

Que peuvent les machines nous apprendre du style ?

Pia BALTAZAR

(Société des Arts Technologiques, Montréal, QC)

Kivanç TATAR

(Chalmers University, Goteborg, SE)

Pour citer cet article :

Pia BALTAZAR et Kivanç TATAR, « Que peuvent les machines nous apprendre du style ? », *Revue Proteus*, n° 20, L'ère numérique du style, Athina Masoura et Anthony Rageul (coord.), 2023, p. 79-95

Résumé

Cet article propose quelques pistes de réflexion sur la notion de style dans sa relations aux outils et méthodes de l'« Intelligence Artificielle » appliquées à la création artistique. Cette réflexion est conduite dans une double perspective : artistique et technique d'une part, et esthétique et épistémologique d'autre part. Elle se concentre en particulier sur la question des taxonomies sur lesquelles se basent les mécanismes de reconnaissance, d'apprentissage et de génération d'artefacts artistiques. Ces taxonomies peuvent être définies de diverses manières plus ou moins autonomes, principalement selon le paradigme connexionniste. Nous nous intéressons en particulier à l'aspect fonctionnel de ce paradigme, le « réseau de neurones artificiel », qui permet de reconnaître et de générer, à partir d'un jeu réduit de critères initiaux, des caractères morphologiques complexes, qui font sens d'une manière intuitive pour l'humain. Nous posons à travers cet examen quelques hypothèses sur les raisons de l'efficacité de ces techniques, en rapprochant ces taxonomies – qui s'élaborent et « se tiennent » sans le besoin d'une définition préalable explicite – de la notion proposée par l'épistémologue Richard Boyd de « grappes de propriétés homéostatiques » et la mettons en relation avec la définition du style que donnent Deleuze et Guattari autour de la notion de « consistance ». Pour conclure, nous proposons quelques interrogations et pistes de réflexion sur les contraintes et conditionnements sur l'expression artistique qu'implique l'usage de ces systèmes dans une pratique artistique.

Intelligence artificielle — Connexionnisme — Apprentissage automatique — Épistémologie — Poïétique

Abstract

This article proposes some lines of thought on the notion of style in its relation to the tools and methods of "Artificial Intelligence" applied to artistic creation. This reflection is conducted from a double perspective: artistic and technical on the one hand, and aesthetic and epistemological on the other. It focuses in particular on the question of the taxonomies on which the mechanisms of recognition, learning and generation of artistic artefacts are based. These taxonomies can be defined in various more or less autonomous ways, mainly according to the connectionist paradigm. We are particularly interested in the functional aspect of this paradigm, the "artificial neural network", which allows to recognize and generate, from a reduced set of initial criteria, complex morphological characters, which make sense in an intuitive way for humans. Through this examination, we put forward some hypotheses on the reasons for the effectiveness of these techniques, by bringing these taxonomies – which are elaborated and "held together" without the need for a prior and explicit definition – closer to the notion proposed by the epistemologist Richard Boyd of "clusters of homeostatic properties", and relate it to the definition of style given by Deleuze and Guattari around the notion of "consistency". To conclude, we propose some questions and avenues of reflection on the constraints and conditioning on artistic expression implied by the use of these systems in an artistic practice.

Artificial intelligence — Connexionism — Machine learning — Epistemology — Poiesis

Que peuvent les machines nous apprendre du style ?

L'usage de l'intelligence artificielle dans la création artistique est en plein essor depuis quelques années, et connaît notamment depuis peu une forte croissance d'intérêt de la part du grand public et des créateur·rices artistiques¹. Cet essor génère également de nombreux et houleux débats², dans lesquels nous ne rentrerons pas dans cet article. En effet, ces débats soulèvent de nombreux enjeux esthétiques, éthiques et même économiques, que nous préférons ne pas aborder pour l'instant afin de nous concentrer sur une tentative de compréhension poïétique des mécanismes au cœur des dispositifs de création artistique fondés sur les techniques d'intelligence artificielle (IA, ou aussi « apprentissage automatique³ »). L'objectif est d'en saisir au mieux le fonctionnement et les effets, ainsi que les moyens de les infléchir afin de tirer le meilleur parti des possibilités esthétiques comme épistémologiques qu'ils ouvrent.

Cet article se fonde pour cette observation des aspects factuels de l'utilisation des techniques d'apprentissage automatique dans l'art, sur une

double perspective : artistique et technique d'une part, et poïétique/épistémologique d'autre part. Le premier point de vue, artistique et technique, s'appuie sur la pratique artistique personnelle d'artiste, compositrice intermedia et *designeuse* interactive de la première autrice, notamment autour de l'application de systèmes de génération musicale par apprentissage automatique. Le co-auteur, pour sa part spécialiste de l'intelligence artificielle pour la musique et les médias interactifs, a contribué de son expertise sur les procédés et sur les paradigmes techniques sur lesquels ils se fondent. Pour ce qui est de la perspective philosophique, elle s'appuie sur la formation universitaire et les lectures en esthétique, poïétique et en épistémologie de la première autrice⁴.

Il s'agira plus particulièrement ici d'interroger, à partir de la conception esthétique du style de Gilles Deleuze, Félix Guattari et Brian Massumi la manière dont le fonctionnement des systèmes d'apprentissage peut-être expliqué à l'aide de la notion de « grappes de propriétés homéostatiques⁵ » de l'épistémologue Richard Boyd. Cette démarche vise à poser quelques hypothèses afin de mieux comprendre pourquoi l'IA est aussi efficace pour modéliser et simuler le style artistique, comme ont pu le montrer les derniers développements en date de celle-ci (notamment avec les dispositifs *Dall-E* et *Midjourney*⁶, pour ne citer que les plus populaires). À travers cette initiative, nous espérons que nos réflexions pourront nourrir le

1. Cf. KIVANÇ TATAR, Petter ERICSON et Kelsey COTTON, « A Shift In Artistic Practices through Artificial Intelligence », 2023, <<https://arxiv.org/abs/2306.10054>>, consultée le 2 septembre 2023, ou Sofian AUDRY, *Art in the age of machine learning*, Cambridge, MA, MIT Press, 2021 pour une présentation des différents courants actuels de l'utilisation de l'AI dans les arts et des enjeux esthétiques associés.

2. Cf. notamment Joanna BRYSON et Martin EISENLAUER, « Artificial Intelligence and ethics », dans *Faster than the Future*, Barcelona, Hertie School, 2021, p. 57-73 et Joanna BRYSON, « The Moral, Legal, and Economic Hazard of Anthropomorphizing Robots and AI », dans Mark COECKELBERGH, Janina LOH *et al.* (dir.) *Envisioning Robots in Society - Power, Politics, and Public Space - Proceedings of Robophilosophy 2018 / TRANSOR 2018*, IOS Press, Vienne, 2018.

3. Pour reprendre la distinction de Bruno Latour, nous préférons dans un premier temps dans le cadre de cet article, en rester au niveau des aspects factuels (« matters of fact ») avant de traiter dans un futur article plus spécifiquement des préoccupations (« matters of concern ») qu'ils entraînent. Cf. à ce sujet Bruno LATOUR, « Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern », dans *Critical Inquiry*, vol. 30, n° 2, janvier 2004, p. 225-248.

4. Les auteur·trices tiennent à remercier Anthony Rageul et Athina Masoura pour leur invitation à intervenir à la journée d'études dont est issue cette publication, et ensuite à la formaliser sous la forme de cet article. Ce travail a par ailleurs été partiellement soutenu par le programme Wallenberg AI, Autonomous Systems and Software - Humanities and Society (WASP-HS) financé par les fondations Marianne and Marcus Wallenberg et Marcus and Amalia Wallenberg.

5. Richard BOYD, « Homeostasis, species, and higher taxa », *Species: New interdisciplinary essays*, 1999, vol. 141, p. 185-205.

6. Ces dispositifs sont accessibles en ligne aux URLs suivantes : <<https://openai.com/dall-e-2/>> et <<https://midjourney.com/>>, consultées le 2 septembre 2023.

champ de l'esthétique en donnant des outils, pour contribuer à formaliser la notion de style, à d'autres auteur·rices de domaines comme la stylistique ou l'histoire de l'art.

