La générativité formelle et statistique, nouvelle donne pour les données

Andreas Pfeiffer et Jean Lassègue

> Si un jour les machines électroniques écrivaient des pièces de théâtre parfaites, peignaient des tableaux inimitables, il y aurait à se poser de sérieuses questions. Si elles se mettaient à aimer, le sort de l'espèce zoologique serait réglé[1].

Les pages qui suivent ont pour objectif de proposer quelques remarques générales sur l'intelligence artificielle générative telle qu'elle est en usage dans les domaines du texte, de l'image et seulement latéralement, du son, pour des raisons qui seront rendues explicites plus loin. Il s'agira donc de mener de front une analyse épistémologique des logiciels d'intelligence artificielle générative et une analyse sociologique de leurs effets. Après voir montré que les trois cas de figure du texte, de l'image et du son se distinguent du point de vue de la compétence dans l'écriture qu'ils requièrent, on tentera la question de l'impact que ces techniques pourraient avoir à terme sur ce que tout un chacun est en droit d'attendre des ressources du langage qui sont très directement visées par ces nouvelles technologies. Nous finirons par des remarques sur la créativité et son hybridation numérique à l'âge de l'intelligence artificielle.

L'avènement de la générativité formelle et statistique

Différence de « littéracie » entre les domaines du texte, de l'image et du son

En observant l'engouement sans précédent pour l'intelligence artificielle générative, et en particulier pour la production textuelle de *chatGPT*[2], on ne peut pas ne pas être frappé par le fait que l'usage de l'intelligence artificielle générative dans le son et la musique ne suscite pas le même intérêt dans le public. Malgré des développements impressionnants (dont nous ne décrirons pas ici

[1] André Leroi-Gourhan, *Le Geste et la parole II. La mémoire et les rythmes*, Paris, Albin Michel, 1965, p. 76.

[2] Ayant atteint 100 millions d'utilisateurs en deux mois, l'interface de dialogue d'*OpenAI* est considérée comme la technologie la plus rapidement adoptée dans l'histoire de l'informatique.

les évolutions faute de place et que d'autres articles du numéro décrivent en revanche en détail), on peut se demander pourquoi l'IA générative dans le son et la musique ne fait pas autant événement et passe presque inaperçue, alors que *chatGPT* échauffe les esprits en remettant en question, ici, l'honnêteté des examens universitaires, là la nécessité des avocats.

La raison principale de la différence entre la génération de texte et celle de la musique semble tenir aux compétences requises dans l'écriture : à peu près tout le monde, dans les sociétés développées au moins, a appris à écrire sa langue mais pas à écrire de la musique. Dès lors, un système informatique qui semble capable d'écrire sans intervention humaine touche les individus de façon autrement plus puissante qu'un système informatique écrivant « tout seul » de la musique. Et on comprend aussi que le fait d'avoir à partager cette capacité d'écriture de la langue avec des logiciels génératifs soit beaucoup plus angoissante que le fait d'écrire ou pas de la musique. C'est donc bien la question de la capacité à écrire qui fait le fond de la question car les logiciels génératifs produisant de la musique sont logés à la même enseigne que les logiciels génératifs produisant du texte dès lors que des questions de propriété sont en jeu : quand il s'agit d'authentifier une production en vue de lui reconnaître une valeur juridique (travail personnel de l'étudiant dans notre exemple), la musique ne se distingue pas des productions textuelles (production par voie générative d'une chanson utilisant sans autorisation le nom d'un artiste par exemple[3]).

Si l'on met donc de côté la question juridique de la propriété et sa financiarisation possible et que l'on s'en tient à celle des compétences dans la « littéracie » où la différence entre texte et musique semble jouer à plein, on doit essayer de comprendre pourquoi la différence de compétence dans l'écriture de textes ou de la musique peut laisser à penser que certains métiers du texte comme ceux d'étudiant ou d'avocat seraient remis en question par l'IA générative mais pas certains métiers de la musique comme celui de compositeur.

La réponse nous semble être que ceux qui savent écrire leur langue se trouvent néanmoins confrontés, du fait du caractère ubiquitaire du langage, à des situations d'expression qu'ils ne maîtrisent pas et qu'ils sont ainsi conduits à vouloir

[3] Le cas de la fausse chanson *Heart on My Sleeve* utilisant la voix du rappeur canadien Drake mise en ligne en avril 2023 sur les grandes plateformes (YouTube, TikTok, Spotify) est éclairant à cet égard : immédiatement écoutée par des millions d'individus, elle a entraîné tout aussi immédiatement une violente réaction des plateformes qui l'ont supprimé de leurs contenus mais aussi de la part du label de l'artiste, Universal Music Group, qui a menacé de rétorsion sévère toute personne s'avisant de produire des chansons en utilisant, au moyen de l'IA générative, la voix des artistes se produisant sous son label. Document en ligne consulté le 13 août 2022 : <https://www.theguardian.com/music/2023/apr/18/ai-song-featuring-fake-drake-and-weeknd-vocals-pulled-from-streaming-services>.

réaliser des tâches *qu'ils ne savent pas faire* (résumer par écrit une situation complexe), *qu'ils ne maîtrisent pas* (traduire une langue inconnue) ou *dont ils voudraient se débarrasser au plus vite* (écrire un texte de style administratif en vue d'obtenir une bourse). Les nouveaux systèmes d'IA générative sont alors plébiscités. Le sentiment de compétence serait ainsi inversement proportionnel à l'universalité du moyen d'expression : l'usage du langage, du fait de la polyvalence de son usage, rendrait le recours à l'aide logicielle générative particulièrement apprécié dans des tâches spécialisées. Au contraire, les nombreux systèmes qui utilisent l'intelligence artificielle dans le contexte musical s'adressent au premier chef à tous ceux qui font *déjà* de la musique et dans toute la diversité de ses aspects. La question revient alors à ceci : pourquoi des millions de personnes veulent créer et partager des textes ou des images alors que les outils pour la musique générative n'attirent que ceux qui possèdent déjà une compétence musicale affirmée ?

