Sens [public]

Une exploration de la photographie vernaculaire par l'IA générative

François Gutherz

Publié le 15-10-2025

http://sens-public.org/articles/1800

Abstract

This essay traces a personal exploration, carried out as an amateur explorer, testing generative artificial intelligence tools through vernacular photography. By interpolating images across the latent representations of AI, this practice reveals the artifacts and hallucinations specific to these technologies. Rather than trying to fix them, it celebrates their poetic potential and calls for an imperfect generative AI, one that drifts and surprises. (FG)

Résumé

Cet article constitue le récit d'une exploration personnelle, menée en amateur et qui met à l'épreuve les outils d'intelligence artificielle générative appliqués à la photographie vernaculaire. Par des interpolations d'images à travers la représentation interne des modèles d'IA, cette démarche documentée met en lumière les artefacts et hallucinations caractéristiques de ces technologies. Loin de chercher à les corriger, elle en défend la puissance poétique et plaide pour une IA générative imparfaite, capable de dériver et de surprendre. (FG)

Mot-clés : IA, espace latent, création, interpolation, Stable Diffusion, MidJourney, hallucination, photographie vernaculaire

Keywords: hallucination, MidJourney, Stable Diffusion, interpolation, latent space, creation, AI, vernacular photography

Table des matières

Une exploration de la photographie vernaculaire par l'IA générative

François Gutherz

Latent Space Cadet, explorer l'espace latent

Face à l'irruption grandissante des intelligences artificielles génératives dans notre quotidien visuel, j'ai tout d'abord fait preuve d'une curiosité distante. Ni technophile enthousiaste, ni opposant farouche, je les ai abordées avec la réserve due à ma formation artistique classique. Ce texte n'est ni un article scientifique visant à évaluer la puissance des modèles, ni un manifeste artistique qui prendrait position pour ou contre l'intelligence artificielle. Il constitue le récit d'une exploration menée en amateur, au sens noble du terme : expérimenter, se laisser dériver, sans chercher à conserver le contrôle total sur cet espace mathématique devenu soudainement tangible et figuratif.

Un vif débat au sujet de l'image générée par IA anime les communautés d'artistes, dont certains exposent leur travail sur des plateformes en ligne et expriment la crainte que leurs travaux n'aient été pillés pour entraîner des modèles rendus accessibles sur abonnement. Au cours de l'année 2022, j'ai vu progressivement apparaître les expérimentations visuelles issues de ces logiciels de génération d'image par intelligence artificielle mis à disposition du grand public. Certaines images produites par quelques pionniers facétieux, à défaut de présenter des qualités esthétiques, donnaient à voir un propos insolite, sinon iconoclaste, comme cette image de « Jésus sortant du tombeau filmé par une caméra de surveillance » (u/KonoDioNo 2022). Les images générées par ces prototypes de DALL · E, bien que très imparfaites, ont attiré mon attention sur les possibilités d'exploration offertes par le *deep learning* dans le domaine de l'image.

FIGURE 1 – *CCTV video of Jesus rising from the tomb*, 2022

Irruption d'une nouvelle forme d'image

Les progrès du *deep learning* et de la puissance de calcul ont conduit l'IA générative d'images à atteindre un degré de ressemblance et d'iconicité très proche d'une photo originale. À l'image du débat autour du *deep learning* et de ses répercussions économiques, j'ai été témoin d'une forme de polarisation dans mon entourage, entre ceux qui sont farouchement opposés à l'utilisation de ces outils et ceux qui s'y sont essayés aussi tôt que possible.

FIGURE 2 – Couverture du JDD Magazine

Bien que relativement indifférent à ceux-ci, sans doute du fait de ma sensibilité, mes certitudes ont vacillé en voyant la couverture du JDD Magazine publié en France en mai 2023. Ce mensuel présentait une série de photographies de l'écrivain Michel Houellebecq au rendu étrange, tant par le grain de l'image particulièrement lisse que par les traits de son visage lui donnant l'aspect d'un vague sosie (fig. 2). Ces photographies accompagnaient un récit intitulé « Image artificielle », opérant une réflexion sur l'image publique que l'écrivain s'est créé au fil de ses prises de paroles et de ses écrits. La couverture mentionnant bien la nature artificielle des images, j'ai trouvé une réminiscence avec l'époque où les images 3D naissantes étaient appelées « nouvelles images » (Hénon 2018), à la différence qu'aujourd'hui, ce ne sont plus des algorithmes modélisant les lois optiques qui produisent des images, mais des réseaux de neurones simulés informatiquement.

Comment un réseau de neurones donne-t-il un caractère précis à une image ? C'est ce que j'ai découvert par hasard, à travers une publication de Xuan Luo *et al.* (2020) qui présentaient une méthode générative pour restaurer les couleurs des portraits photographiques en noir et blanc. Cette technique utilise un *sibling*, image sœur, obtenue en projetant l'original dans un générateur entraîné sur des photos de visages trouvés sur Flickr (Karras, Laine,

et Aila 2019). Les deux portraits, l'un historique et l'autre au rendu actualisé, sont fusionnés dans un espace nommé « espace latent » pour créer une image en couleur du personnage historique, très ressemblante physiquement à l'original en noir et blanc mais contemporaine dans sa picturalité.

La publication scientifique était accompagnée d'une vidéo de présentation dont les résultats, aussi stupéfiants qu'émouvants, montrent comment la photographie noir et blanc, dont la sensibilité au spectre lumineux était à ses débuts plus qu'imparfaite, a pu contribuer à façonner notre imaginaire collectif. Cette publication m'a aidé à me représenter le fonctionnement des IA génératives et à mettre un nom sur cette représentation interne, l'espace latent, dont je supposais l'existence. Cette utilisation de l'IA générative au seul domaine de la photographie m'a également éveillé à la possibilité de fabriquer des images sans revendiquer un acte pur de création et dont la picturalité ne serait pas fondée sur un pillage algorithmique d'œuvres antérieures. J'emploie ici le terme « fabriquer » à dessein : ni tout à fait « créer » (dont l'auctorialité me semblerait mensongère), ni simplement « générer » (ce qui restreindrait le processus à sa seule dimension technologique).

Une dernière découverte marquante dans mon exploration des liens entre IA générative et photographie documentaire a été le travail du designer et artiste danois, Bjørn Karmann. En 2023, il a présenté son appareil numérique portable, le Paragraphica (Karmann 2023), dont la particularité est de produire une image numérique du lieu et du moment où l'observateur se trouve, sans capteur optique ! De la même façon qu'un boîtier photo compact permet de capturer un instant sous la forme d'une photographie, le Paragraphica utilise une puce GPS et une connexion internet « 4G » pour collecter des données numériques contextuelles (lieu, date et heure, conditions météo, points d'intérêt à proximité) et composer un paragraphe descriptif qui est ensuite converti en image par une IA générative. Ce projet m'a amené à réfléchir à la façon dont les machines peuvent « voir » ou « représenter » le monde. Le Paragraphica est à la fois un moyen de visualisation des données et une réflexion sur les couches d'informations numériques géolocalisées que nous avons accumulées, décennie après décennie.

