



HAL
open science

L'accès aux connaissances logiciels : enjeux et perspectives pour l'intelligence artificielle et la Science ouverte

Mélanie Clément-Fontaine

► **To cite this version:**

Mélanie Clément-Fontaine. L'accès aux connaissances logiciels : enjeux et perspectives pour l'intelligence artificielle et la Science ouverte. Propriété industrielle, 2024, Les revues Lexisnexis, 4 (4), pp.9-13. hal-04884965

HAL Id: hal-04884965

<https://hal.science/hal-04884965v1>

Submitted on 13 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

L'accès aux connaissances logiciels : enjeux et perspectives pour l'intelligence artificielle et la Science ouverte.

Mélanie Clément-Fontaine, professeure de droit UVSQ-ParisSaclay.

Revue Propriété industrielle, LexiNexis · 3 avr. 2024, n°4 PP. 9- 13

Dans un paysage technologique en perpétuelle évolution, où le logiciel libre, la Science ouverte et l'intelligence artificielle (IA) se croisent et redéfinissent les frontières de l'innovation et de la connaissance, démêler l'écheveau de leurs interdépendances s'avère à la fois prééminent et complexe. Ces concepts, bien que distincts, partagent une vision commune de l'accessibilité et de la diffusion des connaissances, tout en naviguant dans des eaux souvent troubles, marquées par des débats et des enjeux juridiques non résolus.

1. Dans le contexte actuel, marqué par un rapprochement contrarié entre le mouvement du libre, la science ouverte, et l'intelligence artificielle (IA), nous nous engageons dans une exploration visant à décrypter les influences mutuelles entre ces trois sphères d'innovation. Notre démarche s'inspire de la technique de la 'ligne claire' popularisée par Joost Swarte afin de dessiner ces interactions. Cette approche nous permettra de tracer les contours d'un paysage en évolution, tout en acceptant que certains détails puissent volontairement être omis pour privilégier une compréhension globale des enjeux. Cette exploration intervient à un moment critique, où la Science ouverte cherche encore son assise juridique, et où les entreprises d'IA jonglent entre ouverture pour profiter de l'innovation collective et fermeture pour garantir leur viabilité économique. Ce qui est en jeu est la façon dont ces dynamiques façonnent l'accès à la connaissance et le futur des avancées technologiques.

2. D'un côté, nous avons le mouvement du libre, pionnier d'une ère de collaboration et de partage sans contraintes, principalement motivé par des valeurs éthiques et communautaires plutôt que par des considérations économiques¹. De l'autre, la science ouverte et l'IA se développent à la croisée des chemins de l'accessibilité et de l'utilisation des connaissances, empruntant au logiciel libre tout en naviguant dans des eaux institutionnelles et économiques qui leur sont propres.

3. Ce panorama, bien complexe, offre une opportunité de réfléchir aux moyens par lesquels ces trois domaines peuvent coexister et s'enrichir mutuellement, tout en tenant compte des réalités économiques et juridiques qui les entourent.

I. Science ouverte et logiciel libre : un rapprochement nuancé

4. L'intersection de la science ouverte et du logiciel libre représente un terrain fertile pour la réflexion sur l'évolution actuelle de la recherche et de l'innovation. La Science ouverte, telle que définie dans le cadre du deuxième Plan national pour la période 2021-2024², vise à

¹ R. M. Stallman, S. Williams, C. Masutti, Richard Stallman et la révolution du logiciel libre, Eyrolles, 2010.

² https://www.enseignementsup-recherche.gouv.fr/sites/default/files/content_migration/document/2e-plan-national-pour-la-science-ouverte-2021-2024-7794.pdf : « la diffusion sans entrave des résultats, des méthodes et des produits de la recherche

promouvoir l'accessibilité et la transparence de la recherche financée par des fonds publics. Elle repose sur trois piliers fondamentaux : l'objet, qui concerne la production scientifique subventionnée par l'État ; le principe d'« accès ouvert », favorisant la diffusion large et maîtrisée des résultats ; et la finalité, qui consiste à encourager l'accès à la recherche dans un esprit de transparence, de démocratisation du savoir, et de stimulation économique.

5. Cette définition révèle une dualité intrinsèque à la Science ouverte. D'une part, elle partage avec le mouvement du logiciel libre une volonté de mutualiser la connaissance de manière ouverte et accessible. D'autre part, elle intègre des dimensions supplémentaires, telles que la maîtrise des résultats et les retombées économiques, qui ne sont pas traditionnellement au cœur des préoccupations du mouvement du libre. Ainsi, bien que la Science ouverte et le logiciel libre ont des valeurs communes de transparence et d'ouverture, la Science ouverte embrasse également des objectifs plus larges, reflétant la complexité et les multiples facettes de la recherche contemporaine. Cette nuance souligne l'importance de comprendre le rapport entre ces deux sphères, non seulement au regard de leurs similarités, mais aussi de leurs différences, afin de mieux saisir les enjeux et les occasions qu'ils représentent pour l'avancement de la connaissance et le développement technologique.