Il est difficile de répondre dans l'abstrait sans en passer par une enquête sociologique de terrain sur l'étude des pratiques et des collectifs qui se créent autour de ces pratiques, car il semble que, dans la socialité propre à la musique, il soit tout à fait possible d'échanger des morceaux de musique comme on s'échange des images mais qu'il soit beaucoup plus difficile de *créer* de la musique à l'aide des logiciels numériques quand on n'est pas un peu du métier. La différence d'appréciation des nouveaux logiciels dépend donc bien du niveau de « littéracie » du milieu social dans lequel le logiciel est utilisé car on peut constater que ce sont les mêmes techniques d'apprentissage par la machine, apprentissage dit « profond », les mêmes réseaux neuronaux, qui traitent, selon les buts des développeurs, images, texte, sons et musique. C'est donc la question de la littéracie qui est centrale et pour laquelle trois cas doivent être distingués.

Dans le cas des logiciels de génération de texte, le fait de vivre dans des sociétés massivement alphabétisées induisant une capacité généraliste d'écriture rend difficile de réaliser de nombreuses tâches spécialisées et qui sont pourtant socialement requises. L'usage massif de l'IA générative se double alors d'un profond sentiment de *dépossession* qui se manifeste sur le mode mythologique du dépassement de l'humain par la machine, entendez par un usage mécanique de l'écriture.

Dans le cas des logiciels d'image, le rôle fondamental du *text prompt* (en français une « consigne de texte », c'est-à-dire ce que l'on doit donner comme instructions à l'interface de dialogue pour démarrer une recherche et attendre une réponse) a lui aussi commencé par susciter la crainte de voir les humains dépassés. Mais avec un peu de recul, on s'est vite rendu compte que ces logi-

ciels déplaçaient seulement la question de la compétence : la compétence *linguistique* se substitue à la compétence *gestuelle*. Cela montre bien que l'enjeu est celui de la maîtrise d'une nouvelle forme de « littéracie » et non des techniques de la peinture, ni même d'ailleurs de son contenu, dont on se rend souvent compte, passé le premier sentiment d'ébahissement devant la perfection froide des images, que ce contenu a perdu toute aura picturale. Ce déplacement de compétence vers le linguistique entraîne donc plutôt une compétition sociale entre ceux qui maîtrisent l'art du *text prompt* pour une tâche donnée et les autres : on risque non pas d'être remplacé par un logiciel devenu supra-humain mais par un autre humain, plus compétent.

Dans le cas des logiciels du son et de la musique enfin, on doit remarquer que la situation est complexe car, du fait de leur digitalisation, ces logiciels sont curieusement à la fois *en avance* et *en retard* sur le texte et l'image. En *avance* parce que le numérique a commencé à prendre une place importante dans la production sonore dès les années 1970-1980 avec l'apparition du CD musical et que les outils de composition faisant appel aux techniques de l'intelligence artificielle existent au moins depuis les années 1990[4]. En *retard* cependant quant à leur reconnaissance par le public, reconnaissance qui fait aujourd'hui le succès de *chatGPT* ou de *Dall-E* du fait que la maîtrise de l'écriture musicale reste encore confidentielle et ne pose pas encore le problème du rapport au langage *via* le *text prompt*, même si ce problème se posera dans les années à venir, du fait du développement des logiciels langage-musique. C'est donc moins ce dernier cas qui sera examiné dans les pages qui suivent parce que, pour une fois, nous ne sommes pas en retard d'une technologie puisque celle-ci est encore embryonnaire.

Le fond de l'affaire : de la réécriture formelle de règles à la conversation en ligne

Une fois cette différence entre les cas du texte, de l'image et du son clarifiée du point de vue de leur rapport à l'écriture et à ses conséquences sociales, on voudrait maintenant se situer à un niveau épistémologique au plus près des données pour ainsi dire, avant qu'un traitement ne les transforme en objets numériques, sonores, visuels ou textuels. Les données ne sont, en effet, ni « sonores », ni « visuelles », ni « textuelles », ce sont d'abord, selon le niveau de compilation où l'on choisit de se situer, des suites de nombres et plus profondément encore, dans le modèle de la machine de Turing au moins, des suites de traits ou d'espaces vides sur les cases du ruban de la machine mathématique

[4] Document en ligne consulté le 2 mars 2023 : <https://computerhistory.org/blog/algorithmic-music-david-cope-and-emi/>.

imaginée par Turing. Bref, ce sont des marques graphiques définissant un degré zéro de l'écriture qui ne vise pas à *restituer* quoi que ce soit, son, parole, image, etc. – comme c'était le cas jusqu'à présent pour toute écriture – parce que l'écriture informatique est fondée sur une *séparation stricte des marques graphiques et du sens* qu'elles évoquent, qu'il soit visuel, sonore ou textuel. De ce point de vue, l'écriture informatique est autonome du fait qu'elle ne vise pas à *représenter* une réalité extérieure à elle-même. L'écriture informatique est *muette* mais on se débrouille techniquement *pour faire croire qu'elle peut parler*, c'est-à-dire qu'on fait comme si elle pouvait annexer l'écriture des langues en produisant d'elle-même des énoncés *sensés* – ce qu'elle ne peut en aucune façon sans l'intervention de la capacité humaine d'interprétation. Nous allons nous intéresser ici à la façon dont ce retour au sens est effectué, c'est-à-dire à la façon dont, en s'élevant dans les niveaux de compilation, les données deviennent à proprement parler sonores, visuelles ou textuelles. Mais il faut ajouter un point concernant l'écriture, pour bien comprendre la situation numérique que nous vivons.