J'ai noté que certains observateurs refusent à cet appareil le nom de « caméra » sous prétexte qu'il ne possède pas d'objectif, sans s'interroger sur l'étymologie d'un mot qui ne dit pas explicitement que la lumière soit une partie du processus d'enregistrement. Le terme « caméra » provient de l'ex-

pression latine *camera obscura*, pour « chambre noire », qui est un dispositif d'observation d'un sujet dans le but d'en faire un dessin.

FIGURE 3 – Image issue du *press kit* du Paragraphica (1).

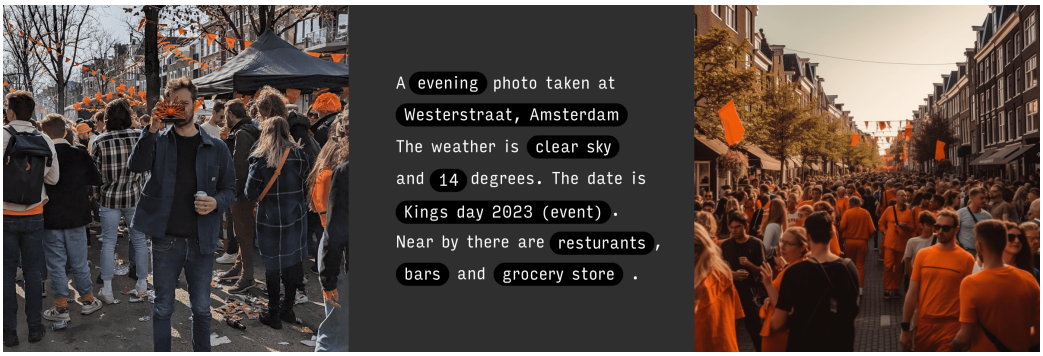

FIGURE 4 – Image issue du *press kit* du Paragraphica (2). À gauche, l'auteur utilisant son appareil au lieu et à la date indiquée dans le cartouche du centre. Le *prompt* généré donne des précisions sur le moment, y compris lorsqu'il est d'une nature particulière (fête nationale, par exemple). À droite, la photographie virtuelle imaginée par son dispositif.

Ainsi, l'argument de Karmann pour utiliser le terme *camera* (traduction anglaise d'appareil photo) pour le Paragraphica tient debout, même sans objectif conventionnel. Son appareil est, dans un sens moderne, une chambre à l'intérieur de laquelle se créent des images basées sur les données et l'intelligence artificielle. Il étend le concept de *camera* pour englober les technologies qui ne s'appuient pas strictement sur la réaction à la lumière mais sur la capture et l'interprétation sensible de données à même de générer une image, ce qui reste fidèle à l'idée d'une « chambre » comme un lieu d'interprétation.

Entrée dans l'espace latent de MidJourney

Ce qui a réellement lancé mes expérimentations, c'est la possibilité de mélanger deux images dans le logiciel MidJourney, non par interpolation de la valeur des pixels, mais en réalisant une moyenne des images sur le plan sémantique, par interpolation de vecteurs dans l'espace latent. Autant l'utilisation du seul *prompt* comme unique moyen de piloter un générateur d'image peut s'avérer aléatoire, autant le contrôle de la trajectoire dans l'espace latent par mélange de sources visuelles présente pour moi un réel intérêt.

Le logiciel MidJourney, disponible sous forme de service en ligne, propose ainsi une série de fonctions dont je retiens les suivantes :

Imagine

fonction de « texte vers image », qui permet à l'utilisateur d'entrer une phrase (en anglais, français, …) qui sera transformée en image. MidJourney fait quatre propositions à partir d'un tirage aléatoire. Chacune se rapproche du *prompt*, montrant toutefois des différences. Depuis le lancement de MidJourney, les modèles génératifs ont connu plusieurs évolutions, jusqu'à la version 7 disponible en 2025, qui tend à gommer de plus en plus les anomalies.

Describe

fonction « image vers texte », permet à l'utilisateur de téléverser une image à partir de laquelle MidJourney génère quatre descriptions textuelles distinctes. Chaque suggestion semble être une variation aléatoire, tirée parmi les réponses les plus probables. La réponse est faite en anglais, formulée d'une manière qui tient plus de la classification que d'une description purement visuelle.

Blend

fonction « images vers image », fusionne le contenu de deux à cinq images différentes afin d'en produire une nouvelle. MidJourney ne se limite pas à un fondu ou *morphing* pixel par pixel dans l'espace colorimétrique RGB, mais réalise une interpolation dans l'espace latent qui conserve les caractéristiques conceptuelles des images sources. Là encore, quatre variations sont proposées, dans un style cohérent mais avec des différences de composition et d'organisation des sujets dans l'espace pictural.

La première image qui m'a aidé à percevoir l'intérêt de la démarche est une photo du « 108 », un tiers-lieu installé au centre de la France, accueillant un *hackerspace* qui accompagne les démarches mêlant création et pratique numérique. J'ai fabriqué une photo imaginaire de la cour du 108, image qui a provoqué une forme de fascination auprès des résidents du lieu. Cette réaction m'a fait réfléchir au rapport que nous entretenons avec l'image et sur notre façon de percevoir et d'interpréter une photographie, laissant la place à un espace d'étrangeté dans une représentation en apparence réaliste.

Dans leur ensemble, les images que j'ai choisies pour cette exploration ne relèvent pas d'une démarche photographique professionnelle ou artistique au sens strict. Il s'agit de clichés personnels, saisis au fil de mes déplacements ou de ma vie quotidienne – ce qu'on pourrait regrouper sous le terme de photographie vernaculaire. C'est précisément cette banalité qui m'a intéressé : d'une part, parce qu'elle limite le risque de reproduire involontairement le style d'un artiste connu *via* les biais d'entraînement des modèles génératifs ; d'autre part, parce que ce qui n'était, au départ, qu'une exploration technique s'est muée peu à peu en une forme de trouble face à ces fausses images de moments réellement vécus.

J'ai tout d'abord testé le processus itératif suivant : je pars d'un cliché issu de mes archives personnelles, que je transforme en description textuelle dans le sabir algorithmique de MidJourney, à l'aide de sa fonction *Describe*. À partir des multiples descriptions qui me sont alors proposées, je choisis celle dont le résultat, à nouveau transformé en image par la fonction *Imagine*, m'évoque le plus possible l'original, pour m'en approcher à l'aide de la fonction *Blend*. Agissant comme une interpolation du vecteur latent, le résultat se situe généralement à mi-chemin. Peut-être est-ce de là que MidJourney tire son nom ?

Pour me rapprocher au mieux de l'image originale, je dois généralement répéter l'opération plusieurs fois de suite, en réinjectant la photo d'origine avec le résultat du précédent mélange, comme détaillé en figure 5. Afin de faciliter la lecture du document, une signalétique aidera le lecteur à distinguer les prises de vue réelle des images crées par IA : un point rouge indique une prise de vue réelle (argentique ou numérique), un point bleu indique une image générée par IA (MidJourney ou Stable Diffusion).