6. Le logiciel libre et la Science ouverte, bien que partageant un idéal commun de diffusion du savoir, émergent de contextes et de motivations fondamentalement différents. Le logiciel libre, issu d'une démarche communautaire et indépendante des institutions, met en avant la création collaborative et la suppression des obstacles à l'utilisation des codes informatiques. Son essence repose sur une éthique de partage inconditionnel, où la valeur intrinsèque du savoir et de l'innovation prévaut sur l'exploitation commerciale.

7. En contraste, la Science ouverte s'inscrit dans une démarche plus institutionnelle, cherchant à conjuguer l'ouverture et la transparence de la recherche publique avec des ambitions économiques. Elle rejoint, en ce sens, le concept d'Open data, visant à rendre les données publiques accessibles et réutilisables, non seulement pour accélérer le progrès scientifique, mais aussi pour stimuler la croissance économique.

8. Les fondations de la science ouverte sont ainsi tracées par des organisations internationales telles que l'UNESCO, qui promeut le partage universel du savoir, et appuyées par des entités économiques et politiques comme l'OCDE, le G8 et l'Union européenne. Ces dernières se concentrent davantage sur le potentiel économique de l'accès ouvert aux résultats de la recherche, soulignant une approche pragmatique qui vise à équilibrer les bénéfices scientifiques et économiques.

9. Ce cadre institutionnel robuste a façonné la Science ouverte, non seulement en tant que mouvement pour la diffusion du savoir, mais aussi comme un levier stratégique pour le développement économique. En 1999, l'UNESCO enjoint les gouvernements à mettre en place des « mécanismes » ouverts d'accès aux résultats de la science³. Mais, cinq ans plus tard, c'est vers l'OCDE que les ministres de la Science et de la Technologie⁴ se sont tournés pour lui

scientifique. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche. Elle permet à la recherche financée sur fonds publics de conserver la maîtrise des résultats qu'elle produit. Elle construit un écosystème dans lequel la science est plus transparente, plus solidement étayée et reproductible, plus efficace et cumulative. Elle induit une démocratisation de l'accès aux savoirs, utile à l'enseignement, à la formation, à l'économie, aux politiques publiques, aux citoyens et à la société dans son ensemble. Elle constitue enfin un levier pour l'intégrité scientifique et favorise la confiance des citoyens dans la science ».

³ Déclaration sur la science et l'utilisation du savoir scientifique, 1999 : Résolution 40 c/63 sur l'opportunité d'une recommandation de l'UNESCO sur la science ouverte - sa résolution 29 C/28, Annexe III, § 63.

⁴ Il s'agissait de tous les membres de l'OCDE auxquels s'étaient joints l'Afrique du Sud, la Chine, Israël et la Russie : Declaration on Access to Research Data from Public Funding : <https://legalinstruments.oecd.org/en/instruments/157>

demander de définir un ensemble de lignes directrices à partir de principes établis d'un commun accord en vue de faciliter un accès au moindre coût aux données numériques de la recherche financée sur fonds publics. Les principes et les lignes directrices élaborés à la suite d'un processus de consultation ont été approuvés par le Comité de la politique scientifique et technologique de l'OCDE en octobre 2006, puis joints à une Recommandation de l'OCDE et entérinés par le Conseil de l'OCDE à la fin de l'année 2006⁵. La recommandation vise les données de la recherche recueillies à l'aide de fonds publics entendues largement puis qu'il peut s'agir aussi bien de chiffres que de textes ou d'images et de sons. Autrement dit, la notion peut couvrir toutes les formes de productions scientifiques dès lors qu'elles constituent une « source principale pour la recherche scientifique »⁶. En outre, la recommandation apporte un premier éclairage sur ce que devrait être le cadre juridique en préconisant que les droits de propriété intellectuelle ne doivent pas faire obstacle à l'accès de ces données⁷. En ce sens, ces recommandations s'alignent sur le projet du mouvement de libre qui, grâce aux licences libres, ont permis que le droit de la propriété intellectuelle ne soit pas un obstacle à l'usage collectif des logiciels. Néanmoins, à la différence du mouvement du libre, l'ouverture de la recherche publique connaît une dimension économique forte : le G8⁸, par exemple, a insisté sur l'importance de préserver un équilibre entre l'accessibilité de la production scientifique financée sur fonds publics et les intérêts commerciaux. Les recommandations de la Commission européenne⁹ s'inscrivent parfaitement dans cette approche de la Science ouverte en déclarant faire « du développement d'une économie fondée sur la connaissance et l'innovation, une priorité en faveur de l'accès aux informations scientifiques et à leur conservation ».

