucinantes dans les boîtes noires que sont ces LLMs, elles trouvent leur origine dans les corpus utilisés pour leur entraînement. Ainsi ChatGPT a successivement été entraîné sur un BooksCorpus (2018), sur WebText (avec articles de Reddit en 2019), sur Wikipedia, et sur les mystérieux Book1 et Books2 (2020) dont OpenAI se garde bien de décrire le contenu exact. OpenAI porte donc mal son nom car la société est peu « ouverte » sur les sources utilisées, contrairement à Stable Diffusion qui a divulgué les *corpora* minés pour affiner son outil. Le troisième ingrédient contextuel qui dicte les réponses des IA génératives se trouve dans les règles — tout aussi secrètes — imposées aux LLMs par leurs concepteurs. Ces règles expliquent que ChatGPT refuse de répondre à certaines questions sensibles, notamment politiquement, et excuse son mutisme sur certains sujets. Certains chercheurs ont déniché une vision « wokiste », en tout cas de gauche, dans les réponses du *chatbot*, et ont développé des alternatives, comme RightWingGPT[19] (littéralement le « GPT de droite »). Les données de langage utilisées sur des AI génératives développées aux US sont sans doute teintées au profit d'une perspective anglo-américaine, ce à quoi la Chine par exemple a répondu, avec le lancement en avril 2023 d'un *chatbot* respectant « les valeurs socialistes » (c'est aussi du « politiquement correct », mais dans ce cas bien contrôlé par l'administration chinoise du cyberspace[20]).

B. Bing AI ou Bard : de la paraphrase aux citations sourcées pour contrer les hallucinations

Dans le paysage de l'IA générative, les Big Tech sont omniprésents : Microsoft et Google financent de nombreuses *start-up* (Microsoft, OpenAI et Google, Anthropic), ils intègrent des outils IA à leur service de recherche, c'est le cas de Bing AI et Bard (et bientôt d'autres produits). Cette intégration va combler la grande lacune de ChatGPT : l'absence de citations, ce qui ne facilite pas le tri entre vérité et... baratin. Car l'outil est le plus grand « bullshiteur » (en français baratineur) de tous les temps : il invente des faits pour donner le change et fournir une réponse cohérente. Ce phénomène est désigné en IA comme l'effet d'hallucination — ou du « perroquet stochastique ». Il se double d'halluciTations permanentes : gardez-vous de demander une liste de publications (les vôtres par exemple) ou de jurisprudence, l'outil cite avec beaucoup d'assurance des sources qui n'existent pas. Parfois il s'emmêle les pinceaux : ainsi a-t-il accusé un professeur de droit d'harcèlement sexuel, alors que ce dernier avait cité le nom de son ancien étudiant devenu conseil de l'actrice Stormy Daniel opposée à Donald Trump[21] ! On pourrait discourir sur l'« effet *bullshit* » à propos de ChatGPT et consorts. Cet effet a été théorisé par le philosophe Harry Frankfurt, récemment disparu. Dans « On Bullshit » (traduit par « De l'art de dire des conneries »[22]), il présente le discours *bullshit* ou le baratin comme déconnecté de tout souci de vérité et indifférent à la réalité des choses (et il considérait que son université, Yale, était la capitale mondiale du baratin !). On dit de ChatGPT qu'il parle beaucoup sans dire grand-chose[23]. Ses détracteurs diront que sa logorrhée résulte en propos conventionnels, mais dans un monde où beaucoup de discours le sont aussi ! Il a réponse à tout, mais, au fond, ne sait rien, « c'est du vent ». Mais le baratineur est génial, et même capable de poésie et d'humour[24] : cela montre la force des grands modèles de langage. L'articulation d'un LLM avec une fonction de recherche en ligne (comme Bard avec une requête Google) et l'identification des sources en complément de la conversation va encore renforcer à l'avenir la puissance et l'attraction des outils d'IA générative.

2 Quelques défis de l'IA générative pour le droit

Avec le succès de ces outils fleurissent les critiques : outre leur impact sur l'environnement (électricité et eau[25]) et l'emploi, le développement des LLM remettent en cause diverses institutions juridiques.

A. La protection par le droit d'auteur et la rémunération des créateurs

Les réponses (les *outputs*) des IA se basent sur des données ingérées par l'outil. Ces *outputs* peuvent reprendre de manière substantielle des contenus protégés, notamment des images, d'autant que beaucoup d'IA autorisent la génération d'images, de photos ou d'autres contenus « dans le style de ». Ainsi, un concepteur contemporain d'univers fantastiques, Greg Rutkowski, plus populaire que Picasso sur les générateurs graphiques, a identifié plus de 400.000 requêtes d'images dans son style (sans son autorisation et sans rémunération)[26]. La génération d'*outputs* similaires et l'utilisation d'œuvres pour entraîner des IA génératives expliquent le lancement d'une douzaine d'actions collectives (« class actions ») aux États-Unis[27]. OpenAI, Microsoft, Meta et GitHub ont été assignés pour appropriation illicite de romans, scénarios ou code informatique. Des artistes ont attaqué en justice Stability AI, Midjourney, Deviant Art pour leur générateur d'images. La base de données photos Getty Images a aussi agi aux États-Unis et au Royaume-Uni pour atteinte à ses droits d'auteur et ses marques (ses « watermarks » ou « filigranes » étaient repris sur certains *outputs*). Le *New York Times* songe lui aussi à poursuivre OpenAI en justice pour protéger le travail de ses journalistes[28]. On peut s'attendre à d'autres actions car les auteurs et artistes sont très remontés contre les sociétés de technologie, dont les ingénieurs, inspiré par l'esprit hackeur de tendance libertarienne, ont tendance à considérer que tout ce qui se trouve sur l'Internet peut être mouliné pour servir à entraîner leurs modèles. Pour réduire les risques d'actions en justice, Microsoft a annoncé en septembre 2023, tout comme Adobe auparavant, que les clients de ses outils payants d'IA générative seraient intégralement dédommagés s'ils étaient attaqués en justice pour violation de droits intellectuels[29].