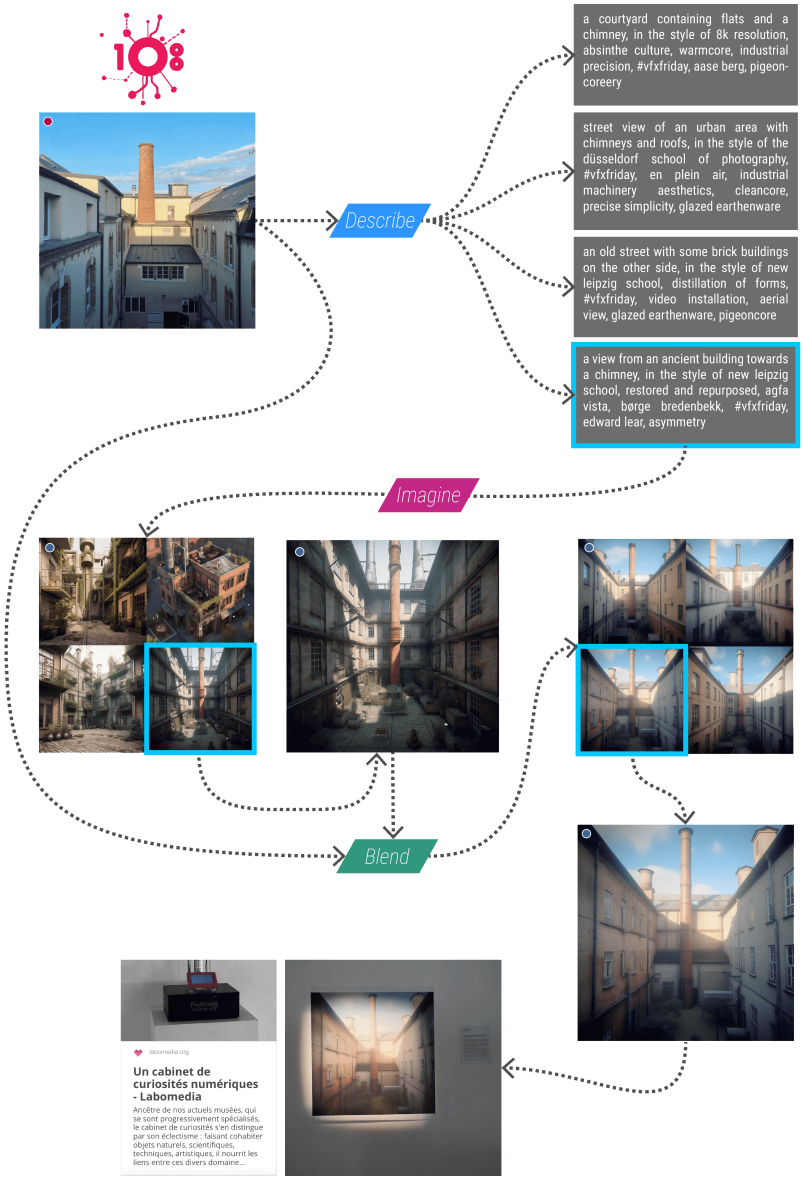

FIGURE 5 – Étapes de génération des images

Dérives, hallucinations et angles morts

Il est notable que si je parviens à me rapprocher de l'original dès les premières itérations, comme si l'opération me permettait de me déplacer dans l'espace latent, il arrive un point où toutes les opérations de *Blend* suivantes contribuent à m'en éloigner (fig. 6). Ce phénomène me fait penser à la trajectoire d'un projectile attiré par la gravité d'un corps céleste mais qui parvient finalement à s'en détacher après en être passé le plus proche possible, comme si l'on était en présence de forces attractives et répulsives à l'œuvre dans l'espace latent qui puissent soit rapprocher, soit éloigner le résultat final de l'image désirée. Il est également possible que l'injection d'une part d'aléatoire projette le vecteur plus loin que ce l'interpolation ne devrait le faire, ou encore que se produise un effet de vase clos qui, à force de réinjecter les composantes du vecteur latent, oriente le modèle vers des zones où la reconstruction devient moins fidèle à l'image de départ. Mon procédé fonctionne par une série de choix sur des critères esthétiques et sur l'intuition de la meilleure trajectoire dans l'espace latent, à la recherche d'un équilibre entre le contrôle et l'imprévisibilité dans ce processus de fabrication déléguée.

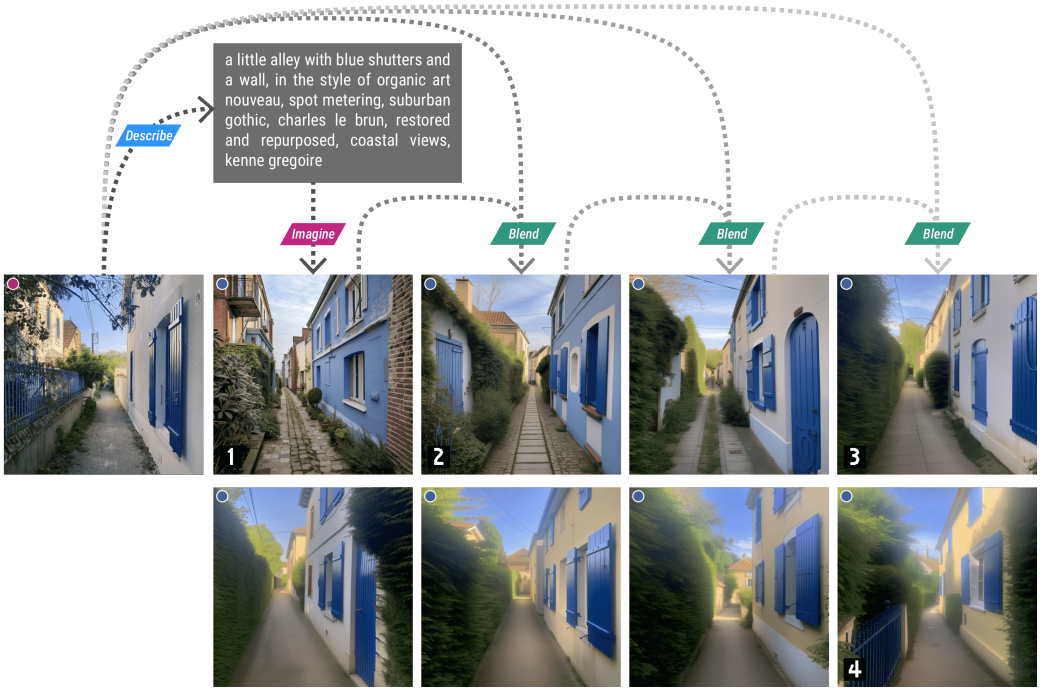

FIGURE 6 – À partir d'une photo, le prompt fabriqué puis réinterprété par MidJourney produit une image à l'aspect conventionnel et plausible (1). À mesure que je réinjecte la photo originale, la brume qui était à peine perceptible sur l'original se révèle progressivement (3), jusqu'à complètement contaminer l'image (4)

L'expérimentation « *Latent Space Cadet* » s'est prolongée sur plus d'un an, me permettant d'observer l'évolution des modèles génératifs, notamment entre MidJourney (versions 5 et 6) et Stable Diffusion. C'est en cherchant une alternative à MidJourney que mon choix s'est porté sur Stable Diffusion XL (SDXL). Cet outil présente un modèle étendu dans l'interprétation du texte, une grande stabilité dans le niveau de détail des images, et il fonctionne sans avoir recours à un serveur distant. Un catalogue ouvert est par ailleurs maintenu par une communauté qui partage activement ses modèles, entraînés, tout comme MidJourney, avec un respect très relatif de la propriété intellectuelle (Tapper 2024).