10. Véritable politique publique menée par l'Union européenne et ailleurs, la Science ouverte consiste aujourd'hui davantage en des lignes directrices qu'en un régime juridique. Ses objectifs se rapprochent en partie de ceux de l'Open data¹⁰, ce qui conduirait éventuellement à l'élargissement du régime des données publiques aux productions de la recherche financées sur fonds publics. Pour autant, elle demeure dans le giron du droit de la propriété intellectuelle et de ses dispositions particulières relatives aux créations réalisées par les chercheurs et chercheuses du secteur public.

11. La Science ouverte répond à l'adage « aussi ouvert que possible, mais aussi fermé que nécessaire ». Dans cet entre-deux, le principe est par conséquent la mise à disposition ouverte et les limites sont, outre celles que l'on retrouve pour les données publiques¹¹, les intérêts commerciaux légitimes. L'hésitation entre l'accès sans entrave et un contrôle de la diffusion est légitime tant les enjeux sont importants. On perçoit cette modulation en particulier dans la seule disposition législative générale relative à la Science ouverte consacrée à ce jour qu'est la directive 2019/1024 du 20 juin 2019¹² concernant les données ouvertes et la réutilisation des informations du secteur public. En son article 10, interprété à la lumière des considérants 27 et

⁵ OCDE, *Recommandation du Conseil concernant l'accès aux données de la recherche financée sur fonds publics* de 2006 modifiée en 2021 : voir le Recueil des instruments juridiques de l'OCDE : <http://legalinstruments.oecd.org>

⁶ Ibid. page 18

⁷ Ibid. page 21

⁸ G8 Science statement (London, June 2013) :

<https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013>

⁹ Recommandations (2012/417/UE) du 17 juillet 2012 relatives à l'accès aux informations scientifiques et à leur conservation.

¹⁰ Il s'agit du régime applicable aux données publiques selon lequel l'administration est, en principe, tenu de permettre l'accès et la réutilisation des données qu'elle recueille ou produit à toute personne qui en fait la demande : Directive n° 2013/37 du 26 juin 2013 modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public. Loi n° 2015-1779 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public du 28 décembre 2015. Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. Directive n° 2019/1024 du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public (refonte).

¹¹ Il s'agit des droits des tiers comme les droits de propriété intellectuelle, la protection des données à caractère personnel, la confidentialité, les impératifs de la sécurité.

¹² Directive 2019/1024 du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

28, le législateur européen enjoint les états à prendre les mesures nécessaires afin de rendre librement accessibles les données résultant de la recherche financée au moyen de fonds publics (« politique de libre accès ») qui respectent le principe d'ouverture par défaut et sont compatibles avec les principes FAIR selon lesquels les ressources intellectuelles doivent être « faciles à trouver, accessibles, interopérables et réutilisables »¹³. Toutefois, si essentielle qu'elle soit la réutilisation par le public des résultats de la recherche pourrait être limitée, voire écartée pour répondre aux nécessités économiques.

12. Il règne un certain flou sur les conditions de mise à disposition des productions de la recherche qui n'est pas favorable au rayonnement de la Science ouverte. Afin de pallier cet état de fait, le gouvernement français, dans le cadre du Plan national de la Science ouverte, a confié à des experts, réunis sous la forme de collèges, le soin d'apporter des éclairages. Le collège *codes sources et logiciels*¹⁴ a fait son apparition à l'occasion du Second plan pour la science ouverte en 2021. Pourtant, l'accès ouvert des logiciels, qui a déjà été largement éprouvé grâce aux licences libres, est également garanti par le régime de l'Open data si bien que l'on peut se demander quel est l'apport de la science ouverte en ce domaine.

13. Pour autant, l'objet de la science ouverte est suffisamment générique et pour accueillir les logiciels comme en témoignent notamment les lignes directrices de l'OCDE¹⁵ ou encore la directive 2019/1024 du 20 juin 2019 qui définit les données de la recherche¹⁶ comme « *des statistiques, des résultats d'expériences, des mesures, des observations faites sur le terrain, des résultats d'enquêtes, des enregistrements d'entretiens et des images* », mais « *également des métadonnées, des spécifications et d'autres objets numériques* ». Il s'agit plus particulièrement des logiciels de recherche financés par des fonds publics, à savoir, selon la définition retenue le collège *codes sources et logiciels*, les « *développés pour répondre à des besoins spécifiques de la science. Ils sont conçus, maintenus, et utilisés par des scientifiques (chercheurs et ingénieurs) et institutions de recherche, éventuellement dans une dimension internationale. Ils peuvent découler de travaux de recherche comme ils peuvent les favoriser, notamment par des publications avant/sur/autour/avec le logiciel* »¹⁷.

