




Dans le regard

de l'IA



Dans le regard de l'IA est un outil d'animation développé par Média Animation, avec le soutien du Conseil Supérieur de l'Éducation aux Médias (CSEM) et de la Fédération Wallonie-Bruxelles.

Réalisation des imagiers :
Emmanuel Blondeau, Daniel Bonvoisin,
Thomas Gilson

Rédaction :
Nicolas Bras, Daniel Bonvoisin,
Yves Collard, Brieuc Guffens,
Chloé Tran Phu

Conception graphique :
Roland Vixen, Louisa Vanhoucke

Remerciements :
Gil Bartholyens et Philippe Descola

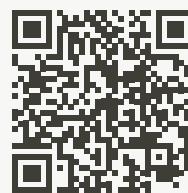
Éditeur responsable :
Anne-Claire Orban de Xivry
– Média Animation

ISBN : 987-2-931139-12-7

Dépôt légal : 2026/3462/1

media-animation.be

Découvrir les imagiers et la démarche d'animation de *Dans le regard de l'IA*



FÉDÉRATION
WALLONIE-BRUXELLES

méd:a
ANIMATION

conseil supérieur
de l'éducation
aux Médias

CSEM

Sommaire

Prompts, étiquetage, data sets et génération prédictive : comment fonctionnent les IA génératives ? 4

De l'AI à l'IA, de l'Artiste Inspiré à l'Intelligence Artificielle 16

Généalogie de l'IA visuelle : entretien avec Gil Bartholeyns 28

Les IA rêvent-elles de patriarcat blanc ? 36

Manières de voir les mondes : entretien avec Philippe Descola 42

Ce que les IA génératives disent de nos cultures visuelles

Les intelligences artificielles génératives produisent des images ou des médias en réponse à une simple commande textuelle (le prompt). Leur argument de vente : permettre à notre créativité de s'exprimer sans limite. Pourtant, les résultats qu'elles génèrent sont conditionnés par de multiples facteurs. Plutôt que de produire des images objectives ou fiables, elles révèlent les représentations du monde statistiquement dominantes dans les

données numériques. Nourries d'une connaissance essentiellement occidentale des images et des représentations, elles portent l'empreinte des systèmes de domination historiques. Elles constituent ainsi un miroir fascinant de nos propres imaginaires visuels et, en creux, de notre méconnaissance des autres manières de représenter le monde.

Cette brochure accompagne l'outil d'animation *Dans le regard de l'IA*, et ambitionne d'équiper tout animateur·rice qui souhaite l'exploiter. **L'objectif de cet outil est double :**

- Faciliter la compréhension d'une technologie d'apparence simple et intuitive, mais qui mobilise une mécanique informatique extrêmement complexe.
- Ouvrir une réflexion critique collective sur les biais nichés dans nos cultures visuelles, et sur la manière dont ils influencent notre perception du monde.

Les textes qui suivent permettent de comprendre le fonctionnement des IA visuelles, complexes et opaques. Ils proposent aussi et surtout d'envisager ces IA au regard d'une histoire globale des images : celle de leur production dans le monde des arts, de leur évolution dans l'histoire moderne, du rôle qu'elles jouent dans la reproduction des stéréotypes et de leur fonction anthropologique.

Prompts, étiquetage, data sets et prédiction : comment fonctionnent les IA génératives ?

Par Nicolas Bras (pour le Centre de Formation d'Animateurs)

On peut avoir l'impression qu'une image apparaît « comme par magie », en réponse à quelques indications adressées à ChatGPT, Gemini ou Copilot. Ces logiciels dont l'usage se banalise sont simples d'utilisation, mais leur fonctionnement est opaque. Comment les « intelligences artificielles générative » interprètent-elles les mots que l'on tape sur notre clavier (le fameux « prompt ») pour nous proposer un résultat visuel ? Explorer le fonctionnement de ces logiciels, interroger la manière avec laquelle ils « s'entraînent » à satisfaire nos attentes, c'est ouvrir une réflexion critique sur ces images et les biais qu'elles contiennent.

Quelques mots d'histoire

Ce que nous appelons intelligence artificielle générative – ou GenAI pour l'anglais *generative artificial intelligence* – est le fruit d'une longue histoire des technologies, de l'informatique et de la robotique. Le fantasme de créer des êtres artificiels dotés d'intelligence remonte, dans nos contrées, à l'ère antique. Golems, automates, robots et autres Frankenstein de tous poils sont quelques-uns des avatars de ce fantasme millénaire.

Pierre angulaire de l'histoire moderne de l'intelligence artificielle : la conférence de Dartmouth. En 1956, alors que l'informatique grand public est une vue de l'esprit, un groupe composé d'une quinzaine de chercheurs¹ en informatique et en psychologie se réunissent pendant deux mois pour travailler sur l'idée de machines capables d'imiter voire de surpasser l'intelligence humaine à l'aune des technologies émergentes de l'informatique. Le terme « intelligence artificielle » naît à ce moment-là pour fédérer chercheurs et financiers autour d'un concept accrocheur. En d'autres mots, la

1 Dans l'histoire de l'informatique, le travail des chercheuses et techniciennes a été largement invisibilisé. La conférence de Dartmouth n'a, ainsi, rassemblé que des hommes. Le choix d'écrire ici « chercheur » sans emploi de l'écriture inclusive rend donc état de cette discrimination.

terminologie « intelligence artificielle » pour délimiter ce champ d'étude trouve son origine dans la nécessité de trouver un slogan accrocheur.

Aujourd'hui, certain-es chercheur-euses préfèrent utiliser le terme d'Informatique Avancée pour démystifier ce que nous nommons Intelligence Artificielle. Cet usage sémantique vise à éviter l'amalgame avec les intelligences humaines tout en accréditant le renouvellement d'une structuration numérique permise par les nouvelles avancées technologiques. L'émergence d'intelligences synthétiques capables de conscience et de sensibilité est un horizon souhaité par certain-es, terriblement craint par d'autres, qui occupe régulièrement l'espace médiatique. Mais, en l'état des recherches sur le sujet, cette hypothèse relève de la science-fiction.

Derrière l'usage du terme « intelligence », le vocabulaire de ces programmes est truffé de termes qui renvoient au vivant : apprentissage, neurones, langage... De la même manière que le stockage informatique de données disponibles via Internet dans des serveurs gigantesques et énergivores est nommé le « Cloud » (le nuage), le champ lexical en lien avec les IA est orienté vers des formes d'anthropomorphisation.

C'est un leurre : les créations générées par IA sont avant tout le résultat de générations probabilistes extraordinairement poussées à partir d'un traitement sophistiqué de milliards de données textuelles et imagées, transformées en chiffres.

Qu'est-ce qui se cache derrière le concept d'intelligence artificielle générative ?

Les GenIA sont des programmes informatiques complexes qui sont en mesure de générer du contenu à partir d'un **prompt**. Il existe deux grands types d'IA génératives grand public : les génératives de textes et les génératives d'images.

Du point de vue des utilisateur-rices, la génération d'une image fonctionne par la réponse d'une machine à la rédaction d'un prompt grâce à un programme dédié. Pour générer des images, les logiciels d'IA génératives doivent avoir accès à une base de données gigantesque d'images entraînées et d'algorithmes sophistiqués qui permettent la transformation d'un texte en image en passant par du code chiffré. Les paragraphes qui suivent précisent chacun des termes évoqués ci-dessus.

Qu'est-ce qu'un « prompt » ?

Un prompt est une suite de caractères soumis à un logiciel d'IA pour générer du texte, des images, des vidéos ou tout autre type de contenus. Quand on tape sur « enter », une première opération déterminante a lieu : la traduction du prompt en langue anglaise. En effet, les logiciels popularisés chez nous ont été développés par des entreprises nord-américaines : elles ne fonctionnent que sur base de commandes et « d'étiquettes » (voir plus loin) en anglais. Déjà peuvent se nicher, à cette étape du processus, une série de biais culturels ou d'erreurs de traduction.

La « traduction » se poursuit : ces caractères se transforment en chiffres afin de générer de nouvelles suites de chiffres qui seront à leur tour traduites en texte ou en image pour fournir une réponse à l'utilisateur·rice.

En d'autres termes : le prompt correspond à la requête effectuée par un·e utilisateur·rice (qui peut-être une machine) pour générer du contenu après une transformation chiffrée.



Prompt: spongecat

(chat éponge)

Généré sur Grok

Qu'est-ce qu'un LLM ?

Un grand modèle linguistique (**Large Language Model** – LLM – en anglais) est un type de modèle d'IA générative entraîné sur de gigantesques ensembles de données textuelles pour produire du texte. En d'autres termes, ce sont des programmes informatiques nourris de milliards d'exemples textuels. Ce qui permet à la machine de reconnaître le langage humain. Ces programmes sont alimentés par les données disponibles sur Internet. Penser aux GenIA textuelles nous renvoie spontanément vers ChatGPT, Grok et autres Gemini. Mais les LLM sont aussi utilisés dans les moteurs de recherche en ligne, sont au cœur de chatbots, sont exploités pour des usages très précis comme des recherches scientifiques autour de l'ADN, ou encore pour des expériences en reconnaissance de sentiments. L'entraînement de ces modèles requiert des dispositifs techniques très puissants et énergivores, nécessitant l'utilisation de processeurs puissantes.

Les IA génératives d'images ont des fonctionnements proches des LLM (entraînement des données, récolte de données imagées en ligne, modèle de fonctionnement gourmand en énergie) mais pour des résultats imagés plutôt que textuels.

L'un des moments clés dans l'histoire de la génération d'image est la création par OpenAI – société à l'origine de ChatGPT et Dall-E – du processus CLIP, le Contrasting Language Image Pre-training. CLIP est un modèle qui a la capacité de générer des étiquettes à partir d'une image qui lui est soumise. C'est donc un modèle de type automatisé qui a permis de créer, à une vitesse record, la source de données – binôme image et texte – nécessaire pour générer des images.

Qu'est-ce qu'un « data set » d'images ?

L'un des éléments centraux de la génération d'images est la constitution de **bases de données** gigantesques (les « data sets »). Ces bases de données rassemblent **plusieurs milliards d'images entraînées pour réagir aux prompts** rédigés par les utilisateur·rices – les inputs – et générer des images – les outputs.

La source des bases de données est principalement la masse d'informations disponible sur Internet. Cette réalité pose un important problème légal. Car les images qui alimentent les bases de données sont purement et simplement capturées sans la moindre demande de consentement des ayants droits et des personnes dont l'image est utilisée. Des ayant-droits ont déjà fait part de leur mécontentement face à cette situation ; d'autres comme la plateforme Getty Image sont partis en procès contre des opérateurs comme Stable Diffusion et Midjourney pour faire valoir leurs droits. De l'aveu même des opérateurs à l'origine des programmes d'IA générative, accumuler les bases de données nécessaires au fonctionnement fluide de ces outils aurait été impayable. Des milliards d'euros sont évoqués.

Le choix des structures à l'origine du développement des IA génératives a donc été d'imposer cette technologie au grand public en faisant fi de plusieurs principes juridiques. Aujourd'hui, certains opérateurs achètent des réseaux sociaux – Elon Musk et X, ex Twitter, par exemple – ou tissent des accords avec d'autres pour disposer d'un accès à leurs bases de données, qui alimenteront les leurs (Grok, dans le cas de Musk).

Après plusieurs années de génération d'images, Internet se retrouve progressivement inondé d'images générées par logiciels d'IA génératives. La corrélation directe est leur présence croissante au sein des bases de données de ces mêmes logiciels.

Qu'est-ce que « l'étiquetage » des images ?

Pour que le programme soit en mesure de transformer ces prompts en images, il faut d'abord qu'aient été « étiquetées » les données (rassemblées dans le « data set »). Les étiquettes, ce sont des mots attribués aux images dans leur ensemble ainsi qu'aux parties qui la composent. Le travail d'étiquetage est central dans **l'entraînement des bases de données** car c'est par ce travail que le logiciel est en mesure de reconnaître l'agencement de pixels en formes, en couleurs, en mouvements, en texture, en matière... L'efficacité et la justesse de la génération d'images dépend en grande partie de la qualité de l'étiquetage.

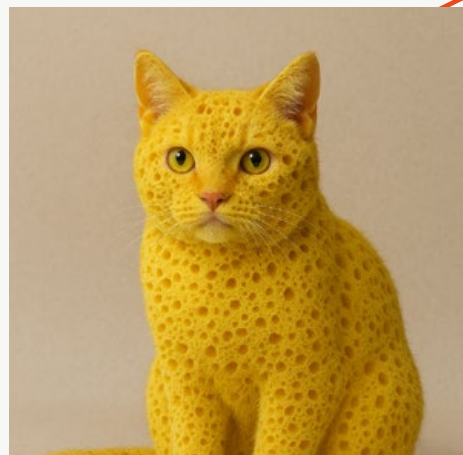
La machine est en mesure de plus ou moins bien générer un « chat éponge » grâce aux millions d'images de chat et d'éponge qui ont auparavant été catégorisées (étiquetées) comme « chat » et « éponge ». Si chaque image de chat avait été étiquetée sous le nom « rhododendron », nous devrions rédiger le prompt « rhododendron » pour que le logiciel génère une image de ce que nous reconnaissons comme un chat². La création de corrélation entre des étiquettes et des images est ce que nous appelons « **l'entraînement** ».