Ma compréhension du fonctionnement de ces outils s'est initialement basée sur l'observation et sur des articles de vulgarisation. J'ai d'abord eu une ré-

vélation quant à la simplicité conceptuelle qui sous-tend ChatGPT en découvrant la publication de Stephen Wolfram, « Que fait ChatGPT et pourquoi est-ce que ça fonctionne ? » Wolfram, reconnu pour ses contributions au calcul symbolique (notamment à travers les logiciels *Mathematica* et *Wolfram Alpha*), adopte ici une posture singulière : il s'intéresse avec rigueur et curiosité à une IA d'un type diamétralement opposé au postulat technique de ses propres outils. Il décrit ChatGPT comme un système statistique qui prédit, mot après mot, la suite la plus probable d'un texte. D'après lui, l'invention de ce type d'intelligence artificielle apporte un éclairage nouveau sur la nature même du langage et de la pensée humaine, en révélant à quel point leur structure peut être modélisée statistiquement :

> Cela suggère quelque chose de scientifiquement très important : que le langage humain et les schémas de pensée sous-jacents seraient, d'une certaine manière, plus simples et plus déterministes que nous ne le pensions. ChatGPT semble l'avoir mis en évidence, de manière implicite. (Wolfram 2023)

L'enthousiasme de Wolfram, s'il est rigoureusement argumenté, n'évoque pas explicitement les erreurs que peut faire ChatGPT. Le débat s'est cependant ouvert sur la capacité de grands modèles de langage à comprendre le sens caché d'un texte et des études en montrent les limites actuelles (Dentella et al. 2024). C'est un des points qu'aborde William Audureau (2024) dans son article sur le fonctionnement des intelligences artificielles génératives. Il rappelle que ces outils ne font pas preuve d'une intelligence sensible à proprement parler, mais fonctionnent par interpolations statistiques sur d'immenses jeux de données pouvant donner lieu à des biais ou hallucinations.

Qu'elles soient issues d'un modèle de langage ou liées à la génération d'images, les hallucinations prennent parfois des formes subtiles. L'une d'elles est apparue lors d'une de mes tentatives de reproduire une photo de coupure de presse (tirée du journal *Le Monde* dans son édition papier) montrant une image du film *Boulevard du crépuscule* avec Gloria Swanson et Erich von Stroheim. La photo originale (1) montrée en figure 7, est accompagnée d'une légende, mais dans l'interprétation qu'en fait MidJourney version 5, le texte de l'article a migré sur la robe, devenant un motif du textile (2). Avec MidJourney dans sa version 6, on voit que le problème a été corrigé. En personnifiant cette IA générative, on pourrait dire qu'elle n'arrive pas à se concentrer sur plusieurs

sujets simultanément, provoquant un accident aussi surprenant que poétique. Ainsi, jusqu'à présent, j'ai plutôt eu tendance à exploiter les anomalies et défauts de MidJourney dans sa version 5. À mes yeux, une reproduction trop fidèle n'aurait plus aucun intérêt, même si c'est sans doute l'objectif poursuivi par les concepteurs de ces outils.

FIGURE 7 – Extrait du journal *Le Monde*, montrant une photo du film *Boulevard du crépuscule* (1), reproduite avec MidJourney v. 5 (2) et v. 6 (3).

À force d'expérimentations, j'ai constaté que la fonction *Describe* de MidJourney, qui transforme une image en texte, produit une description qui n'a réellement de sens que pour la fonction *Imagine*. Cette description est constituée de bribes de phrases, de références multiples et en apparence hermétiques.

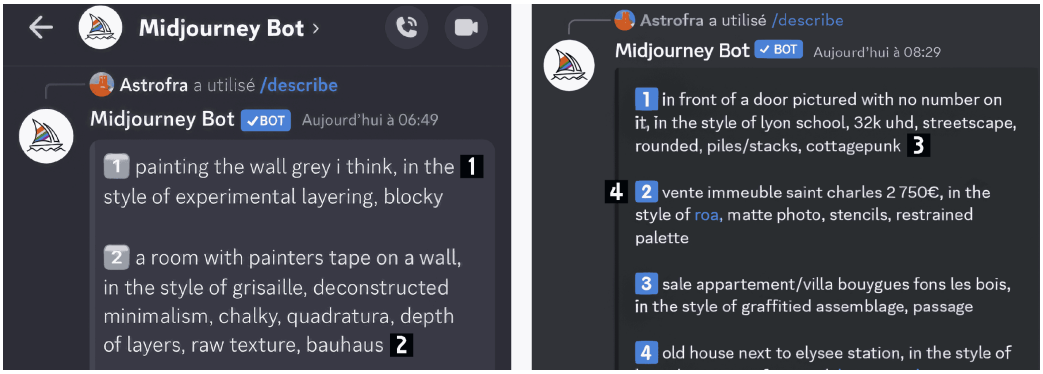

FIGURE 8 – Description d'image faite par MidJourney. Certaines réponses indiquent un apprentissage sur des données annotées oralement (1), des courants artistiques (2), des références à la culture internet (3) ou par aspiration automatique de divers sites web (4)

Si les descriptions proposées par MidJourney sont généralement pertinentes, en cherchant à obtenir un résultat plus fidèle à ma photo originale, j'ai tenté de remplacer la fonction *Describe* en chargeant ChatGPT d'analyser mes images.

Le *prompt* que j'utilise pour extraire une description textuelle depuis une photo est le suivant :

```
« Can you describe this photo as a prompt for Stable Diffusion
IA image generator? The framing, the subject, the image composition,
the lighting and the photo type? As one single prompt? (I just
need the text) »
```

Dans certains cas, je demande immédiatement à ChatGPT de générer l'image qui correspond à ce *prompt*, par l'intermédiaire de DALL · E. Si le résultat visuel ne me convient quasiment jamais, en raison d'une esthétique trop marquée, l'image générée me permet tout de même de m'assurer de la pertinence de la description et du cadrage.

Ainsi, comme le personnage de Rick Deckard dans *Blade Runner* (fig. 9), je demande à l'IA de zoomer sur une partie de l'image au-delà de ce qui est optiquement possible ou de révéler des détails qui n'auraient pas pu être capturés lors de la prise de vue. Cependant, de la même façon que l'agrandissement à l'extrême d'un film argentique ne peut révéler une quelconque vérité objective, les hallucinations d'un modèle génératif errant dans l'espace latent contribuent définitivement à nous en éloigner.

FIGURE 9 – Dans *Blade Runner* (1982, Ridley Scott), Harrison Ford incarne Rick Deckard, un policier chargé de distinguer les créatures synthétiques au milieu des humains, explorant ainsi une frontière sur laquelle il se situe lui-même. La séquence dite *enhance*, convoquant des références comme *Blow-Up* (1966, Michelangelo Antonioni) ou *Les Époux Arnolfini* (1434, Jan van Eyck), préfigure les développements récents de l'IA comme l'agrandissement des images et le *neural radiance field*

Provoquer les hallucinations pour mieux les contrôler ?