14. De plus, ainsi que l'explique François Pellegrini¹⁸, le logiciel occupe une place particulière dans la production de nouvelles connaissances en ce qu'il constitue un « outil immatériel permettant d'exprimer formellement et de façon univoque des traitements de données ». On perçoit dès lors, sans être spécialiste de l'informatique, que le logiciel joue un rôle pivot pour l'innovation. Leur diffusion en Science ouverte est donc déterminante. En effet, la spécificité des logiciels au sein de la Science ouverte est d'être à la fois un outil, le résultat et un objet d'études de la recherche. Ils concentrent à eux seuls toutes les problématiques du libre accès relatives à l'archivage, le référencement, la citation, la valorisation, la pérennisation des projets de recherche¹⁹. Il n'est donc pas étonnant que plusieurs dispositions juridiques

¹³ [Commission d'enrichissement de la langue française, « données FAIR » \[archive\]](#), sur [FranceTerme](#), [ministère de la Culture](#)

¹⁴ <https://www.ouvrirlascience.fr/college-codes-sources-et-logiciels/>

¹⁵ Voir plus haut

¹⁶ Considérant 27

¹⁷ [Rapport](#) : Deuxième Plan national pour la Science ouverte - Généraliser la science ouverte en France 2021-2024

¹⁸ F. Pellegrini, « Qu'est-ce qu'un logiciel de recherche », dossier spécial *La Science ouverte et les logiciels*, M. Clément-Fontaine (dir.), RLDI 2024/210, pp. 39-40.

¹⁹ Clément-Fontaine M., Di Cosmo R., Guerry B., Moreau P., Pellegrini F., Note d'opportunité sur la valorisation des logiciels issus de la recherche, [Rapport de recherche] Comité pour la Science ouverte. 2019, 6 p. hal -03606374

fassent référence aux logiciels ouverts lorsqu'il est question d'outils éducatifs²⁰ ou encore d'accès aux ressources administratives²¹.

15. La place particulière qu'occupent les logiciels de recherche dans la production de nouvelles connaissances et plus généralement dans le fonctionnement de notre société permet d'expliquer le choix de la France d'ouvrir l'accès des logiciels produits par l'administration en application du régime de l'Open data. En effet, après une période d'incertitudes²² auxquelles la Commission d'accès aux documents administratifs²³ ainsi que le juge avaient dû répondre²⁴, le législateur²⁵ est venu confirmer en 2016 que, parmi les documents administratifs soumis au libre accès et la libre réutilisation, figurent les codes sources²⁶. Rien ne contraignait la France à prendre une telle direction. Aucune référence dans la définition de donnée publique n'est faite au logiciel dans la Directive 2013/37/UE du 26 juin 2013²⁷, quant à la directive 2019/1024 du 20 juin 2019²⁸, le législateur, au détour d'un considérant, laisse aux états le choix d'étendre le régime aux programmes informatiques. La France va également se doter de règles particulières²⁹ relatives aux algorithmes³⁰ utilisés dans les prises de décision administrative qui imposent la transparence et partant, l'accès au code source³¹. Il découle de ce qui précède que l'accès et la réutilisation des codes sources et des logiciels est à présent obligatoires dès lorsqu'ils relèvent du champ de l'Open data, c'est-à-dire qu'ils ont été produits par l'administration ou les personnes morales chargées d'un service public. Néanmoins, il ne va pas de soi que les logiciels de recherche dépendent de l'Open data dans la mesure où, les chercheurs et chercheuses du secteur public bénéficient d'un régime dérogatoire de celui des autres agents publics prévu à l'article L.111-1, alinéa 4 du Code de la propriété intellectuelle. Toutefois, cette dérogation ne s'étend pas aux auteurs de codes sources et de logiciels : l'article L111-1, alinéa 4 du CPI ne déroge pas à la dévolution légale des droits à l'autorité hiérarchique prévue aux articles L.113-9 et L.113-9-1, mais uniquement aux articles L.121-7-1 et L.131-3-1 à L131-3-3 du CPI au profit des agents auteurs qui, en vertu de leur statut ou des règles qui régissent leurs fonctions, ne sont soumis à aucun contrôle préalable de l'autorité hiérarchique. Par conséquent, l'institution publique détient les droits sur le code source et les logiciels réalisés par les chercheurs et chercheuses ainsi que par tout agent public (article L.113-9, alinéa 3 CPI)³².

16. Si l'on file ce raisonnement jusqu'au bout, la plupart des logiciels et des codes sources de la recherche financés sur fonds publics doivent être librement accessibles et réutilisables par

²⁰ Article L123-4-1 du Code de l'éducation.

²¹ [Article L300-4](#) Code des relations entre le public et l'administration tel que modifié par la loi n° 2016-1321 du 7 octobre 2016, article 3

²² M. Clement-Fontaine

²³ Avis 20144578 — Séance du 08/01/2015, Direction générale des finances publiques (DGFIP). Avis 20161990 — Séance du 23/06/2016 Institut national polytechnique de Toulouse (INP Toulouse).

