

Titolo: ***InterArtes***

ISSN 2785-3136

Periodicità: annuale

Anno di creazione: 2021

Editore: Dipartimento di Studi Umanistici – Università IULM - via Carlo Bo 1 - 20143 Milano

Direzione: Laura Brignoli - Silvia T. Zangrandi

Comitato di direzione

Gianni Canova, Mauro Ceruti, Paolo Proietti,
Giovanna Rocca, Vincenzo Trione

Comitato editoriale

Maria Cristina Assumma; Matteo Bittanti;
Mara Logaldo; Stefano Lombardi Vallauri;
Marta Muscariello

Comitato scientifico

Daniele Agiman (Conservatorio Giuseppe Verdi Milano); Maurizio Ascarì (Università di Bologna); Sergio Raúl Arroyo García (Già Direttore Generale del Instituto Nacional de Antropología e Historia); Claude Cazalé Bérard (Université Paris X); Gabor Dobo (Università di Budapest); Felice Gambin (Università di Verona); Maria Teresa Giaveri (Accademia delle Scienze di Torino); Maria Chiara Gnocchi (Università di Bologna); Augusto Guarino (Università L'Orientale di Napoli); Rizwan Kahn (AMU University, Aligarh); Anna Lazzarini (Università di Bergamo); Massimo Lucarelli (Université de Caen); Elisa María Martínez Garrido (Universidad Complutense de Madrid); Luiz Martínez-Falero (Universidad Complutense de Madrid); Donata Meneghelli (Università di Bologna); Giampiero Moretti (Università Orientale di Napoli); Raquel Navarro Castillo (Escuela Nacional de Antropología y Historia, Mexico); Francesco Pigozzo (Università e-campus); Richard Saint-Gelais (Université Laval, Canada); Massimo Scotti (Università di Verona); Chiara Simonigh (Università di Torino); Evangelia Stead (Université Versailles Saint Quentin); Andrea Valle (Università di Torino); Cristina Vignali (Université de Savoie-Mont Blanc); Frank Wagner (Université de Rennes 2); Anna Wegener (Università di Firenze); Haun Saussy (University of Chicago); Susanna Zinato (Università di Verona).

Segreteria di redazione

Caterina Bocchi

INTERARTES n.3

AI: ArtIntelligence

dicembre 2023

Francesco Pigozzo – AI: ArtIntelligence. Créativité artistique et automatisation. Introduction.

ARTICOLI

Daniel Raffini – La ridefinizione dell'autorialità nell'AI novel.

Arnaud Mery - Le regard de l'artiste à l'épreuve des modèles texte-image.

Caterina Bocchi – Insegnare con ChatGPT. Esempi di utilizzo di ChatGPT in laboratori linguistici universitari.

Jacques Demange – IA et cinéma: altérités ludiques et nouveaux enjeux de médialité.

Emmanuelle Stock – Beauté métallique dans la science-fiction: érotisation des corps artificiels féminins dans *L'Eve future* (1886) de Villiers de L'Isle-Adam et dans *No Woman born* (1944) de Catherine Lucille Moore.

SEZIONE SPECIALE

“Lingue e linguaggi per la pace”

Elena Liverani – Introduzione.

Stefano Maria Casella – «Yet here is peace for ever new!»: Henry Beston's “philosophy” of peace.

Eloy Martos Núñez, Aitana Martos García – Imaginarios del agua, crisis ambientales y cultura de la paz.

Eduardo Encabo-Fernández – La comunicación literaria como camino para la cultura de la paz. Una aproximación desde la Didáctica de la lengua y la literatura.

Beatriz Durán González, Estíbaliz Barriga Galeano, Alejandro Del Pino Tortonda – La Paz en formas alternativas de Literatura: los paracosmos de Studio Ghibli y sus aplicaciones didácticas.

VARIA

Michela Spacagno – Parler d'une même voix? Étude sur le discours collectif dans les mystères médiévaux.

RECENSIONI

Florjer Gjepali – Coscienza, coscienza artificiale e inconscio artificiale (Emanuela PIGA BRUNI, *La macchina fragile. L'inconscio artificiale fra letteratura, cinema e televisione*, Carocci, 2023)

Le regard de l'artiste à l'épreuve des modèles texte-image

Arnaud MERY

Université de Montréal

Abstract:

I propose an account of my reflexive practice of text-image models in order to better understand how the withdrawal of artists' conscious involvement induced by these *media* possibly renegotiates their capacities for perception, reflection and action. I begin by pointing out that the desire for representation intrinsic to textual queries – known as *prompts* – clashes with the probabilistic and non-representational operations underlying image generation. I then emphasize the fact that the image generation process informs descriptive choices while at the same time anticipating them, which embeds the artist in co-adjustment loops that I qualify, with Mark Hansen, as *feed-forward loops*. I show that these loops imply a withdrawal of the influence of the supervisory consciousness over the forms produced in favor of what I call, following Thierry Bardini, the *formal agency* of these loops themselves. Finally, linking Shane Denson's techno-phenomenology to my ongoing artistic project, I testify that working with text-image models cannot be reduced to an a-subjective generation of forms. Indeed, the adjustments made during these loops still allow the artist to shape representations. The result is what Denson calls a *dis/correlative* dynamic: a tension between, on the one hand, the artist's ability to become involved in a singular and meaningful point of view that obliterates the algorithmic materiality of the images generated; and, on the other, his or her inability to bring his or her gaze into symbiosis with the image-generating process, leading him or her to become aware only of the algorithmic materiality of the images generated.

Keywords:

Visual culture, Art and artificial intelligence, Human-machine interactions, Research-creation, Techno-phenomenology.

Introduction

Récemment popularisés par des applications comme DALL-E, StableDiffusion ou MidJourney, les modèles texte-image sont des systèmes d'intelligence artificielle (IA) qui génèrent des images à partir de requêtes textuelles en langage naturel nommées *prompts* (synonyme d'invite de commande, en informatique). La plupart de ces modèles relèvent d'une technique d'IA nommée *apprentissage supervisé* (elle-même une sous-catégorie de

l'apprentissage profond – ou *deep-learning*)¹ qui consiste à exposer des architectures de réseaux de neurones artificiels² (ANNs) à un grand nombre d'images étiquetées, c'est-à-dire décrites par de courts textes et mots-clés³. Les modèles apprennent alors à auto-ajuster leurs paramètres de façon à minimiser l'écart entre des étiquettes données et la génération de nouvelles images (Le Cun, 2019). Grâce à ce procédé d'auto-ajustement, l'humain n'a aucunement besoin d'inculquer aux modèles des règles ou des heuristiques (Audry, 2021), comme c'est encore le cas avec les IA dites symboliques⁴ (ce pourquoi l'apprentissage profond est quant à lui qualifié de sub-symbolique). A la suite de cet apprentissage, le modèle peut générer de manière autonome des sorties visuelles qui correspondent selon lui le mieux aux *prompts* entrés par l'utilisateur.

Les modèles texte-image sont en train de s'imposer comme un nouveau *medium*⁵ de production visuelle incontournable. Leur démocratisation croissante peut en partie être attribuée à leur simplicité d'usage, couplée à la promesse commerciale d'un alignement quasi instantané des sorties générées avec ce que l'utilisateur-rice a imaginé en rédigeant un *prompt*⁶.

¹ A distinguer des deux autres sous-catégories de l'apprentissage profond que sont l'apprentissage *non supervisé* et de l'apprentissage *par renforcement* voir, par exemple (Audry, 2021).

² De façon générale, chacun des ANNs d'un modèle se compose, au minimum, d'une couche de neurones d'entrée, de trois couches de neurones dites cachées, et d'une couche de neurones de sortie. Par oppositions aux réseaux neuronaux connectionistes des années Quatre-vingt et Quatre-vingt-dix que l'on qualifie aujourd'hui de *shallow learning* (Audry, 2021), les ANNs de *deep-learning* sont composés de très nombreuses couches.

³ Cet étiquetage étant généralement effectué par des humains, et en notamment par les MTurkers, ou Mechanical Turks, ces employés précaires d'Amazon qui complètent des tâches sur demande, dont l'étiquetage des images qui constituent ensuite des bases de données sur lesquelles s'entraînent des ANNs.

⁴ L'IA dite symbolique (IA historiquement prépondérante) employée pour la production d'œuvres visuelles, fonctionne sur la base d'arbres de décisions hiérarchiquement déployés en ramifications de routines de type « Si » / « Alors » / « Sinon » (Kirschenbaum, 2021: 375). Un exemple emblématique de son déploiement dans le cadre pictural est le programme AARON de Harold Cohen, un système expert essentiellement programmé pour générer des desseins et des peintures originales à partir de connaissances sur des agencements de couleurs et des compositions encodées directement dans la machine par Cohen lui-même. Mais comme la plupart des systèmes experts, AARON est incapable d'acquérir de nouvelles connaissances, car tout ce qu'il « sait » correspond au savoir directement codé par l'artiste (Audry, 2021).

⁵ Le mot *medium* – *media* au pluriel – est mis en italique pour souligner son étymologie latine renvoyant à la fois à des milieux (en l'occurrence des environnements technologiques à base d'algorithmes), des moyens de communication, et enfin des opérateurs de médiation (Bardini, 2016).

⁶ A titre d'exemple, dans un article de son blog, l'artiste Grégory Chatonsky est parti d'un des slogans de l'application Stable Diffusion – «Imagine and just type it» – pour remettre en question la promesse selon laquelle les modèles texte-image peuvent traduire, sans perte, les représentations humaines en texte, puis le texte en images. Texte consultable au lien suivant : <<http://chatonsky.net/imagine-something-and-just-type-it/>>.

Toutefois, si cette simplicité apparente et cette prétendue absence de médiation entre l'imaginaire de l'utilisateur et le résultat obtenu peuvent satisfaire un usage superficiel de ces modèles, elles constituent en revanche des obstacles majeurs pour les artistes qui souhaiteraient expérimenter les potentialités et les enjeux esthétiques de ce *medium*. De plus, d'autres problèmes reconnus et documentés semblent compliquer davantage cette exploration artistique. Tout d'abord, les modèles texte-image produisent souvent des clichés, car ils sont entraînés sur de vastes ensembles de données stéréotypés (Salvaggio, 2023), ce qui peut considérablement limiter la possibilité pour les artistes d'élaborer un regard singulier avec eux. En outre, le fonctionnement technique de la génération d'images n'efface jamais complètement l'écart entre les sorties visuelles désirées par l'utilisateur et les sorties réellement générées (Feyersinger et al., 2023), ce qui rend ces dernières relativement imprévisibles. Enfin, les cheminements complexes par lesquels les modèles passent pour proposer ces sorties visuelles, ainsi que leur vitesse de calcul et les volumes de données traitées⁷, excèdent les aptitudes de la conscience phénoménologique humaine (ce qui amène communément à qualifier les opérations des ANNs de «boîtes noires»). Certes, les artistes sont toujours libres de sélectionner un modèle en particulier ainsi que l'ensemble des données sur lesquelles ce dernier sera entraîné. Néanmoins, l'influence effective de leur conscience de supervision (l'ensemble des délibérations capables d'orienter les choix esthétiques et les représentations vers l'achèvement d'une œuvre) demeure exclue des opérations algorithmiques générant de nouvelles sorties visuelles (Terzidis et al., 2023). L'usage de ces modèles semble donc impliquer une remise en question profonde du poids de l'intentionnalité de projet des artistes et, de fait, de leur autorité sur la constitution d'une œuvre.