Qu'appelle-t-on l'« intelligence artificielle » dans l'art ?

Le récent essor d'engouement autour des « créations » générées par les dispositifs d'intelligence artificielle, s'il a attiré l'attention sur ce domaine, a aussi contribué à troubler la compréhension de ce en quoi consistent ces dispositifs, et des mécanismes qu'ils mettent en œuvre. Il convient à notre avis de revenir en premier lieu sur la notion même d'intelligence artificielle. Le *Trésor de la langue française informatisé* donne la définition suivante de cette notion : « Recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains¹ ». À notre sens, le terme particulièrement important à relever dans cette définition est « comparable » : en effet, si un dispositif informatique peut « créer » des artefacts disposant de qualités esthétiques *comparables* à celles qui seraient produites par des humains, c'est avant tout parce qu'il a été entraîné de façon massive et systématique à imiter des créations humaines. C'est pourquoi, dans le reste de cet article, afin d'éviter toute méprise sur une potentielle « intelligence » ou « créativité » informatique, nous utiliserons à sa place le terme technique du procédé informatique sur lequel s'appuient la majorité des dispositifs de création de contenu esthétique (visuels, sonores ou textuels) qualifiés d'« intelligence artificielle » : « l'apprentissage automatique » (*machine-learning* en anglais, abrégé en « AA » dans la suite de cet article).

Le *Vocabulaire de l'intelligence artificielle* du *Journal Officiel* donne comme définition de l'apprentissage automatique :

Processus par lequel un algorithme évalue et améliore ses performances sans l'intervention d'un pro-



Fig. 1 et 2. Deux images générées par le système *Dall-E 2*, sur le script « des ours en peluche en savants fous mixant des produits chimiques pétillants, dans un style *steampunk* »

grammeur, en répétant son exécution sur des jeux de données jusqu'à obtenir, de manière régulière, des résultats pertinents².

Dans le cas qui nous intéresse, les algorithmes d'AA vont être entraînés à l'aide de grands nombres de créations esthétiques humaines, avec de potentielles métadonnées (soient des données caractérisant des informations corollaires à l'œuvre elle-même, comme la date de production, le nom de l'auteur·rice, des techniques de création, le genre, ou toute autre information jugée pertinente) ou d'autres informations associées à l'artefact esthétique. Par exemple, les systèmes de génération d'images par scripts (telles que *Dall-E*, cité ci-dessus) qui font beaucoup parler d'eux au moment de la rédaction de cet article ont fondé leur efficacité sur l'analyse conjointe de très nombreuses images associées à leurs textes descriptifs (qui sont notamment destinés à servir d'assistance pour les personnes non- ou mal-voyantes, ou encore pour la classification automatique). Les algorithmes d'apprentissage vont alors dans ce type de dispositif dégager des corrélations entre des motifs visuels et des motifs textuels dans les descriptifs, qu'ils pourront ensuite exploiter « à rebours » lorsqu'un·e utilisateur·rice saisira un script descriptif, comme par exemple : « des ours en peluche en savants fous mixant des produits chimiques pétillants, dans un style *steampunk* » (voir Fig. 1 et 2³).

1. Dans *La Recherche*, n° 96, vol. 10, janv. 1979, p. 61, cité dans l'entrée « Intelligence » de <<http://atilf.atilf.fr>>, consultée le 2 septembre 2023.

2. Voir : <<https://www.legifrance.gouv.fr/jorf/id/JORF-TEXT000037783813>>, consultée le 2 septembre 2023.

3. Le statut légal des images présentées ici est relativement indéterminé, mais en tout état de cause, les conditions d'uti-

Nous reviendrons plus tard plus en détail sur le fonctionnement concret de ces systèmes d'apprentissage, mais l'important pour l'instant est de retenir que, parmi les critères que reproduisent particulièrement bien ces systèmes d'AA, les notions de « style » et de « genre » jouent un rôle très important.

Quelle approche du « style » ?

Avant d'aller plus loin dans notre réflexion, il nous faut préciser ce que nous entendons par les deux notions de « style » et de « genre », et en particulier celle de « style », qui est l'objet du dossier du numéro de la revue dans laquelle paraît le présent article. À cette fin, nous nous appuyons sur la conception du style proposée par Gilles Deleuze au fil de plusieurs écrits, d'abord seul dans *Proust et les signes*, puis avec Félix Guattari dans *Mille plateaux* et *Qu'est-ce que la philosophie ?*, et plus tard reprise par Brian Massumi dans son commentaire *A user's guide to capitalism and schizophrenia: Deviations from Deleuze and Guattari*.

Dans de nombreux discours, qu'il soient académiques (en sciences de l'art ou en recherche-création, notamment) ou dans le domaine professionnel ou critique, on assimile souvent le style à la notion de singularité, et en particulier à la singularité d'une démarche artistique – dans la lignée de la notion de style de Schopenhauer, Flaubert, Marivaux comme marque d'un esprit individuel original. G. Deleuze et F. Guattari semblent dire quelque chose d'approchant quand ils écrivent qu'« il faut le style – la syntaxe d'un écrivain, les modes et rythmes d'un musicien, les traits et les couleurs d'un peintre¹ – pour s'élever des perceptions vécues au percept, des affections vécues à l'affect² ». Autrement dit, dans cette conception,

c'est en quelque sorte le style qui fait ce qui relève proprement de l'art, qui est le marqueur d'une qualité spécifique à un artefact esthétique, qui apporte ce « quelque chose » en plus qui fait la spécificité d'une œuvre d'art. Cependant, si cette citation semble associer le style à l'auteur·rice, G. Deleuze a déjà précisé, après Proust et notamment dans sa controverse avec Sainte-Beuve³, du fait que ce style n'est pas réductible à l'individu qui l'a produit : « Le style n'est pas l'homme, le style, c'est l'essence elle-même⁴ ».

Cette notion d'essence peut paraître assez mystérieuse pour l'instant – nous y reviendrons – mais l'essentiel est que Proust comme Deleuze ne la considèrent pas comme quelque chose de platonicien, c'est-à-dire d'« essentialiste », tout comme c'est le cas de la conception épistémologique des essences de Richard Boyd, qui présente à notre sens de nombreuses similitudes avec celles de Deleuze et Proust et sur laquelle nous reviendrons également plus tard. Pour l'instant, ce qui compte est que, selon Deleuze, « un style [n'est] pas une création psychologique individuelle, mais un agencement d'énonciation⁵ ». On peut relever plusieurs aspects dans cette courte citation : tout d'abord, au-delà de la singularité de la personne (ou des personnes) qui ont produit l'artefact, ce qu'indique ici Deleuze est que le style « se fait » en rapport à des dispositifs (notamment techniques) et des contextes (langue, genres...). Par ailleurs, au-delà du style d'un artiste, il y a aussi le style d'une œuvre (et entre les deux, d'une « période » d'un artiste), voire d'un élément d'une œuvre, etc. mais aussi des « grappes » de styles approchant, que l'on subsume généralement en genres ou courants. Enfin, on peut aussi considérer dans cette perspective les contraintes et possibilités stylistiques qu'induisent l'usage de telle ou telle langue ou technique de création. Autrement dit, c'est pour chaque œuvre un agencement d'éléments hétérogènes qui donne corps à chaque fois à un

lisation présentées sur <<https://openai.com/terms/>> (consultée le 2 septembre 2023) indiquent la possibilité d'un usage pour publication.

1. On pourrait ici rapprocher cette conception de l'étymologie du terme « style » comme instrument ainsi que du domaine de la stylistique, où il est connexe notamment des règles de grammaires

2. Félix GUATTARI et Gilles DELEUZE, *Qu'est-ce que la philosophie ?*, Paris, Éditions de Minuit, 1991, p. 160.

3. Marcel PROUST, *Contre Sainte-Beuve*, Paris, Gallimard, 1995.

4. Gilles DELEUZE, *Proust et les signes*, Paris, Presses universitaires de France, 1964, p. 62.

5. Gilles DELEUZE et Félix GUATTARI, *Mille plateaux : Capitalisme et schizophrénie*, 2, Paris, Minuit, 1980, « Postulats de la linguistique », p. 123.

style donné. Cet aspect est souligné par Proust lui-même quand il écrit les mots suivants :

La vérité ne commencera qu'au moment où l'écrivain prendra deux objets différents, posera leur rapport, analogue dans le monde de l'art à celui qu'est le rapport unique de la loi causale dans le monde de la science, et les enfermera dans les anneaux nécessaires d'un beau style¹.

