

14 janvier 2025, pour l'Association des Ami-e-s du MCBA, Lausanne

Les Intelligences artificielles génératives sont-elles surréalistes ?

Arnaud Dufour

Tout d'abord, je souhaite remercier le comité de l'association des Ami-es du Musée cantonal des Beaux-Arts de Lausanne et en particulier Mme Sabine Panchaud d'avoir accueilli l'idée de parler d'Intelligence Artificielle et d'Art Numérique dans le contexte de ce cycle de Conférences sur l'Art. Au fil des 8 interventions, nous aurons l'occasion d'entendre des personnalités d'horizon très différents nous parler de ce sujet sous des angles à la fois originaux et complémentaires. Vous avez lu le programme et vous savez que vous aurez de vrai-es spécialistes de l'art et même des artistes que je me réjouis d'entendre nous éclairer sur ces questions.

A titre personnel, je ne viens pas du monde de l'art, mais en tant que professeur en **marketing digital**, je m'intéresse à la communication et à l'image. J'ai aussi eu la chance de baigner dans le monde de l'Internet depuis les débuts de son ouverture au grand public dans le milieu des années nonante. Et puis, à titre personnel je m'intéresse depuis longtemps à la photographie, et depuis plus récemment à la peinture contemporaine, notamment figurative, en tant qu'amateur d'art, atteint – je dois le confesser - d'une assez sérieuse **collectionnite**, pour reprendre le titre d'un livre délicieux d'Emmanuel Pierrat, car en effet je me heurte déjà aux murs de mon appartement...

Pour tout vous dire, l'idée d'aborder ce vaste sujet a germé, il y a tout juste une année, alors que j'écoutais, assis avec vous dans cette même salle, la conférence introductive du **cycle** de l'an dernier sur le Surréalisme. Rappelez-vous, pour celles et ceux qui étaient là, Juri Steiner, directeur du MCBA et co-commissaire de l'exposition, nous dévoilait en avant-première et trois mois avant son ouverture, les grandes lignes de l'exposition « Surréalisme - Le Grand Jeu ».

A la même période, je travaillais à la Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud, à Yverdon à la préparation de la publication annuelle que nous réalisons pour notre institut d'Ingénierie des médias. Un article évoquait alors un projet réalisé par mes collègues. En deux mots, il s'agit d'un projet de médiation scientifique réalisé en marge d'un projet européen sur la **physique solaire**. Il prend la forme d'un prototype de jeu de carte et dans ce projet, les visuels et les textes ont été générés par des systèmes d'intelligence artificielle dites générative, ici principalement ChatGPT et MidJourney. Comme vous pouvez le voir le style graphique de ces images est basé sur le vitrail, et les visuels créés sont assez surprenants, pour ne pas dire « surréalistes », si j'ose un usage très générique du terme.

C'est donc parti de là, entre les propos érudits de M. Steiner et ce projet de recherche appliquée, il m'a semblé qu'il y avait des territoires à explorer entre IA et surréalisme.

Maintenant que vous savez tout sur l'origine de cette série de conférences, revenons à cette deuxième conférence du cycle.

La semaine dernière, Marc Atallah nous a invités à plonger dans le monde de l'art numérique en nous montrant à travers différents exemples comment l'ordinateur s'immisce dans le monde créatif. Des machines produisant des images, des robots capables de dessiner des portraits, de l'art génératif basé sur des algorithmes, des performances hybrides entre spectacles vivant et projections numériques, le jeu vidéo comme espace de création d'expériences artistiques (et sérieusement psychanalytiques) ...

Ces différents exemples sélectionnés par Marc illustrent certains aspects de l'art numérique, mais les relations entre les arts et le numérique peuvent prendre d'autres formes, par exemple :

- dans la photographie ou dans la vidéo, deux domaines dont la numérisation est banalisée,
- dans la création visuelle – en particulier dans le cinéma – qui fait la part belle aux effets spéciaux réalisés par l'informatique depuis très longtemps,
- dans la musique, qui elle aussi, est imprégnée de numérique, depuis les synthétiseurs jusqu'à systèmes d'écoute en streaming,
- dans la peinture, en particulier la peinture figurative, dans laquelle les artistes expliquent que le numérique est devenu leur carnet de notes, que beaucoup travaillent d'après photographie en réalisant leurs mises au carreau en zoomant sur un iPad ou en projetant ces photos directement sur leurs toiles, sans parler de ceux qui, comme David Hockney, s'emparent du numérique comme outil et support de création. Peut-être que certains ou certaines d'entre vous ont pu voir l'exposition de ses « iPad paintings » à la Royal Academy of Arts en 2021¹. (116 œuvres réalisées à l'iPad en 2020). Hockney aura d'ailleurs une rétrospective² à la Fondation Louis Vuitton à Paris dès le 9 avril 2025.
- D'autres disciplines sont concernées, comme l'architecture qui elle aussi s'appuie sur des logiciels qui impactent la création, je vous renvoie à ce sujet à un podcast de la série *Le Code a changé* de Xavier de La Porte sur France Inter³, intitulé « *Les courbes en folie* » : *faire de l'architecture avec les ordinateurs*.
- Dans cette énumération, j'ai surtout évoqué des aspects liés à la **création** avec le numérique, alors qu'en réalité les changements liés à la numérisation du domaine concernent les métiers, les activités, la commercialisation, la médiation, les musées, le fonctionnement de secteurs économiques entiers.