²⁴ Tribunal administratif de Paris, 5ème sec., 2ème ch., jugement du 10 mars 2016.

²⁵ Loi pour une République numérique de 2016, article 2.

²⁶ L'article L300-2 du Code des relations entre le public et l'administration fut modifié en ce sens.

²⁷ Directive 2013/37/UE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 26 juin 2013 modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public.

²⁸ Préc. considérant 30.

²⁹ la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique et le décret n° 2017-330 du 14 mars 2017 : articles L. 311-3-1 et R. 311-3-1-2 du code des relations entre le public et l'administration.

³⁰ Un algorithme constitue « l'étude de la résolution de problèmes par la mise en œuvre de suites d'opérations élémentaires selon un processus défini aboutissant à une solution » (arrêté du 27 juin 1989 relatif à l'enrichissement du vocabulaire de l'informatique).

³¹ L'obligation de transparence des algorithmes a été étendue aux traitements de données personnelles : ordonnance du 12 décembre 2018 modifiant l'article 47 2° de la loi informatique et liberté du 6 janvier 1978.

³² Ainsi que « des personnes qui sont accueillies dans le cadre d'une convention par une personne morale de droit privé ou de droit public réalisant de la recherche créent des logiciels dans l'exercice de leurs missions ou d'après les instructions de la structure d'accueil si elles se trouvent à l'égard de cette structure dans une situation où elles perçoivent une contrepartie et où elles sont placées sous l'autorité d'un responsable de ladite structure » (L.113-9-1 CPI), par exemple un stagiaire.

tous en vertu du régime de l'Open data (articles L311-1 et L321-1 CRPA³³). Plus précisément, selon les termes de l'article L. 323-2 du CRPA, l'administration doit diffuser sous ces licences libres (parmi celles énumérées à l'article D323-2-1 du CRPA³⁴) sauf si elle peut se prévaloir de besoins spécifiques justifiant d'autres conditions de diffusion et qu'elle obtient une homologation de la licence de son choix à l'issue d'une procédure relativement lourde³⁵. Par conséquent, concernant les logiciels et les codes sources de recherche, il semble que le régime de l'Open data suffise pour que la France atteigne l'objectif qu'elle s'est fixée dans le cadre du deuxième plan de la science ouverte de publier en accès libre, d'ici 2030, 100 % des productions.

17. Pour autant, l'intérêt d'appliquer les principes de la Science ouverte demeure entier, car ils tiennent compte des stratégies de valorisation actuelle³⁶ et des modalités économiques de la création de ces logiciels. Toute d'abord, la production de logiciels peut constituer une ressource propre non négligeable pour les institutions de recherche qui, pour cela, se sont dotées de services de valorisation. De plus, il n'est pas rare que les projets informatiques soient menés de concert avec des partenaires privés qui optent pour une exploitation « propriétaire » des logiciels en contrôlant l'accès et l'utilisation des produits. Dans ce contexte, la Science ouverte présente l'avantage d'offrir plus de flexibilité aux institutions de la recherche, par rapport au régime de l'Open data, pour aborder l'ouverture des logiciels. Actuellement, le propos n'est d'ailleurs pas d'imposer les conditions d'exploitation comme celles prévues à l'article L. 323-2 du CRPA qui impose une diffusion sous licences libres. La Science ouverte est un cadre plus souple qui permet éventuellement d'introduire davantage d'options afin de tenir compte à la fois des besoins en ressources propres des institutions de recherche et des partenariats public-privé. Cela se justifie d'autant plus que les logiciels de recherche peuvent présenter une valeur financière importante dans l'économie de l'informatique. Il s'agit finalement de trouver le point d'équilibre exprimé sous la forme de l'adage « *aussi ouvert que possible, mais aussi fermé que nécessaire* ».

II. Intelligence artificielle et logiciel libre : un rapprochement contesté

18. L'IA générative permet de produire des codes informatiques à partir, bien entendu, de codes préexistants. Ces derniers, s'ils sont disponibles en ligne sous la forme de codes sources comme le préconisent notamment les licences libres, constituent donc une ressource pour les services d'IA générative. Quant aux autres logiciels diffusés sous licence propriétaire, la meilleure protection contre leur exploitation à partir d'une IA générative consiste à rendre accessible uniquement le code sous sa forme exécutable, sachant, par ailleurs, que l'exception de décompilation (art. L122-6-1 CPI) ne peut pas justifier leur usage en dehors des conditions d'utilisation prévue par la licence. D'emblée, la relation entre les logiciels libres et l'IA est évidente. Les premiers nourrissent la seconde. Il se pourrait que l'inverse soit vrai également si l'on entérine la thèse selon laquelle les logiciels produits par l'IA, sans apport humain significatif, sont à la libre disposition de tous, autrement dit, ne sont pas protégés par un quelconque droit de la propriété intellectuelle.