Le style, dans cette conception, semble donc tenir à une manière particulière d'agencer des éléments plus ou moins disparates, à les faire « tenir ensemble » dans un certain agencement, dont Deleuze et Guattari vont préciser la nature à l'aide du concept de « plan de consistance ». Brian Massumi reprend et développe cette idée dans l'extrait suivant :

La « consistance », pas dans le sens d'une homogénéité, mais comme un tenir-ensemble d'éléments disparates (aussi connu comme un « style »). Un style dans ce sens, en tant que tenir-ensemble dynamique ou mode de composition, n'est pas quelque chose qui se limite à l'écriture. Les cinéastes, peintres et musiciens ont leurs styles, les mathématiciens ont les leurs, les cailloux ont des styles, tout comme les outils, les technologies et les périodes historiques, même – et particulièrement – les événements ponctuels².

Massumi étend donc ici la notion de style – initialement artistique – à l'ensemble des artefacts humains – et notamment les modélisations scientifiques³ – et même jusqu'aux objets inanimés.

1. Marcel PROUST, *Le temps retrouvé. À la recherche du temps perdu*, tome 8, Paris, Gallimard, 1927, p. 40.

2. « “consistency” —not in the sense of a homogeneity, but as a holding together of disparate elements (also known as a “style”). A style in this sense, as a dynamic holding together or mode of composition, is not something limited to writing. Filmmakers, painters, and musicians have their styles, mathematicians have theirs, rocks have style, and so do tools, and technologies, and historical periods, even—especially—punctual events », c'est nous qui traduisons à partir de Brian MASSUMI, *A user's guide to capitalism and schizophrenia: Deviations from Deleuze and Guattari*, Cambridge, MA, MIT press, 1992, p. 7.

3. On peut ici se référer aux cours de Ian Hacking au Collège de France de 2002-2003 sur les « styles de raisonnement », dont le résumé intitulé « Philosophie et histoire des

Cette conception nous paraît particulièrement fertile pour le sujet qui nous occupe, dans la mesure où les dispositifs d'AA ne font pas de distinction entre un style proprement artistique et le style d'un artefact quelconque. Par ailleurs, on retrouve ici un fonctionnement analogue entre artefacts esthétiques et modèles scientifiques – analogie que pointait déjà Proust dans la citation ci-dessus –, qui va particulièrement s'avérer utile dans le rapprochement que nous allons tenter dans les sections suivantes entre ces aspects esthétiques et des concepts issus de l'épistémologie de Boyd.

Morphologies et grappes de propriétés

Définitions et représentations

Pour pouvoir convoquer les concepts épistémologiques de Richard Boyd, il nous faut revenir brièvement à l'analyse du fonctionnement des dispositifs d'AA dédiés à l'imitation d'artefacts esthétiques : dans la mesure où ceux-ci sont dédiés à l'imitation de formes, contours, profils, matières, textures, etc. il peut être utile de subsumer ces diverses qualités sous le terme générique de « morphologie ». Le *Larousse* définit notamment ce terme comme décrivant l'« étude de la quantité, de la forme, de la répartition et de la proximité des éléments d'une structure macro- ou micrographique⁴ ». Pour préciser cette notion, il peut alors être utile d'observer comment cette étude se fait dans les sciences, et en particulier dans celles qui s'intéressent aux phénomènes les plus complexes que nous puissions observer : le vivant.

Dans cette démarche, il est nécessaire de pouvoir extraire un certain nombre de paramètres objectifs du phénomène étudié, pour pouvoir ensuite classifier ce phénomène vis-à-vis d'autres phénomènes similaires, et enfin élaborer des modèles efficaces permettant de préciser les hypothèses de fonctionnement de ces phéno-

concepts scientifiques » est disponible <https://www.college-de-france.fr/sites/default/files/documents/ian-hacking/UPL4445123752442236773_Hacking2002_2003.pdf>, consultée le 2 septembre 2023.

4. <<https://www.larousse.fr/dictionnaires/francais/morphologie/>>, consultée le 2 septembre 2023.

mènes. Plusieurs modes de représentation sont possibles à partir des données que constituent ces jeux de paramètres : dans la plupart des cas, un grand nombre de paramètres sont extraits de l'analyse des phénomènes étudiés. Une représentation visuelle en deux (voir Fig. 3) ou trois dimensions permet à des agents humains de se figurer une partie de l'espace de paramètres mesurés, mais aussi d'extrapoler la façon dont un traitement informatique peut établir des corrélations entre un nombre bien supérieur de paramètres.

Les architectures d'AA utilisant des réseaux neuronaux artificiels ont la capacité d'organiser des ensembles de données dans des espaces abstraits qu'ils génèrent. Ces espaces sont appelés « espaces latents », car l'architecture d'IA organise des formes similaires plus proches les unes des autres. Par exemple, Fashion-MNIST¹ est un ensemble de données bien connu dans la littérature sur l'apprentissage automatique, qui est souvent utilisé pour comparer différentes architectures d'IA. Fashion-MNIST se compose de 70000 images en niveaux de gris de produits de mode d'une résolution de 28x28 pixels. Les images de l'ensemble de données sont réparties en 10 catégories : T-shirt, pantalon, pull-over, robe, manteau, sandale, chemise, chaussure, sac et botte de cheville. Pour illustrer la manière dont une architecture d'IA organise les formes dans un espace latent, nous avons entraîné une architecture d'apprentissage profond spécifique, appelée auto-encodeur variationnel (AEV). La figure 3 montre comment l'AEV entraîné a organisé les données dans Fashion-MNIST. Chaque point de la figure 3 représente une image de l'ensemble de données Fashion-MNIST projetée dans l'espace

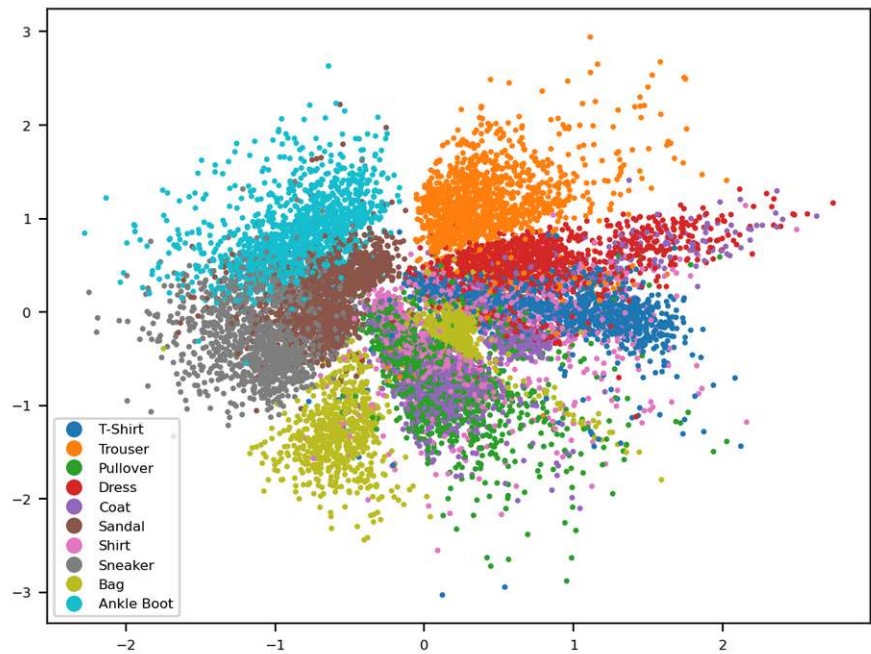


Fig. 3. Un espace latent bidimensionnel généré à l'aide d'un auto-encodeur variationnel (AEV) entraîné sur l'ensemble de données Fashion-MNIST

latent à l'aide de l'encodeur de l'AEV. Le succès de la formation est évident dans la façon dont les données d'image des mêmes catégories sont regroupées dans l'espace latent.