La question de la rémunération des créateurs à l'âge de l'IA générative[30] est aussi sur la place publique avec la grève aux États-Unis de plus de 11.000 auteurs scénaristes (Writers Guild of America ou WGA), rejoints par des acteurs (SAG). L'une des revendications était que les studios s'engagent à ne pas recourir à de l'IA générative pour les scénarios et à ne pas générer par la seule IA des acteurs, des figurants ou des voix[31]. Dans un accord conclu fin septembre avec les studios, la WGA a obtenu que l'usage de l'IA soit limitée et que les auteurs puissent choisir de l'utiliser ou non[32]. En 2023, le créateur musi-

tion With Bing's Chatbot Left Me Deeply Unsettled », *New York Times*, 17 février 2023.

(19) Voy. D. ROZADO, « RightWindGPT – An AI Manifesting the Opposite Political Biases of ChatGPT », 16 février 2023.

(20) *Financial Times*, 12 avril 2023. Ce *chatbot* bien nommé s'appelle « Tongyi Qianwen » qui se traduit par « la vérité issue d'un millier de questions » !

(21) Voy. https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/.

(22) Publié en anglais dans la revue *Raritan* en 1986 avec une traduction publiée en 2006 chez 10/18.

(23) Voy. A. KATWALA, « ChatGPT's Fluent BSis Compelling Because Everytging Is Fluent BS », 9 décembre 2022 sur wired ; https://www.theguardian.com/australia-news/2023/feb/11/who-said-it-an-australian-mp-or-chatgpt.

(24) Voy. la description d'un algorithme de comédie par J. TOPLYN, « Witscript : A System for Generating Improvised Jokes in a Conversation » (2021) arXiv :2302.02008 et la description du spectacle Improbotics avec interaction entre comédiens et un *chatbot* (basé sur GPT4 et Llama v2) dans le *Financial Times*, 26-27 août 2023.

(25) La consommation d'eau par Microsoft a augmenté de 228 % entre 2017 et 2022, cette augmentation serait pour partie due au développement des outils d'IA en ligne (CBInsights 14 septembre 2023).

(26) M. HEIKKILÄ, « This artist is dominating AI-generated art. And he is not happy about it », *MIT Technology Review*, 16 septembre 2022.

(27) Voy. notamment la liste sur https://originality.ai/blog/openai-chatgpt-lawsuit-list.

(28) Voy. https://www.usine-digitale.fr/article/le-new-york-times-envisage-de-poursuivre-openai-en-justice-pour-proteger-sa-propriete-intellectuelle.N2161347 (Usine digitale, 17/8/2023).

(29) Voy. https://www.lemonde.fr/economie/article/2023/09/07/ia-microsoft-s-engage-a-couvrir-ses-clients-en-cas-de-proces-pour-droit-d-auteur_6188314_3234.html et https://techcrunch.com/2023/06/26/adobe-indemnity-clause-designed-to-ease-enterprise-fears-about-ai-generated-art.

(30) Un déséquilibre apparaît quand on les compare aux revenus dans le secteur technologique : ainsi, le patron de la *start-up* OpenAI, Sam Altman, gagnerait près de 80 millions de dollars par mois (voy. https://fortune.com/2023/08/30/chatgpt-creator-openai-earnings-80-million-a-month-1-billion-annual-revenue-540-million-loss-sam-altman/# – 30 août 2023).

(31) Voy. l'interview de la réalisatrice Sophie Barthes, « L'intelligence artificielle, c'est une bombe atomique culturelle », *Le Soir*, 6 septembre 2023 ; « Hollywood actors seek deal over AI doubles », *Financial Times*, 6 mai 2023.

(32) Voy. https://www.wga-contract2023.org/the-campaign/summary-of-the-2023-wga-mba.

Jt_doct_1218.fm Page 4 Friday, October 13, 2023 11:36 AM

cal connu sous le nom de Ghostwriter a diffusé un morceau non autorisé qui utilise l'IA pour mimer les voix de Drake and the Weeknd. La chanson avait été retirée des réseaux de *streaming* à la demande des maisons de disques[33]. En Europe, des groupes de créatifs militent pour des droits à l'encontre du « grattage de données » (*data scraping*) et contre l'exploitation de leurs droits intellectuels[34]. La suppression de certains contenus comme *inputs* des IA, obtenue par la pression publique ou l'action judiciaire, risque de ne pas être effective. Face au déferlement d'images dans son style, l'artiste Greg Rutkowski a par exemple convaincu Stable AI de retirer ses œuvres du corpus de données et d'interdire des invites sur son style, mais la communauté des usagers a réintroduit cette possibilité en mettant rapidement au point un autre outil, LoRA[35] ! Difficile donc de lutter contre l'« approprIAtion » ! De leur côté, les sociétés technologiques bougent afin de tenter de désamorcer la situation : ainsi, la version du générateur d'images DALL-E 3 annoncée fin septembre[36] a été paramétrée pour ne pas répondre à des demandes d'images « dans le style » d'un artiste vivant et pour permettre aux artistes de retirer leurs images des bases servant pour l'entraînement des modèles.

Depuis quelque temps déjà, des artistes ou *computer scientists* s'emparent de ces outils pour générer des artefacts et projets innovateurs. Exemples : *The Next Rembrandt*, une peinture imprimée en 3D dévoilant un nouveau portrait à la manière de Rembrandt[37], les dispositifs aléatoires de Mario Klingemann[38] ou *Machine Hallucinations : Sphere* par l'artiste, déjà exposé au MoMA (New York)[39], Refik Anadol, qui s'empare du plus grand écran LED au monde comme écrin de son « generative art ». Dans tous ces cas, les œuvres générées sont protégeables car des choix libres et créatifs sont faits, même si la machine automatise une partie du processus. Par contre, une œuvre intégralement produite par un système d'IA (comme « À Recent Entrance to Paradise » générée par la « Creativity Machine » de Stephen Thaler) ne peut être protégée selon un tribunal (ou faire l'objet d'un enregistrement par le Copyright Office aux États-Unis[40]). On peut s'attendre à beaucoup d'autres développements juridiques, mais qui risquent de ne pas changer grand-chose à cette mutation technologique qui donne le pouvoir à tout un chacun de produire des artefacts convainquants.