Prompt: spongecat
running in a flowerfield

chat-éponge courant
dans un champs de fleurs

Généré sur Stable Diffusion



Prompt: realistic cat –
textured sponge cat

(chat réaliste – texture d'éponge)

Généré sur Dall-e/Open AI

2 Notons déjà plusieurs éléments sur base de cet essai inoffensif. Aucune image générée n'a emprunté la texture de l'éponge marine. C'est uniquement des éponges de vaisselles qui ont été sollicitées dans la génération d'image. Aussi, traduire le prompt français « chat éponge » en « spongecat », mène certains logiciels à générer des « Bob l'éponge » du dessiné éponyme vaguement « félinisés » (la série se nomme Sponge Bob Square Pants en anglais).

Qui est responsable de l'étiquetage des données ?

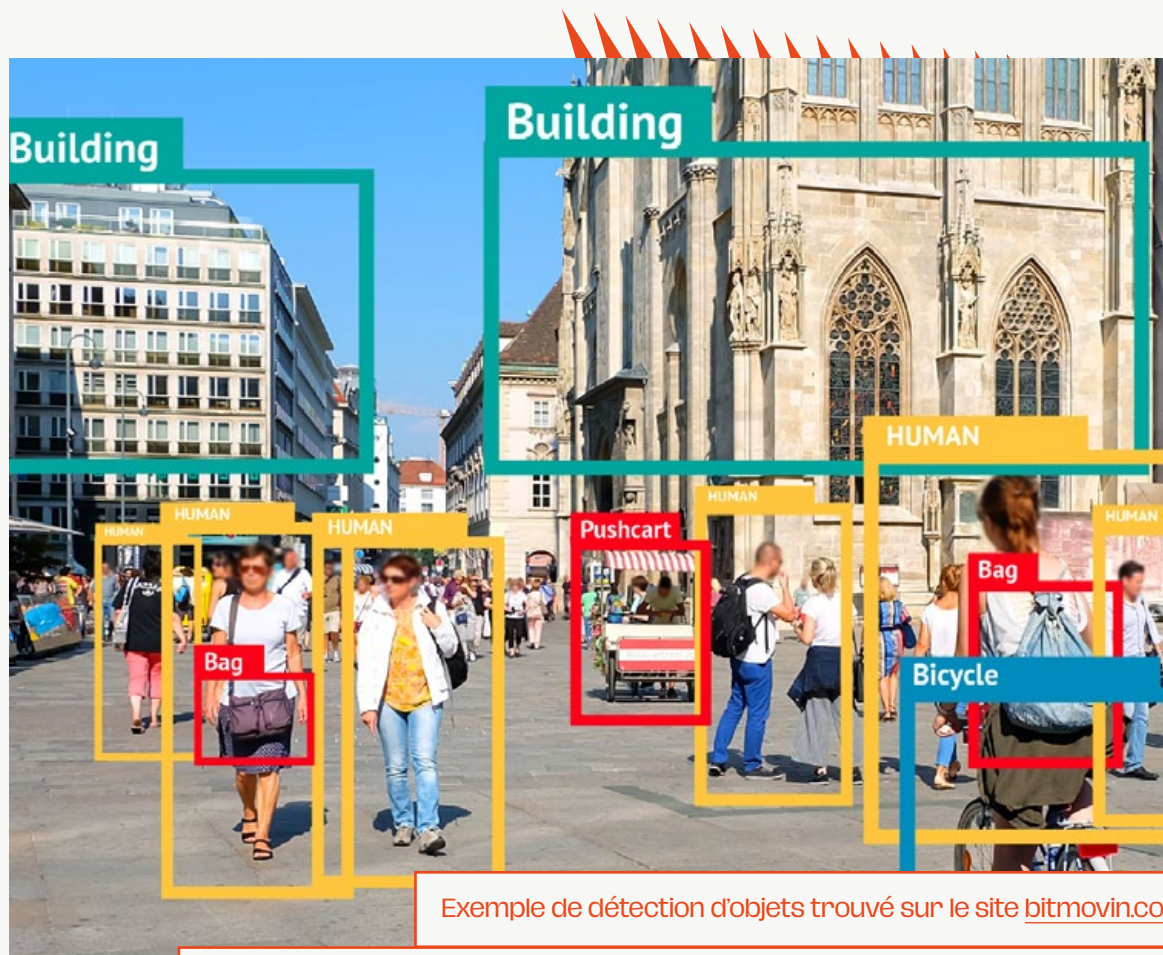
Dans les premiers temps de l'entraînement des modèles, ce sont des personnes humaines qui se sont chargées manuellement de catégoriser chacune des images afin que le logiciel les reconnaisse. Aujourd'hui, l'étiquetage est plus ou moins automatisé en fonction des usages. Voici quelques méthodes d'étiquetage :

- **L'annotation manuelle** : celle qui requiert la présence humaine. Elle est aujourd'hui privilégiée pour la création d'IAG qui nécessitent des étiquettes les plus précises possible. Pour une IA médicale par exemple, l'annotation humaine est nécessaire. Vu le **caractère sensible** de ce type de données, l'étiquetage humain est validé par d'autres annotateur·rices expérimenté·es afin d'assurer un traitement rigoureux des données. Les IA génératives grand public n'entrent pas dans cette catégorie.
- **L'annotation assistée par l'IA** : des modèles d'IA sont **pré-entraînés** pour étiqueter des images via des boîtes englobantes, des classifications et/ou des masques de segmentation (voire le texte concernant les zones d'étiquetage). Des personnes vérifient ensuite les annotations effectuées par l'IA, les adaptent quand c'est nécessaire et alimentent ainsi les bases de données.
- Enfin, **des processus totalement automatisés** sont aujourd'hui à l'œuvre. Les données sont soumises à l'étiquetage d'un modèle d'IA pré-entraîné sur d'immenses ensembles de bases de données afin de produire un étiquetage sans intervention humaine. Ces modèles sont réputés efficaces sur des données standards dans des contextes communs. L'un des résultats de ce type de modèles est de favoriser le **renforcement des standards** – autrement dits, des clichés et stéréotypes – vu que la base de données d'images s'autoalimente à partir des mêmes données. Le programme sera également plus à même de générer des contenus farfelus. Dans un langage de type anthropomorphisant, on parle alors d'**hallucinations** de la machine. Enfin, c'est un processus très gourmand en puissance de calcul et qui est donc énergivore.

Comment fonctionne l'étiquetage des images ?

Les images sont étiquetées en fonction d'un découpage de l'image en zones. Les zones sont de plusieurs ordres et créent des qualités de précisions diverses :

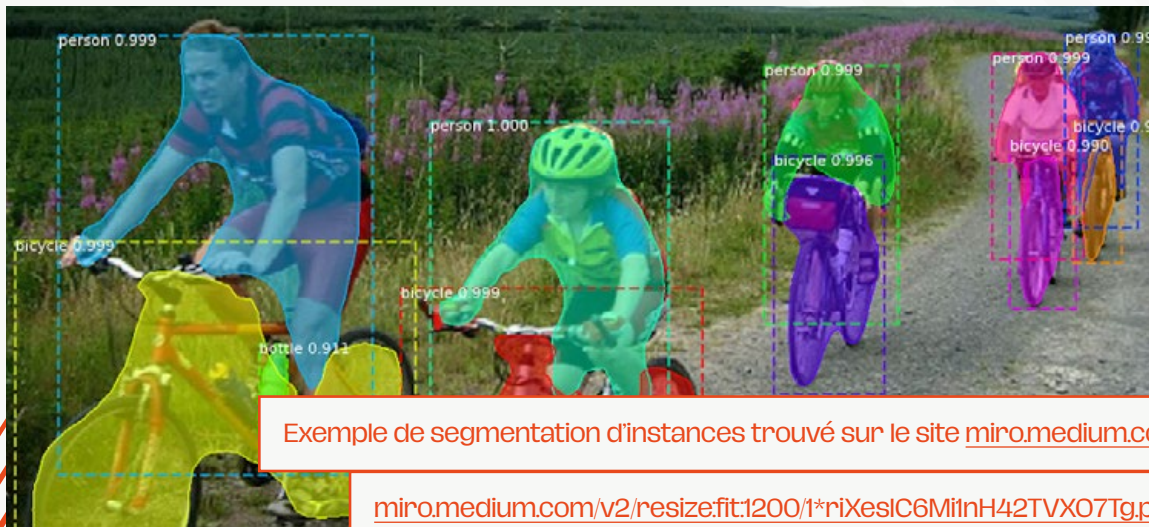
- **La classification d'images** correspond à la méthode la plus simple et la moins précise. L'annotateur-riche – informatique ou humain-e selon que la méthode utilisée soit automatique ou manuelle – annote l'intégralité de l'image avec une ou plusieurs étiquettes.
- **La détection d'objets** est un système par lequel l'annotateur-riche crée des boîtes rectangulaires qui entourent des objets et leur attribue une ou plusieurs étiquettes.



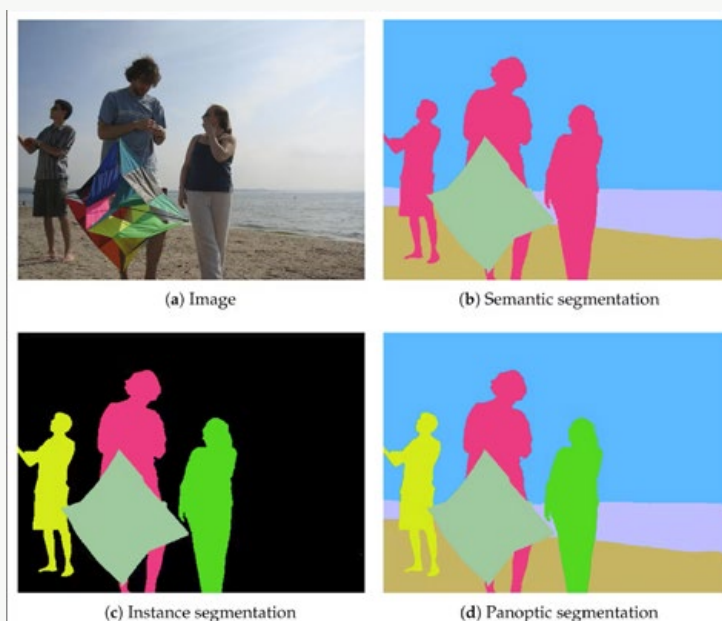
Exemple de détection d'objets trouvé sur le site bitmovin.com

bitmovin.com/wp-content/uploads/2019/08/Object_detection_Blog_Image_Q3_19.jpg

- **La segmentation d'instances** permet de créer des contours précis aux objets détectés. Alors que la boîte de la détection d'objets est un rectangle qui englobe également des éléments parasites, la création de contours précis évite les parasites en englobant exclusivement le contenu concerné par les étiquettes.



- **La segmentation sémantique** est un procédé plus complexe et plus complet qui consiste en une application de masques sur l'image qui apportent des couches sémantiques dans l'analyse de l'image. En plus de sélectionner les éléments à analyser, les couches sémantiques permettent au logiciel d'enregistrer les relations spatiales ainsi que les propriétés des matériaux dans des environnements complexes. Quel est le rôle de ces étiquettes dans la génération d'images ?



Exemple de segmentation d'instances trouvé sur [encord.com](https://encord.com/blog/guide-to-semantic-segmentation/)

encord.com/blog/guide-to-semantic-segmentation/

Ces différentes méthodes permettent de créer des binômes texte/image plus ou moins complexes et pertinents qui alimentent les bases de données. Les images des bases de données sont accompagnées de textes – les étiquettes – qui permettent aux logiciels de trouver les images utiles à la génération d’une réponse – un *output* – pertinent en fonction du prompt – l’*input* – de l’utilisateur-riche. De manière logique, une IA générative sera donc incapable de produire l’image de quelque chose qui n’a pas été étiqueté comme tel dans sa base de données (et improvisera un *output*).

Techniquement, comment le programme génère-t-il des images au départ d’un prompt ?

Le processus de génération d’image utilise un **algorithme – une boucle – que l’on nomme *Diffusion***. Le fonctionnement technique est le suivant :

- Le processus démarre par une **image composée exclusivement de bruit**. Cette « image » est un tas de pixels aléatoire sans structure ni signification. On peut imaginer ce point de départ comme une version numérique du brouillard de l’écran de télévision.
- À partir de ce bruit, **l’algorithme fait surgir l’image** que nous lui avons commandée à partir du prompt. C’est un processus qui fonctionne par étape.
- À chacune des étapes :
 - ❑ L’algorithme **réduit le bruit** en réponse au prompt. Lors de la première étape, le logiciel va **puiser des images qui correspondent au prompt** et va faire émerger une image avec encore 90 % de bruit.
 - ❑ À chaque étape, le logiciel va **confronter l’image obtenue à sa base de données** constituée lors de l’entraînement pour évaluer si ce qui est lisible correspond au prompt rédigé.
 - ❑ En parallèle, le logiciel génère une image avec la même quantité de bruit mais sans se baser sur le prompt. Le logiciel compare alors les deux images obtenues (celle générée avec et l’autre sans prompt), détecte les différences entre elles et renforce les particularités de l’image dotée d’un prompt.
 - ❑ Le programme redémarre ensuite le processus en **réduisant la quantité de bruit sur l’image** obtenue pour arriver à une image qui ne contient plus de bruit.

Ce processus est nommé « diffusion de type arrière » : le point de départ est le bruit, le point d’arrivée est une image. Le programme doit avoir expérimenté au préalable le fonctionnement inverse à savoir la « diffusion avant » pour réaliser une « diffusion

arrière ». Le point de départ de la « diffusion avant » est l'image, le point d'arrivée, le bruit. Les générateurs d'images actuels utilisent tous des modèles de Diffusion avec des variantes qui leurs sont propres.

On le constate : le fonctionnement des IA génératives est complexe, et vulgariser leur fonctionnement est épineux. Mais s'y essayer permet de dévoiler la dimension « probabiliste » du processus. Partant d'un agglomérat de pixels, jouant de comparaison avec des images exploitées comme ressources, le logiciel choisit à chaque étape ce qui lui semble correspondre à la meilleure association d'un mot du prompt avec un résultat visuel.