³³ Code des relations entre le public et l'administration

³⁴ Il s'agit actuellement des licences : « Berkeley Software Distribution License », « Apache », « CeCILL-B » et « Massachusetts Institute of Technology License », « Mozilla Public License », « GNU General Public License », « CeCILL », « European Union Public License » et « Eclipse Public License ».

³⁵ « L'œuvre libre » (refonte), *JurisClasseur Propriété littéraire et artistique*, fasc. n° 1975, 2024

³⁶ N. Jullien ; « La valorisation économique des logiciels libres », in dossier spécial La science ouverte et les logiciels libres, op. cit. pp. 45-47.

19. Nonobstant l'intérêt de cette question, notre attention se porte sur autre point de jonction entre le logiciel libre et l'IA qui intéresse les conditions d'utilisation de la technologie IA : plusieurs de ces technologies sont dites disponibles en libre accès, certaines l'ont été dans un premier temps puis sont devenues « propriétaires », d'autres le sont partiellement. Avant de détailler ces différents modèles de diffusion, arrêtons-nous un instant sur les motivations de la diffusion en libre accès.

20. Officiellement, le choix de diffuser une technologie IA en accès libre a pour objectifs de démocratiser l'IA et de la rendre plus éthique en privilégiant la transparence grâce à un accès libre de la technologie. Les concepteurs d'IA, en optant pour l'accès libre, seraient les bons élèves d'une IA respectueuse des intérêts de l'humanité alors que les inquiétudes s'expriment et que l'Union européenne prône des systèmes d'IA sûrs, transparents, traçables, non discriminatoires³⁷. La diffusion en libre accès se veut donc rassurante. C'est cependant sans compter sur l'art de la rhétorique des concepteurs d'IA qui, pour justifier un changement d'orientation, n'hésitent pas à user des mêmes arguments (notamment les exigences d'éthique, et de sécurité) pour abandonner la diffusion en libre accès au profit d'une exploitation fermée de leur technologie.

21. De manière plus prosaïque, nous pouvons dire simplement que les modèles libre et fermé correspondent à deux modes de valorisation distincts. Ces deux approches de la valorisation peuvent être résumées, de manière très schématique, de la sorte : le modèle libre est un moyen, pour les concepteurs, de profiter de l'expertise d'une large communauté d'utilisateurs qui, parce qu'ils y sont autorisés à le faire, contribuent au développement de l'IA. Par ailleurs, la philosophie du libre, suivant une approche libertaire, conduit finalement à faire reculer le droit de la propriété intellectuelle, ce qui est particulièrement utile pour l'exploitation d'IA générative qui nécessite de très nombreuses ressources intellectuelles à partir desquelles se fait l'entraînement du système IA afin de devenir un outil performant. Inversement, le modèle propriétaire s'appuie sur une valorisation centrée sur une redevance ou un abonnement à la charge des utilisateurs dont la fidélité est assurée par une évolution constante du produit proposé exclusivement par l'éditeur. Qu'en est-il concrètement ?

22. Plusieurs producteurs d'IA ont tenté de concilier ces deux modes de valorisation en adoptant des licences hybrides qui empruntent à la fois aux licences libres et aux licences propriétaires. En voici quelques exemples qui permettent d'illustrer et de comprendre ce phénomène :

23. **L'exemple de Terraform**, la plateforme d'infrastructure as a code promue par HashiCorp. Terraform était initialement disponible sous licence Mozilla (MPL 2.0), mais en août 2023, ce dernier a décidé de changer de licence au profit de la Business Source Licence (BSL³⁸). Les deux licences autorisent la copie, la diffusion et la modification ou, pour le dire autrement, elles permettent de modifier, d'afficher, d'exécuter, de distribuer et d'exploiter le

³⁷ Proposition de règlement du parlement européen et du conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'union, publié le 21 avril 2021, [n°2021/0106 \(COD\)](#) depuis adopté ; Juliette Sénéchal, « l'IA Act déjà obsolète face aux IA de nouvelle génération ? L'exemple de ChatGPT », Actualité Dalloz, 1er février 2023.