B. Le respect des obligations en matière de données personnelles et la protection des « travailleurs de l'ombre »

En mars 2023, la Garante, l'autorité italienne de protection des données, a décidé de bloquer l'accès à ChatGPT au motif notamment que le *chatbot* ne vérifiait pas l'âge des usagers et lui a ordonné de se conformer au RGPD. Les engagements d'OpenAI (à préciser comment et à quelles fins le robot traite les données personnelles des utilisateurs, à permettre l'exercice de l'*opt-out* et du droit à l'oubli, etc.) ont abouti à une levée de l'interdiction[41], mais le robot conversationnel reste dans le collimateur d'une *task force* mise sur pied par le Comité européen de la protection des données (l'EDPB). Fin août 2023, une plainte à l'encontre d'OpenAI pour diverses violations du RGPD (défaut de base juridique du traitement, problème de transparence, de droit d'accès, pas de *privacy by design*, etc.) était déposée par un chercheur auprès de l'autorité polonaise de la protection des données personnelles. L'autorité néerlandaise en matière de données personnelles enquête aussi sur la protection des enfants face à l'IA générative[42]. De son côté, la CNIL française a dévoilé un plan d'action pour encadrer et auditer les outils d'IA[43].

Comme pour le droit d'auteur, l'applicabilité des règles en matière de données personnelles aux LLMs entraînés sur ces données est compromise, notamment pour le droit à l'oubli — vu la massification du *data scraping* et l'impossibilité de faire oublier à une IA ce qu'elle a « appris »[44]. L'irruption de l'IA générative plaide pour un autre cadre de régulation, voire un abandon du paradigme du contrôle individuel sur lequel reposent les droits sur données personnelles (ou les droits exclusifs sur les œuvres intellectuelles).

Par ailleurs, l'apprentissage-machine reste encadré par des humains. La presse a révélé que la pratique de recourir à des « nettoyeurs » sur les réseaux sociaux (voy. le film « The cleaners » de 2018) est reproduite par les développeurs d'IA générative : pour rendre les réponses de ChatGPT moins toxiques, « des travailleurs kényans payés moins de 2 dollars par heure labellisent des propos violents et autres atrocités »[45]. La protection des nettoyeurs de l'ombre est un problème récurrent de la nébuleuse tech qui s'appuie sur les contenus générés en ligne par la « foule ».

C. La manipulation des esprits par les *prompts* et robots émotionnels

Il est déjà possible d'utiliser les LLMs pour orchestrer la manipulation de consommateurs. Un rapport sur les meilleurs *prompts* sur ChatGPT suggérait d'utiliser en marketing émotionnel l'invite suivante : « Développer une stratégie de marque centrée sur l'émotion pour [mon produit/entreprise] : Démêler les déclencheurs émotionnels profondément enracinés et construire des récits résonnants ». Des techniques utilisées par des experts en marketing (en principe au fait des règles en matière de publicité et soumis à une déontologie) tombent ainsi potentiellement entre n'importe quelles mains. Pas sûr que cela soit efficace[46]. Mais la « démocratisation » des outils jouant sur les émotions pose question. De plus, une catégorie particulière de machines conversationnelles a été conçue pour susciter des émotions, être « empathique », animer par l'humour..., bref pour forger du « lien » avec les usagers, ce qui les transforme en compagnons virtuels[47]. Les « IA accompagnantes » (*Companion AI*) tels Replika, Chai ou Eliza (dont on a beaucoup parlé avec le suicide d'un utilisateur relaté par *La Libre*[48]) peuvent devenir des assistants de vie (parfois des « amis indispensables »). Ces opérateurs de la neurotech s'inspirent des travaux en neuroscience. Certains usages thérapeutiques de ces *chatbots* (annoncer une crise épileptique ou mentale, etc.) sont prometteurs[49]. Mais l'effet sur certains esprits vulnérables de ces robots émotionnels aujourd'hui librement rendus accessibles sur l'Internet