Quel programme utiliser ? Comment choisir ?

Les outils de génération d'image les plus populaires tels que Midjourney, Dall-E – le générateur d'images d'Open AI inclus dans ChatGPT –, Crayion ou Grok ont plusieurs caractéristiques communes :

- Ils sont disponibles en ligne.
- Ils imposent à l'utilisateur·rice de se connecter à leur plateforme « mère » (Google par exemple) pour avoir accès aux services.
- Ils proposent des abonnements payants pour bénéficier de l'intégralité des services.
- Ils fonctionnent uniquement à partir de prompts.
- Ils sont, pour la plupart, particulièrement puissants, efficaces, et faciles d'accès.

Ces solutions « clé sur porte » ont évidemment des zones d'ombre :

- Récupération des données des utilisateur·ices qui utilisent ces logiciels.
- Opacité de leur fonctionnement.
- Impact environnemental : ces solutions en ligne sont particulièrement énergivores, car les logiciels font fonctionner leurs infrastructures décentralisées pour générer la réponse à la requête.

En quelques mots : ces programmes en ligne, populaires et faciles d'usage, sont autant de boîtes noires qui récupèrent potentiellement un maximum de données des utilisateur·rices pour alimenter leurs bases de données et leur modèle économique.

Ces outils numériques fonctionnent en osmose avec le reste de l'environnement numérique, les incontournables outils numériques en ligne dont le modèle est

l'extraction et l'exploitation des données des utilisateur·ices (en très bref : les GAFAM³). Par le contrôle, la concentration et l'usage mercantile des données, ces outils sont, structurellement, des menaces pour nos modèles démocratiques. OpenAi, Grok et les autres fournisseurs d'outils d'IA suivent ce même chemin.

Une vigilance vis-à-vis de ces programmes s'impose, et ce malgré les tentatives de régulation de l'espace numérique européen par les institutions européennes...

Notons qu'il existe des réponses étatiques et associatives pour créer d'autres modèles d'IA génératives. Quelques options « open source » se distinguent favorablement du microcosme présenté ci-avant, comme Stable Diffusion, projet imaginé pour, justement, fonctionner en dehors de l'espace commercial.

Peut-on « prendre la main » sur le processus de création d'une image par IA générative ?

Comme souvent dans le registre informatique, pour s'éloigner des solutions proposées par les géants du Web, il faut un peu de temps et d'énergie. Ces solutions « open source » ont l'avantage de permettre à l'utilisateur·rice d'avoir (dans une certaine mesure) la main sur le processus de génération des images.

En mettant à disposition leur code source, elles ont permis le développement de solutions disponibles en téléchargement (afin d'avoir le logiciel sur son ordinateur, et pas en ligne) qui permettent « d'ouvrir la boîte noire » de la génération d'images. À contrepied de la majorité des autres générateurs d'images, un logiciel open source comme Stable Diffusion donne l'accès à de nombreux paramètres au moment de générer une image :

- La possibilité d'orienter la génération d'images en choisissant un modèle. Un modèle est une base de données réduite. Ce choix permet d'orienter l'esthétique de l'image obtenue mais aussi d'en orienter les résultats en matière de représentation par exemple.

3 Acronyme de Google, Apple, Facebook, Amazon et Microsoft. Il devrait, en regard des « rebranding » récents, être modifié en AAMAM (Alfabet, Apple, Meta, Amazon et Microsoft).

- La possibilité de choisir le nombre d'étapes informatiques de génération d'une image, ce qui permet de visualiser concrètement les étapes effectuées par la machine lorsqu'elle affine l'image depuis le « bruit », et de choisir le moment où l'on « arrête » le processus de génération parce que le résultat correspond mieux à ce que l'on cherche.
- La possibilité de pondérer le degré de fidélité au prompt.

L'IA générative: outil de création assistée et poste d'observation des normes

En rendant leur processus « génératif » invisible, les IA génératives masquent la complexe mécanique algorithmique qui est à l'œuvre : elle n'a rien d'intelligent, ou de magique. Ces logiciels « produisent » une traduction visuelle par l'entremise algorithmique, et ne « créent » pas. Ce geste créateur, pour qui utilise l'IA, se bornera probablement à « challenger » le mécanisme pour obtenir un résultat satisfaisant. Si l'on peut considérer les images produites comme fiables, c'est peut-être avant tout pour leur capacité à synthétiser les normes sociales et esthétiques. Malgré elles, les IA génératives facilitent finalement une éducation critique à l'Histoire de nos représentations visuelles.

Pour aller plus loin

- À propos de la notion de « diffusion » : www.guidedlare.com/fr/plateforme/suite-images/technologie/modeles-diffusion-generateurs-images-ia
- À propos de la notion « d'étiquetage » : www.innovatiana.com/fr/post/image-labeling-complete-guide
- À propos du fonctionnement des IA génératives : www.youtube.com/watch?v=47BIShlc4E8

De l'AI à l'IA, de l'Artiste Inspiré à l'Intelligence Artificielle

Par Yves Collard

On aime imaginer l'histoire de l'art comme une fresque héroïque. Une fresque peuplée de génies solitaires, penché-es sur leur toile à la lueur tremblante d'une bougie pour concevoir une œuvre unique. Un Léonard de Vinci plongé dans le sourire de la Joconde, un Van Gogh en sueur devant ses tournesols, un Picasso en bagarre avec la perspective. Cette vision romantique forgée au XIX^e siècle nourrit nos imaginaires et alimente nos manuels scolaires. Pourtant, elle résiste mal à l'examen des pratiques réelles.



«Pourrais-tu me générer une image photoréaliste d'un artiste au travail»

ChatGPT reproduit une vision romantique du travail artistique. Il est opéré par un homme seul, concentré sur son œuvre, dans un atelier vétuste.

La main de l'artiste bohème est dirigée par ce qu'il a dans les tripes.

Derrière la posture de l'A.I. (Artiste Inspiré), on trouve plus souvent un système collectif, parfois quasi mécanique. Rares sont les œuvres, au cours de l'histoire, qui ont été créées d'une seule main, à partir d'un seul cerveau. Elles se fabriquent dans le tumulte des ateliers, organisés en hiérarchies et délégations de tâches. L'artiste ne peint pas toujours lui-même. Il reçoit la commande (au Moyen-Âge, d'un pouvoir ecclésiastique ou économique), organise le travail et prescrit l'œuvre finale. Celle-ci est l'aboutissement d'exécutions multiples au service d'une vision centrale. Le cinéma en offre une illustration contemporaine : dans l'industrie cinématographique, le générique de fin des films rend visible la multitude de métiers au service de l'œuvre.



Le générique d'un film dévoile la multitude de contributions artistiques et techniques, subordonnées à une vision singulière : celle du ou de la cinéaste.

Mais ces contributions sont aussi conditionnées par les enjeux de production et de rentabilité.

The Brutalist, Brady Corbet, 2024

En amont de celle-ci, les courants artistiques offrent des matrices créatives et génèrent des modèles dans lesquels les artistes s'engouffrent : il n'y a qu'à lorgner du côté des ordres grecs anciens en architecture ou des courants picturaux modernes (fauvisme, pointillisme, cubisme, surréalisme...) pour s'en convaincre.



Les artistes entrent en connivence, s'influencent, collaborent et co-exposent.

Ils et elles se rassemblent derrière une vision, esthétique ou politique, parfois ancrée dans un « manifeste ».

Le mécanisme qui consiste à formuler des règles pour que d'autres les matérialisent trouve un écho évident dans les logiques de l'intelligence artificielle. Là où l'apprenti transposait hier les indications du maître, l'algorithme, nourri par un prompt, matérialise aujourd'hui un projet. Vue ainsi, l'histoire de l'art apparaît comme une préhistoire de l'IA : une longue genèse où l'humain apprend à déléguer son regard à la machine. Une variation collective sur un même thème précédant de loin l'invention de ChatGPT.

La préhistoire : corps interchangeables et têtes mobiles

Dès les sociétés préhistoriques, une étonnante plasticité des images se manifeste. Certaines Vénus néolithiques, comme celle du temple de Tarxien à Malte, ont un corps généreux mais celui-ci est dépourvu de tête, et percé de trous d'attache. Tout suggère que des têtes amovibles pouvaient s'y adapter selon les besoins du rituel, faisant varier l'identité de la statue à partir du modèle de base. À Çatalhöyük, en Anatolie, il y a près de dix mille ans, des figurines à tête détachable confirment cette logique de permutation, première tentative de customisation de l'histoire, plusieurs millénaires avant les filtres Instagram, et cela sans 5G.

Changer le visage et garder le corps : le modèle demeure, l'identité fluctue. L'œuvre est une combinaison d'éléments. Une logique qui anticipe celle des systèmes algorithmiques : un prompt fixe génère des variations multiples. Les images produites par une IA matérialisent des résultats différents à partir d'un même cadre.



Cette idole de Malte provient du temple d'Agar Qim, vers 3000 avant l'ère commune.

Tout indique que l'on pouvait y placer la tête de son choix. Une sorte de «face swap» avant l'heure.

Source : Wikimedia. Photo : Gerhard Huber – CC-BY-NC

Le « face swap », banal sur les réseaux sociaux, procède de la même idée : un invariant accueille une infinité de possibles. Loin d'être anecdotique, le détail archéologique révèle combien la logique du prompt plonge ses racines dans des pratiques immémoriales.

L'Antiquité : la matrice et le geste collectif

Dans l'Antiquité, les célèbres bronzes grecs, tels que le *Discobole* de Myron ou l'*Apollon Sauroctone* attribué à Praxitèle, n'ont survécu que grâce aux copies romaines. Ces statues étaient conçues comme des prototypes reproductibles. Chaque exemplaire pouvait varier légèrement selon la commande, dont la taille, mais la posture et la structure générales demeuraient, tout comme certains détails comme les drapés. La sculpture romaine obéit à la même logique : les copies en marbre des chefs-d'œuvre grecs étaient souvent produites en série dans des ateliers spécialisés, où les ouvriers se partageaient la taille du bloc, le lissage et la finition sous la supervision d'un maître.



Internet a généralisé une pratique pourtant millénaire : ici un copié-collé en fonte du *Discobole* présent au Musée du Vatican, pour la Harvard Law School, aux États-Unis.

Source : Wikimedia

La peinture suit le même schéma : à Pompéi ou à Rome, les fresques et mosaïques étaient réalisées à partir de cartons préparatoires. Les mosaïques romaines, avec leurs panneaux géométriques ou figuratifs, permettent de bien se figurer le modèle matriciel : un dessin initial, des mains multiples exécutant selon un code commun.

L'atelier médiéval : la fabrique de l'anonyme

Avançons dans le temps. Au Moyen Âge, l'art s'inscrit dans une perspective religieuse. Il faut fournir aux croyant·es beaucoup d'images de dévotion, ce qui entraîne une rationalisation de la production, déjà proche d'une forme de série. C'est aussi l'époque des moines copistes. Derrière chaque œuvre se déployait un réseau de disciples prenant en charge certaines parties du travail sous la direction du maître. On voit ainsi se multiplier les variations sur le même thème. Le *Triptyque des Sept Sacrements*

d'Anvers ou le *Triptyque de la Rédemption* à Madrid, attribués à Rogier van der Weyde se ressemblent comme deux gouttes d'huile. La main du maître ne touchait pas toutes les surfaces : certains éléments étaient confiés aux assistants qui obéissaient à des consignes précises, ce qui confère à ces œuvres bien des similarités. Les analyses récentes ont confirmé l'intervention de Rogier dans les parties clés comme les visages et les mains, tandis que d'autres zones montraient les interventions subordonnées. Dès cette époque, l'art apparaît comme une production distribuée, presque industrielle, et liée aux impératifs du marché. Une intelligence collective avant l'heure. La logique de production est proche de celle des plateformes numériques : un travail partagé où la valeur finale ne vient pas d'un seul mais d'une multitude.

Les retables fournissent de nombreuses illustrations de ce modèle. Leur compartimentage méthodique traduit la délégation systématisée du travail : le créateur conçoit la structure globale et fixe les règles, d'autres exécutent le canevas. À l'aune de nos préoccupations contemporaines, on peut voir ce fonctionnement comme un précurseur de la production générative : un cadre central régi par des règles précises permet d'accueillir des motifs variés tout en conservant la cohérence globale. Le célèbre polyptyque de *l'Adoration de l'Agneau mystique*, commencé par Hubert Van Eyck et achevé par son frère Jean en 1432, permet une autre lecture plus audacieuse du mécanisme artistique, davantage orienté sur la réception. Au centre de l'œuvre, l'agneau sacrificiel condense et active tout le système de lecture. Il joue le rôle de « mot-clé », comme une indication qui déclenche la compréhension des scènes environnantes et organise tout l'arrière-plan théologique. Un peu à la manière d'une IA, ce motif simple (comme le serait le mot d'un prompt) suscite une interprétation cohérente tandis que les autres détails visuels, périphériques, viennent décorer la scène sans en modifier le sens générique. Le retable n'est pas un simple objet : il est conçu pour provoquer chez le public une série d'émotions spirituelles. Il guide les regards de la communauté des croyants pour générer une expérience mystique vécue collectivement. Les deux dispositifs reposent sur une culture partagée : pour que cela fonctionne, il faut un vocabulaire et une grammaire partagés, un répertoire commun.