La plupart des publications académiques en sciences humaines ne manquent pas de reconnaître le niveau d'autonomie sans précédent des modèles génératifs, suggérant même l'émergence d'un éventuel partenariat co-créatif avec les artistes. Malgré cela, beaucoup de ces publications tendent à ramener la génération d'image à un ensemble d'outils algorithmiques

⁷ Pour donner une idée de l'ordre de grandeur, l'architecture de StableDiffusion déploie environ 890 millions d'interconnexions (Russell, 2022) et le modèle a été initialement entraîné à partir d'une base de données constituée de 5,85 milliards de paires image-texte (Schuhmann et al., 2022).

qui ne seraient que les auxiliaires de finalités proprement humaines (Hertzmann, 2018; Manovich, 2023; Mazzone et Elgammal, 2019; McCormack et al., 2019). Sans pour autant tomber dans des positionnements explicitement dualistes, ces approches reconduisent malgré tout l'idée d'une maîtrise anthropologique de cette technologie.

Or, les témoignages et publications de certains artistes familier·ère·s avec les techniques d'IA génératives vont à l'encontre d'un tel repli instrumental. Ils·elles attestent, au contraire, d'une disjonction entre leurs attentes (ou leur relatif désir de maîtrise, ou tout du moins d'influence décisive, dans la constitution d'une œuvre) et les opérations de génération d'images.

Par exemple, dans la section de son ouvrage *Art in the Age of Machine Learning* (2021) consacré aux artistes faisant usage de modèles à base d'apprentissage profond, l'artiste-chercheur Sofian Audry insiste sur le fait que les artistes peuvent partiellement contraindre, mais non véritablement contrôler, les voies par lesquelles ces modèles vont générer des images, celles-ci étant en dernière instance tributaires des processus d'apprentissage et d'inférence des modèles eux-mêmes. Il précise:

By design, the decision boundaries of machine learning devices are closer to a form of intuition than to a rational, logical, explainable construct. This may seem paradoxical because on one hand machine learning systems are known to be good at making predictions, but on the other hand it remains difficult for a human being to predict the decisions of such learning processes. (Audry, 2021: 59)

Ce constat est notamment partagé par de nombreux artistes qui, comme le note le rapport intitulé «AI and the Arts: How Machine Learning is Changing Artistic Work» (Ploin et al., 2022), apprécient cependant l'habileté singulière des ANNs à générer du nouveau et du surprenant. Dans les entretiens effectués lors de cette étude, ils·elles s'accordent à dire que l'usage artistique des ANNs implique de trouver un équilibre entre une forme d'excitation à la vue de résultats inattendus et une frustration due à un contrôle limité de l'artiste sur ce *medium*. Adoptant une perspective plus radicale, l'artiste Trevor Paglen considère, quant à lui, que la production d'images avec des dispositifs à base d'apprentissage profond contraint notre subjectivité elle-même à se reconfigurer. Ce qu'il nomme les *images invisibles* (2019), renvoient à celles principalement produites ou traitées par des IA à destination d'autres machines, sans nécessiter d'intervention humaine au cours de ce procès. Paglen retourne le

rapport de pouvoir qui assigne habituellement les technologies de vision à des instruments que l'humain modèle pour augmenter ses capacités d'analyse sur le monde. Pour lui, une nouvelle culture visuelle issue de processus qui échappent à notre perception visuelle est en train d'émerger. Elle nous force à reconsidérer ce qui constitue le cadre de notre vision au-delà de nos yeux comme référents, afin de prendre conscience d'un environnement rempli d'éléments tels que «activations, keypoints, eigenfaces, feature transforms, classifiers, training sets, and the like» (Paglen, 2019: 27).

Bien que ces artistes ne fassent pas à strictement parler usage des modèles texte-image, ils-elles m'amènent néanmoins à suggérer que la réalisation d'une œuvre visuelle avec des logiciels à base d'apprentissage profond exige une redistribution de l'agentivité créatrice. Cela signifie que l'idée selon laquelle les artistes seraient le centre névralgique dont la constitution des œuvres dépendrait se retrouve sérieusement compromise. Partant de ce constat, que peut dès lors signifier «faire usage des modèles texte-image»? De quelle façon l'intentionnalité de projet des artistes peut-elle entrer en relation avec un processus génératif qui lui est en grande partie indifférent, et dont les opérations lui sont inaccessibles? Ou, pour formuler plus concrètement ma question de départ: comment l'agentivité des artistes se manifeste-t-elle et opère-t-elle durant la réalisation d'œuvres visuelles avec des modèles texte-image, étant donné la disjonction entre les sorties visuelles désirées par leurs requêtes textuelles, et les sorties visuelles issues des opérations sub-perceptuelles des modèles? Le but de mon étude étant de mieux comprendre en quoi la mise à l'écart de la conscience de supervision des artistes, induite par l'usage de ce *medium*, renégocie éventuellement leurs capacités de perception, de réflexion et d'action.

Ma lentille analytique m'est fournie par les approches techno-phénoménologiques et médiatiques développées par les théoriciens des *media* Thierry Bardini (2016), Shane Denson (2014, 2020, 2023) et Mark Hansen (2015). Celles-ci offrent en effet les moyens d'appréhender conceptuellement les renégociations subjectives en jeu lors de l'usage artistique des modèles texte-image, en se focalisant sur les relations co-constitutives entre la perception humaine et

les opérations sub-perceptuelles des modèles. Pour Hansen, les *media* algorithmiques à base d'apprentissage machine interviennent sur *l'infrastructure causale* (Hansen, 2015: 48) de nos expériences. Nos facultés conscientes et sensorielles sont alors imbriquées au sein de ce qu'il nomme des *bouclages projectifs* (*feed-forward loops*). Ces *bouclages* extraient des données provenant de l'activité humaine, à partir desquelles ils prédisent de nouvelles données qui vont en retour informer, tout en les anticipant, les comportements humains futurs. Cette notion de *bouclages projectifs*, combinée à celle de *cause formelle* telle qu'elle est relatée par Bardini (2016), permet de dégager l'idée d'une *agentivité formelle* mettant en place une récursivité entre la subjectivité des artistes et les images générées par les modèles. La contribution de Denson, enfin, ouvre à l'idée que ces *bouclages* sont en train de nous faire transiter vers une nouvelle forme de subjectivité. Focalisant entre autres ses analyses sur des images générées par apprentissage profond (Denson, 2023: 106-113), il conceptualise une dynamique de *corrélacion / décorrélacion* entre les facultés humaines qui font sens des images générées, et les processus autonomes qui génèrent ces images. Transposé dans les termes de mon projet, cela se traduit par une dynamique au cours de laquelle l'intentionnalité de projet des artistes tente d'entrer en symbiose avec un processus de génération d'image qui, lui, résiste à l'affirmation d'un point de vue singulier.

Ces balises conceptuelles étant posées, toujours est-il que ces auteurs s'en tiennent à une compréhension théorique des processus de médiations qu'ils décrivent, sans pour autant altérer ni reconfigurer par la pratique leur propre habileté à saisir ces dynamiques. Dans cette perspective, la démarche de recherche que je propose consiste à mettre leur approche théorique à l'épreuve d'une *création en tant que recherche* (Chapman, Sawchuk, 2012). Cette méthodologie de *recherche-création* permet ainsi d'éprouver ce que le philosophe des sciences Andrew Pickering (1995) nomme une *danse d'agence*, entre l'engagement subjectif de l'artiste-chercheur-se et les opérations des modèles texte-image. L'analyse des enchevêtrements entre ma subjectivité et l'appareillage des modèles donnera ensuite matière à des entretiens avec d'autres artistes portant sur ce niveau d'expérience.

Le présent article rend compte de la première phase de ma recherche, qui en comporte deux. La première consiste en une composante création (Paquin et Noury, 2018), qui mettra à l'épreuve les enjeux techno-phénoménologiques et médiatiques rencontrés chez Denson, Hansen et Bardini, en prenant appui sur ma propre exploration pratique des modèles texte-image. Il s'agit alors de rendre compte de l'altération et de la reconfiguration de ma subjectivité au contact de ces modèles. La seconde phase consiste en une composante recherche (Paquin et Noury, 2018) qui prendra la forme d'un dialogue soutenu avec un échantillon d'artistes, structuré par une série d'entretiens semi-dirigés centrés sur les points problématisés lors de la composante création. Dans la suite de cet article, j'articule uniquement les premiers éléments de la composante création, en me centrant sur la démarche expérimentale auto-réflexive nécessaire à ma problématisation.

Je propose d'amorcer ma réflexion en identifiant plus précisément le problème initial de ma recherche : le découplage entre les sorties visuelles recherchées par l'artiste et les sorties réellement générées par les modèles. Ce problème va devenir le moteur d'une expérimentation artistique permettant d'affiner mon investigation (Baxter et al., 2008). À partir de mon travail intitulé *An Interview With Onomastika Thanatosis* (2022), je commence par mettre en avant le fait que le désir de représentation intrinsèque aux *prompts* se heurte aux opérations probabilistes et non-représentationnelles qui sous-tendent la génération d'image. Je souligne ensuite le fait que le processus de génération informe mes choix descriptifs tout en les anticipant, ce qui, comme je le montre avec mon expérimentation *Mutant Photograph* (2022), m'imbrique dans des boucles d'ajustements mutuels avec les modèles. M'appuyant sur les travaux de Bardini, j'avance que ces boucles font émerger une *agentivité formelle*, notion que je relie à celle de *bouclages projectifs* chez Hansen. Puis, en articulant les notions de *corrélation* et de *décorrélation* chez Denson avec un nouveau projet intitulé *Disociascope* (2023), je témoigne qu'une transition paradigmatique du regard de l'artiste est en jeu dans ces *bouclages*. Je conclus que la pratique de ces modèles est le lieu d'une redistribution en cours du regard artistique, au sens où les œuvres se font jour au cœur d'une tension entre une intentionnalité de projet – qui mobilise la conscience de supervision de l'artiste et son désir d'affirmer un point

de vue singulier – et une intentionnalité secondaire de sélection – qui ne nécessite pas de conscience de supervision, et qui émerge de la récursivité formelle des *bouclages projectifs*.