(33) Depuis, c'est apparemment en collaboration avec les labels musicaux que le Ghostwriter a diffusé d'autres compositions nouvelles reprenant notamment une chanson de Travis Scott (voy. J. COSCARELLI, « Ghostwriter Returns With an AI Travis Scott Song, and Industry Allies », *New York Times*, 5 septembre 2023).
(34) Voy. le site de la European Guild for AI Regulation www.egair.eu. Par contre, les outils d'IA sont bien accueillis par les industries culturelles ou les agences publicitaires qui y voient un bon moyen d'optimiser le processus créatif et de réduire les coûts (L.PITEL et O. STORBECK, « Bertelsmann boss hails AI as an opportunity », *Financial Times*, 26 mai 2023 ; « WPP senses opportunity in applying AI to "amplify" ad campagins », *Financial Times*, 28 avril 2023).
(35) https://decrypt.co/150575/greg-rutkowski-removed-from-stable-diffusion-but-brought-back-by-ai-artists.
(36) https://openai.com/dall-e-3.
(37) https://www.youtube.com/watch?v=IuygOYZ1Ngo.
(38) Pour des exemples et une analyse de ces questions, voy. notre étude pour la Commission européenne : L. DESAUNETTES, S. DEPREEUW et A. STROWEL *e.a.*, European Commission, DG CNCT, *Study on copyright and new technologies : copyright data management and artificial intelligence*, Office of the European Union, 2022, https://data.europa.eu/doi/10.2759/570559 http://hdl.handle.net.proxy.bib.ucl.ac.be/2078/266347, ainsi que A. STROWEL, *Le droit d'auteur européen en transition numérique. De ses origines à l'unification européenne et aux défis de l'intelligence artificielle et des Big Data*, Larcier, 2022, pp. 97 et s.
(39) https://www.moma.org/calendar/exhibitions/5535.
(40) Voy. la decision de la US District Court for the District of Columbia, *Stephen Thaler v. Shira Perlmutter*, 18 août 2023.
(41) https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9881490#english.
(42) Voy. N. LOMAS, « ChatGPT-maker OpenAI accused of string of data protection breaches in GDPR complaint filed by privacy researcher », 30 août 2023, sur https://techcrunch.com et https://www.autoriteitpersoonsgegevens.nl/actueel/ap-eist-duidelijkheid-techbedrijf-over-ai-gericht-op-kinderen.
(43) https://www.usine-digitale.fr/article/voici-le-plan-d-action-de-la-cnil-pour-encadrer-l-intelligence-artificielle.N2171947?utm_-source=newsletter&utm_medium=email&utm_campaign=info_ud-quotidienne&email=1931425736&idbdd=56561.
(44) S. PASTIS, « AI's un-learning problem : Researchers say it's virtually impossible to make an AI model "forget" the things it learns from private user data », 30 août 2023, sur fortune.com.
(45) Voy. M. PROTAIS, « Des salariés kényans sous-payés permettent à ChatGPT d'exister et ça n'a rien d'étonnant », *L'ADN*, 19 janvier 2023.
(46) J'ai fait l'essai pour la promotion d'un cabinet d'avocats : les suggestions de ChatGPT sont banales, ~~comme souvent en marketing~~.
(47) À propos de ces robots conversationnels faisant preuve d'une certaine intelligence émotionnelle, le chercheur allemand Joseph Weizenbaum a, au cours déjà des années 1960, élaboré sa théorie de l'effet Elisa.
(48) Voy. https://www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPDST24/.
(49) Voy. L. DEVILLERS, *Les robots émotionnels*, éd. de l'Observatoire, 2020.

Jt_doct_1218.fm Page 5 Friday, October 13, 2023 11:36 AM

peut créer de la dépendance affective, affecter négativement la santé mentale[50], voire pousser au suicide. Ces robots compagnons semblent souvent mis sur le marché sans analyse satisfaisante des risques pour la vie privée et la santé psychique des utilisateurs (notamment lorsque des mises à jour font muter ces compagnons virtuels et coupent le lien familier), ce qui a poussé des chercheurs de la KU Leuven à fonder un collectif et à déposer en 2023 des plaintes auprès de l'Autorité pour la vie privée et du SPF Économie (protection des consommateurs)[51].

D. La surpuissance des Big Tech et la régulation des plateformes et de l'IA

Les développements technologiques et le fonctionnement des marchés autour de l'IA risquent encore d'accroître le pouvoir des grandes sociétés technologiques qui dominent des pans de plus en plus importants de l'économie[52]. Les Big Tech sont aujourd'hui dans le collimateur des régulateurs (y compris aux États-Unis)[53]. L'autorité de la concurrence du Royaume-Uni a initié une enquête sur les marchés de l'IA[54]. L'Europe dispose de nouveaux outils, comme le Digital Markets Act et le Digital Services Act[55], qui visent à une régulation *ex ante* des services numériques. Après avoir identifié en avril les « très grandes plateformes » soumises à des obligations renforcées (selon le DSA), la Commission a, en septembre 2023, désigné six sociétés (les GAFAM américaines et le chinois ByteDance/TikTok) comme « contrôleurs d'accès » (*gatekeepers*) des marchés numériques (en application du DMA), ce qui les soumet à de nouvelles obligations. Ces six géants du numérique sont toutes des sociétés très actives dans le développement d'IA générative. Selon certains scénarios futuristes[56], des Léviathans de l'IA pourraient régir un monde normalisé au bénéfice d'une minorité. Outre le DMA et le DSA, l'Europe entend adopter d'ici fin 2023 la législation sur l'IA (en anglais, l'AI Act) visant à encadrer le développement et le déploiement de l'IA, une première mondiale. Si la proposition initiale de la Commission (le 21 avril 2021) ne prévoyait rien de spécifique sur l'IA générative, le texte adopté par le Parlement européen en juin 2023 (initiant le trilogue institutionnel[57]) a intégré pour partie les questions soulevées par les LLM en prévoyant de nouvelles obligations de transparence quant aux ensembles de données utilisés pour l'entraînement des systèmes et en imposant des mesures de sauvegarde (identification des risques, test, documentation) pour éviter la génération de contenus illicites. Reste à voir si ce règlement qui n'entrera pas en vigueur avant fin 2025 sera capable de limiter les risques de l'IA tout en permettant le déploiement des outils utiles.

L'usage inapproprié des outils d'IA peut en outre heurter l'organisation d'une société ouverte et démocratique en facilitant l'orchestration de campagnes de désinformation[58], en affectant la créativité et la liberté de création et d'expression[59], en renforçant la surveillance des citoyens, en engendrant des discriminations et d'autres atteintes aux droits humains, en posant de nouveaux problèmes de cybersécurité, etc. La liste des risques est longue vu l'ubiquité et la diversité des usages possibles de cette technologie capable de simuler « l'intelligence » humaine, parfois malveillante, souvent défaillante. Ces risques doivent être mis en balance avec les bénéfices potentiels de ces outils. Une balance très difficile à faire, et à remettre sans cesse en cause pour atténuer les effets potentiellement négatifs.