Les marques graphiques dépourvues de sens qui sont au fondement de l'écriture informatique peuvent être classées en deux catégories selon qu'il est possible de les concaténer de façon absolument déterministe ou selon que la concaténation s'opère de façon seulement statistique. Dans le premier cas, la concaténation déterministe des marques graphiques s'effectue à partir d'un alphabet fini et définit à proprement parler la notion de calcul. Cependant, à ce niveau pourtant intégralement déterministe, il y a des concaténations de marques graphiques *structurellement inaccessibles au calcul* comme l'ont prouvé les grands théorèmes de limitation des années 30 du siècle dernier (Gödel, Church, Turing). Dans le cas statistique, c'est seulement de façon asymptotique que la concaténation se rapproche de l'idéal déterministe du calcul. Ces deux cas de limitation, celui de la concaténation structurellement inaccessible au calcul et celui de la concaténation statistique ont été rapprochés par l'intermédiaire des modèles informatiques d'apprentissage [5] : l'apprentissage consiste, dans un environnement globalement incertain et muni de ressources limitées de calcul, à exécuter des tâches en vue de produire des résultats en un temps fini. De ce point de vue, le rapport que les modèles informatiques d'apprentissage entretiennent avec le sens est à la fois formel *et* statistique. Il s'agit d'une nouvelle manière d'envisager le niveau graphique, seul niveau où se situe à proprement parler l'écriture informatique : il s'agit soit d'engendrer toute la configuration graphique soit de prévoir la probabilité de la

[5] C'est ce que montre par exemple Leslie Valiant dans *Probablement approximativement correct*, Paris, Cassini, 2018.

configuration graphique suivante à partir d'un énorme stock de ces configurations.

C'est ce rapport formel et statistique qui nous prend souvent au dépourvu parce que nous ne vivons pas du tout notre rapport au monde sur le mode de la conjonction du formel et du statistique qui font tous les deux abstraction du sens. Le retour au sens consistant à interpréter un résultat s'opère donc, dans le cas des logiciels dont nous allons parler, à partir d'une structure très particulière, formelle et statistique, et c'est cette restitution qui vient se heurter à la construction du sens telle qu'elle est pratiquée dans les communautés humaines. Il devient plus facile, à partir de cette remarque, de saisir la façon dont commence à s'opérer l'intégration sociale de ces logiciels formels-statistiques.

L'année 2022 semble avoir marqué un tournant de ce point de vue : après avoir suscité depuis déjà une décennie au moins un très fort engouement dans la communauté des chercheurs après de nombreuses éclipses, les techniques d'apprentissage par *deep learning* atteignent, en quelques mois seulement, le statut d'événement culturel et médiatique. Que s'est-il passé pour que ces techniques sortent des laboratoires et deviennent un phénomène de société ? C'est, à notre avis, la qualité de *leur interface conversationnelle* qui explique cette transformation parce qu'elle « sonne » comme une conversation humaine. Tel logiciel fait fuser les réponses du tac au tac mais s'excuse de ne pas avoir bien saisi une question ou d'avoir répondu de travers et, plein de contrition, promet de s'améliorer : il fait très bien *semblant* de parler. Tel autre produit des images à partir de mots-clés comme si les mots de la langue avaient une contrepartie matérielle. Fin septembre 2022, un premier système de génération d'images à partir d'une simple description en langage naturel est rendu accessible au public[6], suivi de près par plusieurs systèmes concurrents[7], qui tous produisent des images d'un aspect tel que la presse spécialisée affirme péremptoirement, une fois de plus, que nous assistons au dépassement de l'humain par la machine dans le domaine de la création d'images. Accessibles gratuitement ou pour un coût négligeable à tout utilisateur d'internet, les systèmes texte-image attirent un nombre rapidement croissant d'utilisateurs et les images engendrées par l'intelligence artificielle commencent à envahir par millions les médias sociaux. Deux mois plus tard, ces développements sont suivis par la mise sur le marché d'un autre système faisant appel au *deep learning*,

[6] Il s'agit de *Dall-E 2,* système texte-image développé par la société OpenAI. Document en ligne consulté le 13 août 2022 : <https://openai.com/>.

[7] Mis à part *Dall-E 2*, les systèmes le plus utilisés sont Midjourney (https://www.midjourney.com/home/), NightCafé (https://creator.nightcafe.studio/) et Stable Diffusion (https://stability.ai/). Notons que Google dispose d'un système similaire nommé *Imagen*, que la société n'a pour l'instant pas ouvert au public (https://imagen.research.google/).

concernant, cette fois-ci, non pas la génération d'images, mais la réalisation automatisée de textes en tout genre, *chatGPT*, nouvel indice qu'il s'agit moins de données sonores ou visuelles et beaucoup plus de *traitement* de ces données qui fait le cœur du débat. Cette interface de dialogue avec la machine, accessible gratuitement sur internet, est capable de générer des textes d'une grande variété, en plusieurs langues. À partir d'une simple demande écrite, le système peut produire des poèmes, des textes académiques, des recettes de cuisine ou du code informatique (parmi beaucoup d'autres choses), en puisant dans les milliards de données disponibles sur internet qui forment la base d'entraînement du système en question[8]. Malgré des « hallucinations », problèmes occasionnels de fiabilité des résultats produits (dont les concepteurs du système ne se cachent d'ailleurs nullement), la qualité des textes, l'impression qu'ils peuvent donner d'être le fruit d'un effort humain, sont telles qu'en quelques semaines, *chatGPT* devient objet de fascination et de préoccupation intenses, largement relayées par les médias dans les mois qui suivent la mise en ligne du système.