Renaissance et âge d'or des ateliers

À la Renaissance, l'art sort des églises. Il se répand dans les demeures de la haute bourgeoisie. Comme des posters qui décorent les chambres des ados, chacun veut la même peinture que son voisin. L'artiste reste central, mais derrière lui s'active un réseau collectif. Léonard de Vinci, Michel-Ange ou Raphaël incarnent cette tension entre inspiration individuelle et travail d'atelier. Dans les grandes compositions, le maître se réserve l'essentiel, là où se concentre l'émotion et la virtuosité, bref le « génie ». Tout le reste est confié à une armée d'assistants suivant des indications précises ou des esquisses préparatoires. Chez Raphaël, le dispositif atteint son apogée : cinquante collaborateurs travaillent sous sa supervision. Les œuvres pleinement « signées » par Raphaël sont rares, témoignages ponctuels d'une main qui dirige plus qu'elle n'exécute.



Raphaël connaît lui-même les contraintes du système où l'assistant doit se conformer aux indications du maître. Mais il choisit une voie différente : il s'entoure d'artistes compétents pour leur permettre de développer une voix propre tout en respectant l'unité de l'ensemble. Penni et Romano sont ses principaux collaborateurs. Penni met au net les compositions, Romano réalise certaines œuvres presque intégralement. L'atelier est une machine de production aux talents multiples, orchestrée par un personnage central. Le résultat préfigure la logique de systèmes où conception et réalisation sont déléguées mais coordonnées, ce qui sera observé plus tard dans les studios de BD.

En Flandres, cette organisation prend une tournure quasi industrielle. Rubens, superstar de son temps, multiplie les commandes et se retrouve souvent *sold-out*. Pour répondre à l'afflux de travail, il élabore un système automatisé : une esquisse rapide sert de modèle, les assistants agrandissent et complètent. Le maître finalise par des retouches sur l'expression et la lumière. Sa « touche » devient sa marque de fabrique, une signature apposée sur une toile presque achevée, à la façon d'un dialogue entre l'intelligence artificielle et son usager-ère. Chaque toile devient un système modulable, dans lequel le maître n'intervient que pour ajuster le travail. Comme avec un algorithme, c'est la combinaison de directives et d'exécutions multiples qui produit le résultat final.

Pieter Brueghel l'Ancien applique une logique similaire pour ses scènes paysannes. Dans son atelier, la production était si intense et codifiée qu'elle ne laissait place qu'à de petites variations, parfois incongrues. Il fournit à ses collaborateurs des indications écrites sur les thèmes à représenter, ce qui laisse place à l'interprétation. Les copies réalisées par son fils Pieter Brueghel le Jeune respectaient la composition générale, mais pouvaient intégrer des détails différents, et comportaient parfois des bugs de lecture. Une anecdote célèbre mais non attestée, évoque la confusion entre « paardje » (petit cheval) et « paarje » (petit couple) dans l'une de ces copies. Ce glissement lexical illustre à merveille la logique de l'atelier. Celui-ci devient un système génératif analogue aux IA : un même cadre, une infinité de variantes générées par des prompts,



Le dénombrement de Bethléem

Pieter Brueghel l'Ancien (1566 – 1569)

semblables mais jamais identiques. Les ateliers produisent ainsi des œuvres en série comportant, pour chacune d'elles des approximations, voire des erreurs induites par un processus de reproduction presque mécanisé.

Une autre discipline éclaire cette logique prescriptive : la musique écrite. La partition n'est pas la musique elle-même mais un langage codifié permettant à l'interprète de donner vie à l'œuvre. Cette partition fonctionne déjà comme un prompt analogique : elle oriente l'exécution tout en laissant place à l'interprétation du musicien. L'écriture musicale illustre ainsi, bien avant l'ordinateur, la dissociation entre conception et matérialisation.

L'essor de la photographie au ^{xix}^e siècle confirme cette logique. Poser devant l'objectif, choisir un cadrage, une lumière, un décor, revient à formuler un *prompt* visuel. Chaque détail constitue une instruction implicite adressée à la machine, qui fixe ensuite l'image.

Le portrait photographique repose sur des plans et des décors systématisés : la bibliothèque pour le savant, le bureau pour le notable, la mappemonde ou la carte pour l'explorateur. Ces scénographies codifiées rejouent des archétypes sociaux et transforment chaque individu en variante d'un modèle. Le portrait devient ainsi un exercice de conformité visuelle, une fabrique d'identités normées où la singularité se mesure à l'intérieur d'un cadre reproductible.

Comme la partition musicale, la photographie devient une interface entre une intention humaine et sa matérialisation technique. Elle révèle combien les arts, depuis longtemps, reposent sur des dispositifs codifiés, où l'artiste conçoit les conditions de l'apparition de l'œuvre plutôt que l'objet lui-même.

L'artiste comme auteur de règles : de la modernité au conceptuel



La photo de classe, un format éternellement reproduit.

Source: Wikimedia

En architecture, Le Corbusier, avec son *Modulor*, incarne une bascule. Le *Modulor* définit un code qui oriente l'action d'autrui à grande échelle. Il fonctionne déjà comme une sorte de code source : il décrit des proportions universelles qui peuvent être appliquées à une infinité de bâtiments, sans que l'architecte en personne ne pose la moindre brique. On retrouve ici la logique des modèles d'IA qui, une fois entraînés, génèrent d'innombrables résultats à partir d'une même base de règles.

Duchamp radicalise cette logique. Il affirme que l'art ne réside plus dans l'exécution matérielle. L'objet industriel devient œuvre dès que l'artiste le recontextualise. Le geste créatif ressemble à notre interaction avec une IA : nous ne dessinons pas chaque pixel, mais nous orientons et validons le résultat final. L'artiste agit comme un opérateur symbolique, un « éditeur responsable » d'objets ou d'images, qui transforme la matière existante en création artistique, orchestrant l'apparition du produit final plutôt que sa fabrication.

Sol LeWitt pousse la logique plus loin. Ses *Wall Drawings* sont des suites d'instructions textuelles : « tracez une ligne horizontale, ajoutez des cercles de telle dimension... ».



Avec son *Modulor*, l'architecte Le Corbusier a créé une échelle des proportions à appliquer au dessin de chaque pièce, porte, rambarde...

Une sorte de règle d'or pour que chaque espace soit adapté aux gestes humains.

Source: Fondation Le Corbusier

Par exemple, *Wall Drawing #122* (1972), décrit toutes les combinaisons de lignes droites, courbes et brisées, générant 150 variations possibles sur les murs. L'œuvre existe sous forme de protocole, un algorithme humain que d'autres exécutent. LeWitt anticipe ici le fonctionnement de l'intelligence artificielle générative : l'artiste fournit une séquence d'instructions, un *prompt* avant la lettre, et une autre entité, humaine ou machine, réalise l'exécution.

Un pas de plus est franchi avec l'art dit performatif, dont Tino Sehgal est une figure emblématique. Sehgal ne produit aucun objet : ses œuvres sont des « situations construites », confiées à des interprètes humain·es qui en assurent la mise en acte. Par exemple, le public entre dans un musée et se voit interpellé par un gardien qui entame un chant, une danse, ou une conversation déconcertante. Rien n'est improvisé, tout est régi par des consignes orales, transmises de bouche à oreille, car Sehgal interdit toute trace écrite, photo ou vidéo de son travail. Pour « *Cet objectif de cet objet* » (2004), le public est confronté à cinq personnes qui lui tournent le dos. Ces cinq personnes scandent : « L'objectif de cette œuvre est de faire l'objet d'une discussion », et si l'audience ne répond pas, elles s'effondrent lentement. L'œuvre n'existe qu'à travers son exécution par d'autres.



Ce dispositif pousse à son comble la logique prescriptive à l'œuvre chez LeWitt : l'artiste n'exécute pas, il programme une série de comportements. L'analogie avec l'intelligence artificielle saute aux yeux : dans un cas, un ensemble d'instructions active des corps humains, dans l'autre, des lignes de code mobilisent la machine. La valeur de l'œuvre réside moins dans sa matérialité que dans l'actualisation d'une règle initiale. L'art et la création se définissent de plus en plus par la prescription, le protocole et la règle, et de moins en moins par l'exécution matérielle explicite.

L'IA comme atelier contemporain

Le parallèle avec l'intelligence artificielle devient évident. L'IA est un atelier dématérialisé. Elle exécute des instructions, produit des variantes, accomplit des tâches à partir d'un prompt. Le rôle de l'utilisateur n'est plus d'exécuter mais de concevoir la consigne et d'affiner le prompt.

Qui signe ? L'élève, le maître, l'algorithme ? La question de l'auteur·e se pose, un débat ancien : déjà, au XVI^e siècle, on discutait pour savoir si un tableau peint par un élève était encore « un Raphaël ». L'IA ne fait que réactualiser ce vieux problème.

Autre enjeu : la sérialité. Les ateliers produisaient des Madones à la chaîne comme aujourd'hui on peut acheter en ligne des posters des œuvres de Magritte, Folon, Keith Harings ou Andy Warhol, produits de manière industrielle. Les IA génèrent des images en séries jumelles. Rien de neuf, sinon l'échelle et la vitesse.

De la Vénus à tête amovible aux multiples Brueghel, des ateliers médiévaux aux Wall Drawings, l'histoire de l'art se révèle comme une histoire de délégation. L'artiste n'a jamais été seulement une main inspirée : il ou elle est prescripteur-trice et organise les processus. L'intelligence artificielle s'inscrit dans cette continuité. Elle prolonge un geste ancien : focaliser l'attention et l'énergie sur l'idée créative, et en déléguer l'exécution. La machine est un prolongement du geste créateur, sans jamais prétendre à se substituer à l'imagination humaine.

Avec les IA génératives d'images, on peut en quelques clics doter une image d'un style artistique reconnaissable, à la manière des ateliers de la Renaissance.



Ici, le filtre «studio Ghibli».



Les IA rêvent-elles de patriarcat blanc ?

Par Daniel Bonvoisin, Chloé Tran Phu, Brieuc Guffens

Depuis 2022, les intelligences artificielles génératives s'imposent à la planète numérique et nous troublent : l'expression artistique que l'on pensait si humaine serait-elle réductible à une équation informatique ? Est-ce la fin de l'art ? Les machines vont-elles nous remplacer ? En animant les unes des journaux et les débats en ligne, ces fantasmes dystopiques masquent la leçon la plus spectaculaire que nous donne l'IA : notre culture est profondément inégalitaire et structurée autour de représentations biaisées, construites par l'histoire et les dominations. Si les IA génératives d'images traduisent et amplifient ces discriminations sociales, c'est principalement à cause des préjugés qui structurent les données d'entraînement des algorithmes. Mais l'inégalité se situe aussi dans l'usage de ces technologies en apparence élémentaire.

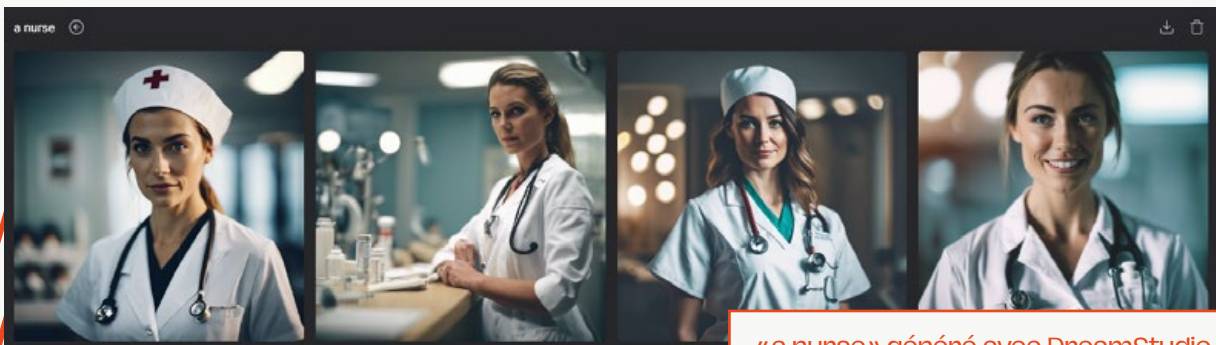


« Photo d'une femme devant un beau paysage ». Cette requête invite l'intelligence artificielle (IA) générative Stable Diffusion à puiser dans les profondeurs de ses algorithmes pour proposer en quelques secondes une image stupéfiante de réalisme. La même instruction confiée à une IA textuelle comme ChatGPT décrit de manière lyrique que sa « peau est douce et lumineuse, caressée par le soleil et imprégnée de la fraîcheur de l'air environnant. Une légère teinte rosée embellit ses joues, témoignant de l'émotion qui l'envahit devant tant de splendeur. » La femme est blanche, le paysage ressemble à un parc national américain. Quelle vision du monde ces créations automatisées traduisent-elles ? Contrairement à ce que dit l'adage, les goûts et les couleurs des IA méritent d'être discutés.