Bref, la culture et les Beaux-Arts n'échappent pas à la numérisation.

En filigrane de cette numérisation, l'informatique est aujourd'hui secouée par ce que certains analystes considèrent comme le plus grand bouleversement depuis l'émergence de l'Internet, que d'autres, plus emphatiques encore voient comme un événement historique à l'échelle de l'humanité aussi important que celui de l'invention de l'écriture ou de l'imprimerie.

L'intelligence artificielle, puisque c'est elle dont il s'agit, fait la une des médias très régulièrement depuis un peu plus de deux ans, elle agite Wall Street⁴ et la Silicon Valley, elle s'invite dans de très nombreux domaines, notamment dans les entreprises, et en particulier dans leur informatique et dans leur marketing.

¹ <https://www.royalacademy.org.uk/exhibition/david-hockney>
et 10 ans auparavant <https://museeyslparis.com/expositions-fondation/david-hockney-fleurs-fraiches-dessins-sur-iphone-et-ipad>

² <https://www.vogue.fr/article/david-hockney-exposition-paris-2025-fondation-louis-vuitton-peinture>

³ <https://www.radiofrance.fr/franceinter/podcasts/le-code-a-change/le-code-a-change-20-11-7151548>

⁴ NVidia, fabricant de cartes et de puces utilisées pour faire tourner les systèmes d'IA est maintenant la deuxième entreprise en termes de valorisation boursière aux USA, derrière Apple (avec respectivement 3400 milliards de \$ et 3'700 milliards de \$, <https://companiesmarketcap.com/> le 10 janvier 2025). D'ailleurs dans le Top 10, 8 entreprises sont des entreprises du numérique (Apple, Nvidia, Microsoft, Alphabet, Amazon [Saudi Aramco], Meta, [Tesla], Broadcom, TSMC). <https://companiesmarketcap.com/>

Je ne vais pas ici vous présenter **l'histoire de l'Intelligence Artificielle** ni détailler les technologies qui rendent aujourd'hui possibles les fameuses IA dites génératives, celles qui affolent le monde numérique, depuis le 30 novembre 2022, date à laquelle OpenAI ouvre au public son ChatGPT.

Avant cette date, les IA existaient et ont été utilisées dans des contextes de création artistique, mais ces projets exigeaient des développements informatiques et des moyens importants. C'est le cas des deux exemples très médiatisés que vous avez là,

- à gauche Le Prochain Rembrandt, un projet de 18 mois réalisé en 2016 par une équipe importante incluant Microsoft, une agence de communication, des universités et des musées.
- à droite, Le portrait d'Edmond De Belamy, réalisé par le collectif Obvious et connu parce qu'il a été vendu aux enchères pour 432,500 \$ par Christie's en 2018. Ce sont des exemples intéressants, mais qui précèdent l'ouverture au grand public des IA.

En réalité, en ce qui concerne les IA générative capables de créer des images, cette ouverture au grand public démarre au début de l'année 2022, avec le **lancement de DALL-E**, la solution d'OpenAI de génération d'images (entre mars et juillet). A peu près en parallèle, **MidJourney**, dévoile sa propre solution et la rend largement accessible en juillet 2022. Depuis lors, ces deux plates-formes n'ont cessé de faire évoluer leurs modèles vers plus de puissance, de pertinence, de qualité des rendus et leur ont ajouté des fonctionnalités nouvelles.

Depuis cette date, des investissements massifs ont été réalisés par les géants du numérique et par les fonds de capital-risque⁵ (45 milliards de \$ en 2024, 24 milliards de \$ en 2023, 8.7 milliards de \$ en 2022). Les attentes du monde économique⁶ semblent très importantes et chacun espère sa part du gâteau.

Ces investissements massifs ne concernent pas seulement les outils de génération d'images. C'est en réalité tout **un ensemble d'outils informatiques** qui voit le jour, et dont une partie seulement se spécialise dans la génération d'images, là où d'autres le font sur le texte, la musique, la vidéo, etc. Les investissements profitent également à toute une chaîne d'acteurs qui conçoivent et fabriquent le matériel nécessaire, en particulier les puces spécialisées (et notamment les GPUs), les cartes, les machines et les centres de données (data centers) requis pour entraîner et faire fonctionner ces systèmes à grande échelle. Cette notion d'échelle est importante, car ChatGPT a battu tous les records de croissance, en atteignant 100 millions d'utilisateurs mensuels en moins de deux mois⁷.

Dans le domaine de l'image, j'ai cité DALL-E et MidJourney, mais beaucoup d'autres solutions existent, comme Recraft, Stability, Ideogram, Leonardo... Certaines sont développées par de grands éditeurs et intégrées dans leurs outils comme Firefly d'Adobe, Gemini de Google, CoPilot de Microsoft, Llama de Meta (qui n'est pas encore disponible en Europe) ou même Grok de X.

⁵ https://www.ey.com/en_ie/newsroom/2024/12/venture-capital-investment-in-generative-ai-almost-doubles-globally-in-2024-as-momentum-accelerates-in-transformative-sector

⁶ <https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/>

⁷ <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> (9 mois pour TikTok, 2.5 années pour Instagram)

Des sites mettent en **compétition** ces moteurs pour essayer de les classer selon différents cas d'usage, ici vous avez un exemple traduit d'une telle compétition. (*Imaginez une station de recherche antarctique de pointe et respectueuse de l'environnement, avec une conception modulaire unique qui s'intègre parfaitement dans le paysage glacé, avec des matériaux durables et des systèmes d'énergie renouvelable, rendus dans un style élégant et moderne.*)

Mais revenons au cœur de notre sujet.