Les interfaces de dialogue dans le cas des systèmes texte-texte et texte-image

Qu'il s'agisse de systèmes texte-image comme *Dall-E 2* ou *Stable Diffusion* ou d'interfaces de dialogue tels que *chatGPT*, ces systèmes font appel à une architecture complexe de réseaux neuronaux spécialisés, dont un des éléments principaux est le modèle de langage sur lesquels ils reposent : c'est bien, et de façon essentielle, de langage qu'il s'agit. Le terme « modèle de langage » est cependant trop flou pour donner une idée de ce dont il s'agit exactement, ni à quoi correspondent les milliards de paramètres qui font la puissance de ces modèles. Sans entrer dans des détails techniques complexes, on pourrait dire qu'un modèle de langage est un encodage de la distance statistique entre des mots contenus dans les données d'entraînement qui forment un corpus de milliards de mots. Chaque paramètre est une valeur, un poids de connexion dans le réseau neuronal. Un modèle de langage n'est donc aucunement une sorte de dictionnaire qui serait basé sur le *sens* des mots et des phrases analysés, mais représente une *analyse mécanisée de leur potentiel de proximité* à d'autres termes, syntagmes ou chaînes de mots. Il s'agit donc *d'une représentation exclusivement construite sur la probabilité statistique* de l'usage de la construction du sens et non pas sur une compréhension des textes analysés. Vu sous cet angle, il devient immédiatement clair que les systèmes basés sur un modèle de langage ne pourront jamais créer des contenus réellement nouveaux, qui, par

[8] Ce point, pourtant crucial, reste tout à fait obscur pour un système comme *ChatGPT*, au point qu'on peut se demander si le secret industriel ne compromettra pas toujours sa clarification.

définition, dépassent ce qui est statistiquement prévisible parce que ce qui est statistiquement prévisible exige la mesure d'un espace des possibles *alors que cette mesure fait précisément défaut quand il s'agit du sens*.

Prenons le cas de la métaphore – que le linguiste Roman Jakobson situait au cœur même de l'activité linguistique[9] – dans une phrase d'*Eugénie Grandet* de Balzac. Celui qui a lu le roman connaît l'histoire du tonnelier Grandet, la fortune qu'il a réussi à accumuler, son ambition pour les mariages de ses filles dans l'aristocratie et la dot qui ferait oublier (du moins l'espère-t-il) à leurs futurs époux leur condition roturière. Balzac écrit : « L'ancien tonnelier rongé d'ambition cherchait, disaient-ils, pour gendre quelque pair de France, à qui trois cent mille livres de rente feraient accepter tous les tonneaux passés, présents et futurs des Grandet. » En lisant cette phrase, le lecteur perçoit immédiatement que le sens du mot « tonneau » concentre en lui-même les notions de richesse, de pouvoir et d'espoir d'ascension sociale : ce sens n'est absolument pas répertorié dans un dictionnaire, il n'aura d'ailleurs aucune postérité et il n'existe que dans ce texte de Balzac : c'est une métaphore « vive » au sens de Ricœur[10]. Voilà ce que peut un texte littéraire et ce pouvoir ne s'intègre absolument pas à une logique statistique parce que la construction de ce sens de « tonneau », qui nécessite à proprement parler tout un roman, était complètement imprévisible. *C'est donc la notion de fonction métaphorique du langage que ce genre de systèmes remet en question* parce que la construction du sens repose sur l'idée d'un espace possible du sens qui serait mesurable à l'avance. Tout se passe comme si la métaphore dépendait donc désormais d'une plus ou moins grande proximité statistique basée sur l'usage passé. On peut légitimement se poser la question de savoir si, à terme, ce mécanisme de la proximité ne risque pas de produire, son usage se généralisant, un *affaiblissement de la capacité métaphorique* car celle-ci consiste à pouvoir inventer un sens nouveau porté par le même signifiant, comme dans l'exemple d'*Eugénie Grandet*. Cela invite aussi à rappeler la distinction merleau-pontienne entre « parole parlante » et « parole parlée », la première étant celle « dans laquelle l'intention significative se trouve à l'état naissant »[11], alors que la parole parlée « jouit des significations disponibles comme d'une fortune acquise » – belle formulation qui s'applique parfaitement au fonctionnement statistique des modèles de langage.

[9] En particulier Roman Jakobson, « Deux aspects du langage et deux types d'aphasie », *Essais de linguistique générale*, Paris Minuit, [1956], 1963, p. 43-67 ; et R. Jakobson, « Linguistique et poétique », *Essais de linguistique générale*, *op. cit.*, [1960], p. 209-248.

[10] Paul Ricœur, *La Métaphore vive*, Paris, Le Seuil, 1975, p. 223.

[11] Maurice Merleau-Ponty, *La Phénoménologie de la perception*, Gallimard, Paris, [1945], 2020, p. 238.

Les interfaces de dialogue tels que *chatGPT*[12] ne construisent donc pas leurs textes basés sur une compréhension des questions qui leur sont posées, mais sur une analyse statistique de la probabilité de combinaison de mots et de syntagmes. Ces interfaces sont, comme le dit Timnit Gebru, de simples « perroquets stochastiques »[13]. Encore faut-il savoir les interroger : la formulation d'une question devient désormais capitale parce que c'est elle qui permet de tirer parti de l'analyse statistique produite par le modèle.

Les systèmes texte-image, quant à eux, se servent également des modèles de langage, mais en y ajoutant d'autres traitements par réseaux neuronaux spécialisés, en particulier le recours à des bases de données composées d'images associées à des termes descriptifs. Schématiquement, la génération d'une image s'effectue en plusieurs étapes, tout d'abord en créant une représentation numérique de l'image souhaitée, qui encode les qualités recherchées (couleur, ambiance, style, etc.) en fonction du modèle de langage utilisé. Cette représentation numérique est ensuite traitée par plusieurs réseaux neuronaux, en utilisant des pairs texte-image disponibles dans les données d'entraînement pour générer l'image produite par le système[14]. Là encore, s'il devient certes possible de créer de « nouveaux Rembrandt » à partir des anciens, comme ce fut le cas dans le projet « The Next Rembrandt » proposé par l'université technologique de Delft[15], ces systèmes d'IA générative fonctionnent « à l'entraînement » et il serait aussi insensé d'attendre d'eux la création d'un nouveau style que d'attendre d'un singe typographe qu'il ne tape sur un clavier l'intégralité de *l'Éneide* de Virgile, pour reprendre l'exemple du mathématicien Émile Borel.