Un désir de représentation confronté aux ressemblances de l'espace latent

A bien des égards, la rédaction d'un *prompt* reconduit le désir de corrélation texte-image qui a façonné l'art pictural Occidental depuis ses débuts (Mitchell, 2005: 258). En atteste par exemple le *DALL·E 2 Prompt Book*⁸, un manuel aidant les usager·ère·s de cette application à rédiger les *prompts* les plus adaptés à ce que *DALL·E 2* pourra générer en termes de style pictural, de type de focale optique, de mise au point sur un objet en particulier, d'intensité de la lumière, ou encore de distributions des couleur etc. La relation autoréférentielle attendue entre la sortie visuelle et l'entrée textuelle fait écho à ce que Mitchell qualifie d'*espoir ekphrastique*, c'est-à-dire «the hope that a verbal description could succeed in bringing forth an image with perfect representational clarity» (Bolter, 2023: 204); ou encore, à ce que la philosophe José Marie Mondzain a qualifié d'apparition des images sur le mode de l'incorporation: à savoir, un régime d'expérience de l'image qui suppose l'absence de distance entre les signaux visuels perçus par les organes sensoriels et le geste par lequel le sujet encode et dénote des figures signifiantes au sein de cette visibilité (ce qui pour Mondzain définit l'idolâtrie (Mondzain, 2002)). Cela étant, le *prompt* n'est qu'un moment de la procédure générant des images, non un mot d'ordre qui commanderait une représentation visuelle à *son image*. En effet, le modèle n'opère pas une traduction visuelle directe du *prompt*, mais cherche des relations de ressemblances probables entre les entrées textuelles de l'usager·ère et des caractéristiques visuelles élémentaires extraites des images étiquetées sur lesquelles il a été entraîné. Pour bien saisir ce point, il faut introduire une notion importante : *l'espace latent* du modèle.

Etant donné que les algorithmes de chaque application divergent dans leur fonctionnement technique, ce que je synthétise dans les lignes suivantes doit être compris

⁸ URL: <<https://dallery.gallery/the-dalle-2-prompt-book/>>.

comme une simplification essayant de dégager leurs principes communs⁹. Considérons une base de données d'images étiquetées. Durant la phase d'entraînement, des ANNs vont extraire des motifs visuels à partir des images auxquelles ils sont exposés. L'encodage de ces caractéristiques visuelles transforme ces images en données dites latentes, au sens où elles s'intègrent à un espace mathématique qui va les lier à d'autres encodages par des probabilités d'être rapprochées (ou éloignées) d'eux (Audry, 2021: 104-108). Parallèlement, d'autres ANNs encodent également en données latentes les suites de symboles – ou *tokens*, pour reprendre le terme d'analyse lexicale – des étiquettes. Au cours de l'apprentissage, chaque encodage d'image et d'étiquette s'ajoute aux précédents, constituant une « carte » multidimensionnelle de leurs rapprochements/éloignements potentiels qui se nomme l'*espace latent* (Figure 1).

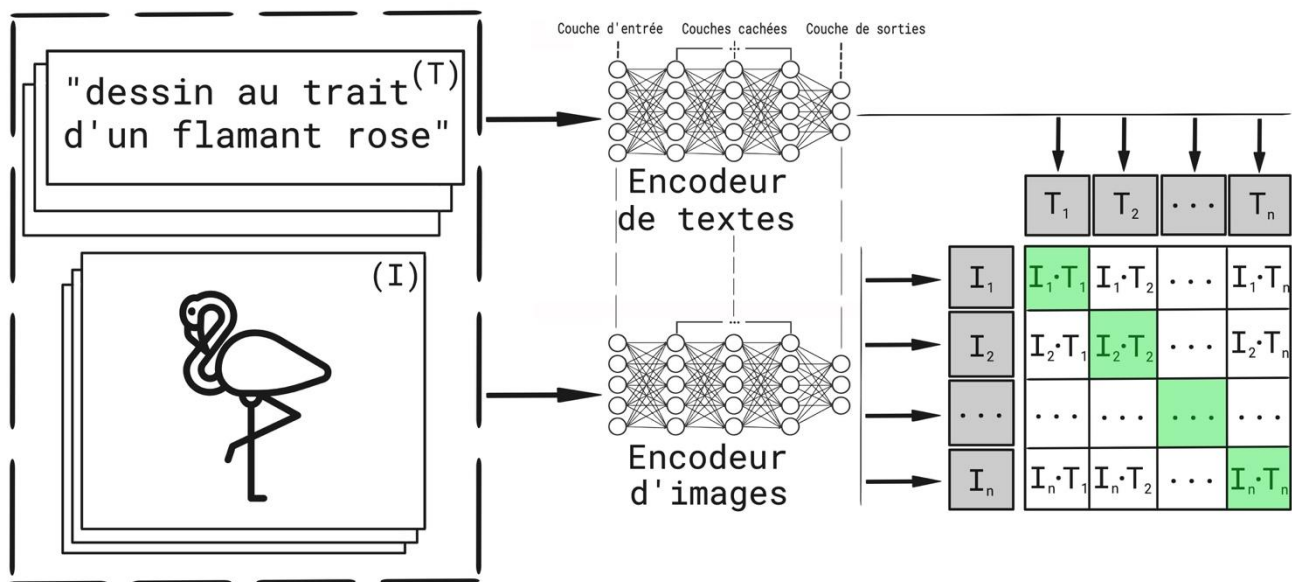


Figure 1. Illustration de l'entraînement du modèle CLIP (Contrastive Language-Image Pre-training), développé par la compagnie OpenAI. Exposé à des paires texte-image, le modèle encode les motifs visuels extraits des images ainsi que les suites de *tokens* des étiquettes pour constituer un espace latent à partir duquel il est capable d'opérer une matrice de similarité. En vert: les prédictions de rapprochements texte-image que CLIP considère comme les plus probables. Adapté de (Radford et al., 2021).

⁹ Par exemple, contrairement à la structure d'entraînement antagoniste du GAN (où un générateur tente de générer des données de façon à atteindre une valeur validée par le discriminateur), StableDiffusion est entraîné en simulant un processus de *diffusion* au cours duquel un filtre de bruit est progressivement appliqué sur l'encodage des images. Le modèle apprend alors à reconstruire l'image en la débruitant.

Une fois la phase d'apprentissage achevée, le modèle peut alors encoder un premier *prompt* entré par l'utilisateur. En un sens, la machine va interroger son espace latent afin de rapprocher la valeur numérique de ce *prompt*, de la valeur numérique d'une distribution de motifs visuels la plus probable de correspondre. Un décodeur permet ensuite de convertir cette inférence statistique en une image visible par l'utilisateur.

Contrairement à des IA symboliques, les modèles texte-image ne génèrent pas de sorties visuelles qui suivraient des règles de composition picturales prédéfinies par des programmeurs, mais trouvent des rapprochements probables entre des fragments de données visuelles et textuelles encodées dans l'espace latent. Cela fait écho à la différence que souligne le philosophe Nelson Goodman entre une image relevant d'une ressemblance et une image relevant d'une représentation (Mitchell, 1986). Là où la représentation opère un référencement symbolique qui assigne l'image à une dénotation (A désigne B), la ressemblance opérable via l'espace latent permet de se détacher d'une compréhension iconographique humaine, pour fonctionner comme une conversion mutuelle d'un ensemble de données en un autre ($A \simeq B$). Un de mes premiers projets intitulé *An Interview with Onomastika Thanatosis – AIWOT* – illustre cette différence. Je suis parti de numérisations de photographies argentiques issues d'un ancien projet personnel, que j'ai importées dans un modèle multimodal du nom de VQGAN+CLIP¹⁰, capable d'encoder chacune de ces images comme une entrée à partir de laquelle d'autres images vont être générées. Une fois un *prompt* entré, le modèle lance une génération itérative de nouvelles images. Au bout d'un certain temps, j'arrête le calcul puis je sélectionne certaines sorties dans l'ordre chronologique de leur génération. J'importe ces sorties dans un modèle d'*interpolation de trame*, dont les ANNs sont capables de générer de nouvelles images entre les images existantes de façon à produire une animation faisant imperceptiblement muter une image en la suivante. Pour ce projet, je suis parti de *prompts* jouant autour du texte «body

¹⁰ Une combinaison de deux modèles – Vector Quantized Generative Adversarial Network (VQGAN) et Contrastive Language-Image Pre-training (CLIP) – où VQGAN génère des images basées sur des codes visuels et CLIP, étant entraîné sur des paires texte-image, fournit une évaluation *rétroactive* à VQGAN afin d'affiner sa génération d'image dans une direction plus proche des descriptions textuelles.

without organs» induisant des résultats à la fois très organiques et surprenants (car réagénant des motifs visuels à partir d'une invite qui ne correspond sans doute pas à des étiquetages très courants de la base de données d'entraînement). Chaque interpolation de trame de *AIWOT* fait donc dériver mes photographies initiales vers des figures organiques mutagènes qui rendent les images de départ méconnaissables (*Figure 2 et 3*).