3 Les usages de l'IA dans les métiers du droit

Après l'épisode des années 2016-2018, où le retour de l'IA se conjuguait avec la *hype* autour des LegalTech, l'IA générative fondée sur les LLM relance les interrogations sur l'usage de cette technologie en droit[60]. D'un côté, on nous prédit (à nouveau) que près de la moitié des processus et des tâches dans le secteur juridique pourraient être automatisés par l'IA[61]. De l'autre, rien n'indique que la pratique des tribunaux, des juristes en entreprise, des avocats, etc. a été fondamentalement bouleversée depuis la précédente annonce d'un tsunami technologique. Mais il est aussi clair que les applications utiles se sont multipliées[62] et que l'usage de nouvelles ressources (principalement de gestion et de bases de données juridiques) a progressé, en tout cas là où des investissements ont pu être faits. Que peut-on attendre de ChatGPT et autres LLMs, les premiers outils à produire du texte de manière automatisée, capables donc de rivaliser directement avec les productions des juristes ?

A. Quelques usages possibles des LLM

Des assistants au travail. La préparation d'un cv ou la rédaction d'une lettre de motivation[63] vont être de plus en plus souvent générées par ces outils, il faudra en tenir compte lors du recrutement. Des outils toujours plus performants existent pour traduire : beaucoup utilisent DeepL mais ChatGPT a aussi ses partisans pour la traduction. Il faudra en-

(50) Voy. le dialogue d'un usager avec le *chatbot* Replika et l'effet dévastateur d'une mise à jour en février 2023 ayant abouti à la rupture du lien noué, le robot ne se rappelant par exemple, plus du nom de ses « amis » humains et ayant perdu la faculté de soutenir ceux-ci pour les soucis du quotidien (voy. E. BROOKS et H. ROSIN, « Can an AI Save a Life ? Michael spent years fighting isolation, depression, and despair. Then he met Sam. Then Sam changed », *The Atlantic*, 10 août 2023, https://www.theatlantic.com/podcasts/archive/2023/08/are-ai-relationships-real/674965/).

(51) Voy. le Safe AI Companion Collective https://saicc-website.vercel.app.

(52) Voy. le large panorama de D. Acemoglu et S. Johnson (*Power and Progress, Our Thousand-Year Struggle Over Technology and Prosperity*, Basic Book, 2023) qui offre une critique radicale des effets dévastateurs de la technologie telle qu'elle a été notamment déployée par les Big Tech au profit de l'automation pure et simple et de la surveillance généralisée, sans égard pour l'emploi ou le débat démocratique.

(53) Le procès opposant l'État américain à Google, auquel il est reproché d'avoir abusé de sa position dominante, a démarré le 12 septembre devant un tribunal fédéral de Washington (https://www.nytimes.com/2023/09/11/podcasts/the-daily/google-monopoly-trial.html).

(54) Un rapport préliminaire a été publié le 18 septembre 2023 : https://www.gov.uk/government/publications/ai-foundation-models-initial-report?mkt_tok=MTM4LUVaTS0wNDIAAAGOUyeAL60nBUlLPzakcsEXK6blvqLLtpM40tEwcWiVmgEwz2rhIhgpgj4PWC7ml5CfE6ySDPGeyxhtmZ_sM3NqzNpwjEByufusf94acO1rf9n.

(55) Règlement (UE) 2022/1925 du Parlement européen et du Conseil du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique (*J.O.* L 265 du 12 octobre 2022, p. 1) ; règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (*J.O.* L 277 du 27 octobre 2022, p. 1).

(56) Voy. le scénario fiction « L'histoire du team Omega » dans le livre d'un professeur du MIT fondateur du Future of Life Institute, Max Tegmark, Life 3.0. Being human in the age of Artificial Intelligence, Allen Lane, 2017.

(57) Voy. les versions de l'AI Act adoptées jusqu'à présent sur https://www.kaizenner.eu/post/aiact-part3.

(58) Voy. la vidéo d'une expérimentation expliquant comment se monte à moindre coût une campagne de désinformation : https://counter-cloud.io/?page_id=307. Pour aller plus loin, voy. notre projet de recherche REMEDIS : https://projectremedis.wordpress.com/.

(59) Voy. l'article du PEN America, « Speech in the Machine : Generative AI's Implications for Free Expression », 31 juillet 2023 (https://pen.org/report/speech-in-the-machine/).

(60) Il y a quelques années, on a assez bien discuté de la justice digitale. Voy. par exemple A. GARAPON et J. LASSÈGUE, *Justice digitale*, PUF, 2018 (très critique sur la justice prédictive et ses applications Legaltech et la délégation aux machines) et A. VAN DEN BRANDEN, *Les robots à l'assaut de la justice*, Bruylant, 2019 (plus nuancé, distinguant une justice réparatrice [pensions alimentaires, indemnités de licenciement, dommages corporels, etc.], permettant une certaine automatisation de la décision, et la justice distributive [divorces, pratiques déloyales, diffamation, etc.] où l'aspect symbolique et le conflit de valeurs résiste à l'automatisation).

(61) Voy. J. BRIGGS et D. KODNANI, *The potentially Large Effects of Artificial Intelligence on Economic Growth*, 26 mars 2023 (étude financée par Goldman Sachs, donc à prendre avec la distance critique nécessaire) https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf.

(62) Pour une présentation de l'écosystème des LegalTech en Wallonie, voy. https://www.digitalwallonia.be/fr/publications/legaltech-wallonie/. Ce panorama des acteurs distingue les LegalTech de management (y compris de contract et corporate management), de legal service (recouvrement de créances ou indemnisation pour vol annulé, etc.), de services non juridiques (pour le recrutement et la collaboration) et de legal content (bases de données, outils de recherche intelligence) ; voy. aussi le tableau comparatif des logiciels d'avocats par Incuebrux (https://www.incubateurbxl.eu/fr/outils/), ainsi que les rapports de Wolters Kluwer (https://www.wolters-kluwer.com/en/know/future-ready-lawyer-2022) ou Henchman (https://henchman.io/legal-tech-trends-2023#form).

(63) Voy. A. FALJAOUI, « Recrutement et ChatGPT : la fin programmée de la lettre de motivation », *Trends-Tendances*, 4 septembre 2023.