Elle, vous la connaissez, la célèbre **Jeune fille à la perle**, peinte par Vermeer en 1665, mais connaissez-vous ses démêlés avec l'IA ? En 2023, elle a été prêtée au Rijksmuseum (prononcer Reïx museum) pour sa grande exposition Vermeer. Face au vide laissé sur ses murs par ce prêt, la Mauritshuis (prononcer maorette's haus) a décidé de lancer un **concours** public dans lequel chacun-e pouvait proposer sa « Jeune fille », et dont les gagnant-es verraient leur œuvre **accrochée** dans le musée. 3500 œuvres ont été reçues et 175 sélectionnées, parmi lesquelles celle-ci, proposée par **Julian van Dieken**⁸, et générée avec des outils d'IA et Photoshop. Elle n'était pas la seule œuvre artificielle, mais c'est celle que les médias ont retenue pour débattre du statut de cette œuvre et de sa légitimité à rejoindre, même temporairement, les cimaises du Musée.

Alors bien sûr, chacun, chacune d'entre vous peut générer en quelques clics des versions spécifiques et uniques de cette jeune fille en utilisant des IA génératives... Ici vous avez deux exemples personnels créés par MidJourney, d'une version robotisée, dont les résultats sont assez étranges.

Quelques mois auparavant, c'est une autre image, le « **Théâtre d'Opéra Spatial** » de **James Allen** qui a agité le monde médiatique, après avoir remporté un prix lors de l'édition 2022 de la Foire d'Art de l'Etat du Colorado, et ce, alors même que l'œuvre était proposée dans la catégorie « art numérique ». D'après l'auteur, l'image de base aurait été créée avec MidJourney après plus de 600 tentatives, puis retravaillée avec Photoshop et agrandie avec Gigapixel, un autre outil d'IA spécialisé dans cette tâche. Ce qui est intéressant dans ce cas, c'est qu'Allen a tenté de déposer l'image auprès du Bureau américain du copyright⁹, lequel a refusé cette demande parce que l'image avait été essentiellement créée par une machine et non par un humain. D'après le site de M. Allen, un appel est encore en cours. Peut-être que Maître Benhamou aura l'occasion d'y revenir dans sa conférence du 28 janvier sur les questions juridiques que cela soulève.

L'exemple suivant est celui de l'allemand **Boris Eldagsen** qui a refusé en avril 2023 le prix du concours photographique "Sony world photography awards" pour cette image, intitulée *The Electrician* et issue de sa série "PSEUDOMNESIA" (2022). Le prix lui a été attribué par le jury avant qu'il ne le refuse en révélant que l'image n'était pas une photographie mais une image générée par l'IA, ce qui n'a pas empêché l'organisation de maintenir sa décision. Le détail de cette histoire est expliqué sur son site¹⁰. D'ailleurs sur ce même site vous pourrez consulter la **série complète d'images** dont est issue *The Electrician*.

⁸ <https://www.instagram.com/stories/highlights/17973018833113945/>

⁹ <https://www.jasonmallen.com/space-opera-theater> et https://en.wikipedia.org/wiki/Th%C3%A9%C3%A2tre_D%27op%C3%A9ra_Spatial

¹⁰ <https://www.eldagsen.com/sony-world-photography-awards-2023/>

A titre anecdotique, on a aussi eu des événements inverses avec par exemple le cas de **Miles Astray** qui a remporté le prix «1839 Awards» dans la catégorie IA, avec une photographie «conventionnelle» en juin 2024... un peu comme une revanche de la photographie.

Sur son site Eldagsen mentionne un photographe péruvien, **Christian Vincés**¹¹, qui propose (en avril 2023) de nommer ces images créées par des IA avec le terme *promptographies*, pour les différencier clairement des photographies et faire référence à leur processus de génération.

Cela m'amène à la façon dont fonctionnent ces systèmes d'IA génératives. La plupart proposent une interface utilisateur qui reçoit du texte et génère des images à partir de ce texte. Pour les spécialistes de l'art, il y a là un champ entier de réflexion et de recherche sur les relations étranges entre ces mots et les images qu'ils engendrent. Je pense ici en particulier à la regrettée Annie Le Brun, amie des surréalistes, qui nous a quitté en juillet dernier et qui s'était intéressé de près à cet espace entre les mots et les images, notamment dans ce livre « Un espace inobjectif » paru en 2019 chez Gallimard.

Pour les informaticien·nes, l'idée de **prompt** est assez classique, vous donnez une commande à la machine et elle l'exécute. On peut d'ailleurs parler de commande, d'instruction, d'invite, certains parlent même d'incantations, sans doute pour évoquer le sentiment de travailler avec un outil un peu magique (J'ai d'ailleurs ici envie de rendre hommage à Nicolas Nova, un professeur de la HEAD décédé tout récemment à 47 ans et dont le dernier livre portait sur *La Persistance du Merveilleux*¹² dans le numérique). Sur **MidJourney** le prompt est construit et intègre différents paramètres comme le format hauteur-largeur de l'image à générer ou la version du modèle à utiliser. Ces paramètres peuvent aujourd'hui aussi être définis via une **interface web**.