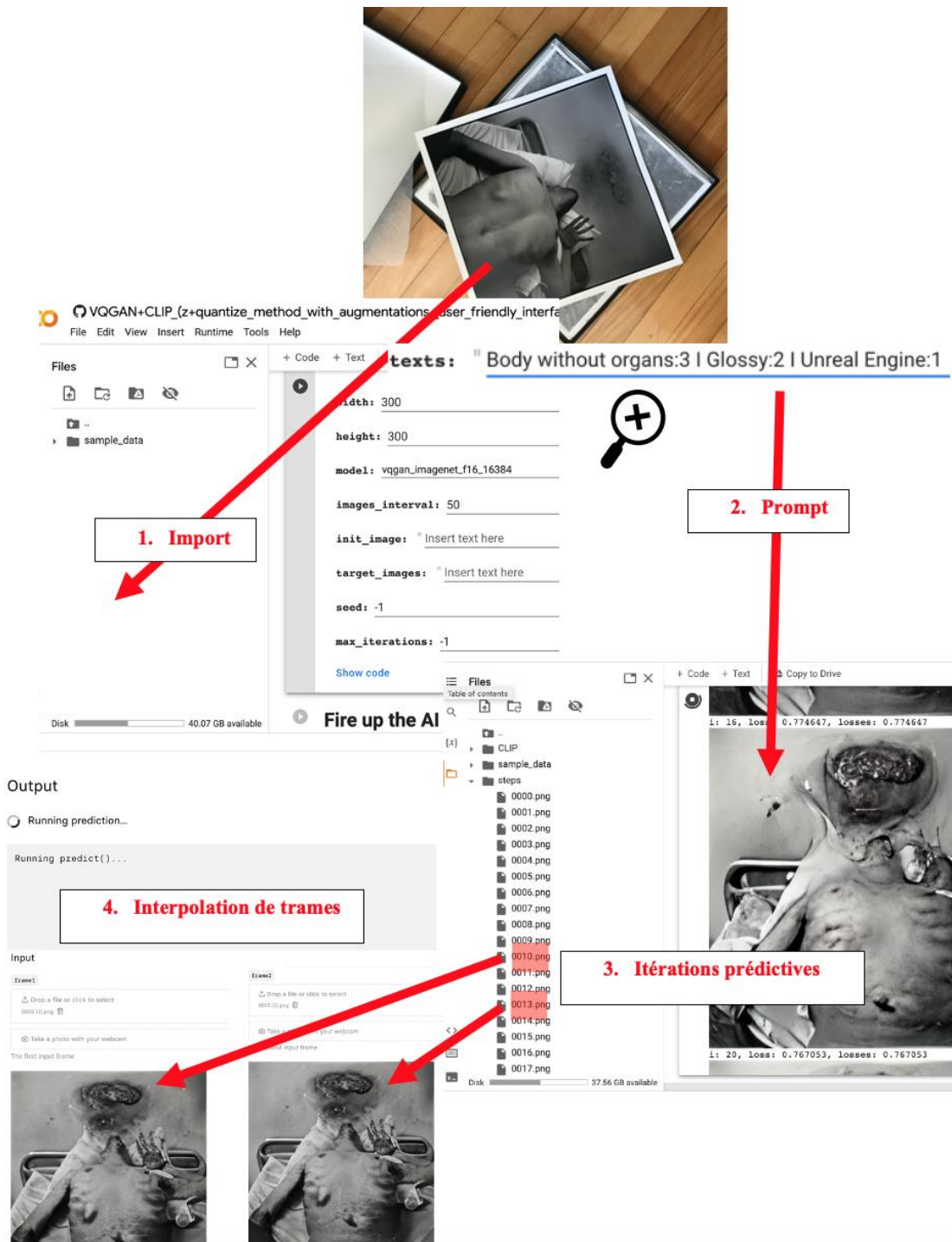


Figure 2. Processus de réalisation de *An interview with Onomastika Thanatosis*.

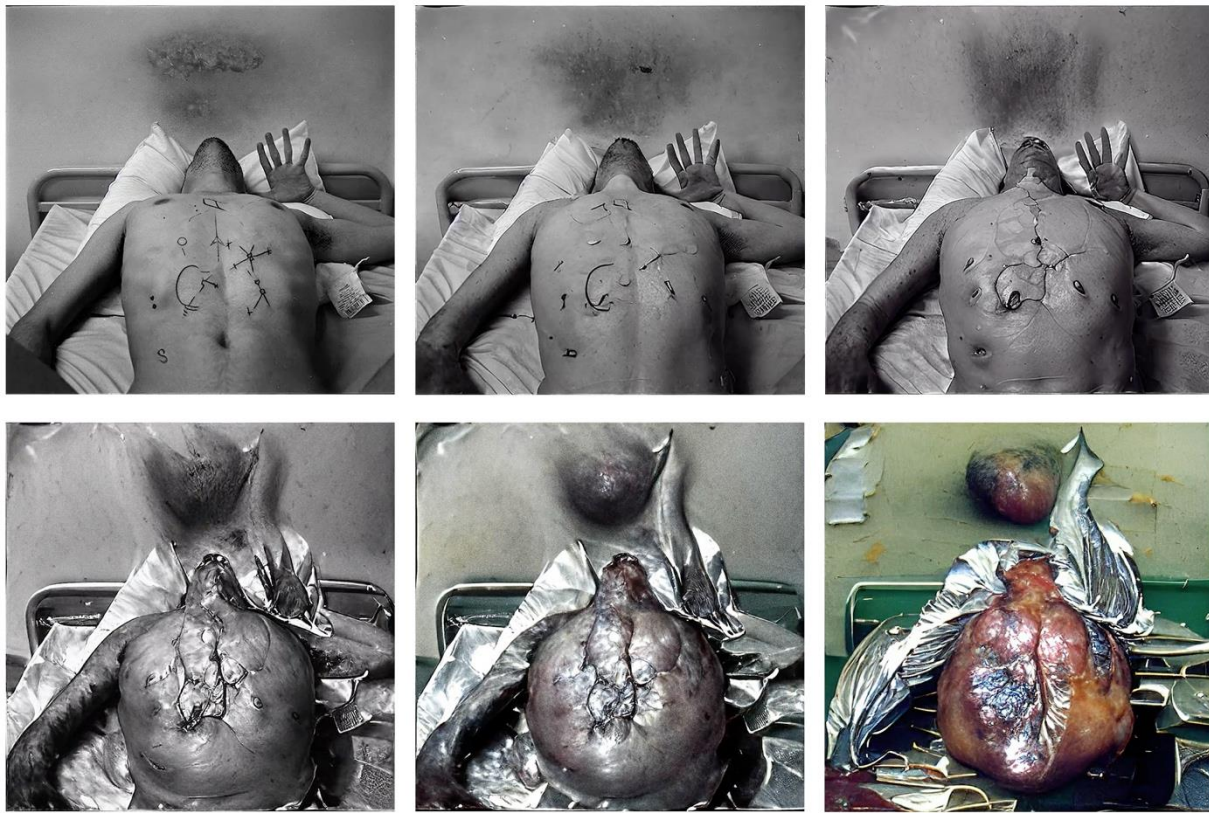


Figure 3. Aperçu de l'effet de *morphing* issu de l'interpolation de trame des sorties successives dans *An interview with Onomastika Thanatosis*.

On remarque alors aisément combien la conversion d'une image en une autre se fait dans une fluidité sans aucun égard pour ce que chaque image générée par VQGAN+CLIP, et a fortiori chaque photographie initiale, peut bien représenter. Bien que quelque chose de l'ordre d'un regroupement sémantique puisse caractériser la distribution des données dans l'espace latent (Bajohr, 2023), le caractère anhistorique de ce regroupement, indifférent à tout ancrage socioculturel, en fait ce que l'artiste et chercheuse Amanda Wasielowski nomme un *formalisme informatique* (Wasielowski, 2023a). Celui-ci n'identifie pas un encodage de concepts, mais comble des écarts entre des valeurs numériques, ce qui n'a pas directement vocation à s'inscrire dans une structure symboliquement signifiante. Cela explique notamment pourquoi de nombreuses incohérences visuelles – souvent qualifiées d'«hallucinations» – continuent de miner la prétention représentationnelle du *prompt*. Comme l'affirme Wasielowski: «DALL-E

and its ilk are able to replicate visual forms but are not 'aware' of or 'familiar' with the referents in the images they produce» (Wasielewski, 2023b: 78).

Naviguer dans l'espace latent: co-ajustements et bouclages projectifs

Ce découplage entre ma représentation textuelle des sorties désirées et les sorties effectivement générées a pour effet de court-circuiter l'influence de ma conscience de supervision. Peu importe le projet artistique que j'entreprends consciemment, mes facultés perceptives et cognitives ne sont en réalité sollicitées qu'après coup, une fois que les sorties sont générées. Toutefois, ce retrait apparent de mon autorité sur l'œuvre me libère d'une intentionnalité de projet pour m'inviter plutôt à un geste improvisateur, au cours duquel prévaut une intentionnalité de sélection spontanée vis-à-vis des sorties visuelles. Par exemple, dans mon expérimentation intitulée *Mutant Photograph*, je suis parti d'une photographie que j'ai prise avec une caméra numérique pour me lancer ensuite, avec l'application DALL·E 2, dans un chaînage continu de rédactions de *prompts* suivies de sélections quasi aléatoires et spontanée de sorties proposées, contribuant ainsi à l'expansion progressive de l'image jusqu'à ce que j'y mette arbitrairement un point d'arrêt. L'image finale n'est pas la résultante d'une somme de représentations issues d'actes de supervisions délibératifs de ma part. Elle est la trace d'une série de réactions sensori-motrices aux sorties visuelles générées (*Figure 4 et 5*).

