# ChatGpt et autres « modèles » d'IA dits génératifs – vers une réponse réglementaire européenne ?

Poullet, Yves

*Publication date:*
2023

*Document Version*
le PDF de l'éditeur

LARCIER INTERSENTIA



**BRÈVE**

# ChatGpt et autres « modèles » d'IA dits génératifs – vers une réponse réglementaire européenne ?

*Yves Poullet*

**Mots-clés**

Modèles de base – Amendements de l'AI Act de la part du Parlement européen – Définition de l'IA – Obligations des fournisseurs de modèles de base.

Foundation models – AI Act proposal amended by the EU Parliament – AI Definition – Obligations of the 'Foundat model' providers



**Résumé**

La brève note tente d'éclaircir la distinction entre les modèles et les applications de systèmes d'intelligence artificielle générative et reprend les débats sociétaux à leur propos. Il décrit la façon dont la régulation de ces modèles exige, selon le Parlement européen, une redéfinition de l'IA mais également une réglementation « sui generis » des risques liés à ces modèles.

The brief note attempts to clarify the distinction between models and applications of generative artificial intelligence systems and takes up the societal debates about them. It describes how the regulation of these models requires, according to the European Parliament, a new definition of AI but also a 'sui generis' regulation of the risks associated with these models.

**Texte**

Le 22 novembre 2022, Open AI lance des applications grand public de son système Chat GpT. Trois ans avant, Sam Altman, CEO d'Open AI, affirmait cependant que son entreprise s'interdisait de promouvoir auprès du grand-public de telles applications au regard des risques liés à ces systèmes dits de « generative AI ». L'initiative connaît un franc succès : un million d'utilisateurs la semaine suivante. Depuis, les applications se sont multipliées et l'angoisse de certains face aux risques liés à ces systèmes, accrue. On connaît la lettre ouverte signée par plus de 35.000 personnes dont nombre d'experts et d'acteurs majeurs du domaine, à l'initiative de l'Institute for the Future of Life, réclame un

moratoire de 6 mois sur les développements en cours, afin de trouver une réponse adéquate aux risques soulevés par ces systèmes. Une autre lettre soulignait l'urgence d'une réaction des décideurs politiques pour sauver la race humaine. Face à ces déclarations alarmistes, notre propos est d'analyser la proposition de réponse récemment avancée par les autorités européennes.

D'emblée, nous souhaitons introduire une distinction entre les modèles et les applications de « generative AI » pour éviter des confusions trop souvent entendues et expliquer les propositions actuellement discutées dans le cadre de la proposition de règlement européen sur l'intelligence artificielle, connue comme AI Act[1] . Notre propos n'analyse que la réglementation des modèles à savoir les systèmes génératifs « génériques », les « foundations models », qui rendent possibles un large éventail d'applications que notre brève n'abordera pas. Pour revenir aux modèles, on parle de « transformers models » qui permet à un système AI d'exploiter des relations entre éléments d'un ensemble de données sans que ces relations soient dictées au préalable par l'humain , de « Large Language models » (LLM) à propos d'agrégations opérées par machine d'éléments textuels, de « multimodal languages », capables de travailler à la fois sur des textes, des images et des sons. Pour ne reprendre que le cas de LLM[2] , ils sont capables de corréler des éléments de langage, présents dans des textes ou des exposés oraux, bref de « comprendre » un texte mais également de produire un document oral ou écrit[3]. Ces LLM s'appuient sur des bases de données multiples regroupant des milliards de données venant tant de sources publiques y compris Wikipedia ou des revues, de sources privées en provenance de leurs propres activités (ainsi Google entend ajouter les données disponibles dans son moteur de recherche ainsi que les requêtes y relatives) que des réponses aux « prompts » (questions) adressés par les clients de ces applications. Le nombre de paramètres permettant les corrélations dépasse, dans les LLM importantes, largement le milliard, ce qui explique la puissance de ces outils mais également leur consommation d'énergie (Une récente recherche reprise par l'étude de l'OCDE estime que le seul entraînement de modèles prod 284 000 kilogrammes of $CO_2$, l'équivalent de l'émission de $CO_2$, de 5 voitures tout au long de leurs vies). Les applications des modèles de langage sont nombreuses et peuvent être le fait tantôt des entreprises ayant dévelo ces modèles, tantôt d'entreprises spécialisées. Elles concernent tantôt la conversion de textes en paroles, tantôt l réponse à des questions (« prompts ») y compris dans le cadre de services de « companion chatbots », tantôt les assistants virtuels par exemple pour la confection de logos et messages publicitaires, tantôt la reconnaissance de la parole...