Pour sonder la pertinence des détails qui peuvent être inférés d'une image par une IA générative, j'ai soumis une photographie originale de la Tour Eiffel parisienne à Topaz Gigapixel, un logiciel exploitant un réseau de neurones explicitement entraîné pour l'agrandissement d'image. J'ai commencé par cadrer le centre de l'image que j'ai demandé à doubler en taille. Puis, en réinjectant dans ce processus chaque étape de l'agrandissement précédent, j'ai pu me rapprocher du deuxième étage de la tour et de son restaurant panoramique vitré. Tout d'abord, c'est la structure industrielle et anguleuse de la tour, caractéristique du style Eiffel, qui s'est transformée en motifs organiques et végétaux. J'attribue cette dérive aux artefacts intrinsèques à la compression JPEG de l'image originale (Omelyanchuk 2013), nourrissant les hallucinations de l'IA. À chaque génération, de nouveaux artefacts hallucinatoires apparaissent ainsi, se consolident et finissent par contaminer toute l'image au point qu'elle en devient complètement abstraite. C'est comme si l'on avait perforé la membrane nous séparant de l'espace latent, pour en révéler les schémas intimes.

FIGURE 10 – Agrandissement par IA, chaque nouvelle vue étant générée à partir de la précédente. La première image est une photo conventionnelle, les suivantes sont des hybrides où la part d'IA est de plus en plus importante

En laissant dériver mon esprit sur cet agrandissement de la baie vitrée, j'aperçois une ville en vue aérienne et ses avenues aux motifs familiers mais étranges. Dans son film *La Jetée* (Marker 1962), le cinéaste Chris Marker imagine les tentatives d'une humanité désespérée pour atteindre une époque future. La voix off, focalisée sur le protagoniste, énonce :

> L'avenir était mieux défendu que le passé. Au terme d'autres essais encore plus éprouvants pour lui, il finit par entrer en résonance avec le monde futur. Il traversa une planète transformée, Paris reconstruit, dix mille avenues incompréhensibles (Marker 1962)

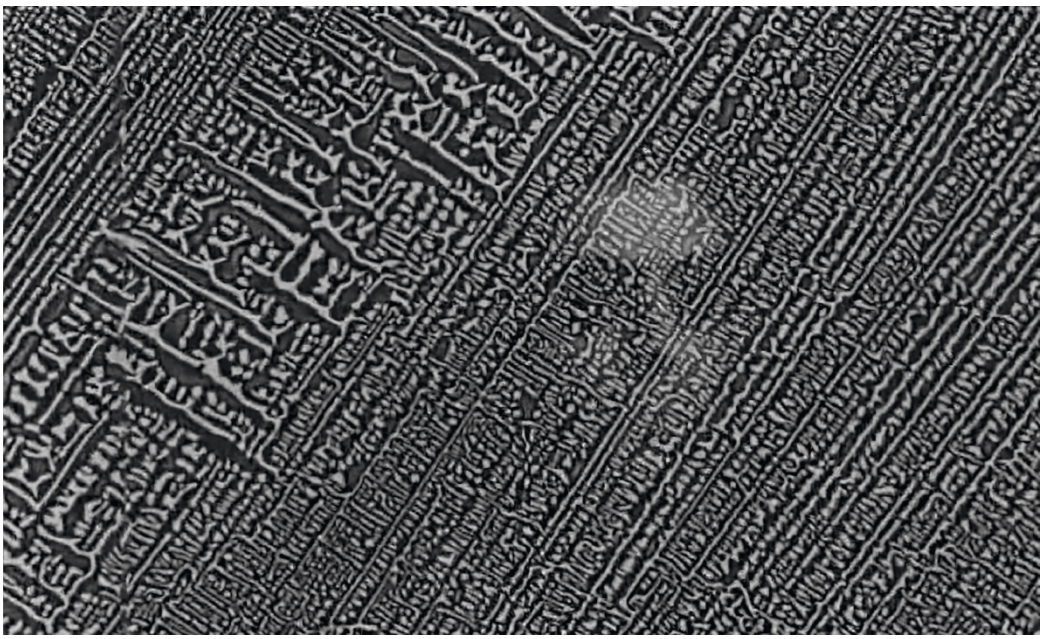

FIGURE 11 – *La Jetée* (1962, Chris Marker), vision d'un Paris du futur, du micro au macro

Pour représenter le Paris de ce lointain futur, Chris Marker a utilisé des images, semble-t-il, obtenues à l'aide d'un microscope électronique à partir d'un matériau dont la structure ordonnée évoque le métal et dessine « dix mille avenues incompréhensibles ». On peut célébrer ici le génie de Chris Marker, dans son économie de moyens et pour le détournement d'images à vocation scientifique.

Au-delà du procédé presque mécanique d'agrandissement par IA, les images fabriquées par mélange de *prompt* et d'interpolation latente sont également propices aux hallucinations. À partir du portrait d'un dormeur assoupi sur un banc dans le parc d'Ueno, à Tokyo, je me suis heurté à une forme d'impensé de l'IA générative. Dans ma photographie originale, seule une portion du visage apparait, laissant transparaître un sentiment paisible. Le reste du corps est invisible mais il n'est pas difficile pour un observateur d'imaginer sa posture. Cet effort d'extrapolation semble, en l'état actuel, difficilement atteignable pour l'espace latent qui provoque une série d'accidents où le corps du dormeur se trouve douloureusement encastré dans le banc ou suspendu dans le vide. Ici aussi, mon exploration par interpolations successives s'est rapidement

arrêtée, MidJourney provoquant des hallucinations de plus en plus éloignées de l'image de départ.

FIGURE 12 – Le dormeur d'Ueno, généré par MidJourney à partir d'une photo cadrée de telle sorte qu'il ne parvient pas à la comprendre

Cette collision s'est produite avec plus de subtilité lors d'une tentative de reproduction d'une photo d'un restaurant situé dans ma rue. Dans la photo originale, on devine l'inscription « Menu du jour » collée sur la vitrine. Dans la version synthétique, en dépit de mes efforts et de la promesse des développeurs de MidJourney concernant la cohérence des textes typographiés, seules quelques lettres de l'inscription originale ont survécu, mêlées au plateau en mosaïque.

FIGURE 13 – Le menu du jour, encastré dans la table

Dans le champ de la photographie argentique, le polaroïd constitue également un matériau idéal : réputé fiable, il exclut théoriquement toute possibilité de trucage grâce à un procédé qui court-circuite les étapes de développement et de tirage. Sa signature visuelle reconnaissable est bien maîtrisée par les IA génératives, ce qui rend d'autant plus insolite la survenue d'accidents statistiques ou d'hallucinations. Ainsi, une simple vue d'immeuble en contre-plongée, une fois passée par les méandres de l'espace latent, ressort toujours aussi banale, mais traversée d'une sphère suspendue, trop lisse pour être un ballon, trop nette pour figurer la lune. L'IA, dotée d'une connaissance implicite des phénomènes lumineux (réverbération, ombrage), en fait un artefact idéalement intégré à la scène : une hallucination parfaite.