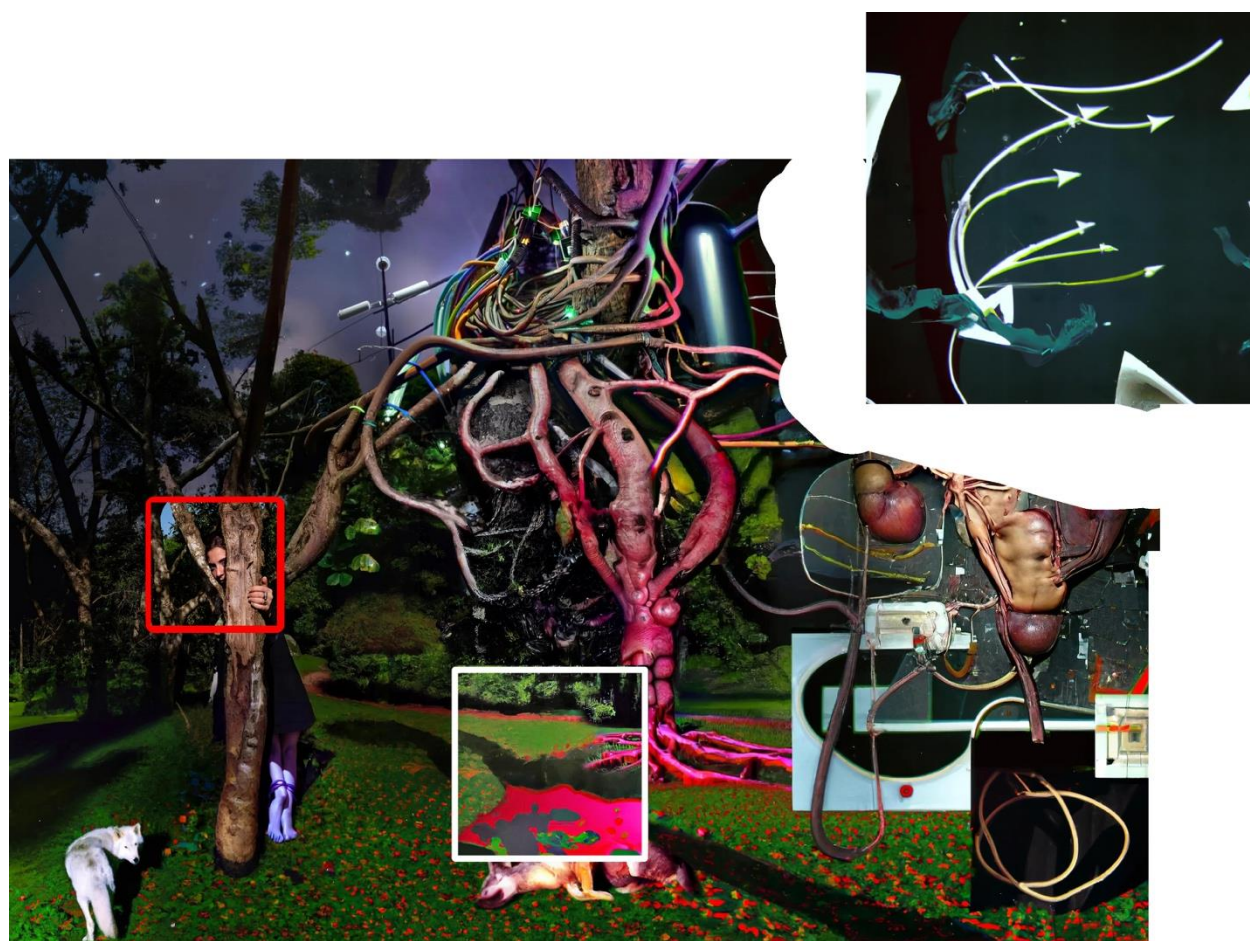


Figure 4. Mutant Photograph: aperçu du processus d'amplification improvisée de l'image avec DALL-E 2; dans le cadre rouge : la photographie de départ; dans le cadre blanc: un exemple de sortie proposée par le modèle; en haut à droite: un exemple de prédiction de l'ANN qui sera rattaché au reste de l'image.



Figure 5. *Mutant Photograph*: image finale.

Tout en produisant un effet de continuité et d'homogénéité de l'image, la génération des sorties peut transiter de façon inconsistante d'un élément visuel à un autre, ou même d'un style pictural à un autre, sans qu'un projet de représentation final soit nécessaire pour organiser une cohérence formelle. Cela se vérifie clairement dans *Mutant Photograph*, où des motifs visuels hétérogènes transitent sans discontinuités de gauche à droite, manifestant la direction que le processus d'improvisation a pris – processus au cours duquel les *prompts* successifs se mêlent à une sorte d'errance formelle. Les requêtes que j'ai lancées prenaient souvent des tournures abstraites que je ne prenais plus le temps de réfléchir, me hasardant à provoquer des simulations de textures ou de couleurs surprenantes que je décide ensuite d'altérer spontanément. Dans ces conditions, les *prompts* agissent comme des coups de gouvernail à l'aveugle au cœur de l'espace latent qui, de son côté, dissout et recycle le texte de mes requêtes dans des calculs probabilistes desquelles ma conscience phénoménale est exclue.

Le processus par lequel les *tokens* du *prompt* sont mis en corrélation avec les données visuelles de l'espace latent, puis transduits en sorties visuelles par le modèle, sorties qui à leur tour sont prises en compte par l'artiste pour rédiger le *prompt* suivant en fonction de l'écart perçu vis-à-vis de la requête initiale, dessine un bouclage auto-amplificateur au fil duquel se matérialise l'œuvre. Dans *Mutant Photograph*, chaque nouveau *prompt* que je rédige spontanément est influencé par les sorties visuelles, d'une façon qui rend le résultat accumulé de plus en plus étrange et re(ou dé-)formée, dérivant d'un style à l'autre. Ma pratique du *prompt* se déploie alors systématiquement autour d'un même mode opératoire caractérisé par des boucles d'ajustements mutuels entre mes descriptions textuelles et les prédictions des modèles¹¹. Ce mode opératoire change ainsi le statut même de l'invite textuelle: celle-ci n'a plus pour vocation de représenter quelque-chose (vocation nécessairement trahie par la génération d'image, comme on l'a vu plus haut), mais de performer¹² un cheminement au sein de l'espace latent, sans finalité apparente. L'espace latent devient donc le champ des possibles auquel je dois m'accorder sur un mode performatif, puisque toutes les images à venir dont je pourrais déclencher la génération via des *prompts* sont déjà en lui à l'état de potentialités, et mineront à chaque itération mon désir de représentation. Autrement dit, tout désir de l'artiste d'agencer certaines formes ou de représenter certaines figures en vue d'un résultat final sera continuellement remis en jeu, ou re-potentialisé, au fil des interactions itératives entre l'artiste et les modèles.

La traditionnelle causalité linéaire de cause à effet amène à penser la médiation artistique en tant qu'inscription technique de l'expérience humaine (Hansen, 2015). Mais cet acte de représentation dirigé par l'artiste se retrouve relativisé, si ce n'est nié, par les boucles de

¹¹ Ce processus de rectification itératif des invites textuelles est ce que l'on nomme communément *prompt engineering* ou encore *ingénierie rapide*.

¹² Au sens où Andrew Pickering l'entend, l'*idiome performatif* suggère une réalité active, en perpétuelle évolution, constamment façonnée et remodelée par les interactions d'agents humains et non humains, contrairement à l'*idiome représentationnel* qui renvoie à la conception traditionnelle de la réalité comme étant représentable par la connaissance (Pickering, 1995). Le terme «idiome» faisant référence au cadre à travers lequel la réalité peut émerger et être vécue par des agents, ce qui a notamment amené le chercheur Daniel Eacho à qualifier de *metaphysical performativity* le paradigme défendu par Pickering (Eacho, 2023).

co-ajustements entre artistes et modèles. Ces boucles mettent plutôt en place la «circularité cause/effet/cause» de la «cause formelle» (Bardini, 2016: 163)¹³: celle de l'espace latent et des sorties visuelles générées qui, paradoxalement, précèdent et informent les descriptions textuelles futures. Autrement dit, les boucles d'ajustements mutuels précédemment décrites sont la structure d'expérience, ou la forme d'usage, que les modèles texte-image habilitent. Et cette forme d'usage a sa propre *agentivité formelle*, découplée de l'*agentivité efficiente*¹⁴ de l'artiste en ce qu'elle contribue fortement à styliser le contenu de l'image et influencer les réorientations descriptives des *prompts* sans que l'artiste puisse prétendre maîtriser ce processus émergent.

Cette redistribution causale peut être appréhendée à l'aide de la notion de *feed-forward loops* – ou *bouclages projectifs* – élaborée par le théoricien Mark B.N. Hansen (2015). Cette notion renvoie à la manière dont un *medium* algorithmique opérant des prédictions de façon autonome inscrit son usager·ère au sein d'une structure d'expérience qui repose sur les données que le *medium* a collecté et traité en amont, puis offert à l'usager·ère (Hansen, 2015). En d'autres termes, les opérations d'apprentissage d'un modèle texte-image à partir de bases de données, la constitution de son espace latent, ainsi que les inférences statistiques générant de futures sorties visuelles, excèdent, précèdent et informent les choix esthétiques que l'artiste peut faire lorsqu'il rédige les *prompts*, ainsi que la conscience que l'artiste peut avoir de l'œuvre en train d'émerger. En comparaison, nos relations avec les *media* visuels précédents tels que la peinture, la photographie ou le cinéma, permettaient d'élaborer des représentations en vue d'un résultat final. Cela passait par des techniques de cadrages de la matière visuelle collectée par les facultés perceptives de l'artiste lui-elle-même (ou de ses prothèses oculaires, comme la caméra). En revanche, dans le cas des *bouclages projectifs* qui structurent la réalisation d'une œuvre

¹³ Bardini réfère ci aux quatre types de causes décrites par Aristote, reprises par Heidegger (1958) et McLuhan (Anton, Logan et Strate, 2017).

¹⁴ Bardini définit l'*agentivité efficiente*, dite aussi motrice, comme «ce qui fait passer une chose à la présence (du non-être à l'être) du fait de ce qui contrôle le mouvement de sa production»: ici l'agent, c'est-à-dire l'artiste. Il définit l'*agentivité formelle* comme «ce qui fait passer une chose à la présence du fait de sa participation à sa forme, c'est-à-dire ce qui la produit comme figure vis-à-vis d'un fond»: ici les boucles de co-ajustements entre les *prompts* et les rapprochements probabilistes de l'espace latent (Bardini, 2016).

visuelle avec des modèles texte-image, ceux-ci mettent l'artiste en relation avec la quasi-exclusion de son intentionnalité de projet (renvoyant aux causalités efficiente et finale qui, depuis Aristote, était communément admises comme motrices des œuvres). Ceci se traduit par une diminution drastique de l'*agentivité efficiente* de l'artiste au profit de l'*agentivité formelle* de ces *bouclages*, liée au traitement algorithmique autonome des données par les modèles. La conscience de supervision de l'artiste est ainsi court-circuitée par une générativité spéculative pouvant fonctionner sans finalité.

Dissolution ou réaffirmation du regard ?

Ceci étant dit, peut-on faire œuvre si les images ne sont que les fruits d'une auto-amplification formelle sans réflexivité, rendant l'artiste incapable d'agencer une représentation, ou d'affirmer un regard qui serait personnel? Les manières dont les sorties visuelles des modèles peuvent être recadrées, leur contenu effacé, modifié et complété via des descriptions textuelles, ne sont-elles pas déjà des façons pour l'artiste d'exercer un pouvoir qui tranche dans la continuité performative des sorties visuelles et des itérations qui les génèrent? Les *bouclages projectifs* ne mettent-ils uniquement en place qu'une récursivité autoalimentée par les sorties visuelles, excluant de fait toute nécessité d'une intervention de la conscience de supervision de l'artiste dans le processus de réalisation d'une œuvre? Ou bien la génération d'image devient-elle, au cours de ces *bouclages*, un substrat pour l'émergence d'une nouvelle forme d'engagement conscient de l'artiste? Comme le note le chercheur en études médiatiques Daniel Eacho, ce n'est pas tant «the animacy of automatic productivity» (Eacho, 2023: 31) qui fait œuvre, que la résurgence d'une différence qui implique la subjectivité de l'artiste en tant qu'*agentivité* réflexive et capable de représentation. Même dans une tentative d'improvisation comme *Mutant Photograph*, il serait forcé de dire qu'il s'agit d'une subordination passive au *formalisme informatique* d'un espace latent dénué d'horizon signifiant.