Lors de leur réexamen de la proposition AI Act déposée par la Commission, le 21 avril 2021, à la fois le Conseil et le Parlement[4] ont saisi l'occasion pour se saisir des questions posées par les « foundations models ». Le premier constat concerne l'inadéquation de la définition donnée par le texte de la Commission pour couvrir ces systèmes d'IA dits de « finalité générale ». En effet, le texte initial définit l'IA notamment par ses finalités spécifiques. Or, comme déjà précisé, les « foundations models » poursuivent une finalité générale qui, certes, permettra l'éclosion de multiples applications, elles, liées à des finalités particulières. Ainsi, tant le Conseil que le Parlement modifie la définition de la notion d'intelligence artificielle et en outre proposent une définition du concept de « fondation model ». Pour reprendre la formulation du Parlement européen dans l'accord de compromis publié le 14 juin 2023, l'article 3.1 définit l'intelligence artificielle en supprimant la référence « for a given set of human-defined objectives » comme suit « ''artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments ; ». Cet élargissement fait explicitement référence aux « foundations models » défini par l'article 3.1 (c): « *''*'foundation model" *means an AI system model that is trained on* broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks;"» suivi par une précision sur la notion de système à finalité générique (article 3.1 (d)):

« "general purpose AI system" means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed ».

Si Conseil et Parlement peuvent s'entendre sur la nécessité de nouvelles définitions et leur contenu, leurs positions en ce qui concerne la régulation de ces modèles, divergent. Le Conseil privilégie leur classement comme systèmes à haut risque et, sous réserve de quelques dispositions dérogatoires favorables à l'innovation que représente ces modèles, leur applique le régime déjà prévu. Le Parlement préfère définir un régime « *sui generis* » pour les fournisseurs de tels modèles, régime plus léger pour les mêmes raisons que celles invoquées par le Conseil. Ce régime ne s'applique pas aux applications des systèmes génératifs qui peuvent être bien évidemment « à haut risque », lorsque leur finalité les fait rejoindre les listes reprises aux annexes 2 et 3, par ailleurs étendues[5] ou entrer dans le champ d'application de l'article 52 qui exige la transparence de système AI, applicable par exemple dans le cadre de « deepfakes ». Notre propos se limitera à un bref examen des propositions du Parlement propres aux modèles, étant entendu qu'à propos des applications, un commentaire sur l'ensemble des dispositions de l'AI Act serait nécessaire.

L'article 4 applicable à tous les systèmes d'IA, qu'ils soient génératifs ou autres, modèles ou non retiendra cependant notre attention, eu égard à son importance. Les principes éthiques affirmés par le groupe d'experts de haut niveau[6] sont enfin traduits dans le texte et s'appliquent également aux *foundations models*: « All operators falling under this Regulation shall make their best efforts to develop and use AI systems or foundation models in accordance with the following general principles establishing a high-level framework that promotes a coherent human-centric European approach to ethical and trustworthy Artificial Intelligence, which is fully in line with the Charter as well as the valu on which the Union is founded ». On connaît ces principes que l'on retrouve dans nombre de textes relatifs à l'éth de l'IA, en particulier la recommandation sur l'éthique de l'IA de novembre 2022[7] : ceux de l'agent et du contrôle humain, de la sécurité, de la vie privée et de la gouvernance des données, de la transparence, de la non-discrimination, du bien-être social et environnemental.