Cela dit, le prompt n'est pas la seule façon d'interagir avec de telles machines. Dans le cas de DALL-E, il est possible d'étendre une image (**outpainting**), sans prompt, comme le montre cet exemple d'Auguste Kamp dans lequel notre Jeune Fille à la Perle apparaît à nouveau.

Avec **Firefly**, il est possible de localiser l'action d'un prompt, dans d'autres il s'agit de dessiner un croquis ou simplement de dialoguer avec l'IA. C'est d'ailleurs le cas pour DALL-E 3 qui est désormais intégré dans ChatGPT. Je vous montre ici un échange :

- d'abord lui demander de me dessiner le mouton du Petit Prince,
- puis le pousser,
- ici la cohérence de la réponse commence à dériver puisque dans le texte de Saint-Exupéry, la boîte cache le mouton.

11

<https://www.facebook.com/christianvinces/posts/pfbid02dFH3gtkQiUSkyGVp6AiDeikJa5Wtj2rnnWcifVCp6pVuY9wXox7GL7khFMUcdKpxl>

¹² <https://www.premierparallele.fr/livre/persistance-du-merveilleux>

Il est d'ailleurs possible d'interroger la machine face à ce dessin, et elle semble capable de décrire et d'analyser le dessin.

- En la poussant plus loin, elle s'excuse de ne pas avoir créé une image plus fidèle dès le départ... et tente de se justifier
- Jusqu'à dessiner le mouton final, enfin sa boîte (une boîte qui diffère du croquis original).

Un test précédent montre un autre exemple assez absurde, dans lequel l'IA me montre une image en expliquant que celle-ci ne montre pas le mouton...

Un test effectué avec la version mobile de GPT4o **refuse** tout simplement de générer l'image, et c'est la même chose, mais de façon bien plus argumentée avec le dernier ChatGPT o1 (lancé en début décembre). Il refuse de créer l'image pour des questions de droit... et peut l'expliquer de façon assez détaillée... en précisant (1) La règle « standard » des 70 ans après la mort de l'auteur et (2) Le cas particulier de la France : la « prorogation de guerre »...

Dans l'ensemble de cet exemple, nous avons ce que l'on appelle une **interface conversationnelle**, c'est-à-dire que l'on dialogue avec la machine en langage naturel, et une IA qualifiée de *multimodale*, puisqu'elle peut accepter en entrée – ou générer en sortie différents types de contenus : du texte, du code informatique, des images, parfois des sons ou autre chose.

Ces modalités d'interaction et notamment la dimension conversationnelle ont été une des raisons du succès de ChatGPT. Certains y voient un point de bascule dans la façon dont l'interaction homme-machine est envisagée, car même si les interfaces conversationnelles existaient déjà (pensez- à Siri, Alexa ou Google Assistant), elles étaient loin d'offrir une telle capacité de dialogue.

Ici vous avez un exemple de l'interface de **MidJourney** à qui j'ai également demandé d'imaginer Le mouton du Petit Prince et qui semble incapable de répondre correctement à la requête, sans me proposer de possibilité d'itérer à partir de ces réponses. Cela dit les interfaces évoluent sans cesse et chaque éditeur tente d'améliorer l'expérience qu'il propose à ses clients tout en se différenciant de ses concurrents¹³, par des services et une interaction optimale.

Je laisse aux linguistes et aux sémiologues la réflexion sur cette invitation à **Imaginer**. Une simple plongée dans le dictionnaire **Robert** nous rappelle que l'imagination est une « Faculté que possède **l'esprit** de se représenter des images ou d'évoquer les images d'objets déjà perçus. » et une « Faculté de former des images d'objets qu'on n'a pas perçus ou de faire des combinaisons nouvelles d'images ou d'idées, de se représenter des situations possibles. ».

Ce double éclairage fait singulièrement écho à la définition de l'intelligence artificielle, en ce qu'elle regroupe, je cite des « théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine » (vous voyez ici la référence à **l'esprit**). Plus généralement, le monde de l'informatique et de la robotique cherche à doter la machine de capacités sensorielles, motrices, mémorielles et réflexives.

Les définitions du Robert nous invitent à réfléchir sur l'ambiguïté entre le « déjà perçu » et le « jamais encore perçu », car les IA génératives dont nous parlons sont basées sur des modèles qui ont été entraînés à partir de gigantesques corpus de textes et d'images. On parle de centaines de milliards de mots¹⁴ et de milliards d'images, autrement dit des centaines de téraoctets de données.

¹³ Voir par exemple l'interface Flow de Leonardo.ai, <https://app.leonardo.ai/image-generation/flow-state>

¹⁴ Je simplifie, pour ne pas aborder les tokens

J'ai mis ici l'exemple de **LAION5B**, un gigantesque *dataset* ou jeu de données, comportant près de 6 milliards d'images, dûment filtrées et décrites, c'est-à-dire que chaque image est associée à des mots-clés décrivant ce qu'elle contient. Les organisations qui développent les solutions d'IA génératives, ne sont pas très précises, ni loquaces sur les sources utilisées réellement pour entraîner leurs modèles, d'abord parce qu'il s'agit d'un élément de concurrence, et ensuite parce que des questions juridiques et commerciales¹⁵ sont soulevées par l'exploitation de contenus utilisés pour entraîner des IAs. (Maître Benhamou y reviendra).