A ce titre, il est important de compléter la définition qu'Hansen donne de la notion de *bouclage projectifs* – intentionnellement tronquée jusqu'ici – qui est avant tout un moyen

d'identifier un nouveau type de déploiement de la conscience et de sa puissance d'agir au-delà des limites immédiatement perceptibles par le corps humain et ses prothèses technologiques pré-algorithmiques. Pour Hansen, ces bouclages ne décrivent pas seulement le niveau sensori-moteur des interactions entre les opérations des modèles et les réactions de l'artiste, mais également le niveau d'une interprétation par la conscience, a-posteriori, des résultats perceptibles issus de ce niveau primordial d'interaction (Hansen, 2015). Pour qu'une œuvre comme *Mutant Photograph* se concrétise, il aura fallu qu'un niveau d'expérience d'ordre supérieur émerge au fil de mes réactions aux sorties visuelles générées, sans lequel je n'aurais pas pu sciemment faire l'expérience de ces *bouclages* ni expérimenter de façon réflexive les conséquences de mon imbrication avec eux. En d'autres termes, les *bouclages projectifs* sont aussi les médiateurs d'une expérience de la différence anthropomorphique et non pas uniquement d'une dérive au gré des spéculations opaques et imprévisibles d'une boîte noire.

S'il est intenable de simplement considérer les modèles texte-image comme des médiateurs de représentations désirées par l'utilisateur, il semble également insatisfaisant de limiter la danse d'agence entre l'utilisateur et les modèles à la continuité d'une performativité sans intentionnalité de projet. Ni l'affirmation d'une souveraineté du sujet par les *prompts*, ni une dissolution complète de la conscience de supervision face à l'*agentivité formelle* des *bouclages projectifs* ne correspondent au régime d'expérience d'un faire-œuvre avec les applications texte-image. A ce stade de l'enquête, mon interrogation de départ doit donc être affinée en reprenant en compte ce qu'il advient de l'artiste en tant que sujet au cours de l'usage des modèles texte-image. Pour reprendre la formulation de Pickering considérant les enchevêtrements performatifs entre humains et technologies du point de vue de l'humain, la danse d'agence entre artistes et modèles semble donc prendre, du point de vue des artistes:

the form of a dialectic of resistance and accommodation, where resistance denotes the failure to achieve an intended capture of agency in practice, and accommodation an active human strategy of response to resistance, which can include revisions to goals and intentions as well as to the material form of the machine in question and to the human frame of gestures and social relations that surround it. (Pickering, 1995: 22)

Je souhaite donc terminer cette contribution en abordant la question de la ré-émergence de l'influence consciente de l'artiste. Une façon plus précise d'aborder cette question est de poser celle de la reconstitution du regard de l'artiste, entendu au sens de sa capacité à (re-)cadrer, modifier et compléter les sorties visuelles des modèles de façon à agencer des représentations malgré la résistance exercée par l'*agentivité formelle* des *bouclages projectifs*.

La réaffectation techno-phénoménologique du regard de l'artiste

Je propose de donner corps à cette interrogation en m'appuyant sur l'approche techno-phénoménologique du théoricien des *media* Shane Denson, puis finir en l'articulant avec un nouveau projet intitulé *Disociascope* (2023). Il me faut tout d'abord évoquer brièvement des aspects clés que je retiens des idées singulières de Denson afin d'expliquer pourquoi elles m'ont permis de concevoir ce projet en lien avec ma problématique initiale. Etudiant la manière dont les opérations du *medium* cinématographique affectent l'expérience incarnée, Denson problématise une contradiction entre ce qu'il qualifie comme un régime d'expérience *corrélatif* et un régime d'expérience *décorrélatif*. C'est-à-dire, une divergence entre ce que le philosophe Don Ihde décrit comme une «symbiosis that channels perception toward the objects of our visual attention» (Denson, 2020: 29), et le fait que la matérialité technologique qui sous-tend la production et la projection d'images prend de court, voire se substitue, aux automatismes perceptifs par lesquels le sujet incarne habituellement un sens visuel de l'espace (Denson, 2020). Il met en avant combien le flux machinal d'images impersonnelles opéré par la machine-cinéma accuse une absence de points d'ancrages auxquels la conscience phénoménologique d'un·e spectateur·rice novice pourrait se rattacher afin de faire sens de ce qu'il voit. Le pauvre investissement narratif, le montage incertain et la fascination pour la matérialité du *medium* qui caractérisaient le cinéma dit «primitif» rendait cette situation d'autant plus manifeste (Denson, 2014). La transition vers un cinéma plus classique et narratif a néanmoins prouvé que l'investissement discursif et représentationnel des spectateur·rice·s a finalement retrouvé une place. Cependant, cela ne s'est pas fait sans une réaffectation de leur habilité narrative au sein même de la matérialité des flux insignifiants et impersonnels du cinéma qui reconditionnent

leurs facultés perceptuelles (Denson, 2014). Autrement dit, il a fallu apprendre à habiter la machine-cinéma, donc à faire avec une altération du mode sur lequel la sensibilité incarnée peut *raconter une histoire* via son interaction avec un nouveau *medium*. L'approche de Denson pense donc un double niveau d'expérience: celui de la subjectivité incarnée qui subit une décomposition de ses facultés perceptives au contact d'un *medium*, et celui de la recomposition de l'expérience subjective via la médiation d'un nouveau régime perceptif par ce même *medium*¹⁵.

Un défi *décorrélatif* bien plus grand encore se présente aujourd'hui aux premiers usager·ère·s des modèles texte-image puisque, contrairement au cinéma, et a fortiori à la photographie, les arts picturaux, ou encore les modélisations 3D, les modèles texte-image génèrent des résultats photoréalistes sans nécessiter d'instruments de capture ni non plus de «compréhension» des règles qui régissent l'optique ou la disposition des corps dans l'espace. Cette absence de vertu épistémique du *medium* (Daston et Galison, 2007) – dont le *deepfake* est l'emblème – compromet toute tentative d'accéder au monde avec eux sur un mode strictement *corrélatif*. Ceci étant dit, embrasser l'écart irréductible entre les *prompts* et les sorties générées, jouer avec les incohérences photoréalistes de ces dernières, ou encore mettre en place une errance improvisatrice au cours des *bouclages projectifs*, pourraient devenir les cadres performatifs de l'affirmation d'un regard qui composerait avec la *décorrélation* entre la subjectivité incarnée de l'artiste et les opérations du *medium* algorithmique. Il est par conséquent important d'expérimenter avec les IA génératives de façon à révéler la *métabolisation*¹⁶ en cours de nos capacités de perception et, par extension, nos capacités de nous représenter ce qu'on perçoit et/ou désire percevoir.

¹⁵ Le monstre de Frankenstein est pour Denson la figure-clé de cette atomisation / ré-animation de l'expérience incarnée conjointement à son imbrication avec la matérialité d'un nouveau *medium* technologique (Denson, 2014).

¹⁶ Denson mobilise le terme *metabolizing* pour désigner les processus qui enchevêtrent matériellement nos organes perceptifs et notre environnement technologique en deçà de ce que la conscience phénoménologique humaine peut percevoir. Le métabolisme désigne alors l'agence transformatrice de cet enchevêtrement qui reconfigure conjointement la nature de notre perception et la nature de notre environnement technologique, sans pour autant relever d'un pouvoir actif ou passif, ni être attribuable à une subjectivité constituée ou bien une intentionnalité de la machine (Denson, 2020).

Disociascope

Afin d'explorer plus concrètement ce point, j'ai tenté une exploration autoréflexive de cette *métabolisation* dans *Discociascope*. Plus spécifiquement, j'ai voulu partir à nouveau de la photographie (*medium* de prédilection de ma pratique artistique) pour rendre compte de la transition d'un regard élaboré conjointement avec la photographie, vers un regard modulé par les *bouclages projectifs* liés à la génération d'images¹⁷. Ce projet se compose d'une série d'images générées au cours d'un chaînage entre une version du modèle Stable Diffusion (via la plateforme Google Colab), un modèle d'*outpainting* disponible sur le site runwayml.com, et un logiciel de retouche photo à base d'apprentissage machine nommé Luminar AI. J'ai commencé par lancer des générations d'images à partir de *prompts* évoquant le stade du miroir, du type: «an analog photography of a bare body making contact with its hand in the mirror, mirror stage, hand reflection in the mirror». Cette évocation d'une relation intime à la specularité se veut un écho thématique au stade du miroir. Au même titre que l'enfant qui touche le miroir apprend à faire tactilement sens de sa présence dans le reflet¹⁸, *Disociascope* tente de décrire – par transposition de ce thème – la façon dont un regard corporellement ancré dans un paradigme photographique tente d'habiter ce que les modèles génératifs lui donnent à voir. Une fois le prompt entré, le reste du processus que je relate ensuite est une improvisation intuitive dont le mode opératoire (similaire à celui de *Mutant Photograph*) a été réitéré pour chaque image composant le projet.

¹⁷ Ce que, en substance, j'avais déjà commencé à faire dans *AIWOT* sans avoir les ressources pour le conceptualiser comme tel.

¹⁸ Dans le stade du miroir, la relation *décorrélative* en jeu entre la technicité spéculaire du miroir et la vision incarnée de l'enfant impose un décalage entre ce qui est vu par l'enfant via la médiation du miroir et sa sensation tactile d'être ancré dans un corps qui est le sien (Denson, 2023). L'enfant apprend à faire sens de cette dissociation en comprenant que ce reflet peut tout aussi bien entrer dans le domaine de l'identification tactile de son propre corps.

J'ai dans un premier temps interrompu les itérations très tôt, minimisant ainsi mon engagement dans des *bouclages projectifs* et lui substituant le sentiment d'une prise de vue photographique, ou le développement d'une impression argentique: en fixant une capture et s'inscrivant dans une temporalité mémorielle, l'idée est de se démarquer du flux anticipateur de la génération d'image (Denson, 2020). Les résultats obtenus sont alors le reflet des premières itérations: des images dont la génération est encore en gestation, un effet de vision floue, une abstraction de motifs sans profondeur. Sans parcourir toutes les sorties proposées, je sélectionne quelques images parmi celles qui se présentent immédiatement à moi en prenant pour seul critère qu'elles soient à cheval entre abstraction et figuration. Je garde ainsi la trace de ce que l'archéologue des *media* Wolfgang Ernst caractériserait comme leur «electro-technical embodiment» (Ernst, 2020: 14), c'est-à-dire l'étoffe algorithmique qui sous-tend leur formalisme. Je fais ensuite appel à un modèle d'*outpainting* entraîné sur des images photographiques pour générer, hors-cadre, des éléments visuels qui vont leur conférer profondeur et photoréalisme. L'*outpainting* permet en effet d'aboutir à un résultat visuel dont la distribution pixelisée des couleurs, des contours, des textures, des ombres et de la lumière est homogénéisante, donnant donc l'impression que les images ne sont pas le fruit d'un tortueux rapprochement de motifs visuels via des algorithmes, mais bien une photographie réellement prise par une caméra dans un lieu unique, en présence de mon corps. La dernière phase consiste à importer cette image étendue par l'*outpainting* dans un logiciel basé sur un modèle d'apprentissage machine capable d'appliquer des filtres et des retouches de façon à donner un effet de pellicule argentique noir et blanc. Cela renforce également la simulation d'un ici-et-maintenant capturé par un objectif et fixé sur du celluloid qui fait, à première vue, croire aux spectateur·rice·s que l'image finale est une réelle photographie argentique. Mais à y regarder de plus près, elle est partout minée d'incohérences visuelles et de surcharges de procédures algorithmiques (*Figure 6*).