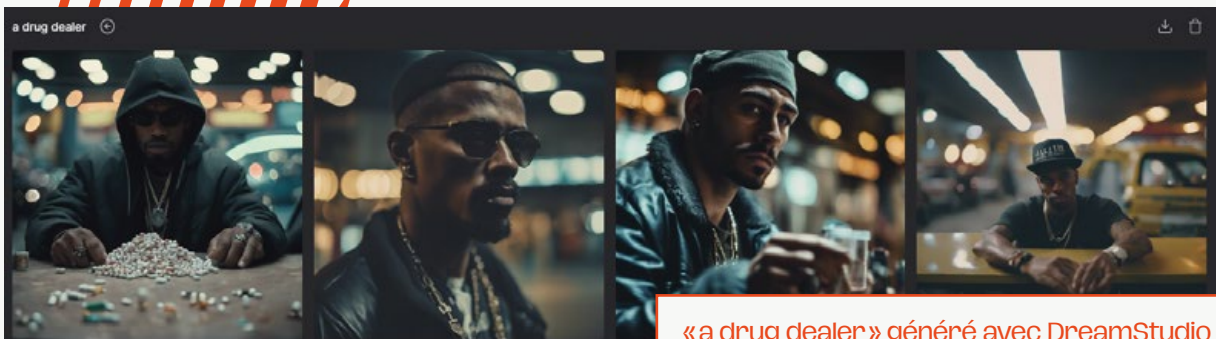
Des données d'entraînement, une classification du monde

Stable Diffusion, Dall-E ou Midjourney sont des logiciels qui permettent de générer une image à partir d'une description textuelle (un prompt). Si vous demandez à une de ces IA de créer un « personnage guerrier de jeu vidéo », elle vous proposera plusieurs résultats spectaculaires de réalisme ou de qualités esthétiques. Mais, malgré le genre neutre propre à l'anglais, notre « *video game warrior character wielding a sword* » est systématiquement un homme. Idem pour « *A lawyer* » (un ou une avocat-e en français). « *A nurse* », en revanche, est une femme infirmière. « *Drug dealer* » (« trafiquant de drogue ») est un homme à la peau noire ; « *a terrorist* » (« terroriste »), un homme basané portant une barbe noire et un turban... Pourquoi ces machines reproduisent-elles ces clichés de manière si grossière ?

Pour être capables de générer des images, ces IA sont entraînées sur des jeux de données appelés *data set* : à chaque image est attribuée une étiquette textuelle. Les algorithmes sont entraînés à répéter des classifications jusqu'à être capable d'effectuer l'opération en dehors du *data set* de référence, et créer ainsi une image inédite qui synthétise les caractéristiques « apprises ».



« a nurse » généré avec DreamStudio



« a drug dealer » généré avec DreamStudio

Quand il s'agit de classifier des images d'humain·es, l'étiquetage à grande échelle s'avère problématique¹. Le jeu de donnée *UTKFace* a tenté d'obtenir une certaine diversité en distinguant Blanc, Noir, Asiatique ou Indien sans parvenir à caractériser toutes les ethnies. Pour s'en approcher, les ingénieurs d'IBM ont, dans leur jeu de données *Diversity in Faces*, mis au point des calculs prenant en compte forme du crâne, symétrie faciale... permettant ainsi en quelques clics de générer des portraits s'appuyant sur les techniques de classification anthropométrique qui ont servi de base aux théories raciales nées au XIX^e siècle. Cette simplification s'empêtre aussi dans les ambiguïtés de la catégorisation « homme » et « femme » à l'heure où bon nombre de personnes ne se retrouvent pas dans cette binarité ou ont des caractéristiques physiques qui ne correspondent pas au genre qui leur a été assigné à la naissance.

Le mécanisme d'apprentissage de ces logiciels fonctionne sur la stéréotypie des idéaux-types culturels : l'IA générative la plus efficace va vouloir faire correspondre le résultat de notre recherche à l'image la plus communément admise d'un « personnage guerrier de jeu vidéo ». Elle fonctionne sur la prédiction : elle va représenter un homme et pas une femme car elle a été entraînée par les images masculines qui prédominent dans l'univers du jeu vidéo.

Les images pour grossir les discriminations

Les banques de données d'images regorgent de personnes blanches dans des situations très variées, ce qui facilite la reconnaissance automatisée des visages blancs. Par contre, d'après l'expérience de la chercheuse Joy Buolamwini, les IA sont moins entraînées sur des visages noirs. Les modèles algorithmiques les plus utilisés (ceux d'IBM, de Microsoft et de Face++) ont 34 % de risques de faire des erreurs lorsque l'individu est une femme noire. Cette sous-représentation numérique rend les produits commerciaux de reconnaissance faciale et les algorithmes de recherche moins adaptés à ces populations : déverrouillage de son smartphone moins efficace, erreur judiciaire due à une mauvaise reconnaissance sur vidéosurveillance...

Une enquête sur plus de 5 000 images créées par Stable Diffusion² montre que les images générées dépeignent plus volontiers des hommes blancs pour représenter des PDG, avocats, politiciens, ingénieurs, et que les femmes sont surreprésentées dans les professions mal rémunérées ou moins valorisées par la société (travailleuses sociales, domestiques, enseignantes) même quand cela ne correspond pas à la réalité. Les femmes ne représentent qu'une infime partie des images générées pour le mot-

1 Mathilde Saliou, *Technoféminisme, Comment le numérique aggrave les inégalités*, Grasset, 2023, p. 172.

2 Leonardo Nicoletti et Dina Bass, *Humans are biased. Generative AI is even worse*, 2023. www.bloomberg.com/graphics/2023-generative-ai-bias/

clé « juge » – environ 3 % – alors que 34 % des juges américains sont des femmes, selon l'Association nationale des femmes juges et le Centre judiciaire fédéral. Pour les mots-clés « détenu », « trafiquant de drogue » et « terroriste »³, le modèle a amplifié les stéréotypes en générant presque exclusivement des visages racisés.

Les IA génèrent des images qui reflètent des inégalités sociales et les accentuent en les reproduisant sans aucune nuance, au risque de contribuer à leur « naturalisation ». Très tôt identifié, ce risque est aujourd'hui largement dénoncé par de nombreuses associations⁴ et lors des débats qui ont animé l'adoption par le Parlement européen de l'IA Act, réglementation supposée encadrer le déploiement de ces technologies, sans pour autant aboutir à une interprétation juridique contraignante⁵.

Incorrigibles *datasets*

Selon Netcraft⁶, plus de la moitié des serveurs internet sécurisés du monde se trouvent aux États-Unis, qui comptent aussi le plus grand nombre de sites web enregistrés. Pour les entreprises américaines, locomotives du marché de l'IA, l'accès à des données et des images essentiellement nord-américaines est facilité, et l'anglais est la langue prédominante pour étiqueter les images. C'est dans LAION-5B, le plus grand ensemble de données image-texte librement accessible au monde (plus de 5 milliards d'images et de légendes venant d'Internet) que Stable Diffusion puise ses données brutes. Bien que Stability AI, la société qui développe ce générateur d'images, prétende avoir filtré le contenu avant d'utiliser les données de LAION, une part significative des images proviennent des bas-fonds du web et sont problématiques : images dégradantes, contenus violents, haineux, pornographiques... Si ces créations sont discriminantes et sont ensuite intégrées aux données d'entraînement, les modèles texte-image des prochaines générations pourraient devenir toujours plus biaisés.

L'histoire visuelle de l'humanité qui a été jugée digne de numérisation est presque strictement occidentale et marchande, issue d'une production à échelle industrielle. Quid des esthétiques non-occidentales ? Des traditions picturales séculaires ou contemporaines qui n'épousent pas les canons dominants ? Absentes ou invisibilisées dans les jeux de données, elles n'ont que peu de chance d'influencer la synthèse

3 Lorsqu'on lui a demandé de générer des images d'un « terroriste », le modèle a systématiquement représenté des hommes à la pilosité faciale foncée, portant souvent un couvre-chef – s'appuyant clairement sur les stéréotypes des hommes musulmans. Rappelons que selon un rapport de 2017 du Government Accountability Office, les extrémistes islamiques radicaux ont commis 23 attentats terroristes meurtriers sur le sol américain depuis le 11 septembre 2001, tandis que les extrémistes d'extrême droite, y compris les suprémacistes blancs, en ont commis près de trois fois plus au cours de la même période.

4 Comme s'en inquiète également Unia dans ses recommandations au regard des élections belges de 2024 : www.unia.be/fr/legislation-et-recommandations/recommandations-dunia/egalite-lutte-contre-les-discriminations-droits-humains-elections-2024

5 Voir le site de la loi européenne sur l'intelligence artificielle, <https://artificialintelligenceact.eu/fr/>

6 Netcraft est une entreprise spécialisée dans les technologies Internet, connue pour ses sondages automatisés d'Internet par nom de domaine à la recherche de serveurs HTTP, donc de sites web.

à l'œuvre dans les processeurs. La génération d'images autour de l'esclavage est particulièrement significative de cet effet de distorsion culturelle et historique. Le prompt « photo of a slave » (« esclave ») renverra systématiquement à une personne noire et l'esthétique de l'image rappelle celles du ^{xix}^e siècle. D'où vient cette inspiration ? Les chatbots Gemini de Google ou Copilot de Windows nous en donnent une idée assez claire. Sollicités sur le thème⁷, ces robots conversationnels inventent des histoires d'esclaves, nommés « Moussa » ou « Elena », forcé-es de travailler sous les ordres d'un maître brutal dans des champs de coton ou de cannes à sucre, et qui réussissent à s'échapper. Si on sait la reconnaître, l'iconographie et la littérature nord-américaine servent de sources premières.



« slave » (esclave) généré avec le modèle sd-xl-turbo de stability-ai

Face aux protestations d'associations et d'utilisateur-rices face à ces clichés, les entreprises cherchent à corriger le tir. À défaut de pouvoir renouveler les sources visuelles et les biais qu'elles contiennent, les logiciels sont modifiés pour produire des résultats qui ne font pas polémique. Au sujet de l'esclavage, thème hautement sensible aux États-Unis, plusieurs générateurs en ligne, dont celui de StabilityAI, bannissent le mot « slave ». Du côté de Google, depuis 2019, l'entreprise proclame dans ses « AI principes » vouloir « éviter de créer ou de renforcer des biais injustes⁸ », notamment sur caractéristiques relatives à l'ethnicité, le genre, l'orientation sexuelle, la religion, etc.

7 Avec le prompt « décris moi une personne mise en esclavage ».

8 <https://ai.google/responsibility/principles>

Les instructions d'inclusivité données à Gemini, un service ouvert au public début 2024, ont produit un moteur d'images qui décline dans toute la gamme ethnique (forcément limitée) les prompts injectés. Au point de susciter la fureur des conservateurs contre cette IA qui crée sans vergogne des images de vikings ou des Pères fondateurs américains à la peau... noire⁹, obligeant Google à suspendre le service et à s'excuser platement.

Cet épisode illustre les tensions autour d'outils soumis à des exigences contradictoires. Pour les uns, ils doivent produire des contenus conformes à la « vérité », et donc au récit (notamment visuel) historique produit par une société inégalitaire ; pour d'autres, elles doivent éviter de reproduire les inégalités et les dominations. L'IA devient un nouvel objectif tactique pour une bataille culturelle largement à l'œuvre dans le champ des représentations culturelles des industries médiatiques.

Vers une fracture numérique culturelle ?

Sans spéculer sur les révolutions technologiques ou sociétales promises par les industries de la Silicon Valley, en à peine une année les services d'IA génératives se sont démultipliés dans les interfaces numériques. Moteurs de recherche, logiciels de création, chats en ligne : partout elles s'offrent à un usage quotidien, professionnel, communicationnel ou récréatif. Si beaucoup craignent que l'inondation d'Internet par des objets graphiques, audiovisuels ou textuels artificiels trouble un peu plus l'accès à un savoir ou une information objectifs, la simplicité apparente de l'utilisation des IA masque l'inégalité des utilisateur·rices.

D'une part, comme on l'a vu, le « spectateur idéal » de ces contenus synthétiques est celui-là même qui profite des inégalités historiques : l'Occidental en général et l'homme cisgenre blanc hétérosexuel en particulier. Les utilisateurs de cette catégorie obtiendront facilement des résultats conformes à leur vision du monde et à leurs attentes. Pour tous·tes les autres en revanche, les productions des IA constituent un espace de domination culturelle où il faudra, là aussi, batailler avec les entreprises pour obtenir des résultats ajustés à son identité, sa culture et ses aspirations.

D'autre part, les IA génératives n'expriment leur potentiel de mixage/collage culturel qu'à la mesure des indications qu'on leur donne. Exprimer son désir, parfois vague, passe donc par une compétence numérique nouvelle : savoir parler à la machine, via un prompt. Or, le prompt renvoie à la complexe architecture lexicale qui est le produit

⁹ Google suspend la possibilité de générer des images d'humains par son Intelligence artificielle Gemini, 22 février 2024, RTBF, www.rtb.be/article/google-suspend-la-possibilite-de-generer-des-images-dhumains-par-son-intelligence-artificielle-gemini-11333708

de l'entraînement des logiciels. Il faut présumer les termes qui fonctionnent et les effets qu'ils génèrent. Autrement dit, il faut maîtriser un degré d'abstraction et des terminologies complexes pour se faire obéir. Si l'on cherche à influencer le logiciel par des ambiances, des styles, des références, pour obtenir le « à la manière de » pour lequel il est entraîné, il faut disposer d'un bagage culturel étendu, et pouvoir le traduire dans son prompt. Les IA génératives ne classent donc pas que les sources qui les ont entraînées, elles produisent aussi un effet discriminant sur leurs utilisateur·rices. Elles valorisent ceux et celles qui maîtrisent le plus finement l'immense capital culturel avalé par la numérisation et restitué par les IA au détriment de la population qui n'envisage pas la culture comme une encyclopédie de références.

Ce capital culturel étant lui-même porteur de discriminations et reproducteur des dominations qui ont structuré son édification historique, l'usage de l'IA impose une prise de conscience critique de ces profonds déséquilibres dont la société prend doucement conscience. Au-delà du spectacle de singe savant que nous offrent ces outils, peut-être que leur plus grand intérêt réside dans l'opportunité qu'ils offrent de regarder en face les angles morts de la culture médiatique dominante... pour mieux les éclairer ?



Généalogie de l'IA visuelle : entretien avec Gil Bartholeyns

Propos recueillis par
Daniel Bonvoisin

Gil Bartholeyns est historien, enseignant-chercheur au département des humanités de l'Université de Lille. Ses travaux portent sur les images, les objets et les techniques dans une perspective d'anthropologie du quotidien. Ancien rédacteur en chef de Techniques & Culture, il est également chroniqueur sur La Première RTBF et écrivain.