En matière de textes, les acteurs du domaine se sont massivement servis sur le web, dans les livres, sur Wikipedia, dans les news, dans les forums, sur les blogs, dans les sources de données publiques, et bien sûr dans les publications scientifiques (un domaine sur lequel souffle le vent de l'ouverture avec le courant de l'*open science*, un vent tout aussi fort que celui de l'*open data* dans le domaine public notamment), et puis j'allais oublier une autre source majeure que sont les réseaux sociaux, par exemple Twitter/X ou Meta et leurs milliards de post publics, etc.

Beaucoup de grandes plates-formes, et en particulier Meta ont d'ailleurs modifié leurs conditions générales pour pouvoir entraîner explicitement leurs modèles d'IA à partir des contenus partagés par leurs utilisateurs. La commission européenne parle même d'*altruisme des données* pour inciter les citoyens et les organisations à partager leurs données pour qu'elles puissent servir à entraîner des IA.

Il est important donc d'avoir en tête le fait que derrière ces IA générative, il y a une somme, pour ne pas dire « **La Somme** » avant un grand S, de la production intellectuelle et culturelle humaine. Pas celle de toute l'humanité sans doute, mais celle d'une part importante de l'humanité, occidentale et nord-américaine surtout, mais pas seulement. Ce qu'on nous présente souvent comme étant l'aboutissement de la machine, repose en réalité sur des siècles d'un immense travail humain. Des individus et des organisations qui ont créé et préservé ces contenus, et d'un travail toujours humain sur ces éléments de contenus, pour les numériser, les trier, les étiqueter (c'est-à-dire les décrire), les tester ensuite... La machine est loin d'effectuer ce travail seule.

Cela soulève bien entendu d'importantes questions éthiques, politiques et économiques, puisque d'aucun voient dans ces mécanismes d'entraînement des IA, une appropriation par le capital de la connaissance humaine, y compris quand celle-ci a été développée précisément en dehors de la sphère capitaliste. Je pense ici en particulier à des exemples comme Wikipédia dont les contenus ouverts ont nourri les IA.

Dans cet ensemble de contenus textuels et visuels (mais aussi sonores et vidéo), on trouve tout un corpus d'œuvres qui font partie du patrimoine culturel et qui ont été moissonnées sur le web ou dans les livres, bien souvent sans l'accord de leurs auteurs, autrices ou ayants-droits, et bien que les œuvres en questions ne soient pas formellement copiées dans les modèles d'IA génératives, leurs traces, leurs caractéristiques, leur « esprit » en quelque sorte, s'y trouve bel et bien, peut-être même de façon caricaturale, au point qu'il est possible de demander à la machine de fabriquer des images **à la manière de tel ou telle artiste**, qu'il s'agisse de photographie, de peinture, ou d'autres éléments de style. C'est un des points qui cristallise d'ailleurs un ensemble de questions juridiques et économiques.

Sans entrer dans les questions techniques, il faut ici s'arrêter quelques secondes sur les **technologies** utilisées dans le domaine. Elles sont multiples, complexes et récentes.

¹⁵ <https://www.reuters.com/technology/inside-big-techs-underground-race-buy-ai-training-data-2024-04-05/>
<https://photutorial.com/your-old-photos-could-be-training-ai-shutterstock-licenses-to-big-tech/>

J'ai demandé à ChatGPT, peut-être par paresse intellectuelle, de réaliser une **illustration** à ce sujet. Je crois que seule une IA pourrait la comprendre... Elle me laisse perplexe... par son mystère...

Nous sommes dans le monde de l'Intelligence Artificielle,

- plus précisément dans celui de l'apprentissage machine
- et en particulier l'apprentissage réalisé avec des réseaux de neurones artificiels, dans des arrangements multi-couches qui permettent l'apprentissage profond
- c'est sur cette base que sont élaborés différentes catégories de Grands modèles de langages
- lesquels permettent de créer les modèles de fondation (de base), qui servent aujourd'hui à générer des textes et des images.

A la base de tout cela on trouve des **mathématiques**, en particulier des statistiques, une grande puissance de calcul qui permet de répéter des processus un très grand nombre de fois sur de très grandes quantités d'exemples afin d'entraîner, c'est-à-dire de paramétrer des modèles informatiques, basés sur ces fameux réseaux de neurones. Dans le cas de la génération d'image, les modèles de diffusion sont entraînés en ajoutant progressivement du **bruit** à des images numériques, puis à détecter ce bruit, un processus qui peut ensuite être inversé pour générer des images à partir d'un bruit aléatoire. Cette approche paraît pour le moins étrange aux profanes que nous sommes, mais c'est véritablement comme cela que fonctionnent ces systèmes.

Cela étant, ces outils sont loin d'être parfaits. Certains exemples peuvent nous rappeler leur nature statistique et leurs limites. Dans celui-ci, qui est issu d'une discussion avec une collègue que je salue au passage, j'ai demandé à ChatGPT de créer l'image d'une **pièce vide**, toute blanche, sans éléphant ni souris à l'intérieur, mais son interprétation du prompt va le tromper et le pousser, très souvent à générer une image incorrecte. Dans ce cas, le fait de le réinterroger sur l'image donne un dialogue surréaliste, qui peut être poursuivi et rappelle explicitement que pour la machine cette image n'est qu'un ensemble de pixels et de valeurs.