Après avoir organisé les données des ensembles de données de formation dans des espaces latents, les réseaux d'IA peuvent utiliser ces espaces pour générer de nouvelles formes. La figure 4 illustre les caractéristiques génératives de l'IA. Les axes x et y représentent les dimensions de l'espace latent 2D, et nous choisissons une plage arbitraire de [-1, 1] pour les deux dimensions. En effectuant une recherche de grille dans ces plages, nous avons généré une grille d'images dans la figure 4. La grille montre comment les dimensions continues de l'espace latent peuvent être utilisées pour générer des formes qui se situent entre les formes de l'ensemble de données d'apprentissage. La figure 4 montre également que les formes similaires sont organisées plus près les unes des autres dans l'espace latent.

On peut rapprocher ces grappes de points de données du concept de Boyd de « grappes de propriétés homéostatiques », en ce qu'elles permettent de caractériser des morphologies à partir de mesures de jeux de paramètres. En effet, dans cette conception, Boyd « identifie une classe de

1. Han XIAO, Kashif RASUL et Roland VOLLGRAF « Fashion-MNIST : A Novel Image Dataset for Benchmarking Machine Learning Algorithms », *arXiv e-prints*, 2017, <<https://arxiv.org/abs/1708.07747>>, consultée le 2 septembre 2023.

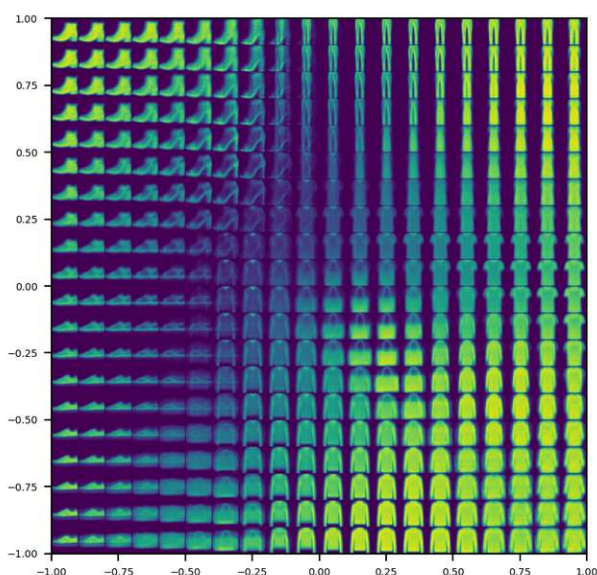


Fig. 4. Une image de grille générée à l'aide du même AEV entraînée que dans la figure 3. Les axes x et y représentent les dimensions de l'espace latent représentées dans la figure 3.

types, propriétés et relations naturelles, dont les définitions sont fournies, non pas par un jeu de conditions nécessaires et suffisantes, mais plutôt par un regroupement «homéostatiquement» soutenu de propriétés ou de relations ». Il précise plus tard qu'« il y a toujours une certaine indétermination ou «approximation» (*vagueness*) dans leurs extensions¹ ». Il s'agit de sa part d'une approche épistémologique pragmatique, qui ne cherche pas à expliquer la nature sous-jacente des choses, mais essaie de trouver une méthode efficace pour la catégoriser ou la simuler. Cette méthode s'avère particulièrement efficace dans son exploitation informatique, surtout depuis que des puissances de calculs considérables sont disponibles à tout un chacun à l'aide des importants progrès techniques réalisés ces dernières années par le calcul parallèle, notamment dans les cartes graphiques et les processeurs dédiés à l'AA, qui équipent maintenant tout ordinateur grand public.

1. Richard BOYD, « Homeostasis, species, and higher taxa », art. cit., p. 185, c'est nous qui traduisons « I identify a class of natural kinds, properties and relations whose definitions are provided, not by any set of necessary and sufficient conditions, but instead by a "homeostatically" sustained clustering of properties or relations [...] there is always some indeterminacy or "vagueness" in their extensions. »

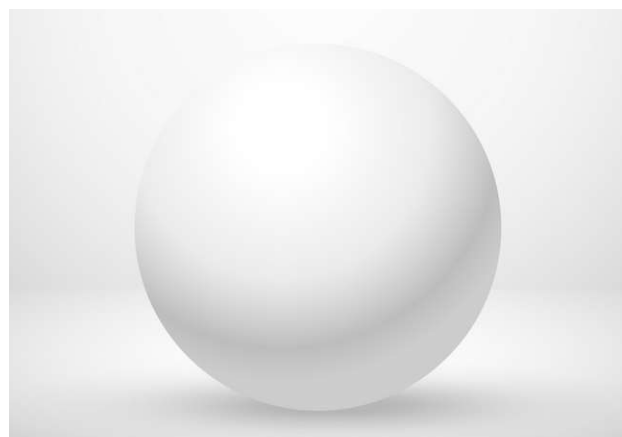


Fig. 5. Représentations d'une projection en deux dimensions d'une sphère et de son ombre

Espaces latents et modes d'apprentissage

Dans le domaine de l'intelligence artificielle, les procédés d'AA organisent ces espaces de données ou grappes de propriétés dans un espace abstrait – les « espaces latents » – où les points de données similaires sont représentés plus près les uns des autres, comme c'est le cas de manière simplifiée dans la figure 3, sous une forme graphique et intuitive, en définissant des zones de l'espace pour un très petit jeu de données. Puisqu'il s'agit d'AA, ce processus doit être mis en œuvre de manière automatisée, et ce sur des jeux de données composés d'un très grand nombre d'éléments. Les procédés permettant d'automatiser ces discriminations se basent généralement sur des techniques de « réduction de la dimensionnalité », dont une métaphore simple est représentée en figure 5². Il s'agit avec ces techniques de réduire le nombre de dimensions des phénomènes : dans le cas de la figure 5, une image en trois dimensions est rendue « lisible » dans un espace en deux dimensions par le truchement d'une projection de lumière virtuelle qui fait apparaître des ombres et donne ainsi une impression de sa volumétrie. Cependant, certaines informations sont perdues (pour la perception, dans le cas de cette métaphore),

2. Image de Starline, <https://fr.freepik.com/vecteurs-libre/sphere-blanche-dans-chambre-vide_824956.htm>, consultée le 2 septembre 2023.

puisque, bien que la sphère soit rendue de manière satisfaisante, il nous est impossible de savoir si un autre objet est situé derrière elle.

D'un point de vue morphologique, ces procédés de réduction de la dimensionnalité permettent donc de pouvoir reconstruire plus ou moins précisément – puisqu'il y a toujours, par définition, une « fonction de perte » permettant la réduction de dimensionnalité – les contours ou consistances, c'est-à-dire les morphologies des phénomènes analysés et ainsi de les classer et de les positionner dans les espaces latents produits par les processus d'AA. Un des points cruciaux, au niveau opérationnel, pour la constitution de ces « espaces latents » concerne l'établissement des taxonomies. Il s'agit pour cela de déterminer la façon dont est interprétée la réalité, ou autrement dit, la façon dont sont déterminées les grilles d'analyse à partir desquelles se fera la constitution de ces grappes multidimensionnelles de données.

Pour créer ces taxonomies, on utilise divers types d'algorithmes d'AA. L'apprentissage profond (*deep learning*) est une sous-catégorie de ces algorithmes d'apprentissage automatique, qui sont souvent appliqués pour créer les espaces latents des morphologies. Les algorithmes d'apprentissage automatique utilisent un ensemble de données d'apprentissage pour créer des espaces latents dans lesquels les données d'apprentissage sont organisées selon des mesures de similarité (ou de dissimilarité) définies. Ces mesures sont définies comme des formules mathématiques sous la forme de fonctions de perte. Le processus d'entraînement de l'algorithme d'AA applique une fonction de perte pour déterminer le succès de l'entraînement tout au long de celui-ci. En poursuivant l'optimisation et par la mise à jour de ses paramètres entre les étapes de son entraînement, l'algorithme d'AA constitue progressivement l'espace latent au long du processus d'entraînement, en fonction de la fonction de perte définie par l'utilisateur, c'est-à-dire de la mesure de similarité.