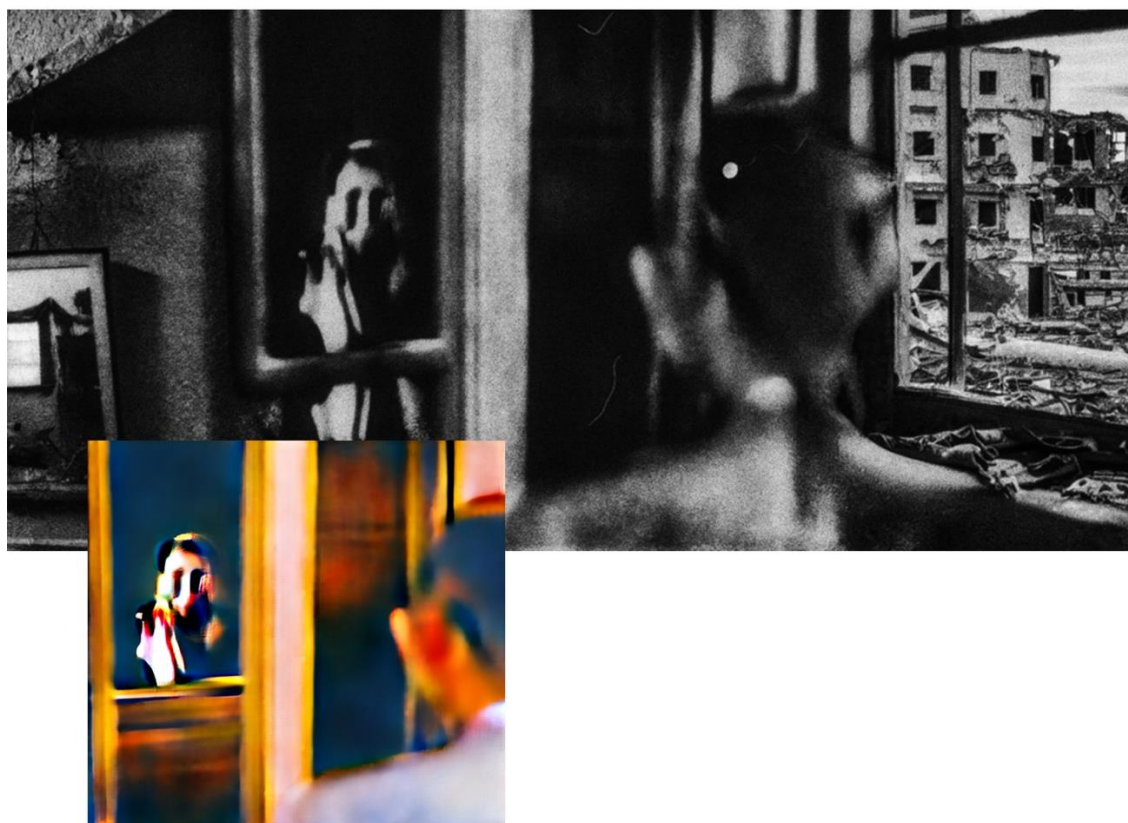
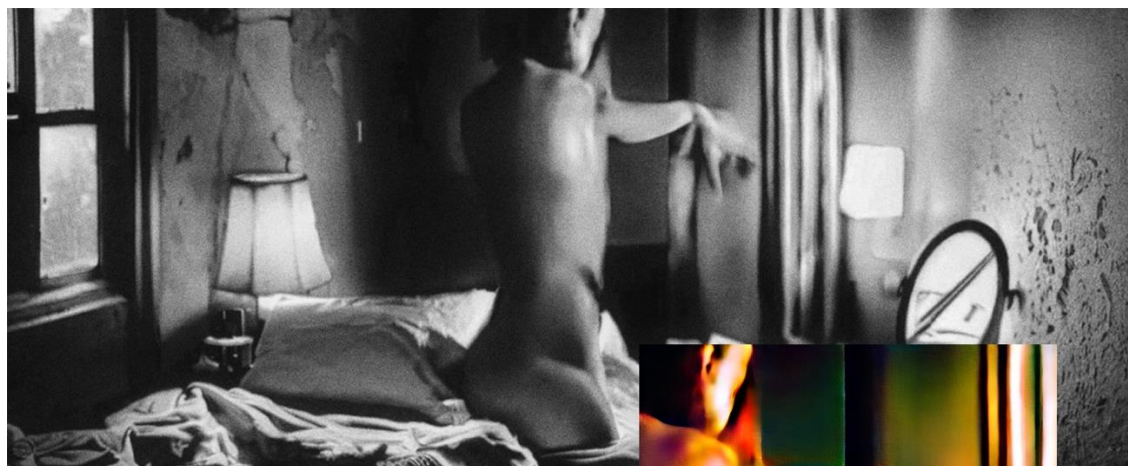


Figure 6. Extraits de *Disociascope*: l'image initialement issue des premières itérations est en couleur, l'image finale est en noir et blanc.

Disociascope joue donc littéralement sur la transformation d'une *image latente* – initialement issue des premières itérations du modèle – en centre d'un point de vue photographique, point de vue qui à son tour dépend d'une simulation algorithmique permise par la pratique intuitive de l'*outpainting*. Cette mise en tension peut être conceptualisée selon les termes du chercheur en études médiatiques Alexander Galloway, tel que repris par Denson, qui décrit l'opposition entre une relation *corrélative* qu'il nomme un *contrat photographique*, et une relation *décorrélative* qu'il nomme un *contrat informatique*¹⁹. Par «contrat» Galloway entend l'établissement d'une relation entre un *medium* et un régime d'expérience particulier qu'il *médie* (Galloway, 2021). Le *contrat photographique* implique un point de perception unique – centré sur l'œil du photographe, ou l'objectif en tant que prothèse oculaire de l'agent – que Galloway compare à un cône ou à une corne, qui engage sa présence corporelle et accède au monde à partir d'un point focal unique et riche. Le *contrat informatique*, en revanche, est plus proche de la modélisation architecturale ou de la sculpture que de la photographie. La perception ne se fait pas à partir d'un seul point, mais plutôt à partir de plusieurs dimensions qui sont potentiellement toutes prises en compte, et qui participent à l'émergence d'une figure vis-à-vis d'un fond à travers un ensemble de différences significatives (plus que signifiantes). Ce *contrat* est décrit par Galloway comme un cône inversé, saisissant le monde à partir d'une base large qui se réduit progressivement à un objet spécifique sur lequel va se porter l'attention. Galloway insiste:

Computational vision takes it as a given that point of view is not necessary for seeing. The issue is not simply that vision has become abstract, not simply that vision has been unloosed from its subjective mooring, [...] The crux of the issue is that seeing no longer requires a point. (Galloway, 2021: 52)

Bien plus que les modélisations informatiques qui continuent de s'aligner sur les règles de l'optique et de la physique des corps, l'espace latent des modèles texte-image incarne

¹⁹ Denson mobilisant également cette opposition dans son ouvrage *Post-Cinematic Bodies* (Denson, 2023).

radicalement cette vue-de-toute-part. En effet, si la génération d'image peut imiter la présence tangible d'une capture photographique, ce n'est qu'au prix d'une atomisation des images en une dispersion de données latentes – données qui pourraient être tout à fait mobilisées pour générer autre chose que des imitations de photographies. Le *contrat informatique* d'un modèle texte-image est en ce sens ce qu'on pourrait baptiser une *prévisualisation latente* qui englobe toutes les images sur lesquelles il a été entraîné et toutes les générations d'images à venir que les *prompts* déclencheront.

La réalisation de *Disociascope* suppose donc d'accepter une redistribution paradoxale de l'agentivité construisant un regard. D'un côté, mon intentionnalité de projet est limitée par les opérations sub-perceptuelles et sub-symboliques des modèles, ce qui compromet l'affirmation d'un point de vue qui serait mien, ou sous ma supervision. De l'autre cependant, l'interruption intentionnelle de l'algorithme, le chaînage entre différentes applications et le maintien de la thématique du stade du miroir au cours du processus de génération, consistent tout de même en l'affirmation consciente d'un point de vue subjectif en vue d'un résultat final (même indéterminé). Ma conscience de supervision compose donc avec la *prévisualisation latente* tout en essayant de reprendre la main sur l'orientation du projet. *Disociascope* permet ainsi d'explorer à tâtons une nouvelle grammaire visuelle caractérisée par une agentivité plurielle et distribuée, qui est aussi une expérience dissociative du regard: d'une part un regard capable d'affirmer une représentation, et d'autre part, un regard capable de prendre ses distances avec son désir de dénnotations, au profit d'un agencement performatif de formes sans finalité. Même si ma conscience de supervision trouve dans le *contrat photographique* la possibilité de faire sens à travers une dénnotation – assignant les formes visuelles à la thématique du stade du miroir – la métabolisation de ce point de vue au cours des *bouclages projectifs* de la génération d'images (en un sens, le point de vue s'y perd, s'y noie), mine cet effort et ouvre à une re-potentialisation formelle. Il convient donc d'insister sur le fait que l'émergence d'un regard n'est, dans ce contexte, ni exclusivement attribuable à une causalité efficiente localisée chez l'artiste (un *contrat photographique* qui signerait l'autonomie absolue du sujet regardant), ni non plus à une causalité efficiente des opérations des modèles (l'absolue autonomie d'une

vue-de-toute-part), ni non plus à une double causalité efficiente anthropo-algorithmique (un assemblage cyborg qui agirait également comme un sujet absolu). Ce regard participe plutôt d'une intentionnalité émergente au cœur d'une récursivité formelle (celle des *bouclages projectifs* au cours desquels les sorties visuelles générées redirigent la rédaction des *prompts*, et ainsi de suite). L'intentionnalité de projet de l'artiste doit alors sans cesse faire avec l'intentionnalité secondaire de sélection à laquelle l'agentivité de l'artiste tend à être cantonnée. L'artiste peut toujours se donner l'illusion que ce regard distribué est, en dernière instance, le sien. Malgré tout, la conscience de supervision de l'artiste ne prend pas pour autant les commandes de l'œuvre, et la *décorrélation* artiste-modèles persiste.