Il convient de souligner un point important. Les processus employés sont d'une complexité considérable et le fruit de longues années de recherche. Cependant,

[12] *ChatGPT* développé par *OpenAI*, est loin d'être le seul système de ce genre. Étant donné les enjeux financiers énormes liés à ces nouvelles technologies, de nombreuses sociétés sont en train de mettre sur le marché leurs propres systèmes : *Microsoft* qui a investi dix milliards de dollars dans la société *OpenAI*, vient d'intégrer ces technologies non seulement dans une nouvelle version de son moteur de recherche *Bing*, mais aussi dans Microsoft Office. Quant au concurrent direct de *chatGPT*, *Google Bard,* il est désormais également accessible librement.

[13] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major et Shmargaret Shmitchell, « On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? », *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, mars 2021, p. 610-623.

[14] Pour une discussion détaillée des techniques employées par Dall-E et d'autres systèmes texte-image, voir entre autres :« DALL·E: Creating Images from Text », OpenAI, 5 janvier 2021. Document en ligne consulté le 20 janvier 2023 : <https://openai.com/blog/dall-e/ et Aditya Ramesh *et al.*, « Zero-Shot Text-to-Image Generation », *arXiv*, 26 février 2021. Document en ligne consulté le 20 janvier 2023 : <http://arxiv.org/abs/2102.12092>.

[15] Document en ligne consulté le 13 août 2022 : <https://www.nextrembrandt.com/>.

si le fonctionnement et les processus mis en œuvre sont évidemment connus et largement documentés dans les textes de recherche, il n'en est pas de même pour les résultats obtenus. Ou pour le dire autrement : si les chercheurs à l'origine de ces techniques peuvent expliquer les mécanismes en jeu pour arriver à un résultat, par quelles étapes de calcul et de traitement il est passé, il est *totalement impossible de savoir comment le système est arrivé à un résultat précis* et en particulier quelles sources il a utilisé pour constituer une image ou un texte. À la différence des méthodes purement déterministes reposant sur des algorithmes où il est possible de comprendre, au moins en droit, comment l'algorithme arrive à un résultat donné ou même quelquefois pourquoi il n'y parvient pas, les méthodes stochastiques employées par les systèmes génératifs sont dans l'incapacité d'expliciter comment un résultat spécifique a été obtenu. Autrement dit : les systèmes génératifs ne produisent pas un résultat précis déterminé par une norme mais une approximation basée sur la seule probabilité statistique.

Pour conclure cet aperçu rapide, soulignons un aspect essentiel en ce qui concerne les systèmes texte-image : il s'agit avant tout de productions langagières. Quel que soit l'intérêt des images et des textes produits, la puissance des systèmes vient des modèles de langage employés et des interfaces conversationnelles mises au point. *In fine*, ce qui semble remarquable, ce ne sont pas les productions visuelles ou textuelles, qui ne sont au fond rien de plus qu'une restitution combinatoire de la mémoire visuelle ou textuelle telle qu'elle est accessible sur internet. Le point important à retenir est que nous nous trouvons face à *une représentation langagière* qui *simule* la restitution du sens à partir de données formelles déconnectées de tout sens, qu'il soit visuel, sonore ou textuel. Autrement dit, c'est le statut si particulier de l'écriture informatique qui fait problème, précisément parce qu'elle ne vise pas à restituer quoi que ce soit, comme nous le faisions remarquer plus haut. De ce point de vue, nous sommes confrontés à des systèmes qui pourront non seulement créer textes ou images, mais aussi servir à une restitution de la parole, de la musique, ou, pourquoi pas, d'environnements virtuels : des prototypes de systèmes engendrant des images animées à partir de textes existent déjà chez plusieurs sociétés, dont *Google* et *Meta* ; d'autres, permettant la génération de modèles 3D à partir de textes, sont en développement entre autres chez *OpenAI*, les concepteurs de *chatGPT* et de *GPT3*. Cependant, comme nous l'avons déjà souligné, toutes ces « créations » se feront sans la moindre compréhension de ce qui est généré, sans la moindre intentionnalité structurante.

Pour grossir le trait, on peut postuler que tout ce que l'on peut trouver sur internet ou dont il existe une description numérique pourra à terme être synthé-

tisé de cette manière, grâce à des systèmes de réseaux neuronaux de plus en plus complexes *mais aussi de plus en plus spécialisés* : là encore des résultats limitatifs, dans le champ des réseaux de neurones cette fois-ci, le prouvent[16]. Loin du rêve consistant à remonter *en amont* à une intelligence générale, la réalité des développements est tout autre et s'inscrit, *en aval*, dans la longue histoire sémiotique de la transformation des signes et des artefacts. Sans grande surprise, ce que l'on trouve à la sortie, dans le cas qui nous occupe, n'est pas l'« esprit humain » mais un amalgame stochastique des *produits* de ce dernier qui ont une influence sur ce que tout un chacun peut attendre du langage, même si la mesure précise de cette attente fait cruellement défaut.

L'insertion des systèmes formels-statistiques dans la production collective du sens

Ce sur quoi nous voulons plus particulièrement nous interroger porte moins sur les critiques que l'on peut formuler à l'égard de ces modèles formels-statistiques que ce qu'ils changent dans la façon *de produire collectivement du sens* quand on leur fait suffisamment confiance pour les intégrer aux interactions collectives. Nous savions par exemple déjà combien le fait d'envisager l'interaction linguistique sur le modèle d'un émetteur et d'un récepteur recelait comme perte d'intelligibilité concernant la notion de dialogue mais cette représentation, tout aussi inadéquate qu'elle fût, n'en avait pas moins des effets réels sur la pragmatique du discours en ce qu'elle la formatait différemment. Dans la même direction, nous voudrions nous interroger sur ce qu'une représentation de type formel-statistique, même si elle est très loin de nos modes habituels de constitution du sens, peut avoir comme conséquences sur nos propres anticipations quand cette représentation est intégrée à nos interactions collectives. C'est en particulier sur le cas des logiciels de génération d'images et de textes que nous voudrions réfléchir parce qu'ils instaurent directement un nouveau type d'anticipation quant au rapport entre écriture, langage et image.