FIGURE 14 – Accident subtil depuis un polaroïd

À la recherche de ce type d'accident, j'ai poursuivi l'exploration de mon stock de vieux polaroïds. L'un d'eux, pris à Otaru en 2001, montre l'entrée d'une brocante débordant d'objets hétéroclites avec, au premier plan, des clients de dos. Après l'avoir soumis à ChatGPT pour en extraire un *prompt*, j'ai généré une première image avec DALL · E. Trop éloignée du style d'un polaroïd, présentant une accumulation obsessionnelle de détails, elle m'a tout de même servi à préciser le cadrage. Stable Diffusion, plus réaliste, en a livré une interprétation proche de l'original. Mais c'est en augmentant la résolution que le processus a complètement déraillé. Le modèle de diffusion génère l'image

par cellules indépendantes : au-delà de 1024 pixels, pris individuellement, les détails sont cohérents à l'intérieur d'une cellule. Mais une fois assemblés, ils produisent des corps distordus, des membres en trop, des jambes sorties de nulle part. Cette limite est bien documentée :

> Les modèles existants de diffusion à grande échelle sont limités à la génération des images en 1K, ce qui ne répond pas aux besoins du marché. L'échantillonnage direct d'images à plus haute résolution donne souvent des artefacts comme les distorsions et la répétition excessive d'objets (Kim et al. 2024).

D'autres travaux soulignent que « les personnages [...] se retrouvent gigantesques une fois les cellules juxtaposées, car Stable Diffusion n'a aucune notion de perspective globale » (McKeag 2023). Tous ces travaux semblent d'ailleurs appeler à l'amélioration des modèles et à l'optimisation de leurs performances.

FIGURE 15 – Voyage à Otaru, Japon, puis à travers de multiples IA génératives, livrant chacune sa propre interprétation de la scène de départ

Résister à l'optimisation continue

Le fait que ces outils de génération d'images en IA soient disponibles en ligne, nous laissant à la merci d'optimisations décidées unilatéralement, m'a amené à explorer les alternatives. J'ai ainsi installé Stable Diffusion sur mon ordinateur personnel, pour le faire fonctionner et évoluer uniquement lorsque j'en ai besoin. Cet usage « domestique » m'offre une plus grande autonomie : il me donne non seulement une maîtrise de l'outil et de sa consommation électrique, mais il me permet également de déborder du contexte d'utilisation pour en exploiter, sinon en chercher, les défauts.

En 2025, les IA génératives m'apparaissent encore imparfaites mais créativement fertiles. C'est en contrariant leurs spécifications techniques que se fissure le cadre normatif dans lequel elles sont supposées fonctionner, ouvrant accidentellement un imaginaire potentiel. Mais il faut nous hâter d'en tirer parti, car l'objectif de l'industrie numérique est de lisser ces anomalies, d'effacer les hallucinations. Le moment où les machines rêvent encore, où « les androïdes rêvent de moutons électriques », est peut-être déjà derrière nous.

Comment ne pas *prompter* ?

Au tout début de mon expérimentation, j'avais une volonté de fabriquer des dérivés de mes propres photos, comme pour prolonger une expérience et compenser le fait que je n'aurais pas ramené assez d'images de mes voyages. Cette intuition sensible se vérifie, puisqu'il suffit que je montre à ma compagne l'une de ces photos de vacances artificielles pour qu'elle reconnaisse sans hésitation le lieu et le moment.

Pour m'épargner la préparation du *prompt* initial, j'avais tout d'abord tenté de fournir deux fois la même photo à MidJourney en lui demandant de les mélanger. Mais l'opération est rendue impossible par un mécanisme défensif qui dissuade l'utilisateur en le menaçant d'un bannissement de la plateforme !

J'ai ensuite tenté de recadrer l'image pour en fabriquer une fausse variante, mais cette approche me semblait intellectuellement peu satisfaisante. Ainsi, pour une grande partie de cette expérience, j'ai initié mes mélanges avec une photo originale d'une part, et une image synthétique issue d'un *prompt*, lui-même généré par MidJourney ou ChatGPT d'autre part.

Le procédé étant relativement long et assez complexe, j'ai fini par trouver une façon de leurrer MidJourney : je mélange ma photo originale avec une image sémantiquement inerte (fig. 16) qui présente deux avantages. Elle n'est pas détectée comme problématique et la déviation occasionnée dans l'interpolation latente est très limitée.

Pour obtenir cette masse latente inerte, j'ai préparé une texture d'un gris parfaitement moyen (0x7F7F7F) marquée par un authentique grain obtenu en isolant les plus hautes fréquences d'une photographie argentique.