Les notions d'apprentissage supervisé et non supervisé font référence aux deux types d'apprentissage dans les algorithmes d'AA. Dans l'apprentissage supervisé, les données de des jeux de données d'entraînement sont associées à leurs étiquettes. Par exemple, un ensemble de données d'images peut contenir des étiquettes d'objets, tels

que voiture, maison, avion, etc. L'algorithme d'AA devient un outil de classification, dans ce scénario, et le processus d'apprentissage vise à généraliser la reconnaissance de différentes catégories à partir de données d'entrée. Le succès de l'algorithme d'AA est étudié en fonction de sa précision dans la reconnaissance correcte des catégories. En comparaison, l'apprentissage non supervisé utilise un ensemble de données ne comportant pas d'étiquettes. Dans ce type d'apprentissage, l'algorithme d'AA applique une mesure de similarité (c'est-à-dire une fonction de perte) pour créer un espace latent dans lequel les morphologies sont organisées en fonction de leurs similarités. Les modèles d'AA créés à l'aide de cette deuxième approche peuvent générer des formes qui ne figuraient pas dans l'ensemble de données d'apprentissage, en manipulant les paramètres des dimensions de l'espace latent.

Une fois structurés à l'issue de ces processus d'entraînement, ce sont ces espaces latents qui vont ensuite être utilisés pour générer de nouveaux artefacts, en fonctionnant de manière inverse, c'est-à-dire en choisissant des valeurs de paramètres, et en « projetant » du bruit (soit une génération de valeurs aléatoires) sur les zones visées de ces espaces latents, afin de reproduire des phénomènes similaires – quoique nouveaux et singuliers – à ceux qui avaient été analysés.

Essences et processualité

Mise en perspective épistémologique

Afin de mieux saisir le paradigme épistémologique qui sous-tend cette méthodologie de l'AA, il est important de la replacer dans une perspective historique : une première ébauche de cette méthodologie avait été initiée à l'origine des recherches en intelligence artificielle, dans les années 1960, mais a rapidement été supplantée par l'approche dite « cognitive ». Cette dernière cherchait à déterminer des règles explicites et symboliques sous-jacentes sur lesquelles se baseraient les systèmes de reconnaissance ou de génération automatique. Cette approche cognitive s'est finalement avérée ne livrer que des résultats assez décevants, mis à part dans certains domaines très particuliers. En

revanche, un champ de recherche indépendant, le « connexionnisme » s'est développé à partir du milieu des années 1980, en se distinguant nominalement de l'intelligence artificielle, et plus généralement de la recherche en informatique, dominées à l'époque par le cognitivisme. De manière schématique, on pourrait dire que, là où le cognitivisme cherche à dévoiler les règles sous-jacentes et donc à expliquer de manière extensive les mécanismes de causation, l'approche connexionniste va davantage se baser sur la mise à jour de corrélations statistiques pour détecter la proximité entre phénomènes, notamment à l'aide du paradigme informatique des « réseaux de neurones artificiels », dont est issu son nom. C'est cette approche qui est au cœur de l'AA et à l'origine de ses modèles.

En prenant un peu de recul sur la spécificité de ces deux démarches et en suivant Varela *et al.* (qui subsume le connexionnisme sous le courant de « l'émergence », voir Fig. 6), on peut renvoyer la distinction entre ces deux démarches de l'intelligence artificielle à l'opposition classique entre les courants épistémologiques de l'essentialisme et de l'empirisme, qui reposent respectivement sur l'idée que les structures conceptuelles qui sont l'objet des sciences (« les Lois de la Nature ») sont :

– soit au cœur de la réalité, dans les choses elles-mêmes, dans le cas de l'essentialisme.

– soit purement conventionnelles, voire arbitraires, dans le cas de l'empirisme.

Face à cette dichotomie entre essentialisme et empirisme, Varela propose le modèle de l'énaction, qui permet d'impliquer le corps et plus généralement le rapport au Monde et à autrui dans la cognition. D'une manière assez similaire, Boyd propose dans l'article que nous citons un modèle épistémologique qui cherche à trouver une troisième voie épistémologique en prenant en compte la processualité et le contexte social dans le fonctionnement concret de la production scientifique.

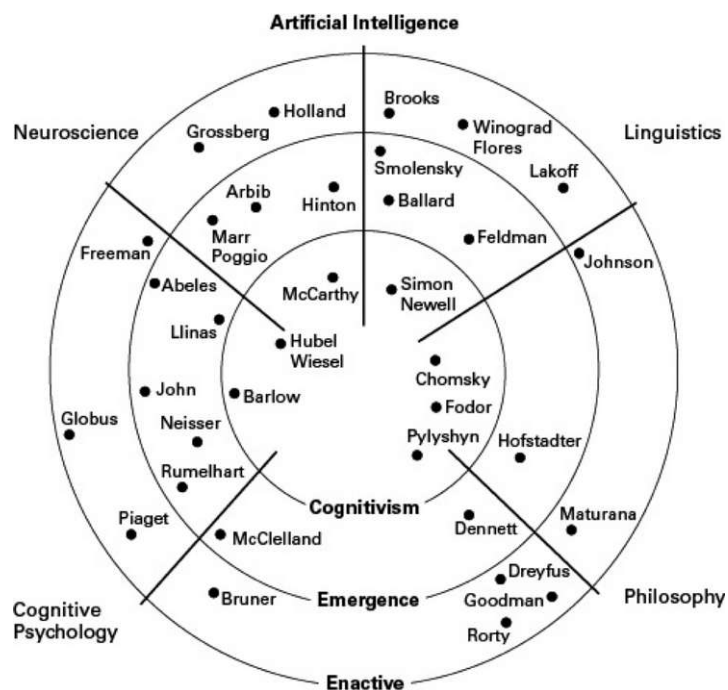


Fig. 6. les trois courants des sciences cognitives, présentés comme des cercles concentriques, avec leurs principaux représentants dans leurs différents champs scientifiques, selon Francisco J. VARELA, EVAN THOMPSON et ELEANOR ROSCH, *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA, USA, MIT press, 2017 [1992].

Essences et sens

En l'occurrence, Boyd rejette la conception essentialiste qu'il y aurait des « principes nécessaires et suffisants » dans le réel, qu'il s'agirait alors de révéler. Autrement dit, dans cette perspective les relations qu'établit l'activité scientifique ne sont pas *a priori*, mais *a posteriori*. Il retient cependant de l'essentialisme la critique de l'empirisme et de son nominalisme (notamment à partir de Locke), selon lequel toute conception scientifique ne serait que conventionnelle, voire arbitraire, ce qui interdirait aux sciences tout effet possible de prévision ou d'action sur le monde :

Ce qui est essentiel à la critique essentialiste du nominalisme lockien sur les types [...], c'est que les types d'une pratique scientifique (et quotidienne) réussie ne peuvent être définis par des « essences nominales » *a priori* purement conventionnelles. Au contraire, ils doivent être compris comme définis par des essences réelles *a posteriori* qui reflètent la nécessité de s'en remettre, dans nos pratiques classificatoires, à des faits concernant les structures causales du monde¹.

1. *Ibid.* p. 187, c'est nous qui traduisons : « what is essential

Cependant, ces essences ne sont plus dans ce cas des propriétés des choses « en elles-mêmes », mais un artefact humain, produit au contact des choses, dans une pratique donnée (une « matrice disciplinaire », pour reprendre ses termes) : « Les types naturels sont des artefacts sociaux. C'est pourquoi se demander si un type existe indépendamment de notre pratique est la mauvaise façon de s'enquérir de sa réalité. Aucun type naturel n'existe indépendamment de la pratique¹ ».

On retrouve une conception assez proche chez Deleuze – du côté de la création et plus de l'analyse – lorsqu'il explicite la notion d'essence chez Proust :

Mais l'essence, de son côté, n'est plus l'essence stable, l'idéalité vue, qui réunit le monde en un tout et y introduit la juste mesure. [Elle] n'est pas quelque chose de vu, mais une sorte de point de vue supérieur. [...] Le point de vue reste supérieur à celui qui s'y place, ou garantit l'identité de tous ceux qui y atteignent. Il n'est pas individuel, mais au contraire principe d'individuation².