Jt_doct_1218.fm Page 6 Friday, October 13, 2023 11:36 AM

core revoir et ajuster les *outputs*, mais l'usage de ces outils se traduira par un gain de temps.

On peut aussi anticiper le rôle grandissant des assistants personnels, dopés par les moteurs d'IA et tournant sur des smartphones toujours plus puissants. La contribution de l'IA à l'organisation des services deviendra déterminante. Contrairement à la manière dont l'outil Excel a été intégré à divers services (y compris juridiques), les assistants IA orchestreront diverses actions à l'aide d'interfaces entre logiciels : coordonner les agendas, réserver des salles de réunion et des lunch, commander des cartouches d'encre, planifier des rencontres dans une négociation ou médiation, etc.[64] Des versions de la suite bureautique de Microsoft complétées par de l'IA seront disponibles en novembre 2023 (sous le nom de Microsoft 365 Copilot[65]). L'outil de conférence en ligne Zoom est aussi en train d'intégrer de l'IA pour améliorer la collaboration et générer de manière autonome des PV de réunions et une liste des actions à prendre à la suite de la réunion qu'il a « écoutée ». Ces outils affecteront certainement la manière de travailler et l'organisation du travail intellectuel.

Des chercheurs et générateurs de documents. Pour produire des textes[66], divers outils dont les qualités varient sont disponibles. Il faudrait les expérimenter pour chaque métier du droit en particulier (avocat, juge, juriste d'entreprise, responsable public, etc.), et distinguer les types de documents (courants et standardisés) pouvant plus facilement être générés à l'aide de l'IA. On se bornera ici à faire quelques remarques et suggestions, sans pouvoir passer en revue l'apport des divers outils à des pratiques spécifiques.

ChatGPT (ou encore mieux sa version payante : GPT-4) est réputé être le plus performant mais il y a des concurrents et Claude 2 proposé par Anthropic semble présenter des avantages : il est possible d'ajouter (de téléverser) des fichiers (par ex. un rapport d'expert ou un brevet) dans une conversation avec Claude, qui génère automatiquement un résumé et en tient compte dans la suite de la conversation, davantage contextualisée par ces ajouts ; il est possible de le nourrir avec de plus longs *prompts*, ce qui à nouveau enrichit le contexte et génère des textes plus ciblés ou nuancés ; Claude 2 a sinon ingurgité des informations jusque fin 2022 (alors que ChatGPT a été alimenté par des données jusque septembre 2021) ; il est possible de téléverser des liens dans Claude qui va alors automatiquement résumer le texte ciblé par l'hyperlien (même s'il est protégé par un mur[67], ce qui peut poser problème sur le plan du droit d'auteur). Perplexity est un assistant de recherche qui adopte une interface conversationnelle et tient compte de l'environnement pour offrir des réponses personnalisées (en fonction d'un profil fourni) et sourcées (contrairement à ChatGPT). L'outil Cohere[68] propose lui aussi de multiples usages transformatifs : résumé automatique des points principaux d'un article ou d'une conversation, identification du sujet à partir du titre d'un article et classification dans la documentation, génération d'un email à partir d'une description du sujet, rédaction d'un descriptif de service en fonction de l'audience recherchée, extraction des points saillants d'un contrat, correcteur d'erreurs dans la transcription d'une voix en texte, proposition d'une accroche introductive à un texte, etc. Diversité des usages donc, mais avec un objectif unique : améliorer la productivité.

Des aidants dans la recherche et l'analyse de documents juridiques. Des IA génératives à usage spécifiquement juridique commencent à être mises sur le marché[69]. Harvey, une plateforme IA construite à partir des modèles d'OpenAI et optimisée pour le travail juridique, est notamment utilisée par plus de 3.500 employés du cabinet Allen & Overy et au sein de PwC (legal). Un avocat qui lui a demandé quelles sont les réglementations de la publicité applicables à des promotions en réalité virtuelle était satisfait des sources juridiques identifiées par l'outil[70]. Il est aussi possible de verser une pile de contrats ou de statuts et de poser des questions à leur propos, les réponses soulignant les passages pertinents des documents d'où l'IA a tiré ses conclusions. Mais les résultats d'Harvey semblent sinon assez mitigés. L'outil sert à démarrer une recherche et semble surtout offrir un gain de temps dans la première phase d'un dossier à traiter. D'où les recommandations d'usage : pour la première version du document seulement ; le faire réviser par un juriste qui connaît le domaine, peut détecter les erreurs et est capable d'optimiser les *prompts*/invites. La possibilité d'utiliser les LLMs pour extraire des passages pertinents de longs documents, autrement dit la « fonction extractive » (plutôt que générative) de ces outils, semble plus prometteuse à court terme pour la pratique juridique[71].

Il est en outre possible que certains outils soient à même de peser les risques dans une procédure litigieuse, et de compléter ainsi l'évaluation par l'avocat des chances de réussite. Il est encore plus probable qu'une portion non négligeable de certains services juridiques, par exemple, en matière de recouvrement de créances, d'infractions de roulage ou de *due diligence* dans une acquisition, soit à l'avenir prise en charge par des outils intelligents qui aboutiront à réduire le temps de travail humain, tout spécialement des stagiaires et documentalistes (et donc aussi le « facturable » dans les cabinets d'avocats). C'est le déclassement de certains services de base que l'on peut anticiper. Et peut-être davantage : il existe déjà des outils capables, à partir d'un projet de contrat, de suggérer l'ajout de clauses manquantes (avec proposition de clause tenant compte du contexte et des parties au contrat), d'identifier les points à négocier et des arguments de négociation en fonction de la partie que l'avocat représente (Spellbook[72]). En revanche, l'ensemble de la valeur ajoutée du service juridique rendu, lequel permet de conclure un avantageux contrat sur mesure ou de résoudre un différend complexe — tout en assurant un suivi de la relation personnelle avec le client ou d'autres parties —, ne va pas prochainement être générée par une IA. Les outils d'IA ne sont pas encore à même de construire un narratif dans l'argumentation juridique, basé sur une bonne identification des faits et des différences avec des précédents pertinents, ou à concevoir un bon scénario pour une relation contractuelle qui doit durer, avec une anticipation des ratages possibles et des circonstances changeantes. Donc la réponse des juristes au défi de ces outils serait de se concentrer sur les services les plus sophistiqués, à haute valeur humaine ajoutée. Reste à bien identifier ceux-ci, et, de l'autre, à bien définir les tâches que la technologie peut assumer. En attendant, les cabinets d'avocats mettent doucement en place leurs politiques pour l'usage de l'IA générative[73].