³⁸ The Licensed Work is (c) 2023 HashiCorp, Inc. Disponible sur le site : <https://www.hashicorp.com/bsl>

logiciel³⁹. La différence entre les licences tient aux restrictions d'usage du logiciel. La MPL 2.0 ouvre la possibilité pour l'utilisateur de réaliser toute exploitation sans restriction à titre gratuit. En ce sens la MPL 2.0 est qualifiée de licence libre. Sa particularité tient essentiellement à la possibilité de diffuser sous deux licences différentes les combinaisons du logiciel avec des éléments extérieurs⁴⁰. La BSL est beaucoup plus restrictive puisqu'elle ne permet pas l'usage à des fins de production, ce qui limite considérablement l'usage commercial. Il faut en ce cas obtenir une licence supplémentaire (§1). Les modifications qui sont apportées par les licenciés sont soumises entièrement à la BSL (§5). La rupture avec la diffusion libre par HashiCorp n'est pas si nette qu'il n'y paraît. HashiCorp entend, avec la BSL, se ménager un temps suffisant pour renforcer l'assise commerciale du projet sans nécessairement renoncer à une diffusion libre. Aussi, la licence prévoit que les restrictions relatives à usage commercial sont lever au bout de quatre ans à compter de la date de modification ou de la première distribution publique d'une version spécifique du travail sous licence BSL (§2). De plus, on retiendra que la diffusion dans un premier temps de Terraform sous la licence libre a permis à une communauté de développeurs de créer une version dissidente (« de forker ») au sein du projet OpenTF. Celle-ci est disponible sous la licence publique Mozilla et elle est chapeautée par la fondation Linux et la CNSF (Cloud native software foundation).

24. **L'exemple de Mistral** (Mistral 7B et Mixtral 8x7B). La société Mistral propose des modèles d'intelligence artificielle qui ont été présentés comme une alternative crédible aux leaders américains tel que OpenAI ou Meta. Outre le fait que Mistral AI est une startup française, ses concepteurs se targuent que leur technologie est disponible gratuitement et sans restriction pour les développeurs. Sur le site de l'entreprise, deux modèles d'IA, Mistral 7B et Mixtral 8x7B, sont effectivement disponibles sous la licence Apache 2.0 (APL 2.0), tandis que les autres versions sont disponibles sous des licences propriétaires. L'APL 2.0 a pour objectif de permettre aux utilisateurs d'utiliser le logiciel à n'importe quelle fin à condition toutefois de conserver les mentions relatives à l'attribution, et ce, dans le monde entier. Peu contraignante, l'APL est une licence libre de type domaine public⁴¹ ou dite permissive. Elle autorise l'utilisation, la diffusion, et la modification du logiciel dans sa version initiale et ses versions modifiées. Si, à défaut de précisions contraires, les modifications réalisées par le licencié sont soumises aux termes de la licence (section 5), une fois modifié, le logiciel peut être redistribué sous une autre licence avec pour seule contrainte de conserver les mentions relatives à l'attribution (section 4 *in fine*) de sorte que les versions dérivées peuvent être distribuées sous licences propriétaires. A notre connaissance, la société Mistral n'ajoute pas d'autre restriction à l'usage qui peut être fait de Mistral 7B et Mixtral 8x7B. En revanche, les autres versions de leur IA générative sont disponibles selon les termes de licences propriétaires.

25. **L'exemple de Meta** (à propos d'une série de quatre modèles de type LLMs sous le nom de LLaMA). La licence LLaMA⁴², sous laquelle sont disponibles ces modèles, est parfois présentée comme une licence *open source* bien que cette qualification soit contestée. En effet, si cette licence permet l'accès au code source, la diffusion et la modification, ces libertés sont toutefois écartées dans deux hypothèses. Tout d'abord, l'utilisation du matériel Llama ou toute

³⁹ Voir le premier paragraphe de la BSL et la clause 2 de la MPL.

⁴⁰ Le résultat de cette combinaison est appelé « Larger Work », voir les clauses 1.7 et 3.3.

⁴¹ « L'œuvre libre » (refonte), *JurisClasseur Propriété littéraire et artistique*, fasc. n° 1975, 2024

⁴² Llama 2 Version Release Date : July 18, 2023

production ou résultat du matériel Llama n'est pas permise pour améliorer un autre modèle de langage étendu à l'exclusion de Llama 2 ou des travaux dérivés (clause 1.v) ; ensuite, la licence ne couvre pas l'hypothèse où, à la date de sortie de la version Llama 2, le nombre d'utilisateurs actifs mensuels des produits ou services mis à disposition par ou pour le licencié, ou les affiliés du licencié, est supérieur à 700 millions d'utilisateurs actifs mensuels au cours du mois civil précédent. Autrement dit, dans ce cas, l'utilisateur devra obtenir une licence particulière de Meta. Ces restrictions ont pour finalité de ne pas fournir aux concurrents le bénéfice de la technologie. Pour autant, certains considèrent que s'il est légitime de choisir une licence qui soutient un modèle économique, il est, en revanche, décevant de parler de licence *open source* lorsque les autorisations accordées par la licence sont en définitif très restrictives.