Quelle place occupe l'image dans nos sociétés ?

La réponse est je crois sans mystère. Nous regardons et nous produisons sans cesse plus d'images que nous partageons sans cesse avec plus de monde jusqu'à un sentiment d'excès et de perte de contrôle relayé par les spécialistes eux-mêmes. Face au « déluge » d'images submergeant la réalité, certains sont inquiets et pragmatiques : il faut donner aux jeunes générations les moyens de comprendre, car notre monde n'est pas seulement saturé d'images, il est désormais produit par elles. C'est la critique de la « société du spectacle », du redoublement de la réalité par les images. Dès les années 1950, avec le développement du cinéma et de la télévision surtout, acquérir une « culture visuelle » à côté d'une « culture littéraire », devient une préoccupation éducative. Dans les années 1990, un « tournant visuel » (*visual turn*) est diagnostiqué simultanément par des universitaires nord-américains et européens. Les uns s'attachent alors à décrire le développement historique d'une culture visuelle mondialisée à travers les industries culturelles et les technologies numériques alors en plein essor. Les autres analysent les nouveautés de la vieille alliance du voir et du pouvoir, de la vidéosurveillance au drone. Ils décryptent les dispositifs et représentations autoritaires contre lesquels s'arme un activisme citoyen producteur de contre-visualités. L'image et le visible sont alors considérés comme fondamentalement politiques dans la mesure où ils déterminent les façons de vivre ensemble et de se rapporter à la chose publique. Les IA visuelles arrivent donc dans un champ critique bien aiguisé, mais cela ne veut pas dire qu'elles ne posent pas des nouveaux défis, pour les inscrire dans une histoire longue et surtout pour en comprendre les effets.

Les images elles-mêmes ont-elles évolué dans un sens ou l'autre ?

Nous vivons aujourd'hui dans un monde d'images plates, principalement rectangulaires et du genre « photographique ». Or ce paradigme est une singularité historique et culturelle. Dans la majorité des sociétés du passé ou traditionnelles, les images sont en relief. Ce sont des corps que l'on touche, que l'on manipule. Elles sont aussi souvent attachées à un lieu, comme la sculpture ou la fresque. Nos images sont au contraire essentiellement des surfaces, voire des dispositifs, comme la projection.

À ce titre on parle souvent de dématérialisation des images. Les images n'ont pourtant jamais été aussi dépendantes de techniques souvent « lourdes ». Une image correspond généralement à un objet (une carte postale, une statue), mais depuis le ^{xviii}^e siècle, avec la lanterne magique ou le stéréoscope au ^{xix}^e siècle, le lieu de l'image « flotte » entre l'instrument, le support et l'observateur, elle est difficile à situer. Je veux dire que la nouveauté radicale est moins à chercher du côté d'une soi-disant immatérialité ou instantanéité que dans le lien entre image et objet. Désormais certains objets ont la capacité de supporter une infinité d'images. On peut les appeler « polyvisuels », pour caractériser cette rupture fondamentale, à savoir être passé de l'image-objet à l'objet-écran : l'écran devient le lieu transitoire de toutes les images. C'est un des nœuds à retenir si on veut faire une archéologie de l'image contemporaine jusqu'aux images générées par intelligence artificielle.

Si on s'attache à faire cette généalogie de la visualité générative, il faut aussi souligner deux moments – sans devoir nécessairement remonter à l'« icône » grecque qui est évidemment le prototype de l'image telle qu'on la connaît, au point que son hégémonie nous engage à sa critique historique et même postcoloniale. Premier moment : en Europe, entre le ^{xiii}^e et le ^{xvii}^e siècles, on assiste à l'avènement de la forme tableau, contre la peinture murale et le retable d'autel. Le panneau mobile devient, si je puis dire, le cadre de l'œuvre d'art et du marché des images. Le manuscrit médiéval enluminé (« illustré ») joue son rôle dans la « réduction » de l'image à ses formes et supports actuels. Et cette normativité s'accroît avec le passage au livre imprimé et aux techniques de gravure. Second moment : l'invention des procédés photographiques au ^{xix}^e siècle. Le développement de la photographie, puis sa pleine démocratisation à parti des années 1870-1880, viennent en quelque sorte sceller le destin de l'image...

Quelle place occupent les IA génératives visuelles dans cette histoire ?

Eh bien, on peut poursuivre avec la photographie. Ce que la photographie apporte à cette histoire, et que l'on doit absolument souligner si l'on veut penser l'image synthétique, c'est l'idée que l'image nourrit un lien presque physique avec la réalité. L'image photographique (écriture par la lumière) duplique presque magiquement le réel. Ce paradigme visuel de la modernité est appelé « indiciaire » : l'indice est un signe qui présente un lien matériel avec sa source, telle une empreinte de pied. On retrouve cette idée jusque dans la théorie des images : c'est « le photographique » de Rosalind Krauss, c'est le « ça a été » de Roland Barthes des années 1970-1980, c'est la peur, chez les meilleurs spécialistes des années 1990, que la pixellisation des images va

marquer la fin de la croyance en la réalité de ce qu'elles montrent. Si Platon faisait de l'image une fausseté redoutable et Aristote un moyen de connaissance ou d'édification pour Grégoire le Grand au Moyen Âge, la modernité visuelle a fait de l'image le lieu de la vérité, en tout cas une preuve ou un témoignage, et au minimum un gage d'existence, à un moment donné, des choses, des personnes, des situations qu'elle donne à voir.

Or depuis les années 80 avec l'image de synthèse d'abord et aujourd'hui avec les images par IA générative, il y a un décrochage, mais un décrochage qui est singulier car, au fond, la peinture menait déjà ce décrochage par rapport à la réalité – et l'on peut même dire que c'est dans le travail de l'écart que se situe l'art, la manière de faire œuvre. Seulement, la peinture reprend des éléments du réel pour composer une image mais en s'écartant de la ressemblance immédiate et directe, alors que l'intelligence artificielle reprend des éléments du réel pour composer une image naturelle. Si bien qu'on rejoue le vieux débat de la vérité de l'image mais d'une manière particulière. Cette fois-ci la défiance envers les images est différente en ceci que le trouble traditionnel envers le signe en général, c'est-à-dire de rendre présent ce qui est absent, devient celui de rendre présent (réel) ce qui est inexistant.



À gauche, une image générée avec Midjourney par le photographe Boris Eldagsen, qui avait trompé le jury du Sony World Photography Awards 2023.

À droite, *Flamingone* de Miles Astray, une photographie conventionnelle prise pour une création d'intelligence artificielle — et primée comme telle lors d'un concours dédié aux images générées par IA.

Par des démarches opposées, ces deux artistes ont réactualisé les débats éternels sur le statut de l'image photographique : son rapport au réel, le rôle des outils dans la définition de l'acte créateur, et la question de la légitimité artistique.

Crédits :

Boris Eldagsen - *Pseudomnesia : The Electrician* - 2023

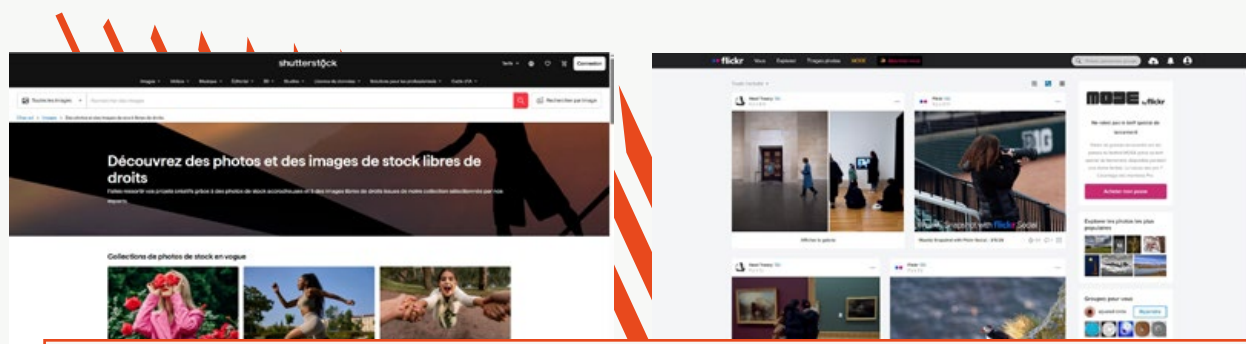
Miles Astray - *Flamingone* - 2023



Si on cherche les points de singularité, on remarquera que l'image de synthèse faisait déjà la même chose. Sauf qu'elle était dévolue aux effets spéciaux et qu'elle a révolutionné le dessin d'animation : ce sont des formes visuelles qui s'écartent de la réalité du sens commun. Ce n'est pas le cas des usages « populaires » des IA visuelles. Pour toutes ces raisons, on a tendance à dire qu'elles marquent une rupture historique, un point de non-retour... Je vois surtout combien elles ont été rendues possibles par des bouleversements techniques et sociaux antérieurs.

La somme des images numérisées, et organisées en base de données, est-elle le résultat d'une certaine histoire des images ?

Oui, on peut même dire que non seulement des changements techniques et sociaux en marche à partir des années 1960 ont rendu possible l'IA générative, mais que ces changements ont aussi conditionné l'état du stock d'images et par conséquent les résultats potentiels de tout prompt visuel. Il y a d'abord la « numérisation » des images, qui est un double processus. D'une part, les appareils de production d'images deviennent numériques. Qu'on pense à la toute première image numérique qui est celle de Mars survolée par Mariner en juillet 1965. Que l'on pense aux premiers appareils photo grand public dotés de capteurs CCD au début des années 1990. C'est une histoire longue mais récente et décisive. D'autre part, les images analogiques existantes sont progressivement numérisées, en masse, par des institutions et d'abord par des initiatives privées. On se souviendra qu'en 1989 Bill Gates crée une société, Interactive Home Systems, rebaptisée Corbis, avec l'idée de numériser des fonds d'images pour ensuite les commercialiser. Les entrepreneurs rêvent alors d'un marché mondial des images entre leurs mains. Mais ni Gates ni personne n'anticipe l'arrivée – quelques années seulement après les débuts d'Internet – de plateformes « libres » dédiées aux images fixes et animées, telles que Flickr, créée en 1997, Photobucket en 2003 ou YouTube en 2005. Le bouleversement touche les modes de production, de diffusion et de consultation des images. De fait, les images ne sont plus uniquement fabriquées et diffusées par des professionnels et des artistes, photographes, journalistes, cinéastes, scientifiques... C'est la « révolution des amateurs », concomitante au modèle dynamique de la technologie Internet.



Les services numériques ont transformé notre rapport aux images. D'un côté, des sites de médias de stock — comme Getty Images ou Shutterstock — ont contribué à homogénéiser la communication visuelle et à figer l'association entre un concept et sa représentation. De l'autre, des plateformes comme Flickr ont amorcé une « révolution des amateurs », amplifiée par des réseaux sociaux tels qu'Instagram ou BeReal.

Les IA réduisent-elles alors les possibilités des manières de représenter ?

Justement, si on s'attarde un instant sur les grandes bases de données comme Shutterstock ou Getty Images, qui sont souvent populaires et en même temps documentaires puisque certaines, comme Flickr, hébergent les fonds de plusieurs dizaines d'institutions publiques (de la bibliothèque du Congrès aux musées de province), elles contiennent un biais initial : celui de la surreprésentation de la visualité anglo-américaines, et plus généralement occidentale, et quand ce n'est pas le cas, les fonds portent la marque de l'impérialisme culturel, celui des nombreuses collections « exotiques ».

Les biais imputés aux IA génératives sont en réalité des biais structurels, systémiques. Et ces biais sont bien connus ! Ce qu'il faut alors noter, ce sont leurs effets : le renforcement des standards visuels et l'invisibilisation, notamment des cultures visuelles non européennes et, au sein de la culture européenne, des minorités. Cela n'est pas propre aux IA génératives. La télévision et le cinéma sont emblématiques de ces phénomènes. Les « limites » des IA génératives sont en fait les limites de la culture dominante, qu'il s'agisse des corpus sur lesquels on les entraîne et de la façon dont leurs items sont étiquetés, ou de nos façons de prompter et de notre capacité à « percevoir » les biais et à se corriger. En attendant, il y a entre l'IA et l'utilisateur une correspondance culturelle problématique, voire une équivalence des ignorances.



Les IA génératives, si elles ne sont pas « challengées » par l'internaute, reproduisent les clichés visuels présents dans les données dont elles sont nourries. Ici, en réponse à la requête « une image représentant l'amour », ChatGPT produit une romance convenue, digne d'un roman à l'eau de rose ou d'une pub pour du parfum, mettant en scène un couple conforme aux normes de représentation occidentales dominantes.