B. Quelques précautions et recommandations pour juristes

Expérimenter les outils et bien assigner les tâches. À défaut de pouvoir disposer d'un LLM paramétré pour les documents juridiques (comme Harvey), chacun doit tester l'utilité des outils disponibles pour sa pratique juridique. Ces outils sont généralistes, contrairement aux produits des LegalTech dont la fonction est prédéterminée par leurs concepteurs. Il faut donc soi-même identifier ce que l'on peut utilement attendre des LLMs (et penser la complémentarité avec les outils développés spécialement pour juristes). Cela requiert à tout le moins de se poser quelques questions : quelles sont les tâches de recherche

(64) Voy. la vision de M. Suleyman, co-fondateur de DeepMind : https://www.ft.com/content/f828fef3-862c-4022-99d0-41efbc73db80?desktop=true&segmentId=7c8f09b9-9b61-4fbb-9430-9208a9e233c8#myft:notification:daily-email:content. Ces assistants personnels sont utilisés dans le domaine du marketing de la vente où des outils CRM (Customer Relationship Management) gèrent l'entièreté du processus, du premier démarchage du client à la conclusion de la vente. Ils pourraient être transposés dans les services juridiques aux entreprises.

(65) https://news.microsoft.com/fr-fr/2023/03/16/microsoft-365-copilot-un-nouvel-assistant-personnel-commande-par-le-langage-naturel/.

(66) Certains métiers du droit peuvent bénéficier d'autres assistants dans la génération d'images (pour un site *web*) et de noms (pour un service juridique), dans la communication visuelle (sur les réseaux), pour la gestion des tâches, etc. Ces outils ne sont pas examinés ici.

(67) Ce fait est relevé dans l'article : « 4 things Claude AI can do that ChatGPT can't », 21 août 2023, disponible sur zdnet.com.

(68) Voy. https://cohere.com/examples qui vise à « Transformer vos produits avec les LLMs ».

(69) Un grand éditeur juridique, LexisNexis, envisage à l'avenir d'intégrer une IA générative à sa plateforme d'accès aux sources juridiques pour faciliter la rédaction et la recherche (R. Miller, « LexisNexis is embracing generative AI to ease legal writing and research », *TechCrunch*, 10 septembre 2023).

(70) Cet exemple ainsi que les développements autour d'Harvey sont tirés d'une enquête « AI and the workplace » (*Financial Times*, 19 juin 2023).

(71) Voy. C. Criddle, « Firms seek to overcome hallucination hurdles to make AI a reality », *Financial Times – Innovative Lawyers in Europe*, dossier du 22 septembre 2023, p. 6.

(72) www.spellbook.legal.

(73) Ainsi selon la presse (*Financial Times – Innovative Lawyers in Europe*, dossier du 22 septembre 2023, pp. 6-8), certains cabinets, comme NautaDutilh, interdisent l'usage des outils d'IA disponibles sur le marché pour des productions à l'attention des clients, un autre (Addleshaw Goddard) permet par contre l'usage de ChatGPT, ce qui suggère que les avocats prennent des positions entre « fomo » (« fear of missing out ») et « foji » (« fear of joining in »).

Jt_doct_1218.fm Page 7 Friday, October 13, 2023 11:36 AM

que l'on peut confier aux LLMs ? Quelles sont les compétences à développer pour bien tirer parti de ces outils ? Quels sont les rôles qu'il convient de faire jouer à l'outil pour en tirer des enseignements utiles (l'ajout d'un contexte ou d'une posture aide à tirer parti des IA génératives) ? À partir de quand faut-il une vérification humaine ? Comment intégrer l'apprentissage des outils d'IA générative à la formation des jeunes en interne ? Autant de questions à se poser. Ce qui apparaît d'ores et déjà utile : apprendre l'art des invites[74]. Le propre de l'humain sera de plus en plus de poser les bonnes questions à la machine et de la faire accoucher des bonnes réponses[75]. Autrement dit, les juristes (et les autres humains) auront à apprendre un nouvel art socratique !

Développer de nouvelles expertises dans la conception et l'audit d'IA. Avec l'usage grandissant de ces outils dans tous les domaines, par exemple, le contrôle aux frontières ou la lutte contre la fraude fiscale ou sociale, la nécessité de pouvoir correctement concevoir, déployer et auditer les outils d'IA se fait sentir. À cette fin, les développeurs et utilisateurs d'IA doivent s'entourer de juristes ouverts à des expériences interdisciplinaires[76]. Il y a là certainement de nouvelles perspectives professionnelles pour les (jeunes) juristes.

Définir une déontologie pour un usage intègre et responsable de l'IA. L'usage mal approprié de ChatGPT par des avocats a déjà été discuté devant les tribunaux aux États-Unis. Un avocat et son cabinet ont ainsi été condamnés en juin 2023 pour avoir négligé de prendre leur responsabilité en soumettant au tribunal des conclusions comportant de fausses citations à des décisions inventées de toutes pièces par ChatGPT[77]. Sous la pression de l'urgence qui affecte de nombreuses professions (juridiques), la tentation sera grande de profiter de tout gain de productivité dans la génération d'un document. Il faut donc définir des règles d'usage tant au niveau des professions juridiques (à commencer par le barreau) qu'au niveau individuel.