26. **L'exemple d'OpenAI** (à l'origine de l'agent conversationnel [chatbot] [ChatGPT](#) et des modèles de langage GPT-3 et [GPT-4](#)). A ses débuts en 2015, OpenAI, qui est une organisation à but non lucratif, expliquait vouloir une « recherche ouverte », et revendiquait une philosophie open source. On peut encore lire sur les pages de son site : « *Les chercheurs seront fortement encouragés à publier leurs travaux, que ce soit sous forme d'articles, de billets de blog ou de code, et nos brevets (s'il y en a) seront partagés avec le monde entier* ». En 2019, OpenAI dévoile la création d'une nouvelle société : OpenAI LP⁴³ qui doit lui permettre d'obtenir de nouveaux investissements en particulier provenant de la société Microsoft. S'agissant de la licence applicable à ChatGPT, le choix initial s'est porté sur la licence Apache 2.0⁴⁴. Toutefois, OpenAI a signé en 2020 avec Microsoft une *licence exclusive du modèle de langage GPT-3* ». Aujourd'hui, GPT-4 n'est utilisable qu'avec un mode d'accès contrôlé, via une API fournie par OpenAI. L'abandon de la diffusion en *open source* par OpenAI a été très remarqué et justifié de manière, peut-on dire, maladroite par Ilya Sutskever, cofondateur de la structure. Ce dernier a fait valoir les dangers qu'entraîne une diffusion libre de ces outils IA : il souligne le risque qu'ils tombent entre les mains de personnes malveillantes de sorte qu'ils nuisent à la société ou encore bousculent certaines professions. On peut toutefois raisonnablement penser que la recherche d'un modèle économique viable a motivé l'abandon, au moins provisoirement, de l'accès ouvert.

27. **Le grand modèle de langage Falcon 180B** de l'Institut d'innovation technologique des Émirats arabes unis. La licence sous laquelle le LLM est disponible est [la licence TII Falcon 180B, version 1.0](#) qui est inspirée de la licence Apache 2.0⁴⁵. Elle contient les conditions d'utilisation, de redistribution et de modification que l'on trouve habituellement dans les licences libres de logiciels. En revanche, elle est présente deux restrictions originales. La première consiste à exiger que les modifications fassent l'objet d'une soumission au concédant pour être incorporées à l'œuvre (art. 1). La seconde a pour objet d'interdire l'utilisation à des fins d'hébergement (art. 9). En conséquence, une telle utilisation est conditionnée à l'octroi d'une licence spéciale (art. 9.3). L'institut d'innovation technologique des Emirats arabes unis semble, par ce mode de diffusion semis ouvert, tenter de trouver un juste équilibre entre l'ouverture qui permet d'associer des programmeurs extérieurs, et la fermeture dans l'éventualité de débouchée économique qui le nécessiteraient.

28. A travers ces exemples, nous pouvons percevoir la difficulté que rencontrent les entreprises à élaborer un modèle économique à partir d'une technologie disponible sous licence libre dès lors que cette dernière, en l'occurrence l'intelligence artificielle, nécessite des

⁴³ <https://openai.com/blog/openai-lp>

⁴⁴ <https://github.com/jianzhnie/open-chatgpt/blob/main/LICENSE>. Licence présentée dans le paragraphe consacré à Mistral.

⁴⁵ Notre contribution : « Pourquoi le grand modèle de langage Falcon 180B est-il en accès libre ? », Dalloz Actualité IT, mars 2023.

investissements financiers considérables et s'inscrit dans un secteur très concurrentiel. La modulation de l'accès et de la réutilisation des logiciels n'est pas un phénomène nouveau. Nonobstant l'effort que l'on peut déployer pour classer les licences de logiciels en catégories, les nuances sont infinies et l'exemple de conditions d'accès des technologies IA est un exemple de plus. Le flou qui en découle est parfois nourri par les entreprises à des fins de communication. L'écueil d'une telle situation est qu'elle induit en erreur les utilisateurs. La dilution des notions de licence libre ou de licence en *open source* peut également être redoutable pour la sécurité des entreprises qui ont recours aux logiciels soi-disant disponibles, mais dont les restrictions d'usage entravent considérablement l'exploitation économique.

29. Il ne faudrait pas que la Science ouverte contribue à cet imbroglio. Or le risque est réel car après tout, les exemples cités révèlent que les entreprises adoptent l'adage qui préside la Science ouverte, en retenant un mode de diffusion « aussi ouvert que possible » afin de bénéficier des contributions extérieures et de l'aura du libre, mais « aussi fermée que nécessaire » pour se parer contre les assauts des concurrents et attirer les investisseurs. Ce constat me permet de conclure qu'il existe potentiellement de nombreuses modalités de mise à disposition des résultats de la recherche qui répondent au concept de la Science ouverte qui couvre le spectre allant des licences libres aux licence telle que la Business Source Licence⁴⁶. L'écueil peut être évité si la Science ouverte, dont le régime n'est pas encore consacré, s'ancre à la croisée du libre et de l'Open data, et contribue à clarifier les modalités de diffusion ouverte au moins dans le secteur de la recherche.

⁴⁶ Voir le paragraphe consacré à Terraform.