Il semble en effet que le problème posé par l'ensemble des technologies du *deep learning*, et en premier lieu par les modèles de langage qui les sous-tendent, ne peut se résumer à la seule question de la qualité des textes ou des images produits, ni à celle des usages que ces développements modifient, en particulier celui consistant à apprendre à poser les bonnes questions à une machine. Il n'y a aucun doute que ces technologies représentent une étape importante dans le processus de numérisation progressive de notre rapport au monde, ni qu'elles soient – surtout dans le cas des textes génératifs – d'une

[16] Shai Ben-David, Pavel Hrubeš, Shay Moran, Amir Shpilka et Amir Yehudayoff, « Learning can be undecidable », *Nature*, Machine Intelligence, vol. 1, janvier 2019, p. 44-48.

utilité immédiate dans de nombreux contextes personnels et professionnels, et ceci quelles que soient les réserves que l'on peut formuler par rapport aux usages en question. Mais si l'on peut être confiant dans le fait que les problèmes d'usage que posent ce genre de systèmes dans le cadre, par exemple, de l'éducation où il favorise la fraude et le plagiat, trouveront sans doute une parade, il n'en est pas de même pour des problèmes fondamentaux concernant l'utilisation de ces technologies dans la production de contenus en tout genre. L'un de ces problèmes peut être formulé sous la forme d'une question : « comment ces nouveaux systèmes risquent-ils de modifier profondément les pratiques sociales entourant la production des textes et des images ? »

Il paraît clair, même si notre recul est encore trop limité, que ces développements auront un impact durable et diversifié sur la société, et ceci dans de nombreux domaines, allant de la création visuelle ou architecturale à la recherche académique en passant par les médias sociaux et la rédaction de textes administratifs, d'autant plus qu'il s'agit de technologies évoluant très rapidement[17]. Il paraît tout aussi évident qu'il est pour l'instant quasiment impossible d'en prédire l'impact social et les ramifications à moyen terme, pas plus que les possibles utilisations malveillantes. Nous allons donc nous focaliser ici sur deux aspects seulement, l'un épistémologique et lié à cette nouvelle « sémantisation statistique » du langage, l'autre esthétique et lié à la « créativité de la machine » et à l'hybridation numérique de l'expressivité humaine.

Précisons en outre que nous avons délibérément écarté l'épineuse question des biais parfois dangereux contenus dans les données d'entraînement des modèles de langage comme celles d'éventuelles utilisations malveillantes actuelles ou futures liées à ces technologies. Il s'agit là de problèmes cruciaux mais dont l'analyse dépasserait largement le cadre de ce texte.

Aspect épistémologique

Il nous paraît important de souligner que c'est moins le thème si rabâché de « l'ordinateur dépassant l'humain » – thème qui reste prisonnier d'une idée de l'intelligence conçue comme intériorité privée – qui fait véritablement question que ce genre de technique fait à l'usage du matériau public que sont les signes de la langue. Deux questions semblent pouvoir être posées à ce propos : qu'est-ce qui est modifié et surtout sera modifié à terme dans l'usage de la langue par

[17] Il s'agit en effet d'une progression extrêmement rapide : Le modèle de langage GPT2 d'OpenAI, lancé en 2018, disposait de 1,75 milliards de paramètres ; *GPT3*, qui arrive en 2020, est à peu près cent fois plus grand, à 175 milliards de paramètres. (Dans les systèmes d'apprentissage par la machine, un paramètre correspond à une valeur apprise lors du processus d'entraînement ; le nombre de paramètres est donc une mesure des capacités du système.)

ces logiciels et leurs interfaces conversationnelles au fur et à mesure de la généralisation de leur usage ? Quel état de la langue ce genre d'outil propose-t-il implicitement ? Tentons d'en dire quelques mots.

Pour ce qui est de ce que ces logiciels font mais surtout feront à la langue, on sait que les productions langagières que ce type de logiciels rend possible sont dictées par un seul et même but : continuer par tous les moyens statistiques à disposition ce qui a été proposé au départ comme configuration graphique hors de toute considération de sens. Nous avons déjà vu qu'une des conséquences possibles était l'affaiblissement de la capacité métaphorique consistant à se donner les moyens de préparer la langue (dans notre exemple : écrire *Eugénie Grandet*) en vue de produire un (ou plusieurs) sens nouveau. Si on pousse ce raisonnement à la limite, peut-on aller jusqu'à dire que *l'idée de métaphore fait déjà partie d'une époque révolue de la langue* ? On peut au moins avancer deux remarques pour contribuer à répondre à la question.

D'une part, le renforcement d'une connexion ayant déjà été avéré ne peut que faire diminuer le potentiel de nouveauté présent à son maximum dans la métaphore. Faire ainsi un usage généralisé de ce type de structure formelle-statistique fait partie d'une numérisation plus globale du rapport au monde fondé sur le désir collectif de limiter l'imprévisibilité, même si celle-ci n'est pas complètement éliminable. Ce sont donc les trois instances temporelles du passé, présent et futur qui se trouvent reconfigurées par le biais de cette structure. Mais cette reconfiguration diffère de l'étape logique-algorithmique qui la précédait. Celle-ci tentait de parvenir à produire un résultat dans le cadre général du calcul en imaginant ce résultat seulement futur comme déjà présent : puisqu'il allait être produit nécessairement par le présent, le futur n'en différait pas essentiellement. Dans cette nouvelle étape formelle-statistique en revanche, on considère plutôt le *futur comme déjà du passé* : puisque le futur n'est qu'un renforcement du passé, le présent devient finalement évanescent et perd toute consistance.