L'article 28 B énonce les obligations particulières de tout « fournisseur des modèles, qui souhaite mettre sur le marché son produit peu importe qu'il soit incorporé ou non dans un système ou un robot, qu'il soit en *open sour* ou non. La première obligation est essentielle, elle découle du principe d'*accountability* qui impose à celui qui développe un produit à risque à la fois de prendre les précautions nécessaires à la fois en évaluant les risques liés à son produit et aux produits dérivés envisageables et à la fois de démontrer l'utilisation de divers moyens de les réduire. Elle suppose la mise sur pied d'un organe de gestion de la qualité. Le texte mentionne les risques à prendre en compte et ce bien au-delà des seuls risques relatifs à nos libertés individuelles, ainsi les risques encourus par la santé, ceux collectifs de discrimination ou d'atteinte à la justice sociale mais également les risques de non-respect de la règle de droit. Sans doute, cet élargissement des risques à prendre en compte est-il affirmé pour tous les systèmes IA mais son rappel ici répond à la constatation de l'ampleur des risques liés aux « generative AI » : sans être exhaustifs, la possibilité pour chacun de nous de produire de « fausses vérités » ou des *deepfakes* constitue un risque majeur pour nos démocraties ; l'exclusion de certaines langues et donc de certaines cultures, à défaut d'être représentées dans les *big data* à prédominance anglo-saxonne voire américaine en est un autre, enfin, on a déjà signalé la consommation massive d'énergie liée à la gourmandise de ces modèles. Le devoir d'évaluation va jusqu'à documenter les risques résiduels non susceptibles d'être couverts. Les moyens de réduire sont énoncés : le *design* des systèmes mais surtout le *testing* des systèmes afin d'éviter les biais et de s'assurer de la qualité, de la représentativité et de la complétude des données. Enfin, toutes ces opérations nécessitent, le cas échéant, le recours à des experts.

Au-delà, le même article impose que le design du modèle et son développement permettent de s'assurer tout au

long de la vie du système de son efficacité, de la possibilité de prévoir, d'interpréter et de corriger les résultats. Il exige de n'utiliser que des bases de données sujettes à des règles de gouvernance et appropriées. L'obligation de documenter le système et de mettre à disposition des utilisateurs (en particulier aux développeurs d'applications fondées sur ces modèles) les instructions nécessaires à leur propre respect de leurs obligations. Enfin, l'obligation de veiller à la réduction de la consommation d'énergie et à l'efficience des systèmes est expressément prescrite. On ajoutera trois obligations additionnelles applicables tant aux fournisseurs de modèles qu'à ceux qui développent des applications d'AI generative en matière de sons, d'images, de textes: la première est l'obligation d'informer leurs utilisateurs de l'utilisation de telles applications; la deuxième est de veiller au respect des libertés individuelles, en particulier la protection des données et la liberté d'expression ; la troisième est de publier les sources de données protégées par les droits de propriété intellectuelle, utilisées lors des procédures d'entraînement.

Concluons, les autorités européennes ont découvert, avec le déferlement des Chat GpT et autres applications, les enjeux majeurs des modèles qui autorisent le foisonnement de leurs applications. La réaction proposée, dont on note qu'elle suit celle chinoise précurseur en la matière[8], tente de concilier innovation, compétitivité européenne et devoir de précaution. La réaction est-elle trop hâtive ? Les autorités européennes répondent par l'urgence de répondre aux risques majeurs de cette technologie, dans le même temps, elle se montrent conscientes du fait que le débat ne fait que commencer, elles reconnaissent l'absence de standards[9] en la matière à la différence de ce qui existe à propos d'autres types d'intelligence artificielle. Par ailleurs, elles s'en remettent à l'AI Office, institution nouvelle proposée par le Parlement européen, rassemblant des experts européens chargés de suivre les évolutions de l'IA et de conseiller les autorités européennes sur les mesures à prendre[10]. Les débats sur les modèles génératifs ne font que commencer et on peut souhaiter que ceux-ci soient ouverts, multidisciplinaires et *multistakeholders* à hauteur des enjeux de ces technologies...