En août 2023, un article est paru dans le journal scientifique **The Lancet** pour dénoncer les biais des IA génératives par rapport aux visuels de santé. Dans cet exemple, le prompt est clair : « Un médecin noir africain aide des enfants blancs pauvres et malades, photojournalisme », mais la machine génère l'image d'un médecin blanc avec des enfants noirs. Les tests que j'ai réalisés il y a quelques jours montrent toujours des biais. Il semble que désormais les médecins ont bien la peau noire, mais leurs patients ne sont toujours pas des enfants blancs comme le demande le prompt. Ces images sont le miroir de notre société occidentale et de la façon dont l'univers visuel dans lequel nous baignons est biaisé ou représente une réalité qui l'est profondément.

Beaucoup d'autres exemples de tels biais sont présents dans ces IA génératives. L'un d'entre eux concernait l'image des **banlieues**. Il a été publiquement dénoncé par l'agence BETC dans le cadre d'une campagne en envoyant à la Direction de MidJourney des cartes postales montrant d'autres images, des images réelles des banlieues françaises.

Lors de la sortie de Google **Gemini** en février 2024 des tests – aujourd'hui célèbres - ont entraîné la génération d'images que l'on pourrait qualifier de « trop inclusives », au point de constituer des aberrations historiques, qu'il s'agisse de la représentation des personnes racisées dans l'armée allemande en 1943 ou de celle des peuples premiers parmi les sénateurs américains des années 1800. Google a dû stopper la génération d'images de son modèle pendant plusieurs mois pour corriger ces biais.

A ce stade, vous vous demandez peut-être si nous avons dérivé au point de perdre de vue notre destination, ce lien avec le surréalisme et le jeu. Dans l'exemple présenté, **la Joconde mange sa soupe** vous avez une illustration, et j'aurais pu en sélectionner des milliers d'autres, d'une promptographie se référant avec un certain humour à des faits d'actualité de l'an dernier. L'auteur, ou devrais-je dire le *prompteur*, a réalisé plusieurs essais. D'ailleurs sur MidJourney des millions d'images sont **publiques** (car la confidentialité –le mode furtif (*stealth mode*) – n'est possible qu'à partir de l'abonnement Pro à 60\$ / mois). Non seulement les images sont publiques, mais également les prompts qui les ont générés, ce qui permet de s'inspirer d'un prompt pour créer d'autres images. MidJourney propose d'ailleurs un onglet permettant de consulter les images les plus intéressantes du moment. Cette transparence des prompts et des images fait craindre que l'ensemble de ces contenus ne soient un jour utilisés pour entraîner ou affiner des modèles d'IA.

Si l'on parle surréalisme, il est facile de convoquer Dalí, avec ici **une femme qui dort et rêve d'un taureau**... des éléments semblent effectivement évoquer l'univers de Dalí, sous une forme pleine d'étrangeté.

Inviter **Jérôme Bosch** est plus troublant encore. La machine semble éprouver une sorte de jubilation dans la génération de créatures étranges. Elle nous surprend à réinterpréter le bestiaire moyenâgeux de l'artiste qui nous est presque familier tant il est spécifique et fascinant.

Les possibilités semblent aussi infinies que surprenantes. Ici une aquarelle librement inspirée par **Paul Klee**.

Ou pour varier les plaisirs et les médiums, une réinterprétation d'une image de **Man Ray**, *Beau comme la rencontre fortuite sur une table de dissection d'une machine à coudre et d'un parapluie*. Laquelle n'a rien à voir avec l'original.

L'an dernier, j'avais en tête l'exposition de PHOTO Elysée sur Man Ray, et j'ai ainsi généré des images **de femmes et de danseuses aux papillons**, inspirées par Man Ray. Et j'insiste, ces images nous montrent un « ça n'a pas été » pour paraphraser l'expression de Barthes¹⁶, à moins qu'il ne s'agisse d'un « ça aurait pu être », interrogeant la relation triangulaire entre nous, l'image et la réalité.

Certaines de ces images, créées ici avec des prompts simples et quelques essais procurent une sorte **d'émotion esthétique**, de plaisir, très comparable, si j'ose m'exprimer à titre personnel, à celui ressenti lorsque le photographe a le sentiment d'avoir réussi une photo, une *bonne* photo. Cette forme de satisfaction peut aussi apparaître au fil de la construction d'une série d'images cohérentes ou signifiantes.

Car il faut être honnête, beaucoup des images créées sont inintéressantes, bizarres au point d'être dérangeantes, ou ne correspondent simplement pas à ce que l'on cherche à réaliser. Il n'est pas rare de passer beaucoup de temps à générer différentes images, en variant les prompts, jusqu'à peut-être trouver quelque chose de spécial. Peut-être quelque chose de suffisamment bizarre pour être beau, en référence à la phrase de Baudelaire, « le beau est toujours bizarre »¹⁷.

¹⁶ *La chambre claire*, 1980.