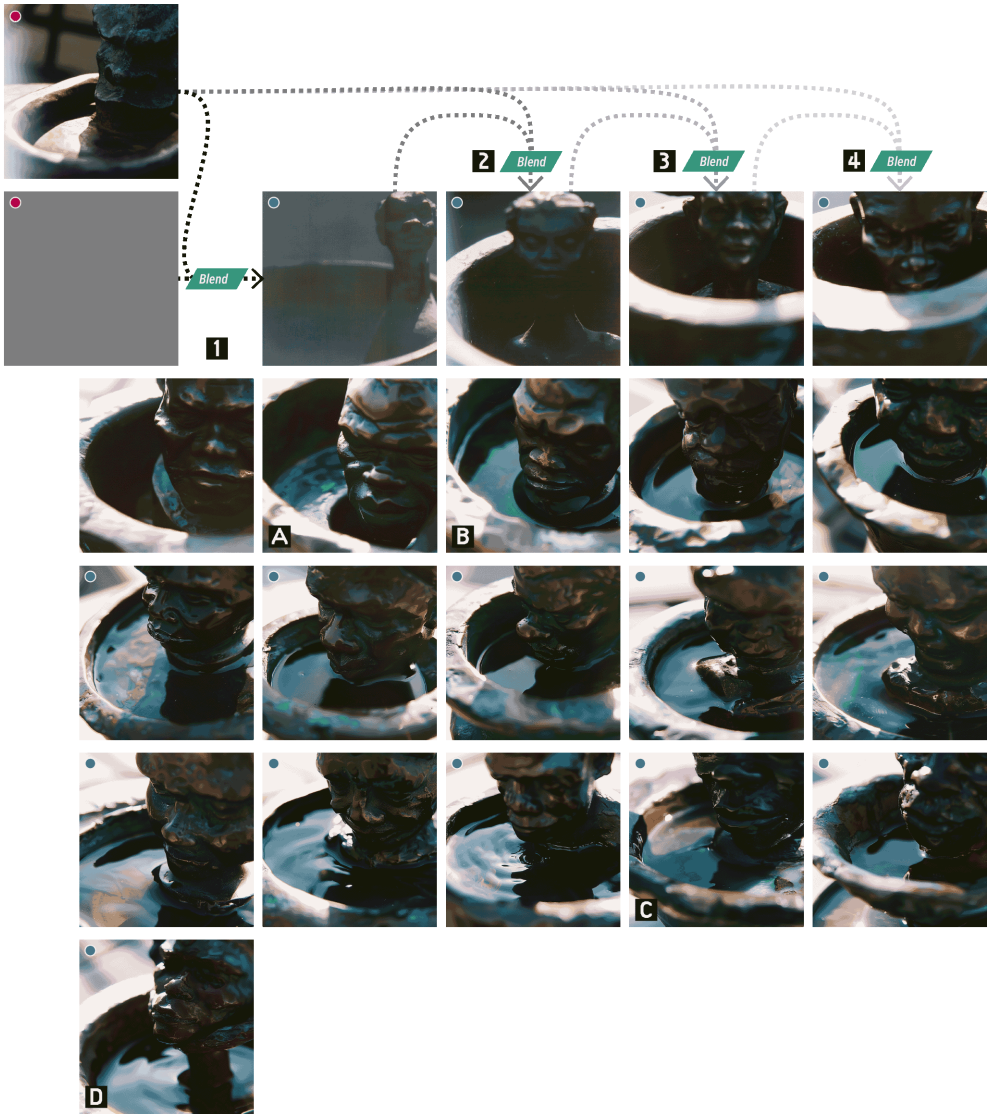

FIGURE 16 – Exemple d'interpolation dans l'espace latent entre une photographie conventionnelle et une image complètement grise (habillée d'un léger bruit RGB). MidJourney réagit positivement et par itérations successives parvient à s'approcher de l'original

L'interpolation latente fait progressivement émerger le sujet de l'image originale, comme une silhouette sortant du brouillard.

Du flou optique à l'indécision statistique

À la différence de MidJourney, Stable Diffusion fonctionne systématiquement à partir d'un *prompt*, seul ou accompagné de paramètres de contraintes. Bien que l'architecture de SDXL ne s'appuie pas réellement sur l'interpolation dans l'espace latent, elle permet de faire dériver très finement la *seed* (le bruit initial) à partir de laquelle l'image prend forme.

FIGURE 17 – Réinterprétation d'une photo argentique de Prague en 1993. La photo a été « lue » par ChatGPT qui a pu en déchiffrer les principaux éléments en dépit d'une mise au point approximative. Les variations obtenues sous SDXL en manipulant la *seed* peuvent donner un sens nouveau au flou de l'image finale

Pour mes expérimentations sous SDXL (fig. 17), le *prompt* d'origine est généré à l'aide de ChatGPT que je charge de décrire la photo originale (1). En l'absence de fonction *Blend*, j'utilise la *seed* pour avancer dans les multiples représentations possibles à partir d'un *prompt* donné. Partant toujours de la même valeur de *seed*, j'y ajoute de minuscules variations afin d'obtenir des

versions presque identiques du même sujet. Il est remarquable que, si une grande partie de l'image reste invariante, certains détails changent de forme, d'aspect ou se déplacent dans l'image (2).

Une fois ces dizaines d'images générées, j'ai écrit un programme en langage Python qui se charge de les mélanger à proportions égales tout en calculant l'écart type de valeur RGB (3) pour chaque pixel. Cette comparaison entre chaque image de la série me donne une carte d'intensité que j'exploite pour appliquer à chaque pixel un flou gaussien dont le rayon est fonction de l'écart type.

J'obtiens ainsi une image (4) dont les zones de flou, ni optique, ni analogique, indiquent le niveau d'incertitude de l'image, comme le produit d'un désaccord statistique entre des milliers d'hallucinations locales. Cette capacité des IA génératives à « halluciner », en fréquence comme en intensité, des milliers d'images semblables, est comme un écho à la masse d'images numériques que nous avons produite depuis l'avènement de la photographie numérique et des réseaux sociaux.

Expositions et échanges avec le public

En novembre 2023 et avril 2025, j'ai été invité dans le cadre d'événements organisés à Orléans par des acteurs institutionnels à exposer certaines images du projet « *Latent Space Cadet* ». Une vingtaine de paires constituées de la photographie et de l'image générée ont été exposées, accompagnées d'un texte racontant les circonstances de l'instant photographique. Jouant volontairement sur l'ambiguïté des ressemblances entre IA et réalité, j'ai ainsi laissé aux visiteurs le soin de distinguer les vraies images des images synthétiques. Certains détails, comme la présence de texte difforme et illisible dans une image générée étant, à ce stade de la technique, des marqueurs d'artificialité.

FIGURE 18 – Photo de famille personnelle (ma mère et moi). L'IA a inversé les motifs de nos vêtements, mais quelque chose dans le regard a persisté

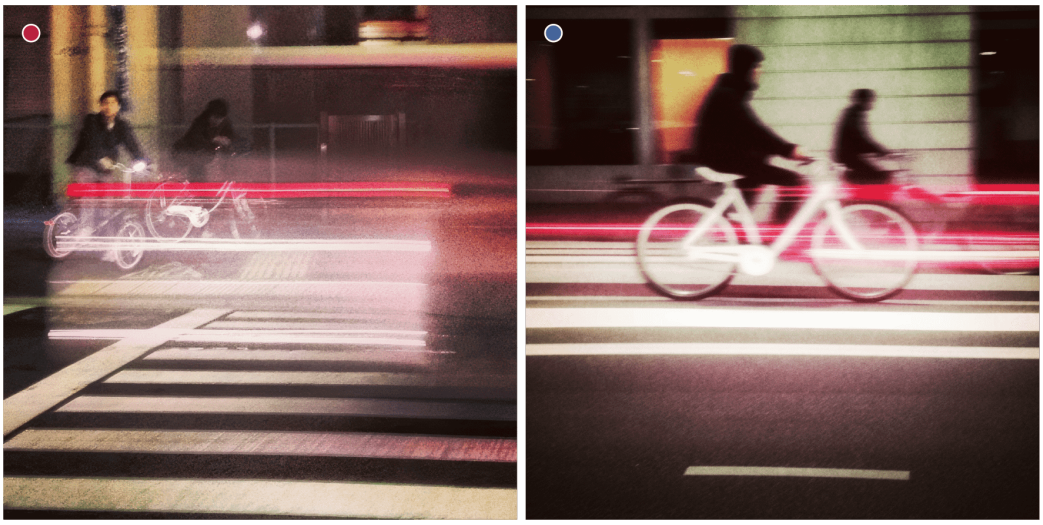

FIGURE 19 – Photographie prise sur un smartphone de 2015. La version synthétique, bien qu'un peu simplifiée, en conserve l'essentiel

FIGURE 20 – Photographie prise dans le métro d'un couple qui semblait avoir besoin de lunettes à verres progressifs. Le sujet et l'expression des visages en font un cas d'école pour l'IA

Lors de ces événements, j'ai pu également échanger avec le public, ce qui m'a poussé à chercher une définition de l'espace latent. J'en ai trouvé une description synthétique proposée par Dave Bergmann (2025), comme espace mathématique multidimensionnel où chaque point représente une version possible d'un contenu généré (image, texte, etc.). Les variables implicites de notre représentation du monde y ont été compressées puis rendues manipulables par un modèle. Dans une tentative de vulgarisation, je décrirais l'espace latent comme gigantesque et multidimensionnel, à l'intérieur duquel sont incubées toutes les réalités qu'une IA générative peut fabriquer. Chaque point dans l'espace latent correspond à une version de l'image. Lorsque nous avançons dans cet espace, chaque pas que nous faisons modifie légèrement l'image, ajoutant ou enlevant des détails, changeant les formes et les couleurs, en modifiant l'ambiance ou la picturalité. Pour se représenter ce qu'est un espace multidimensionnel (à N dimensions pour $N > 3$) on peut convoquer la séquence finale du film *Interstellar* de Christopher Nolan dans laquelle le héros navigue à travers un tesseract, un cube à N dimensions qui représente l'espace et le temps de façon continue.