Dans ce passage, Deleuze tire les conclusions des positions précédentes en affirmant le perspectivisme qu'il implique, et la valeur épistémologique comme esthétique de ce perspectivisme – au-delà de l'opinion individuelle, ou de la convention entre quelques pairs.

to the essentialist critique of Lockean nominalism about kinds [...] is that the kinds of successful scientific (and everyday) practice cannot be defined by purely conventional a priori "nominal essences." Instead they must be understood as defined by a posteriori real essences which reflect the necessity of our deferring, in our classificatory practices, to facts about causal structures in the world » – nous faisons le choix dans cet article de traduire le terme « natural kinds » par le terme « types ». En effet, ce terme est généralement traduit par « espèces naturelles » dans les écrits sur la biologie, qui sont le domaine d'origine de cette notion, mais le terme anglais de « kind » a une acception et une étendue bien plus larges, qui peuvent notamment s'appliquer aux sujets que nous traitons ici et s'appliquer à des démarches taxonomiques plus générales.

1. *Ibid.*, p. 200, c'est nous qui traduisons : « natural kinds are social artifacts. That's why asking whether a kind exists independently of our practice is the wrong way to inquire about its reality. No natural kinds exist independently of practice ».

2. Gilles DELEUZE, *Proust et les signes*, *op. cit.*, p. 133.

Deleuze développe cette question de l'adéquation entre les essences et les artefacts grâce au concept de « signe », à travers notamment l'exemple – que reprend Massumi, et qu'il reprend lui-même dans *Mille Plateaux* – d'une menuisier qui va construire une table en confrontant son geste à la singularité du bois, en construisant le sens de l'artefact à partir des signes rencontrés dans l'interaction avec la matière : « Le sens n'est pas dans la genèse de la chose, ni dans la pensée de cette genèse, ni dans les mots écrits ou prononcés à son sujet. Il est dans le processus qui conduit de l'un à l'autre³ ». Dans cette perspective (comme en un certain sens chez Boyd), le sens n'est préexistant, ni chez l'artiste, ni dans la matière (en l'occurrence la planche de bois), ni bien évidemment dans quelque idée platonicienne de l'objet visé « en soi » (ici la table), mais dans la façon dont il s'élabore au contact des choses, de la matière, dans le processus de création comme processus d'individuation, au sens de Simondon :

une telle individuation ce n'est pas la rencontre d'une forme et d'une matière préliminaire existant comme des termes séparés déjà constitués, mais une résolution qui surgit dans la piste d'un système méta-riche de potentiels : forme, matière et énergie préexistantes dans le système [...]. Le vrai principe d'individuation est la médiation et non l'individu ou l'élément singulier⁴.

Processus et opacité

Le principe d'individuation implique que les productions artistiques (ou scientifiques) sont l'effet de processus, et ont elles aussi une historicité – notamment les « périodes » de tel ou telle artiste – et gagnent donc à être pensées comme processus, tout comme les « types naturels » selon Boyd ne sont pas forcément des principes éternels – contrairement à la façon dont on les pense

3. Brian MASSUMI, *A user's guide to capitalism and schizophrenia*, *op. cit.*, p. 15, c'est nous qui traduisons : « Meaning is not in the genesis of the thing, nor in the thought of that genesis, nor in the words written or spoken of it. It is in the process leading from one to the other ».

4. Georges SIMONDON, *L'individuation psychique et collective*, Paris, Aubier, 1989, p. 13.

depuis Parménide¹. Afin de pouvoir prendre en compte ce caractère processuel, Boyd développe ce qu'il appelle sa « thèse de l'accommodation » : que ces essences ne sont valables que dans certains cadres théoriques, et certains contextes, mais qu'elles n'en gardent pas moins une efficacité pour l'induction et l'explication scientifiques : « Le groupe de propriétés est individualisé comme un objet ou un processus historique : certains changements dans le temps (ou dans l'espace) dans le groupe de propriétés ou dans les mécanismes homéostatiques sous-jacents préservent l'identité du groupe déterminant² ».

Boyd insiste par ailleurs sur le fait que le caractère « vague » de sa thèse et de sa notion de « grappe de propriétés » n'est pas un problème mais un atout – ce qui lui permet de s'adapter à la réalité qui est fondamentalement « désordonnée » (*messy*) : « À la lumière de cette thèse sur l'accommodation, j'explique pourquoi la référence à des types de groupes de propriétés homéostatiques “vagues” est souvent essentielle à une pratique inductive et explicative réussie dans les sciences³ ». Si cette conception épistémologique est, comme nous en posons l'hypothèse, un modèle pertinent pour comprendre le fonctionnement des systèmes de génération esthétique par AA – en particulier avec l'apprentissage profond, c'est à dire avec l'apprentissage non supervisé –, il y a probablement un parallèle intéressant à faire entre l'opacité, voire l'obscurité au sein desquelles ils fonctionnent, et l'insaisissabilité du processus de création artistique humaine. En effet, comme l'écrit Deleuze, toujours à propos de Proust

les essences vivent dans les zones obscures, non pas dans les régions tempérées du clair et du distinct. Elles sont enroulées dans ce qui force à penser, elles ne répondent pas à notre effort volon-

taire ; elles ne se laissent penser que si nous sommes contraints à le faire⁴.

Rappelons que le principe de l'AA consiste à chercher à reproduire des exemples en imitant le résultat visé, plutôt que de chercher des règles explicites qui permettraient d'expliquer le processus de leur genèse – comme cherchait à le faire l'approche cognitive. Cette question renvoie au « problème d'explicabilité », qui est souvent discuté dans les considérations théoriques et critiques sur l'apprentissage automatique : on ne sait pas comment les contenus sont concrètement produits, et cela fait qu'il est difficile de pouvoir contrôler leur qualité en amont de la production. Autrement dit, il est important de « garder à l'esprit [que ces systèmes] apprennent sans comprendre, donc qu'ils décident sans comprendre⁵ ». Il est à notre sens pertinent de rapprocher cet aspect du processus de création artistique humain, dans la mesure où, en tant qu'artiste, on ne « comprend » pas tout ce que l'on produit pendant le processus de création, et où une partie importante de ce qui est produit l'est de manière intuitive, semi-consciente, avec une part importante d'entraînement et d'*habitus* issu de la pratique artistique.

Quelle place pour les humains ?

L'inconvénient, dans le cas des systèmes de création par AA, est que ce processus n'est majoritairement pas conduit par des humains, mais de manière automatique, ce qui pose de nombreuses questions aussi bien esthétiques qu'éthiques ou sociales, que nous allons maintenant brièvement aborder. La problématique récurrente parmi ces questionnements revient à s'interroger sur la place des humains dans ces évolutions techniques, parfois sur la base de craintes plus ou moins fondées (par exemple sur la disparition des artistes à cause

1. Cf. à ce sujet Paul FEYERABEND, *Philosophie de la nature*, Paris, Le Seuil, 2014.

2. Richard BOYD, « Homeostasis, species, and higher taxa », art. cit., p. 187, c'est nous qui traduisons : « the property cluster is individuated like a [...] historical object or process: certain changes over time (or in space) in the property cluster or in the underlying homeostatic mechanisms preserve the identity of the defining cluster ».

3. *Idem*.

4. Gilles DELEUZE, *Proust et les signes*, op. cit., p. 62.

5. Laurence DEVILLERS, citée dans Benoît GEORGES, « Le talon d'Achille de l'intelligence artificielle », *Les Échos*, 15 mai 2017, <<https://www.lesechos.fr/2017/05/le-talon-dachille-de-lintelligence-artificielle-168099>>, consultée le 2 septembre 2023.

de l'efficacité croissante de ces systèmes), que nous n'aborderons pas en elles-mêmes ici. En revanche, il peut être intéressant de se demander quels sont les endroits où l'intervention humaine reste nécessaire, voire cruciale dans le fonctionnement des systèmes existants d'AA.

Établissement des taxonomies

Pratiquement, dans le cas qui nous intéresse, de l'usage de l'AA pour les arts, le perspectivisme dans lequel nous nous plaçons souligne l'importance, avant toute autre chose, du choix de la grille d'analyse qui sera mise en œuvre afin d'extraire des phénomènes présentés au dispositif des caractéristiques mesurables, qui seront utilisées pour constituer les modèles – et qui aura donc une incidence cruciale sur la pertinence du résultat qui en sera produit. C'est également le point de vue de Boyd, lorsqu'il écrit que « Quine avait raison de dire dans *Natural Kinds*¹ que la théorie des types naturels concerne la manière dont les schémas de classification contribuent à la formulation et à l'identification d'hypothèses projectibles (au sens de Goodman²)³ ».