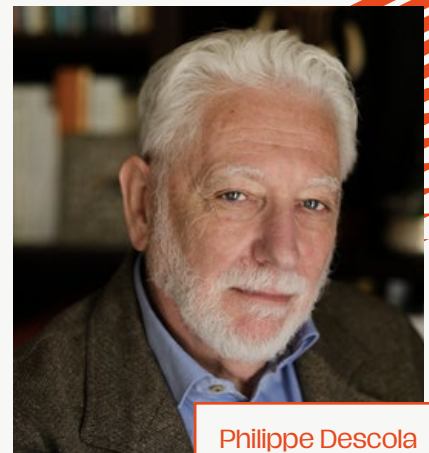
Si vous demandez à une IAG de vous dire si ses résultats sont objectifs, neutres, etc., elle vous répondra que non, dans la mesure où elle tire son information d'une certaine culture et que cette culture – si on lui demande de préciser – comporte toute une série de biais : des biais raciaux, du moins « ethniques », des biais de genre (le fameux male gaze), des biais relatifs aux valeurs, comprenant des biais « d'évitement éthique » sur la violence et la nudité, des biais de classe et même une normativité bourgeoise. L'IA vous dira que par conséquent les images qu'elle produit sont porteuses, incidemment, de tous ces biais, auxquels il faut ajouter des biais esthétiques (elle s'inspire surtout de la photographie commerciale, de la mode et du cinéma) et des biais d'environnement (moderne et propre, urbain plus que rural, de « style Pinterest/Instagram »), et qu'en somme il revient à l'utilisateur de s'interroger d'abord sur le monde dans lequel il vit s'il souhaite l'utiliser, en conscience, comme outil créatif ou comme instrument de connaissance. Retour à la case départ de l'éducation, de l'enseignement et des sciences sociales.

Manières de voir les mondes

Entretien avec Philippe Descola

Propos recueillis par Daniel Bonvoisin

Pour l'anthropologue Philippe Descola, la perspective occidentale, héritée des Lumières, qui conçoit la nature comme une entité spécifique et la culture comme le propre de l'humain, n'a rien d'universel. D'autres sociétés entretiennent un tout autre rapport au monde. Dans *Par-delà Nature et Culture*, à partir de sa propre expérience auprès des Indiens Achuar d'Amazonie et d'une abondante documentation ethnographique, Descola identifie quatre grandes perspectives, qu'il appelle des formes de mondiation.



Philippe Descola

Dans **les mondes animistes**, arbre, animaux, humain-es ou objets se distinguent les uns des autres par des formes différentes mais sont animés d'un esprit qui peut voyager, se transformer et agir sur le monde. Pour **les sociétés totémiques**, les êtres de toute sorte ont un lien avec un prototype (le totem) dont ils sont issus et se répartissent en différents groupes qui cohabitent mais se distinguent par cette appartenance. Dans **les sociétés analogiques**, chaque être se différencie des autres non par des caractéristiques propres mais par la position qu'il occupe dans le système global dont établir la cohérence et le fonctionnement est une difficulté en soi. Enfin, **les sociétés naturalistes**, qu'incarne la culture occidentale depuis sa prétendue modernité, identifient des règles biologiques et physiques, établies par la science, qui s'appliquent pareillement à tous les existants mais où l'humain-e se distingue par la spécificité de son intériorité. Dans ces différentes sociétés, les questions existentielles s'articulent autour de préoccupations tout à fait spécifiques : quel est l'esprit qui habite cet oiseau qui me regarde ? Ce kangourou est-il de ma famille ? Quelle est la place que j'occupe dans le cosmos ? Ne suis-je qu'un amas fugace de particules ?

Cette peinture aborigène d'un kangourou nous montre un univers ordonné selon le modèle du prototype **totémique**: cet animal est le monde pour ceux et celles dont il est le totem.



Peinture d'un kangourou, Kunwinjku, Alligator River, terre d'Arnhem, Australie, vers 1915

Ce masque Yup'ik, peuple amérindien, illustre d'une part que les transformations typiques des sociétés **animistes** trouvent dans les masques une manière d'exprimer des images. Celui-ci montre les différentes formes que l'esprit peut prendre et suggère les perspectives que chacune de ces formes offrent sur le monde.



En outre, ce masque illustre des manières d'agir des images. Destiné à être porté, un masque existe différemment dans le rituel où il est utilisé que sous forme d'une photo dans un catalogue de musée ou vaguement reproduit par une IA générative, en supposant qu'elle serait capable de générer quelque chose qui ressemble à la tradition Yup'ik.

Source: rawpixel.com



Le célèbre Agneau Mystique des Frères Van Eyck traduit une approche visuelle caractéristique d'un monde **analogiste**: celui où tous les êtres sont en relation dans un grand ensemble (à gauche) où ils occupent une place précise. Les mondes aztèques, moghols et européens médiévaux partagent des manières de voir similaires d'un grand ordre cosmique. Mais cette peinture flamande montre également une transition vers le **naturalisme** occidental (à droite). Des éléments paysagers y sont ordonnés selon des perspectives «réalistes» qui annoncent le développement de l'art «objectif» caractéristique de la «modernité».

Source: Wikimedia

La synthèse ambitieuse que propose Descola amène deux conséquences cruciales. Elles expliquent l'influence de son œuvre dans les domaines de l'anthropologie et de l'ethnographie. Sa réflexion irrigue également des milieux plus militants, notamment des luttes environnementales. D'une part, il faut acter que les représentations naturalistes que notre culture « moderne » nous a inculquées ne sont pas les seules explications qui donnent du sens au monde et à l'existence, même si elles ont pris le pouvoir dans les manières de gouverner. Et d'autre part, il faut prendre conscience que cette relation culturelle à l'environnement facilite l'exploitation des ressources dites « naturelles » — séparées de l'humanité car réduites à des matières dépourvues d'intériorité — que l'industrialisation a transformé en catastrophe aux conséquences planétaires.

Quinze ans après *Par-delà Nature et Culture*¹, Philippe Descola prolonge sa réflexion en s'intéressant aux images dans *Les Formes du visible*². C'est à la suite de ces travaux de réflexion sur les cultures visuelles qu'il nous est apparu pertinent d'entendre l'anthropologue sur les enjeux contemporains des IA visuelles. En filigrane de cet échange plane un enjeu culturel spécifique à un monde globalisé, indifférent aux survivances des patrimoines culturels, des modes de vie alternatifs et de visions du monde potentiellement en voie d'extinction. Si l'IA générative s'impose comme l'outil de référence des représentations visuelles, sa nature d'outil statistique participerait à l'effacement des traditions et des imaginaires non compatibles avec le format numérique et non conformes à la manière de voir le monde qui s'est imposée à la faveur des colonisations et de la globalisation. Si la vocation des images est de donner à voir ce qui ne se voit pas, l'IA contribuerait à appauvrir les mondes invisibles et à faire disparaître des imaginaires qui appartiennent au patrimoine humain.

Après votre travail sur les ontologies et la notion de nature, pourquoi vous êtes-vous intéressé aux images ?

C'est une forme de vérification de la pertinence des propositions que j'avais pu faire concernant les formes de mondiation. J'avais travaillé essentiellement sur la base de documents ethnographiques, à partir de textes philosophiques, de traités médicaux, de livres de toutes sortes qui me donnaient des informations sur d'autres façons de faire monde que les nôtres mais également sur celles qui nous sont familières. Et je me suis dit que s'il y a une pertinence à ces propositions, on devrait pouvoir en trouver la trace dans les images et dans l'organisation d'espaces visuels, c'est à dire dans ce qui était montré ou n'était pas montré dans l'image. J'ai donc commencé à travailler sur cette question par curiosité intellectuelle et parce que je m'intéresse aux images depuis longtemps, en amateur d'histoire de l'art, disons. J'ai commencé par monter une exposition au musée du Quai Branly il y a une douzaine d'années et qui m'a permis, pour la construire, de sélectionner un certain nombre d'images que je souhaitais mettre en regard les unes des autres. Et pour cela, il fallait aussi que ces images soient parlantes.

1 Philippe Descola, *Par-delà Nature et Culture*, Gallimard, 2005. Voir également Philippe Descola, *Diversité des natures, diversité des cultures*, Les petites conférences, Bayard, 2010, qui vulgarise les concepts abordés.

2 Philippe Descola, *Les Formes du visible*, Seuil, 2021

Je souhaitais qu'un public qui n'était pas familier de ces mondes et des civilisations extra-modernes d'où ces images provenaient, et même pas nécessairement familier d'une partie de la peinture occidentale par exemple, puisse repérer, identifier des contrastes entre ces images de façon à ce que le message puisse passer. Je me suis beaucoup amusé.

J'avais sélectionné ces images parmi certaines de telle ou telle civilisation, en prenant celles qui correspondaient à ce que j'avais déjà identifié comme étant d'un régime plutôt animiste, analogiste ou totémique, etc. Dans un second temps, l'expérimentation a consisté à détecter dans ces ensembles d'images des dispositifs qui correspondaient à certains de ces régimes, et je les ai suivis à la trace avec l'ambition de mettre en évidence trois choses.

La première, c'est en quoi elles révèlent certains aspects du monde qui ne sont pas nécessairement visibles ou exprimées par les discours. Par exemple pour l'animisme, ce qui a retenu mon attention, c'est l'effet de commutation, ce que dans les analyses ethnologiques de l'animisme, on appelle la métamorphose. Il s'agit du basculement de point de vue entre la perception de l'intériorité d'un être et la perception de sa physicalité. Ou bien, pour l'analogisme, la mise en évidence de formes de méta-relations, de toutes sortes d'englobements, de correspondances, de hiérarchies entre les êtres.

Le deuxième aspect, c'était l'aspect formel : comment se construit l'espace visuel ? Avec notamment, et on aura l'occasion d'y revenir quand on parlera des images produites par l'intelligence artificielle, le fait qu'une des caractéristiques de la figuration naturaliste, c'est le mono-perspectivisme, c'est à dire l'émulation de la vision humaine : la tentative de reproduire le plus fidèlement possible ce que perçoit un humain lorsqu'il est face à une scène, un personnage. Cette caractéristique tranche de façon considérable sur d'autres façons de figurer et de construire l'espace visuel. Dans d'autres régimes de figuration, c'est le polyperspectivisme qui domine : on va donner d'un être ou d'une scène des points de vue différents, de façon à multiplier les informations sur cet être ou sur cette scène, y compris de façon contre-intuitive par rapport à la vision humaine. Cela a beaucoup troublé les Européens lorsqu'ils ont été confrontés à des images de civilisations extra-modernes vues comme étant malhabiles, contre-intuitives, magiques, etc. alors qu'elles adoptaient une perspective qui consistait à déplier les points de vue et les différentes faces d'un être de façon à ordonner sur cette image le plus grand nombre d'informations possibles sur cet être.

Cet autre masque Yup'ik datant du début du xx^e siècle est exemplatif du polyperspectivisme : selon l'angle par lequel on l'approche, il offre une interprétation différente : c'est une facette spécifique de cette entité qui se révèle.



Source: Wikimedia

Le troisième aspect, c'est le fait que les images ont une puissance d'agir qui leur est propre. Cet aspect a retenu l'attention des historiens de l'art depuis plusieurs décennies mais aussi de l'anthropologie. Et l'hypothèse que j'ai souhaité défendre, c'est que les formes d'expression de la puissance d'agir des images varient en fonction des régimes ontologiques propres à chacun de ces modes de figuration. Il n'y a pas un mécanisme unique qui donne une puissance d'agir aux images, mais différents types de mécanismes d'animation de ces images qui sont en relation avec les circuits d'attente, avec la manière dont ces images sont accueillies.

Entraînées sur des corpus massifs de données occidentales, réalisées par des outils occidentaux, les IA génératives incarneraient l'omniprésence de l'image sur une surface plane (l'écran). Qu'est-ce que cette caractéristique technique vous inspire au regard de la diversité des images et de ce qu'elles donnent à voir ?

Même quand c'est un carré plat, le fait que ces images soient numérisées la plupart du temps, soit parce que ce sont des photographies numériques, soit parce que ce sont des peintures numérisées, aboutit à gommer le support, essentiel dans la puissance d'agir des images. Lorsqu'on regarde un retable dans une chapelle latérale d'une église mal éclairée avec des bougies, le lieu pour lequel il a été fait était déjà pris en compte par son concepteur. L'effet qu'il va exercer sur le spectateur est en partie défini par ces circonstances, qui sont toutes invisibles sur une surface en deux dimensions numérisées, même si l'image est déjà une surface en deux dimensions. Donc le mode d'exposition général, c'est-à-dire ce que j'évoquais en parlant des circuits d'attente ou des circuits d'accueil des images, disparaît complètement et uniformise des images qui sont en deux dimensions mais dont les formes d'exposition ne sont pas les mêmes selon les contextes culturels.

Le second aspect, c'est que beaucoup d'images dans le monde ne sont pas en deux dimensions. Soit parce que ce sont des images en trois dimensions, c'est à dire des sculptures, ou parce que ce sont des corps humains qui sont transformés en images [par des tatouages, des costumes, des masques...]. Et c'est cette diversité que l'IA et la généralisation des images numériques nous rend de plus en plus difficile d'imaginer. Ces différentes formes de figuration qui semblent extrêmement diverses ont pourtant le même sens : rendre présentes des choses invisibles ou absentes. Il faut en outre mettre l'accent sur le fait que ce qu'on appelle maintenant les arts premiers ont été pour l'essentiel des figurations en trois dimensions. Lorsqu'on regarde les revues spécialisées dans ces arts, elles sont pour l'essentiel consacrées à des sculptures, des effigies, des masques, etc. Il y a donc une sorte de tradition du fétiche qui subsiste et qui tend à garder à distance ces formes de figuration des civilisations extra-modernes. Car nous avons pris l'habitude de regarder les images comme étant des dispositifs en deux dimensions.

Donc ça veut dire qu'on aurait beau pousser la capacité à reproduire des images issues de toutes ces cultures, à partir du moment où on les désincarne de l'endroit où elles sont supposées être exposées, on a déjà perdu une partie de leur rôle et de leur valeur.