Conclusion

Au cours de cette pratique réflexive, j'ai pu mettre à l'épreuve le cadre conceptuel puisé chez Thierry Bardini, Mark Hansen puis Shane Denson en l'articulant sur des observations pratiques issues de ma propre exploration du potentiel esthétique des modèles. Une démarche expérimentale permettant de saisir les enjeux phénoménotecniques de ces modèles est en train de se dessiner, et servira de guide pour la conduite ultérieure d'entretiens semi-dirigés avec un échantillon d'artistes visuel-le-s travaillant avec des logiciels à base d'apprentissage profond.

J'ai mis en évidence que loin d'être une simple traduction déterministe du texte à l'image, l'usage artistique de ces modèles amène à s'inscrire dans des *bouclages projectifs*, qui ne sont pas sans altérer ni reconfigurer la subjectivité de l'artiste. En effet, ces *bouclages* impliquent un retrait de la conscience de supervision au profit de co-ajustements itératifs avec les modèles d'une façon qui informe en retour les intuitions et les représentations de l'artiste. La danse d'agence dans laquelle il-elle se retrouve alors imbriquée impose de faire avec un enchevêtrement de son intentionnalité de projet avec l'*agentivité formelle* des *bouclages projectifs*. L'artiste a bel et bien un pouvoir d'interruption de la continuité performative de la génération d'images, ainsi que celui de recadrer et modifier les sorties visuelles pour orienter la

génération d'image vers des figures reconnaissables et porteuses de sens pour lui-elle. Mais paradoxalement, il n'en demeure pas moins que la subjectivité de l'artiste reste imbriquée dans la structure d'expérience rendue possible par les *bouclages projectifs*. A l'image de mon projet *Disociascope*, l'œuvre émerge donc d'une tension entre l'*agentivité formelle* des *bouclages projectifs* – qui déplace et re-potentialise continuellement les représentations mises en jeu par l'artiste – et l'*agentivité effective* de l'artiste – dont la conscience de supervision interrompt le flux des *bouclages* et tente d'affirmer un point de vue.

Bibliographie

- ANTON Corey, LOGAN Robert K., STRATE Lance (2017), *Taking up McLuhan's Cause. Perspectives on Media and Formal Causality*, Bristol UK, Intellect Publishing.
- AUDRY Sofian (2021), *Art in the Age of Machine Learning*, Cambridge (MA), MIT Press.
- BAJOHR Hannes (2023), «Dumb Meaning: Machine Learning and Artificial Semantics», *IMAGE*, n. 37/1, pp. 58–70.
- BARDINI Thierry (2016), «Entre archéologie et écologie», *Multitudes*, n. 62/1, pp. 159–168.
- BAXTER Kristin, LÓPEZ Hugo Ortega, SERIG Dan, SULLIVAN Graeme (2008), «The necessity of studio art as a site and source for dissertation research», *International Journal of Art and Design Education*, n. 27/1, pp. 4–18.
- BOLTER Jay David (2023), «AI Generative Art as Algorithmic Remediation», *IMAGE*, n. 37/1, pp. 195–207.
- BUSCH Kathrin (2009), «Artistic Research and the Poetics of Knowledge », *Art & Research: A Journal of Ideas, Contexts and Methods*, n. 2/2. URL: <<http://www.artandresearch.org.uk/v2n2/busch.html>>.
- CHAPMAN Owen, SAWCHUK Kim (2012), «Research-Creation: Intervention, Analysis and “Family Resemblances”», *Canadian Journal of Communication*, n. 37/1, pp. 5–26.
- CHATONSKY Grégory (2021), «Les deux techniques», <http://chatonsky.net/deux-techniques/>
- DASTON Lorraine, GALISON Peter (2007), *Objectivity*, New-York, Zone Books.
- DENSON Shane (2014), *Postnaturalism: Frankenstein, Film, and the Anthropotechnical Interface*, Bielefeld, Transcript Verlag.
- DENSON Shane (2020), *Discorrelated images*, Durham, Duke University Press.

- DENSON Shane (2023), *Post-Cinematic Bodies*, Lüneburg, Meson press.
- EACHO Douglas (2023), «Performativity without theatricality: experiments at the limit of staging AI», *Theatre and Performance Design*, n. 9/1-2, pp. 20-36.
- ERNST Wolfgang (2020), «Existing in Discrete States. On the Techno-Aesthetics of Algorithmic Being-in-Time», *Theory, Culture & Society*, n. 38/7-8, pp. 13-31.
- FEYERSINGER Erwin, KOHMANN Lukas, PELZER Michael (2023), «Fuzzy ingenuity: creative potentials and mechanics of fuzziness in processes of image creation with AI-Based Text-to-Image production», *IMAGE*, n. 37/1, pp. 135-149.
- GALLOWAY Alexander (2021), *Uncomputable. Play and politics in the long digital age*, London, Verso Books.
- HANSEN Mark (2015), *Feed-Forward. On the future of Twenty-first century media*, Chicago, University of Chicago Press.
- HEIDEGGER Martin (1958), *Essais et conférences. La question de la technique* [1953], Paris, Gallimard, pp. 9-48.
- HERTZMANN Aaron (2018), «Can Computers Create Art?», *Arts*, n. 7/2. URL: <<https://tspace.library.utoronto.ca/bitstream/1807/89089/1/arts-07-00018-v3.pdf>>.
- KIRSCHENBAUM Matthew (2021), «Spec Acts: Reading form in Recurrent Neural Networks», *ELH*, n. 88/2, pp. 361-386.
- LE CUN Yann (2019), *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*, Paris, Odile Jacob.
- MANOVICH Lev (2023), «AI Image Media through the Lens of Art and Media History», *IMAGE*, n. 37/1, pp. 34-41.
- MAZZONE Marian, ELGAMMAL Ahmed (2019), «Art, Creativity, and the Potential of Artificial Intelligence», *Arts*, n. 8/1. URL: <<https://www.mdpi.com/2076-0752/8/1/26>>.
- MCCORMACK Jon, GIFFORD Toby, HUTCHINGS Patrick (2019), «Autonomy, Authenticity, Authorship and Intention in Computer Generated Art», *Lecture Notes in Computer Science*, n. 11453, pp. 35-50.
- MITCHELL William John Thomas (1986), *Iconology - Image, text, ideology*, Chicago, University of Chicago Press.
- MITCHELL William John Thomas (2005), «There are no visual media», *Journal of Visual Culture*, n. 4/2, pp. 257-266.
- MONDZAIN Marie-José (2002), *L'image peut-elle tuer?*, Paris, Bayard.
- PAGLEN Trevor (2019), «Invisible Images. Your pictures are looking at you», *Architectural Design*, n. 89/1, pp. 22-27.
- PAQUIN Louis-Claude, NOURY Cynthia (2018), *Définir la recherche-crédation ou en cartographier les pratiques?*, ACFAS, Recherche et Création. URL:

<<https://www.acfas.ca/publications/magazine/2018/02/definir-recherche-creation-cartographier-ses-pratiques>>.

PICKERING Andrew (1995), *The mangle of practice. Time, agency, and science*, Chicago, University of Chicago Press.

PLOIN Anne, EYNON Rebecca, HJORTH Isis, OSBORNE Michael A. (2022), *AI and the Arts: How Machine Learning is Changing Artistic Work*, Oxford, UK, report from the Creative Algorithmic Intelligence Research Project, Oxford Internet Institute.

RADFORD Alec, KIM Jong Wook, HALLACY Chris, RAMESH Aditya, GOH Gabriel, AGARWAL Sandhini, SASTRY Girish, ASKELL Amanda, MISHKIN Pamela, CLARK Jack, KRUEGER Gretchen, SUTSKEVER Ilya (2021), «Learning transferable visual models from natural language supervision», *Proceedings of Machine Learning Research*, n. 139, pp. 8748–8763.

RUSSELL Angus (2022), «How to use Stable Diffusion to generate images from a text prompt — no coding or technical knowledge required», *Medium*. URL: <<https://medium.com/nightcafe-creator/stable-diffusion-tutorial-how-to-use-stable-diffusion-157785632eb3>>.

SALVAGGIO Eryk (2023) «How to Read an AI Image. Toward a Media Studies Methodology for the Analysis of Synthetic Images», *IMAGE*, n. 37/1, pp. 83-99.

SCHUHMANN Christoph, BEAUMONT Romain, VENCU Richard, GORDON Cade, WIGHTMAN Ross, CHERTI Mehdi, COOMBES Theo, KATTA Aarush, MULLIS Clayton, WORTSMAN Mitchell, SCHRAMOWSKI Patrick, KUNDURTHY Srivatsa, CROWSON Katherine, SCHMIDT Ludwig, KACZMARCZYK Robert, JITSEV Jenia (2022), «LAION-5B: an open large-scale dataset for training next generation image-text models», *Advances in Neural Information Processing Systems*, n. 35. URL: <https://papers.nips.cc/paper_files/paper/2022/file/a1859debf3b59d094f3504d5ebb6c25-Paper-Datasets_and_Benchmarks.pdf>.

SHAVIRO Steven (2010), *Post cinematic affect*, Winchester, Zero Books.

TERZIDIS Kostas, FABROCINI Filippo, LEE Hyejin (2022), «Unintentional intentionality: art and design in the age of artificial intelligence», *AI & SOCIETY*, n. 38/4, pp. 1715–1724.

WASIELEWSKI Amanda (2023a), *Computational formalism*, Cambridge, MA, MIT Press.

WASIELEWSKI Amanda (2023b), «“Midjourney Can’t Count”: questions of representation and meaning for text-to-image generators», *IMAGE*, n. 37/1, pp. 71–82.

Come citare questo articolo:

Arnaud Mery, «Le regard de l'artiste à l'épreuve des modèles texte-image», in *InterArtes* [online], n. 3, “ArtIntelligence” (Francesco Pigozzo ed.), dicembre 2023, pp. 14-44, <<https://www.iulm.it/wps/wcm/connect/iulm/885536a7-0ec7-484b-9649-4236274105af/03+MERY+.pdf?MOD=AJPERES>>.