D'autre part, la structure formelle-statistique semble fixée une fois pour toutes puisqu'il ne s'agit que de la retrouver hors de toute temporalité : aussi est-ce moins le fonctionnement de la langue que cette structure exprimerait que celle des faits, c'est-à-dire l'état, fixé une fois pour toutes, du monde. C'est en particulier le cas des logiciels produisant des images à partir de mots-clés qui donnent l'impression de pouvoir produire immédiatement des images sans en passer par le relais du sens, comme si chaque mot avait une contrepartie immédiate sous forme d'image. Ce faisant, il n'est pas impossible qu'implicitement, ces logiciels contribuent à entretenir une certaine confusion ontologique quant au rapport du langage à la réalité. D'où la question-limite suivante : les dis-

tances entre les mots dans les modèles de langage ne sont-elles pas aussi vécues comme des distances entre les choses fondées sur la répétition de leur rapprochement ? Autrement dit, les rapprochements effectués par les modèles n'ont-ils pas un poids ontologique au sens où ils manifesteraient le monde tel qu'il est censé être ?

Il s'agit évidemment ici de tendances à prendre en compte sur le long terme, à partir de la façon dont nous nous représentons la situation aujourd'hui. Dans le même ordre d'idées, il faut s'arrêter sur la façon dont, à partir de l'usage à grande échelle de ces logiciels, la question de la créativité a été renouvelée.

La question de la créativité

Malgré le fait que l'émergence des systèmes texte-image et des interfaces de dialogue comme *chatGPT* coïncident presque, les questionnements et angoisses que les deux ont provoqué dans les médias sont d'ordre différent, comme nous le disions en commençant. En effet, si *chatGPT* peut impressionner par la facilité de générer des textes qui semblent écrits par un humain, personne n'a pour l'instant déclaré la fin de la littérature à cause d'un *chatbot*, et *chatGPT* n'a pas suscité la comparaison de ses productions avec les écrits de Flaubert, James Joyce ou Apollinaire. Aujourd'hui, c'est le *fait littéraire quotidien* qui semble menacé par l'intelligence artificielle, non pas la *créativité littéraire* elle-même, même si la capacité métaphorique semble, dans un futur plus lointain, pouvoir être mise à mal.

Dall-E 2 et ses compétiteurs comme *Stable Diffusion* ou *Midjourney*, ont tout d'abord fait surgir le spectre du dépassement de l'humain par la machine dans le domaine de l'image. On a pu lire que « le meilleur artiste sur terre est une intelligence artificielle », et on a tenté de montrer que même les conservateurs de musées n'étaient pas toujours à même de distinguer une œuvre d'art de la production de la machine[18]. À tort ou à raison, les systèmes texte-image ont été crédités de créativité là où *chatGPT* est perçu avant tout comme apportant de la facilité.

Cependant, l'affirmation de la présumée « créativité de la machine », souvent évoquée par des commentateurs impressionnés par l'exploit technologique, ne peut s'appliquer qu'à condition de se cantonner à une conception très réductrice de la créativité, c'est-à-dire à une définition qui se limite à l'analyse de

[18] Jo Lawson-Tancred, « Is this by Rothko or a robot? We ask the experts to tell the difference between human and AI art », *The Guardian*, janvier 2023. Document en ligne consulté le 15 janvier 2023 : <https://www.theguardian.com/artanddesign/2023/jan/14/art-experts-try-to-spot-ai-works-dall-e-stable-diffusion>.

quelques caractéristiques visuelles de l'objet créé[19]. Si, en revanche, on est de l'avis, comme c'est notre cas, qu'il n'est pas possible de séparer l'analyse de la créativité des aspects sociaux, phénoménologiques ou sémiogénétiques, et du *processus créatif* dans son ensemble, la question de la créativité de la machine ne semble pas se poser, étant donné la nature particulière des modèles de langage discutée plus haut.

La question de l'hybridation du geste créatif

Si l'on prend pour guide l'histoire des innovations techniques et la façon dont les humains ont intégré ces innovations dans des gestes nouveaux qui façonnent autrement notre manière de nous exprimer et d'être au monde, on peut supposer que la créativité humaine s'appropriera également cette dernière innovation – nouvelle occasion offerte par le modèle de la machine de manifester cette créativité. On peut même, sans trop s'avancer, affirmer que les aspects novateurs des développements décrits plus haut s'inscrivent dans une histoire de l'hybridation du geste créatif et qu'ils représentent une suite logique de développements commencés dès les débuts de la numérisation des pratiques. L'hybridation des gestes avec le numérique commence donc avec l'arrivée de l'ordinateur, et, par la suite, avec l'émergence de l'informatique graphique dans les années 1980. En effet, ces développements nous fournissent déjà la trame comportementale et sociale que l'on retrouve aujourd'hui dans les réactions des utilisateurs aux outils dits « génératifs ». On constate la même utilisation fascinée par la nouveauté, dont le but premier est l'exploration des limites du système ; on retrouve la même relation quasi symbiotique entre utilisateurs passionnés et développeurs ; enfin, quoiqu'à une échelle beaucoup plus étendue, ce sont les mêmes socialités déjà propres à l'informatique graphique qui animent les échanges entre utilisateurs passionnés[20].

En revanche, une question déjà présente dans l'utilisation d'outils de création numérique acquiert avec les systèmes texte-image une importance prépondé-

[19] Cette approche est largement répandue parmi les défendeurs d'une vision computationnelle de la créativité. On peut citer entre autres Margaret Boden, qui définit la créativité ainsi : « Creativity is the ability to come up with ideas or artefacts that are new, surprising, and valuable. » (M. A. Boden, *Creativity and art: three roads to surprise*, Oxford University Press, 2010, p.29-30). Ajoutons néanmoins que si les productions des systèmes texte-image peuvent sous certaines conditions répondre aux deux premiers qualificatifs, la question de la valeur des images produites reste à clarifier.