En effet, ces « espaces latents », constitués de grappes de points de données, nécessitent de passer le réel dans une grille d'analyse, et un exemple simple, présenté dans les figures 7a et 7b, permet d'en mieux saisir l'importance. Dans cet exemple, la représentation d'un même phénomène peut se faire selon deux types de coordonnées – cartésiennes (X et Y) ou polaires (distance vis-à-vis

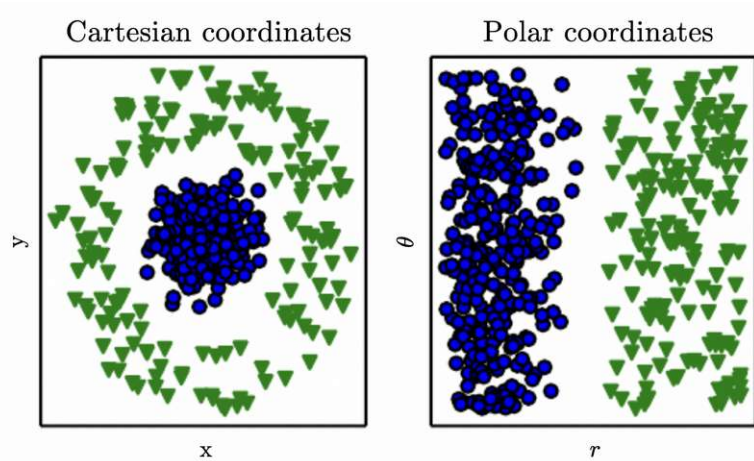


Fig. 7a et 7b. Deux modes de représentation, en coordonnées cartésiennes et polaires, d'un même jeu de données (source : Ian GOODFELLOW *et al.*, *Deep Learning*, Harvard, MIT Press, 2016)

d'un « pôle » et angle à partir d'une origine, d'un « méridien ») – afin de discriminer entre deux grappes de propriétés. Comme on peut le constater à la vue de cette illustration, le mode de représentation d'un même jeu de donnée aura un impact considérable sur la façon dont les données pourront être interprétées par un système de classification automatique.

Supposons que nous disposions d'un système de classification par apprentissage automatique sous la forme $\text{dimension}_2 = \alpha * \text{dimension}_1 + \text{bêta}$. Ce système d'AA utilise cette formule pour classer les points de données d'entrée en deux catégories. Étant donné un point de données (fI, gI), si $gI < (\alpha * fI + \text{bêta})$, les données d'entrée se situent en dessous de la limite tracée par le système d'AA et sont donc prédites comme appartenant à la classe 1, c1. De même, si $gI > (\alpha * fI + \text{bêta})$, les données d'entrée se situent au-dessus de la limite tracée par le système d'AA, et sont donc prédites comme appartenant à la classe 2, c2. La formation du système d'AA recherche progressivement les meilleures valeurs pour alpha et bêta afin que le plus grand nombre de données de l'ensemble de données de formation soient classées correctement.

Il est possible de représenter les données d'entrée de différentes manières. Les figures 7a et 7b illustrent le même ensemble de données d'apprentissage avec deux classes, dans deux systèmes de coordonnées différents, cartésien et polaire. Supposons que l'ensemble de données soit initiale-

1. Willard Van Orman QUINE, *Relativité de l'ontologie*, Paris, Aubier-Montaigne, 1977.

2. Nelson GOODMAN, *Faits, fictions et prédictions*, M. Abran (trad.), Paris, Minuit, 1985.

3. Richard BOYD, « Homeostasis, species, and higher taxa », art. cit., p. 188, c'est nous qui traduisons : « Quine was right in "Natural Kinds" (1970) that the theory of natural kinds is about how schemes of classification contribute to the formulation and identification of projectible hypotheses (in the sense of Goodman 1973) ». Nous avons indiqué dans les deux notes précédentes les traductions françaises de la bibliographie indiquée par Richard Boyd.

ment représenté en coordonnées cartésiennes. Les dimensions de nos systèmes d'AA seraient $d1 = x$ et $d2 = y$. Dans ce scénario, la précision maximale atteinte par le système d'AA est de 50 %, en traçant une frontière de classification à l'aide de $y = \alpha * x + \beta$, quelles que soient les valeurs d' α ou de β trouvées pendant la formation. Toutefois, si nous utilisons une représentation différente des données d'entrée, nous pourrions obtenir une plus grande précision avec le même système d'AA. Si nous projetons les données d'entrée des coordonnées cartésiennes aux coordonnées polaires, les dimensions des systèmes d'AA seront $d1 = r$ et $d2 = \theta$. Cette étape de projection est appelée prétraitement ou extraction de caractéristiques dans la littérature sur l'AA. Dans le scénario des coordonnées polaires, notre système d'AA pourrait atteindre une précision de 100% en utilisant la formule $r = \alpha * \theta + \beta$; plus précisément, $\alpha = 0$ et β est une valeur quelconque dans la région vide située entre deux classes dans la figure 7b.

Une fois les critères de la grille d'analyse définis, ils vont généralement être évalués à l'aide d'implémentations informatique de la modélisation par « réseaux de neurones », constitués d'un ensemble connecté de modules informatiques répondant au nom de « perceptron ». Le modèle du perceptron est représenté sur la figure 8 : ses différentes entrées de paramètres a_n sont chacune pondérées par un coefficient w_n , afin de pouvoir donner des importances différentes, et donc de hiérarchiser les différents paramètres pour l'apprentissage et la classification des morphologies. Le résultat est représenté par la formule sur la droite du schéma. C'est ce système simple qui est répété de nombreuses fois pour de nombreux paramètres, à travers la fonction d'activation σ , sur

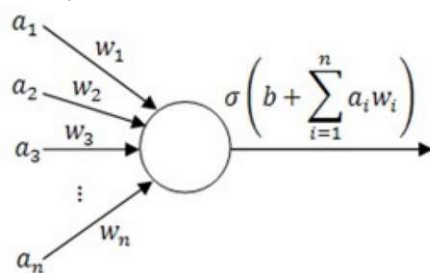


Fig 8. Représentation schématique du fonctionnement du perceptron

des « couches » de complexités croissantes : une première permettra par exemple, dans le cas d'une reconnaissance de forme, de reconnaître les bords des motifs présents dans une image, quand les couches supérieures pourront se charger de paramètres morphologiques plus complexes.

C'est la combinaison de ces deux étapes qui va permettre de laisser l'algorithme d'AA constituer des espaces latents en les appliquant à de grands ensembles d'entrées. Comme on l'a déjà vu plus tôt, l'une et/ou l'autre de ces deux étapes peuvent être automatisées dans le cas d'apprentissages non supervisés (ou semi-supervisés), mais il n'en reste pas moins qu'une action humaine aura la plupart du temps lieu à un ou plusieurs stades du processus : au début, en cours de route ou à son issue. Comme l'écrit Stephen Wolfram dans sa très éclairante et pédagogique présentation du fonctionnement du système *ChatGPT* : l'entraînement des réseaux de neurones « est essentiellement un art¹ ». L'action humaine sur ces dispositifs est l'objet de la prochaine sous-section.

Homéostasie et régulation

En effet, en ce qui concerne les procédés d'AA, même si l'on peut être tenté – notamment par cause des nombreux discours de fascination, qu'elle soit technophile ou technophobe – de parler de « créativité machinique » voire d'assister à un détournement de la notion d'« autopoïèse² »

1. Stephen WOLFRAM, « What Is ChatGPT Doing... and Why Does It Work ? », février 2023, <<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>>, consultée le 2 septembre 2023 – ce passage en particulier permet de mieux comprendre comment cet entraînement de réseaux de neurones se fait en relation à leurs versions successives : « De plus en plus, il ne s'agit pas de former un réseau *ex nihilo* : au contraire, un nouveau réseau peut soit incorporer directement un autre réseau déjà formé, soit au moins utiliser ce réseau pour générer davantage d'exemples de formation pour lui-même » – c'est nous qui traduisons : « increasingly one isn't dealing with training a net from scratch: instead a new net can either directly incorporate another already-trained net, or at least can use that net to generate more training examples for itself ».