C'est le problème d'ailleurs des musées d'ethnographie en général. Quels que soient les efforts qu'on peut faire pour mettre en contexte, des images, on ne pourra jamais y parvenir complètement. Ça ne veut pas dire que ce soit une tentative invalide ou désespérée. Mais j'ai toujours pensé que les images avaient plusieurs vies en fonction des milieux et des médias au moyen desquels elles étaient rendues visibles. Leur apparition initiale pour une fonction précise dans un contexte précis est la plus importante. Mais je n'entends pas pour autant considérer que les autres sont complètement négligeables. Et la vie d'une image dans un musée, c'est une vie tout à fait singulière comme celle d'un masque rituel qu'on va trouver dans une vitrine, mais qui sera complètement dissociée à la fois du rite et de l'incarnation pour lesquels il a été créé, avec tous les effets de présence, de contexte, la musique, l'atmosphère générale, la chaleur, la poussière, le froid... tous ces éléments qui jouent un rôle très important dans les effets que les images produisent sur nous.

Certaines images qu'on a l'habitude de voir dans des églises sont pour moi indissociables de l'odeur de l'encens par exemple. Leur efficacité est liée à cette odeur. Ce sont des choses dont il est très difficile de rendre compte dans un musée. On peut projeter des petits films où on voit par exemple un masque rituel utilisé dans le contexte pour lequel il a été conçu. Mais, on a beau avoir une petite explication, c'est quand même encore très loin de permettre de comprendre quels sont les contextes dans lesquels ces images étaient exposées et étaient actives.

Au fond, je dirais que la façon dont les images sont montrées en deux dimensions par l'intelligence artificielle n'est qu'un prolongement d'un statut des images dans les musées, toutes présentées à l'identique qu'elles soient en deux ou trois dimensions, comme des témoignages d'un mode de figuration artistique d'un certain type que l'on a appris à considérer comme étant digne d'intérêt. Les images produites par l'intelligence artificielle occupent une certaine place dans une succession de monstrations des images qui a déjà commencé avec les musées.

La numérisation des images impose à beaucoup de représentations d'être aplaties, réduites à deux dimensions. L'expérience de lecture d'une image est pourtant, dans beaucoup de cas, ancrée dans son contexte d'exposition.

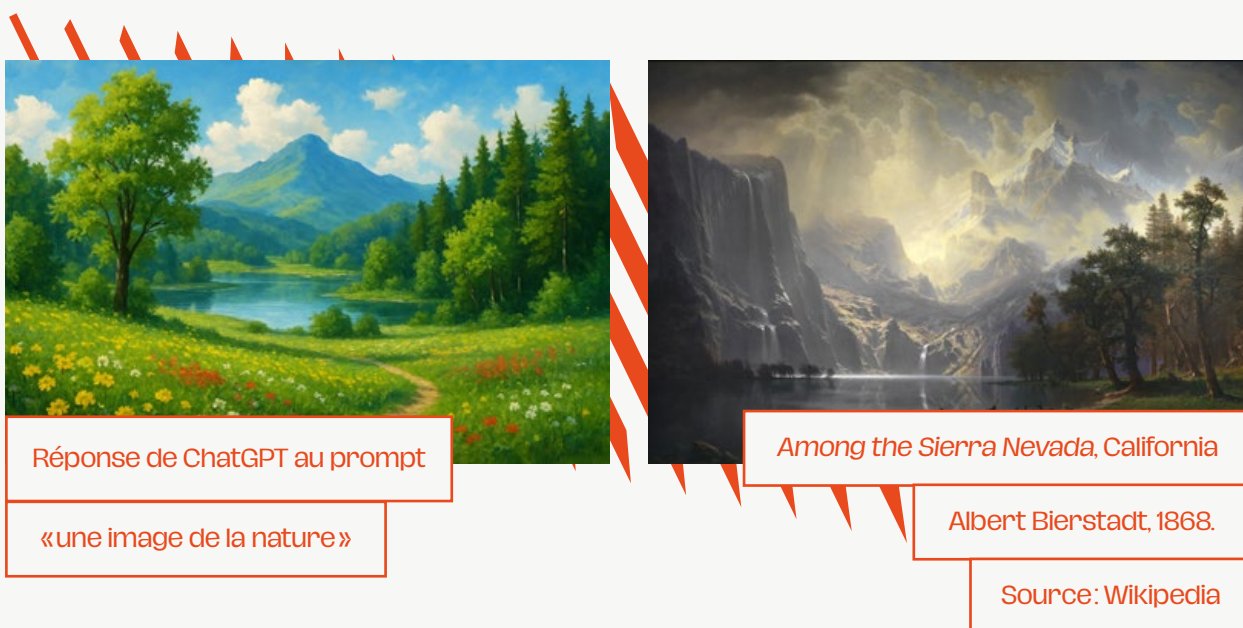


Le retable des Frères Van Eyck, évoqué plus haut, est encore aujourd'hui exposé à Gand dans la chapelle pour laquelle il a été conçu: sa présence physique dans ce lieu permet d'en saisir des effets qui seraient absents d'une reproduction ou d'une consultation numérique.

Source: FlickrR – John Treacy – CC BY-NC-SA 2.0

Ce qui est frappant avec les IA génératives, c'est qu'elles sont conçues comme un outil qui promet de visualiser toutes les représentations possibles. Sur base d'une instruction, vous pourriez générer n'importe quelle image. Et elles sont entraînées à partir de masses d'images énormes qui sont étiquetées en fonction de ce qu'elles montrent. Selon vous, est-ce qu'on pourrait le voir comme une sorte de summum de cette espèce de domestication par le regard de tout, de tout ce qui peut se traduire en images du moins ?

Les premières figurations scientifiques au ^{xvi}^e siècle, jusqu'aux rayons X et à toutes ces techniques de dévoilement, montrent bien les différentes étapes de la construction de l'objectivité naturaliste. Celle-ci a été constituée par les images avec les différents choix qui ont été faits à différentes époques pour révéler telle ou telle chose, c'est à dire soit des types soit au contraire les particularismes du réel. Les cristaux de neige par exemple ont d'abord été figurés de façon prototypique, en fonction des quelques modèles qui étaient identifiés, notamment par le microscope. Et puis peu à peu, on s'est aperçu qu'il y avait une plus grande diversité et on s'est attaché précisément à montrer plutôt la diversité que la prototypicalité. Ce qui est caractéristique des images produites par l'intelligence artificielle, c'est qu'on en est encore au stade de la prototypicalité : elles viennent prolonger des stéréotypes très anciens.



Prenons par exemple l'image de la nature avec une montagne au fond et un lac devant, produite par une IA. C'est un stéréotype qu'on retrouve dans la peinture du sublime aux États-Unis avec des artistes comme Albert Bierstadt (1830-1902). On peut retracer la transmission de cette image dans d'autres supports, notamment les publicités. Cette représentation est vraiment un stéréotype mental et visuel et je pense que les historiens de l'art pourraient s'intéresser à ces images produites par l'intelligence artificielle pour établir des généalogies de stéréotypes qui sont

tout à fait caractéristiques et qui ignorent complètement d'autres modalités de figuration. En particulier le poly-perspectivisme qui propose des images inhabituelles qui veulent montrer simultanément plusieurs points de vue comme dans la peinture cubiste par exemple. Mais cette manière de représenter va tellement à l'encontre de nos modalités naturalistes habituelles, qui cherchent à émuler la vision humaine, qu'on en trouve peu de manifestations.

La stéréotypie visant à l'objectivité scientifique vient de là. De la réduction de la complexité de la diversité des spécimens observables à un type idéal qui existe dans la tête du savant qui demande à un artiste de le reproduire. D'une certaine façon, ce processus est déjà ancien. Pour les images que j'ai pu voir produites par l'intelligence artificielle, elles relèvent plutôt d'écarts empruntant à l'imaginaire. Mais c'est un imaginaire très attendu par ailleurs. Pour une image de création du monde, on voit un créateur. C'est un concept complètement méditerranéen, le démiurge, le potier, qui façonne la matière à partir de ses intentions, ce qu'on trouve à la fois dans la Bible et chez Platon. Mais c'est une vision tout à fait singulière. Pour beaucoup de cultures, le monde n'a pas été créé. Il a toujours existé, ou alors il se transforme par des événements contingents. Ces images perpétuent des mythes anciens et occidentaux.

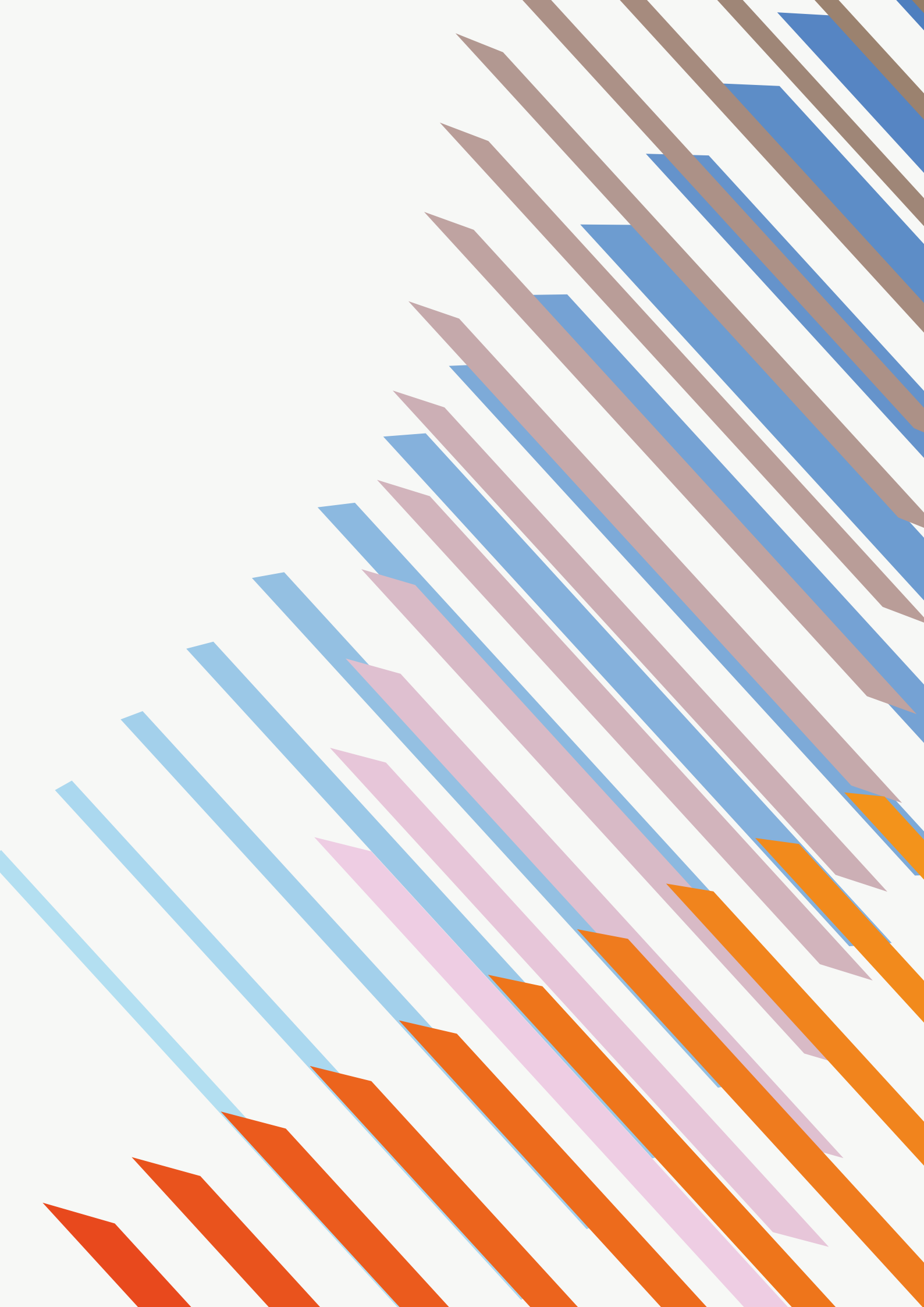


Réponse de Gemini au prompt


«La création du monde»

Est-ce que vous craignez que tous ces modes de représentation que l'IA ne génère pas en viennent à disparaître ?

Pour qui en viendraient-ils à disparaître ? Si c'est pour nous, Occidentaux, elles n'ont jamais été très présentes. On en a des échos dans les musées d'ethnographie ou dans ce que des artistes qui ont été intéressés par ces images ont pu en dire, notamment au début du siècle dernier. Mais il y a un danger possible : celui de l'affadissement, si je puis dire, ou de la naturalisation des images. Le risque est que là où elles sont encore vivaces, d'autres formes de figuration disparaissent sous les assauts de la figuration naturaliste. Une chose qui me frappe en Amazonie, dans toutes les petites villes, dans des baraques en bois suintantes d'humidité, vous allez pour manger un morceau, et il y a des paysages au mur. Ce sont des peintures de châteaux de la Renaissance, de jardins à la française, ou de paysages des Alpes suisses avec des chalets et des pots de géranium. Donc la stéréotypie gagne dans les interstices de la civilisation — si on peut parler de civilisation dans ces villes frontières — mais ces images sont très appréciées par ceux qui peuplent ces villes et qui colonisent la forêt. Ces colons s'efforcent de détruire un certain rapport à la forêt qui les inquiète et les contrarie. Ils préfèrent un monde naturaliste, plus policé et plus maîtrisable d'une certaine façon.







Plutôt que de produire des images objectives ou fiables, les intelligences artificielles génératives révèlent les représentations du monde statistiquement dominantes dans les données numériques.

Nourries d'une connaissance essentiellement occidentale des images et des représentations, elles portent l'empreinte des systèmes de domination historiques.

Elles constituent ainsi un miroir fascinant de nos propres imaginaires visuels et, en creux, de notre méconnaissance des autres manières de représenter le monde.

media-animation.be

méd:a
ANIMATION



FÉDÉRATION
WALLONIE-BRUXELLES

conseil supérieur
de l'éducation
aux Médias
csew