[20] Pour une analyse détaillée des hybridations propre à l'informatique graphique, voir Arnaud Pfeiffer, *L'Émergence de l'informatique graphique et de la praxis créative hybride*, mémoire de master EHESS, 2022. Document en ligne consulté le 14 février 2023 : <https://www.researchgate.net/publication/364815008_A_Pfeiffer_L'emergence_de_l'informatique_graphique_et_de_la_praxis_creative_hybride>.

rante : il s'agit de la question du *pouvoir prédictif et prescriptif de la machine*, ou, formulé autrement, de *la tendance de l'objet technique à prédéterminer en partie ou même dans la totalité* ce que les utilisateurs produiront grâce aux potentialités offertes par la machine. S'agit-il alors, quand ces possibilités augmentent soudainement, d'une libération ou d'un enfermement ? Le déplacement de l'horizon du possible qui fascine tant les praticiens, augmente-t-il seulement les potentialités – ou opère-t-il en même temps un effacement, une invisibilisation de possibilités auparavant employées couramment ?

Les systèmes texte-image ont créé l'événement en montrant que n'importe quel utilisateur d'internet peut produire n'importe quelle image qui était, du moins *potentiellement, déjà présente* dans la combinatoire des données, des images et des styles de traitement qui ont servi à l'entraînement du système. Que cette potentialité, une fois perçue, donne le vertige n'a rien d'étonnant. Ce qui inquiète, en revanche, c'est qu'elle efface de la panoplie expressive disponible un aspect fondamental : l'émergence du sens *par le geste corporel*. L'acte créatif principal dans l'utilisation des systèmes texte-image est *d'ordre langagier* : la traduction d'une image intérieure non pas en gestes et techniques picturaux, mais directement en une description dans une sorte de métalangage qui se situe entre description, référentiel esthétique et code.

Quelles perspectives ?

Ce qui semble en revanche certain, c'est que les systèmes texte-image présentent des capacités qui invitent à les explorer et surtout à les détourner. Si la très grande majorité des images créées à l'aide de ce système *n'a aucun autre intérêt que l'automatisme de leur génération*, tout dépendra du type d'utilisation qui en sera faite : le processus lui-même peut inspirer des explorations tout à fait inattendues. Mais dans ce cas, la créativité ne se trouve plus dans ce que produit la machine, mais dans la créativité de l'opérateur humain capable, souvent dans un processus itératif long, d'indiquer le chemin d'une exploration que la machine elle-même n'aurait jamais pu trouver, pour la bonne et simple raison que la machine est *incapable de juger ses propres productions*, ne peut pas différencier une image intéressante d'une autre qui l'est moins, et ne dispose d'aucune sorte d'intentionnalité pourtant indispensable à toute exploration. C'est donc *l'irruption de l'imprévisible humain* qui permet alors à la machine de générer quelque chose qui dépasse la simple surprise éphémère.

Que l'on se rassure, il n'y a aucune raison de croire que le progrès technique enlèvera à l'espèce humaine son besoin de créer et d'agir : l'intelligence artificielle ne nous transformera pas en robots écervelés subjugués par la machine.

En revanche – et c'est bien là le grand danger des développements du *deep learning* – cela nous rendra peut-être de moins en moins enclins à distinguer, dans nos interactions avec le monde, ce qui, dans le flot incessant d'informations de tout genre qui se présente à nous (et avec lesquelles nous sommes obligés d'interagir, sur lesquels nous sommes obligés de baser nos actions), est le fruit d'un *raisonnement* ou d'une *initiative créatrice humaine*, de ce qui n'est simplement qu'un mirage reconstitué d'une analyse statistique de plus en plus puissante.

Nous sommes à l'évidence entrés dans une nouvelle phase de l'hybridation numérique, *celle des imaginaires et des subjectivités motivantes*. Nous vivons désormais dans un monde où la probabilité statistique cherche de plus en plus souvent à remplacer l'intentionnalité structurante. Une nouvelle fracture commence donc à émerger, moins immédiatement perceptible que la « fracture numérique », celle de l'hybridation du langage humain avec les « artefacts de sens » produits par la machine[21]. Est-ce pour autant une fracture ? Tout dépend à quel niveau on se place. La fracture numérique se résume à une question d'accès à l'outil numérique. Dans le cas de la créolisation du langage, ce n'est pas la disponibilité de l'outil qui est en cause : quiconque dispose d'un accès à internet peut utiliser ces nouveaux systèmes. La fracture se situe donc davantage dans *la nature de l'utilisation* qui sera faite de cet outil : étant donné la facilité offerte par les systèmes de *deep learning* de produire textes et images sans le moindre effort, en déléguant l'intuition structurante à un automate stochastique dont la plupart d'entre nous ignorent le fonctionnement et les limites, la fracture sociétale qui pourrait se faire jour sera celle entre, d'un côté, ceux qui comprennent l'importance de leurs propres efforts en vue de réaliser des actes sémiogénétiques et savent se servir des nouvelles facilités sans se soumettre à la machine d'un côté, de l'autre ceux qui ne se posent pas de questions et se laissent envahir par les artefacts de sens dont ils ne connaissent ni la nature ni les limitations.

[21] Pour une analyse approfondie de cette nouvelle créolisation du langage, voir Frédéric Kaplan, « Les vingt premières années du capitalisme linguistique », Anne Alombert, Victor Chaix, Maël Montévil, Vincent Puig (dir.), *Prendre soin de l'informatique et des générations*, Limoges, FYP éditions, coll. « Nouveau monde industriel », 2021.