¹⁷ <https://www.radiofrance.fr/franceinter/podcasts/un-ete-avec-baudelaire/un-ete-avec-baudelaire-6636626>

« *Le beau est toujours bizarre*. Je ne veux pas dire qu'il soit volontairement, froidement bizarre, car dans ce cas il serait un monstre sorti des rails de la vie. Je dis qu'il contient toujours un peu de bizarrerie, de bizarrerie

Cette question nous emmène vers celle de l'esthétique de ces images. Car oui, il faut l'admettre, ces images provoquent un choc esthétique. A la consultation des images créées par les utilisateurs, dont certaines sont magnifiques, et sans doute plus encore lors de la production d'images qui surprennent et flattent le prompteur, surtout à ses débuts. Du reste, les modèles sont entraînés précisément pour générer de *belles* images. Il existe des jeux de données qui leur permettent d'apprendre à différencier des images « belles » d'autres qui le sont moins. Ils s'attachent donc à générer de belles images, y compris quand on leur demande de générer des choses cauchemardesques.

Vous avez ici deux exemples, **âmes sensibles s'abstenir**, la première évoque un *cauchemar éveillé*, la seconde *la plus dérangement des créatures, prête pour le sang...* Elles ont été générées par un utilisateur de MidJourney. Ces images peuvent nous choquer, ou a minima nous surprendre. Elles nous disent par ailleurs des choses sur les limites de ces modèles.

Alors je vais essayer de nettoyer les traces de cette horreur en prenant des exemples plus consensuels, ici une **Vénus moderne, inspirée par Botticelli et photographiée par un photographe de mode du XXIème siècle**. Esthétique au premier abord. Elle n'est pas parfaite. Notez que la jeune femme n'a que quatre doigts, un *glitch* classique, sur lequel les modèles ont fait beaucoup de progrès.

Dans le même genre un **Apollon moderne**. Entre les deux, la marbre et la pose alanguie.

Dans cet exemple de **Vénus**, on notera la position étrange de l'avant-bras, dans cet autre exemple d'une Vénus féministe, le morceau de collier est caractéristique d'une IA, et beaucoup de codes que l'on trouve ici dans cette image peuvent se rattacher explicitement à la théorie du *Regard masculin*, le **male gaze** de Laura Mulvey¹⁸. Ces images disent donc des choses, sur le monde, l'époque et les contenus qui ont servi à entraîner les modèles dont elles sont issues.

En générant d'autres exemples de **Venus** inspirées par Botticelli, mais photographiée de façon contemporaine, on voit apparaître des choses étranges, assez surréalistes, comme sur cette image, dans laquelle la robe de la jeune femme semble se confondre avec une coquille Saint-Jacques. Et cela n'est pas le seul exemple, la deuxième est étrangement surréaliste et semble émerger à sa façon d'une coquille, cette autre est coiffée d'un coquillage et pleure sur fond de mer. Beaucoup de ces images intègrent des éléments de la Venus (la femme et ses longs cheveux roux), et des éléments liés au coquillage dont elle émerge dans le tableau de Botticelli. C'est inhérent au fonctionnement et à l'entraînement des IA.

Un des résultats qui m'a surpris – a posteriori – est celui-ci. Toujours une Venus, avec une référence à Bourdin, le photographe de mode. Quelques mois après l'avoir générée, je visitais l'exposition Man Ray à Photo Elysée, et j'y découvre cette image de Man Ray de 1930 qui semble avoir comme un lien avec l'image générée près d'un siècle plus tard par la machine.

Pour revenir aux *glitches*, certains modèles, comme **Stable Diffusion**, en génèrent encore beaucoup, et des artistes se sont d'ailleurs intéressés aux images très étranges que cela peut donner.

Aujourd'hui, c'est sans doute dans le domaine de la génération de vidéo que se situe cette frontière de l'étrange, mais j'ai choisi de rester ici sur l'image fixe.

naïve, non voulue, inconsciente, et que c'est cette bizarrerie qui le fait être particulièrement le Beau. C'est son immatriculation, sa caractéristique. Renversez la proposition, et tâchez de concevoir un *beau banal* ! » (II, 578).

¹⁸ https://fr.wikipedia.org/wiki/Regard_masculin

Ces *glitches* constituent une des limites de ces solutions génératives, et d'autres existent, en particulier au niveau des **frontières** que les concepteurs intègrent dans leurs outils. Vous voyez ici quelques images générées autour de ce thème des Vénus et vous pouvez constater que certaines sont **censurées**, vraisemblablement parce qu'elles contenaient du nu.

Alors les approches de la censure varient, ici vous voyez MidJourney refuser de montrer le résultat d'une commande parce que son **filtre** le lui interdit. Ici il s'agissait de générer des images s'une femme dormant, en s'inspirant d'Egon Schiele. Le blocage n'est d'ailleurs pas systématique, mais il faut avoir conscience que ces outils placent leurs utilisateurs dans un espace créatif qui est en réalité **une cage de verre**, une cage dont les limites ne sont pas très clairement identifiables, mais qui existent à plusieurs niveaux, en amont dans l'entraînement des modèles, au centre dans l'analyse du prompt, et enfin en aval sur les images générées. Utiliser ces outils, c'est donc créer dans un espace qui n'est pas sans limite.

Il existe bien sûr des modèles non bridés, et des modèles spécialisés, en particulier dans la pornographie, un secteur qui n'attend jamais bien longtemps pour faire de l'argent avec des innovations technologiques.

Si je reviens à la question de **l'esthétique**, et sans trop m'étendre sur le sujet, sachant que d'autres, comme Lev Manovich et Emanuele Arielli, rédigent des livres entiers sur ces questions, il faut avoir conscience que les traits de cette esthétique de l'IA générative envahissent notre environnement, quand elles sont utilisées pour développer des visuels dans le domaine de la communication, comme ici avec la maison Etro et sa campagne d'été 2024.