***1***. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 f 21 avril 2021, COM(2021) 206 final, disponible en ligne sur https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2021:206:FIN (consulté le 26 mai 2021).*

***2***. *Sur ces LLM, lire l'excellente étude proposée par l'OCDE, « AI Language Models Technological, Socio-Economic and Policy Considerations », OECD Digital Economy Papers, avril 2023, n° 352.*

***3.*** *Sans être exhaustif, loin de là, Google a ainsi développé BERT (« Bi-directional Encoder Representations from Transformers ») ; ChatGpt (Generative Pre-Trained Transformer) a été développé par Open AI et est appuyé par Microsoft ; Baidu, une Bigtech chinoise développe ERNIE (« Enhanced Representation through kNowledge IntEgration »); Meta, « Open Pre-Trained Transformer » (OPT-175B) et LLamA.*

***4.*** *La position du Conseil des ministres européens date du 6 décembre 2022. Elle considère les « Foundation models » comme des systèmes à haut risque auxquels sauf exceptions s'appliquent les mêmes dispositions. Les parlementaires européens ont voté les amendements sur l'AI Act (règlement sur l'intelligence artificielle) le 14 juin 2023 et considèrent par contre les modèles comme des systèmes nécessitant une réglementation particulière. Nous n'avons pu trouver le texte en langue française. Les références au texte sont en langue anglaise.*

***5.*** *Ainsi, la création d' avatars, basés sur des personnalités réelles comme proposée par META dans son projet Metavers, pourrait relever des IA à haut risque, si on suit la proposition du Parlement (voy. l'amendement n°215 de l'article 5 §1, a) qui considère que constituent des systèmes à haut risque, les IA ayant « pour objectif ou […] pour effet d'altérer substantiellement le comportement d'une personne ou d'un groupe de personnes en portant considérablement atteinte à la capacité de la personne à prendre une décision éclairée, l'amenant ainsi à prendre une décision qu'elle n'aurait pas prise autrement »*

***6.*** *HLGE (High Level Group of experts) on AI, Lignes directrices en matiere d'ethique pour une IA digne de confiance, 8 avril 2019, n° 67,*

*texte disponible sur le site : Ethics guidelines for trustworthy AI - Publications Office of the EU (europa.eu)*

***7.*** *Recommandation sur l'éthique de l'IA, UNESCO, novembre 2021. Cette recommandation a été adoptée par 193 pays. Un rapport récent de l'UNESCO (« Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence, UNESCO, Juin 2023 ») analyse comment cette recommandation pourrait s'appliquer aux systèmes génératifs.*

***8.*** *Les « Interim Administrative Measures for Generative Artificial Intelligence Services » adoptées lors de la 12e réunion de l'administration du Cyberespace, le 23 mai, ont été approuvées par divers ministères et commissions et ont pris effet le 15 août. Aux modèles européen et chinois de réglementation publique, on opposera celui américain fondée plus sur une corégulation née du dialogue entre les autorités publiques et les entreprises privées (voy. à cet égard le « T*he Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People *», publié par le White House Office of Science and Technology Policy en octobre 2022).*

***9.*** *A noter que les amendements du Parlement européen réclament que les organisations de standardisation élargissent leurs préoccupations au-delà des seules questions techniques et s'ouvrent à des représentants d'autres disciplines.*

***10.*** *« As foundation models are a new and fast-evolving development in the field of artificial intelligence, it is appropriate for the Commission and the AI Office to monitor and periodically assess the legislative and governance framework of such models and in particular of generative AI systems based on such models, which raise significant questions related to the generation of content in breach of Union law, copyright rules, and potential misuse » (Considérant 60 f, amendement 100 de la proposition du Parlement européen du 14 juin 2023) .*



## Service & Contact