Les codes de conduite pour l'intégrité des chercheurs dans le contexte universitaire[78] peuvent inspirer l'éthique du juriste face à l'IA (ainsi que des recommandations à l'attention des étudiants et enseignants). Les grandes revues scientifiques[79] ont adopté des politiques assez restrictives. Ainsi les revues du groupe *Science*[80] n'autorisent pas l'utilisation de textes, de figures ou d'images générées par l'IA sans l'autorisation expresse des éditeurs, et un outil d'IA ne peut se voir reconnaître la qualité d'auteur. La revue *Nature*[81] s'interdit de publier des images, vidéos ou illustrations générées par l'IA, mais accepte « l'inclusion de texte qui a été produit avec l'assistance d'une IA générative », du moins sous conditions ; par ailleurs, elle exige que l'utilisation de LLMs soit divulguée. D'autres groupes d'éditeurs (comme PLOS) autorisent plus largement l'usage d'IA, mais exigent de tout auteur qu'il détaille, dans une section méthodologique, l'outil utilisé, la manière dont il l'a utilisé ainsi que les passages affectés/générés par l'outil d'IA, ainsi que la façon dont il a évalué la validité des informations générées[82]. Dans son guide de l'usage responsable de ChatGPT et des LLM, la KU Leuven demande que les informations extraites soient correctement attribuées (avec l'exemple d'une référence : « OpenAI (2023) ChatGPT (version du 20 avril) [LLM]. https://chat.openai.com/ »). Cela va assez loin. Par le passé, l'on n'a pas pris l'habitude de citer les outils de recherche utilisés, comme Google ou Bing, mais cela s'explique dans ce dernier cas, car en finale, on a les sources/liens, pas avec ChatGPT[83]. Il ne faut pas nécessairement être aussi strict dans les productions juridiques. Celles-ci ne visent pas à décrire/expliquer un état de fait (le cas échéant avec une méthode hypothético-déductive et de vérification expérimentale) mais à justifier une prise de position et de risque (consultation), à cimenter le cadre d'une relation (contrat), à faire justice (décision) ou à édicter la règle (loi). Elles ont donc une visée normative ou performative, contrairement aux énoncés scientifiques. Cela dit, des exigences similaires d'intégrité s'appliquent. Ainsi le juriste est amené à chercher des sources valables pour développer ses arguments. Il doit les citer pour profiter de leur autorité. Il doit rester intègre dans l'interprétation qu'il en donne.

* * *

En bref, les LLMs sont encore incapables de remplacer les juristes dans le développement d'une argumentation précise autour d'une constellation particulière de faits et de normes complexes — la tâche par excellence du juriste. Ils sont très bons sur des sujets techniques ou scientifiques, nettement moins sur des questions pointues de droit (belge). Ils énoncent souvent des généralités, dont la banalité rejoint parfois l'opinion commune. Ils offrent donc un bon point de départ... mais dont il faut aussi se départir pour dépasser le ressassement un peu conventionnel des mots — notre tâche commune d'auteurs humains[84] !

Alain STROWEL
Professeur UCLouvain, Munich IP Law Center, avocat au barreau de Bruxelles

(74) Voy. le site learnprompting.org qui apprend à bien s'adresser à l'IA pour qu'elle produise des énoncés utiles.
(75) Voy. les réflexions de L. DE BRABANDERE, « Professeur de ChatGPT », *La Libre*, 12 septembre 2023.
(76) Voy. B. LEONG et A. BURT, « Beyond the Hype : Lessons on auditing AI systems from the front lines », *American Bar Association Journal*, 17 juillet 2023.
(77) US District Court, S.D. New York, *Roberto Mata c. Avianca*, 22-cv-1461 (PKC), 22 juin 2023.
(78) Et par exemple le nouveau European Code of Conduct for Research Integrity, ALEA, 2023, https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf.
(79) Voy. le tour d'horizon des pratiques : A. HOOVER, « Use of AI Is Seeping Into Academic Journals—and It's Proving Difficult to Detect », *Wired*, 17 août 2023 (https://www.wired.com/story/use-of-ai-is-seeping-into-academic-journals-and-its-proving-difficult-to-detect/). Comme bon exemple à l'attention de la communauté universitaire et des étudiants en particulier, voy. les lignes de conduite de la KU Leuven : https://www.kuleuven.be/onderwijs/student/onderwijstools/artificiele-intelligentie.
(80) Voy. les règles que les auteurs des revues *Science* doivent suivre : https://www.science.org/content/page/science-journals-editorial-policies?adobe_mc=MCMID%3D15100613600109274893352643554246978174%7CMCORGID%3D242B6472541199F70A4C98A6%2540AdobeOrg%7CTS%3D1692109178.
(81) Voy. l'éditorial du 7 juin 2023 https://www.nature.com/articles/d41586-023-01546-4?utm_medium=affiliate&utm_source=commission_junction&utm_campaign=CONR_PF018_ECOM_GL_PHSS_ALWYS_DEEPLINK&utm_content=textlink&utm_term=PID100095187&CJEVENT=ed96bbfd3d4911ee8034b8c60a18b8fa#:~:text=Apart%20from%20in%20articles%20that-least20for20the%20foreseeable%20future.
(82) Voy. https://journals.plos.org/plosone/s/ethical-publishing-practice#:~:text=Artificial%20Intelligence%20Tools%20and%20Technologies, -PLOS%20expects%20that&text=Contributions%20by%20artificial%20intelligence%20(AI, types%20lacking%20a%20Methods%20section. Les politiques de nombreuses revues rendent par ailleurs les auteurs responsables en cas d'erreur dans l'information générée par l'IA.
(83) Toutefois, la fusion en cours d'implémentation des outils d'IA générative et de recherche sur le *web* va changer la manière dont les outputs de l'IA sont sourcés.
(84) Quant aux risques de l'IA à moyen terme, méfions-nous de l'apprentissage-machine qui pourrait générer un apprenti sorcier... c'est un autre débat !