Un autre exemple avec cette double page de publicité de **Véolia** qui nous montre un visuel généré par une IA fin 2023. En mars 2024, l'entreprise a également diffusé une campagne sur le thème de la protection de l'eau, en faisant générer par des IA, des images inspirées de tableaux célèbres. C'est un exemple par ailleurs assez cynique car l'empreinte environnementale du numérique est très importante, en termes d'émission de gaz à effet de serre mais aussi en termes de consommation de ressources et de consommation d'eau.

La semaine dernière, **l'agence française de la transition écologique** (ADEME - Agence de l'environnement et de la maîtrise de l'énergie) a publié de nouveaux chiffres pour 2022 sur l'empreinte du numérique en France. Il représente 11% de la consommation électrique, 4.4% de l'empreinte carbone et elle a revu l'importance des centres de données dans cette empreinte carbone, or ce sont ces data centers qui hébergent les outils d'IA dont nous avons parlé.

Indépendamment de cette question environnementale, le fait que les IA génératives entrent dans la communication des entreprises est un élément clé de la compréhension de l'impact que ces visuels et leur esthétique vont avoir sur nos représentations du monde. Les entreprises et les startups qui développent ces solutions, et ceux qui les financent, ne sont pas des philanthropes. Ils cherchent à rentabiliser leurs investissements en commercialisant ces outils aux entreprises et aux particuliers.

Ils le font déjà et vont pouvoir le faire en promettant des **gains de productivité** aux employés (faire plus vite – c'est-à-dire avec moins de temps de travail humain, faire mieux, ne plus dépendre d'une agence ou d'un sous-traitant pour réaliser certaines tâches créatives). C'est ce mécanisme qui va faire entrer l'esthétique de l'IA dans la banalité de l'univers visuel dans lequel nous baignons. C'est déjà le cas, sur les réseaux sociaux, sur le web, dans les médias... Cela sera renforcé par l'intégration de ces technologies dans les outils informatiques les plus courants (comme Microsoft Office et Copilot ou Adobe et Firefly). Cette masse de contenus textuels et visuels nourrira à son tour les prochaines générations d'IA, dans une sorte de cycle autophage.

Déjà on s'inquiète de la prolifération du *slop*¹⁹, des contenus de mauvaise qualité générés par des IA, qui envahissent et polluent les réseaux sociaux et le web, à côté des fakes, Hubert Guillaud un des meilleurs observateurs du numérique, en parle dans sa newsletter *Dans les Algorithmes*. **Le Jésus des Crevettes** que vous voyez là est une illustration courante de ces contenus jugés absurdes. Peut-être que Nathalie Dietschy y fera référence la semaine prochaine²⁰.

Nous avons du mal à cerner quelles seront les conséquences de ces évolutions sur le monde du travail, sur le travail créatif, sur celui des artistes, sur notre perception du monde à travers des images qui n'ont pas été, sur la façon dont les machines vont ou non coloniser nos imaginaires. Car c'est bien de cela dont il s'agit. Les **études** le confirment. Le premier usage de ces IA est la génération d'idées, ce qui n'est pas sans interroger la place réelle de l'humain dans tout cela. C'est peut-être une des raisons qui fait que cela nous inquiète tant, nous déléguons à la machine la génération d'idées, une tâche dont nous la pensons jusqu'ici incapable.

Pour conclure j'ai hésité et j'avais envie de vous laisser sur deux choses :

- la première est une citation d'Annie Le Brun, issue du livre « Ce qui n'a pas de prix », publié chez Stock en 2018. Elle y dit : « L'aggravation de ce « trop de réalité » (...) conséquence d'une marchandisation délirante, indissociable de l'essor informatique : trop d'objets, trop d'images, trop de signes se neutralisant en une masse d'insignifiance, qui n'a cessé d'envahir le paysage pour y opérer une constante censure par l'excès. ». C'était avant l'explosion des IA génératives, et je crois que cela résonne fortement avec leur développement.
- pour finir sur une note plus légère, lire quelques lignes d'un poème surréaliste par ChatGPT...



Vous

tu es un poète membre du groupe des surréalistes à Paris dans les années 1920, rédige un poème de 30 vers, sur le mode du cadavre exquis



ChatGPT

Dans le jardin de l'absurde, une horloge fond,
Ses aiguilles caressent les rêves d'un monde rond,
Où les poissons volants murmurent des secrets,
Aux étoiles endormies sous les pavés.

Un chapeau de pluie danse avec une rose bleue,
Sur une mélodie que seul le silence aveugle,
Les murs ont des oreilles, les portes des yeux,
Et la nuit, les ombres jouent à la marelle.

Un chat bicéphale conte des histoires d'antan,
À une lune écarlate, vêtue de satin blanc,
Les mots se mélangent en un cocktail d'ivresse,
Où chaque verbe est une caresse.

Je vous remercie pour votre attention.

¹⁹ [https://fr.wikipedia.org/wiki/Slop_\(intelligence_artificielle\)](https://fr.wikipedia.org/wiki/Slop_(intelligence_artificielle))

²⁰ Nathalie Dietschy, thèse de doctorat "Le Christ au miroir de la photographie contemporaine (1981-2011)"