Mutations et perspectives

Le nom de « *Latent Space Cadet* » (Latent Space Cadet 2022) fait volontairement référence à la littérature de science-fiction populaire du XX^{e} siècle. J'ai joué sur la traduction anglaise d'espace latent, le terme de « cadet » soulignant mon approche de novice à la découverte d'un espace potentiellement infini, où s'ouvrent des brèches et d'où surgissent des hallucinations.

Ce qui avait commencé comme un simple jeu d'imitation s'est transformé en démarche exploratoire. De 2023 à 2024, je me suis fixé un rythme d'une image par jour, publiant sur un compte Instagram anonyme une photo originale et sa réinterprétation générée par IA. J'ai fait ce choix de l'anonymat en anticipation des vives réactions que suscite l'IA dans le domaine de l'image, un débat qui m'a rappelé les tensions en école d'art dans les années 1990. L'enseignement numérique se focalisait sur la publication assistée par ordinateur, dans ce qu'Edmond Couchot (2003) décrit comme « la salle des Macintosh » et l'image de synthèse n'était pas identifiée par les étudiants comme un véritable médium.

Ce débat sur la légitimation d'une technique est évoqué par Pierre Hénon (2018) dans son livre sur l'histoire de l'animation numérique quand il cite Christian Guillon au sujet du débat qui a animé le monde du cinéma au tournant des années 1980 et illustre bien le procès en illégitimité qui a été fait aux pionniers de l'image de synthèse :

> Il y avait aussi une raison psychologique : la plupart des gens qui faisaient du cinéma étaient effrayés par l'idée qu'on puisse faire des images sans capter le réel. Pour eux le cinéma est basé sur le contrat formulé par André Bazin : le cinéma qu'on voit est le témoin d'une réalité. C'est le fondement même du cinéma qui est attaqué par le concept d'image de synthèse. […] Et puis il y avait une raison technique aussi : les images de synthèse étaient de mauvaise qualité, au moins en termes de définition, et n'était pas compatibles techniquement avec le cinéma analogique de définition supérieure ; […]. D'un côté les gens des nouvelles images prétendaient qu'ils pouvaient produire des images qu'on n'avait jamais vues, de l'autre côté les gens du cinéma répondaient : oui, mais c'est moche […]. (Hénon 2018)

De même que les pionniers de l'animation numérique ont fait face à des résistances, la démocratisation de la photographie à la fin du XIX^{e} siècle, perçue comme une technique purement mécanique ne laissant pas de place à l'interprétation personnelle, a également bousculé la peinture académique. Pourtant, des peintres comme Pierre Bonnard ont accepté les possibilités uniques de la photographie (Barreteau 2022) pour enrichir et élargir leur propre expression picturale.

Au fil de cette exploration, j'ai progressivement découvert que l'usage des IA génératives pouvait ouvrir un territoire créatif. C'est dans cet esprit que j'ai choisi d'illustrer la conclusion de cet article par un arpentage poétique (fig. 21), évoquant à la fois l'Ostende embrumée de Harry Gruyaert (2015) et mon quotidien fait de déambulations photographiques. En partant d'un simple *prompt* généré à partir de mes photos, j'ai utilisé Stable Diffusion XL pour décliner des dizaines de variations du front de mer flamand. Il ne s'agit pas ici d'une cartographie : les rues ne se connectent pas, les bâtiments ne forment pas un plan. Mais quelque chose tient, une hallucination cohérente se construit, comme dans un rêve récurrent dont je reconnaîtrais les rues familières.

Plutôt que d'opposer les images artificielles à celles réputées authentiques, peut-être est-il encore temps de défendre ces IA génératives imparfaites ? Tant qu'elles produiront des artefacts et des textes illisibles s'ouvrira un imaginaire dans lequel il sera possible de s'engouffrer.

FIGURE 21 – Promenade sur le front de mer et multiples vues d'une Ostende imaginaire

Bibliographie

Audureau, William. 2024. « Tout Comprendre à l'intelligence Artificielle, Cette Technologie Source de Nombreux Malentendus ». https://www.lemonde.fr/les-decodeurs/article/2024/04/20/tout-comprendre-a-l-intelligence-artificielle-cette-technologie-source-de-nombreux-malentendus_6228954_4355770.html.

Barreteau, Pierre. 2022. « Willy Ronis et Pierre Bonnard : Quand Des Images Se Rencontrent ». *Paris Perdu*.

Bergmann, Dave. 2025. « What Is Latent Space ? » https://www.ibm.com/think/topics/latent-space.

Couchot, Edmond, et Norbert Hillaire. 2003. *L'Art Numérique. Comment La Technologie Vient Au Monde de l'art*. Paris : Flammarion.

Dentella, Vittoria, Fritz Günther, Elliot Murphy, Gary Marcus, et Evelina Leivada. 2024. « Testing AI on Language Comprehension Tasks Reveals Insensitivity to Underlying Meaning ». *Scientific Reports* 14 (1) :1-11. https://doi.org/10.1038/s41598-024-79531-8.

Gruyaert, Harry. 2015. « Rétrospective Harry Gruyaert ». Exposition.

Hénon, Pierre. 2018. *Une Histoire Française de l'animation Numérique*. Paris : EnsAD.

Karmann, Bjorn. 2023. « Paragraphica ». https://bjoernkarmann.dk/project/paragraphica.

Karras, Tero, Samuli Laine, et Timo Aila. 2019. « Flickr-Faces-HQ Dataset (FFHQ) ».

Kim, Younghyun, Geunmin Hwang, Junyu Zhang, et Eunbyung Park. 2024. « DiffuseHigh : Training-free Progressive High-Resolution Image Synthesis through Structure Guidance ». https://arxiv.org/abs/2406.18459.

Latent Space Cadet. 2022. « Latent Space Cadet – Instagram Account ». https://www.instagram.com/latentspacecadet/.

Luo, Xuan, Xuaner Zhang, Paul Yoo, Ricardo Martin-Brualla, Jason Lawrence, et Steven M. Seitz. 2020. « Time-Travel Rephotography ». *CoRR* abs/2012.12261. https://arxiv.org/abs/2012.12261.

Marker, Chris. 1962. « La Jetée ». Science-Fiction.

McKeag, Brendan. 2023. « Why Altering the Resolution in Stable Diffusion Gives Strange Results ». *RunPod*.

Omelyanchuk, Andrey. 2013. « Eiffel Tower and Champ de Mars in Paris, France ».

Tapper, James. 2024. « "We Need to Come Together" : British Artists Team up to Fight AI Image-generating Software ». *The Guardian*, janvier. Londres.

u/KonoDioNo. 2022. « CCTV Video of Jesus Rising from the Tomb ». *Reddit*. https://www.reddit.com/r/weirddalle/comments/ve53sa/cctv_video_of_jesus_rising_from_the_tomb/.

Wolfram, Stephen. 2023. « What Is ChatGPT Doing … and Why Does It Work ? » https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work.