2. Cette notion est utilisée notamment par Varela et Maturana pour décrire l'organisation autonome des systèmes, en particulier vivants: selon eux, une « machine autopoïétique engendre et spécifie continuellement sa propre organisa-

– c'est-à-dire d'autonomie des machines dans la production esthétique de tels dispositifs – il est particulièrement rare de ne pas faire intervenir d'humains à un stade ou un autre du processus – et en particulier au début et à la fin.

En particulier, du fait du caractère itératif de la plupart des processus d'apprentissage, cette action humaine peut être réinjectée dans le système entre ses itérations successives et ainsi agir comme une régulation du processus. C'est d'ailleurs depuis l'introduction de tels mécanismes de rétroaction (souvent nommés « *feedback* », selon le terme anglais) dans les années 1980 que les systèmes d'AA ont pu faire un bond qualitatif, notamment à partir de la possibilité de procéder à une propagation inverse des retours de rétroaction issus de l'évaluation humaine : en substance, il s'agit à l'aide de ce mécanisme de propager une modification de manière récursive sur toutes les couches successives de réseaux de neurones afin d'en améliorer les résultats à tous les niveaux.



Fig. 9. Représentation schématique du phénomène de rétroaction dans l'homéostasie

Ce principe de rétroaction est à la base de la cybernétique, telle qu'énoncée par Norbert Wiener¹, et qui a depuis fait les preuves de sa pertinence, en particulier pour tout ce qui concerne les interactions entre les humains et les machines. En effet, même si les systèmes actuels les plus avancés de génération de contenu par AA semblent à

première vue se passer d'action humaine (au-delà de la requête initiale de génération, bien entendu, et en omettant le fait qu'ils sont le résultat cumulatif de plusieurs générations de développement, avec à chaque fois des ajustements humains), ils sont en réalité la plupart du temps soumis à des phases de régulation, au moins partiellement encadrées par des interventions humaines. Un des cas les plus « spectaculaires » de cette régulation concerne le cadrage, notamment éthique, des systèmes de génération de textes ou d'images par AA, qui sans cela présentent très rapidement de très forts biais, par exemple racistes ou sexistes². En effet, les systèmes d'apprentissage fondés sur le principe des grands modèles de langage (tel que c'est le cas de *ChatGPT*) se contenteraient sans cela d'imiter sans le moindre discernement le corpus textuel sur lequel ils ont été entraînés, qui a été constitué de manière à être le plus massif, sans la moindre considération éditoriale. Il est donc normal qu'un tel système reproduise alors les biais qui se trouvaient dans une grande partie des textes utilisés pour l'entraînement : ces systèmes ne font ainsi que refléter les mécanismes de domination existants dans le monde humain, et le fait qu'ils se basent sur des textes passés tend à créer un effet d'inertie sur les représentations sociales, et à leur faire « tenir des propos » qui ne sont plus tenus pour acceptables de nos jours.

Cette régulation de biais est en général effectuée à l'aide de procédés d'« apprentissage par renforcement », qui se combinent aux procédés d'apprentissage supervisé ou semi-supervisé. Cet apprentissage par renforcement fonctionne à l'aide d'un mécanisme comprenant plusieurs agents (informatiques, notamment les « chaînes de Markov cachées »), un interpréteur (ou aussi discriminatoire) et un environnement. Le principe est de récompenser l'agent en fonction de l'efficacité

tion », cf FRANCISCO J. VARELA, *Autonomie et connaissance : essai sur le vivant*, Paris, Seuil, 1989. En effet, notre thèse est ici que les machines d'apprentissage automatique, bien qu'engendrant leur propre organisation technique interne, ne produisent pas un contenu esthétique autonome, car celui-ci est toujours conditionné par l'imitation de phénomènes générés par des humains, et surtout destinés à leur perception et à leur jugement.

1. NORBERT WIENER, *Cybernetics or Control and Communication in the Animal and the Machine*, Cambridge, MIT press, 2019 [1948].

2. Cf. par exemple Fabian OFFERT and Thao PHAN, « A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space », *arXiv e-prints*, octobre 2022, <<https://arxiv.org/abs/2211.06323>>, consultée le 2 septembre 2023, et Abeba BIRHANE, Vinay Uday PRABHU, et Emmanuel KAHEMBWE, « Multimodal datasets: misogyny, pornography, and malignant stereotypes », *arXiv e-prints*, octobre 2021, <<https://arxiv.org/abs/2110.01963>>, consultée le 2 septembre 2023.

des décisions qu'il prendra sur l'environnement, jugées par l'interpréteur en fonction de critères d'efficacité préalablement définis par les humains. C'est à l'aide d'un tel procédé, intitulé *InstructGPT*¹ que le système de génération automatique *ChatGPT* que nous avons mentionné plus haut a pu aboutir à l'efficacité remarquable dont il fait preuve, en intégrant les préférences humaines en retour aux résultats produits. Cependant, comme cela est indiqué sur le site de *ChatGPT* lui-même², le résultat qu'il produit peut avoir une apparence de véracité, sans que le contenu réel n'ait aucun intérêt ni pertinence. En quelque sorte, ce système est avant tout un générateur de sophisme, dont le ton particulièrement crédible rend difficile, même à des spécialistes du domaine abordé, de détecter les biais, erreurs ou contre-sens sur lesquels il se base pour construire des discours plausibles quoique faux³. C'est pourquoi, à notre sens, il est crucial de penser les systèmes d'AA comme des outils d'aide à la production intellectuelle ou artistique plutôt que de substituts à celles-ci : la responsabilité de produire un contenu sensé reste celle des humains, et il est pour ainsi dire déraisonné de faire confiance aveuglément à ces systèmes et de les prendre pour argent comptant.

Usages artistiques et énonciation

En ce qui concerne notre sujet – la création artistique – si les résultats produits par les systèmes de génération par AA sont ici aussi de plus en plus impressionnants de véracité, le problème est en quelque sorte le même : comme dans la publicité pour une certaine boisson gazeuse canadienne, si ces résultats ressemblent de plus en plus à de l'art, on peut néanmoins légitimement se demander s'il y a au final une quelconque « force » ou singularité esthétique propre aux artefacts que ces systèmes

produisent. Tout comme cette dernière interrogation, les considérations qui vont suivre à ce sujet, ne sont que des hypothèses et interrogations, et demandent à être mises au défi de l'inventivité technique des chercheur·euses et développeur·euses du domaine de l'AA créative, dans les années à venir. Il peut notamment être intéressant d'observer les limites ou faiblesses consubstantielles de ces dispositifs, notamment en regard des pratiques artistiques qui s'y confrontent, ou qui se sont confrontées dans le passé à des questionnements similaires.

Le point le plus fondamental qu'il nous semble utile d'interroger concerne la question de l'imitation : en effet, bien que la pratique artistique implique que les humains s'appuient sur l'existant, et apprennent eux aussi par imitation, il est ainsi admis qu'un·e artiste dépassera toujours ses influences et son histoire personnelle pour amener une singularité, ou autrement dit quelque chose de radicalement nouveau dans sa création. C'est sur cette position que se base par exemple Proust, et après lui Deleuze, dans leurs considérations que nous avons préalablement citées. De même, de l'avis de l'auteur et chanteur Nick Cave, tel qu'il le formule dans une réponse à une question posée sur son site lui demandant ses impressions sur une chanson générée par *ChatGPT* « dans le style de Nick Cave » :

Ce qui fait la grandeur d'une grande chanson n'est pas sa ressemblance étroite avec une œuvre reconnaissable. Écrire une bonne chanson n'est pas du mimétisme, de la réplique ou du pastiche, c'est le contraire. C'est un acte de meurtre de soi-même qui détruit tout ce que l'on s'est efforcé de produire dans le passé⁴.

L'affirmation de Nick Cave peut donner l'impression d'une posture romantique, convoquant la figure discutable du Génie, mais au-delà de cette

1. Long OUYANG *et al.*, « Training language models to follow instructions with human feedback », mars 2022, <<https://arxiv.org/abs/2203.02155>>, consultée le 2 septembre 2023.

2. Cf. par exemple Arvind ONARAYANAN et Sayash KAPOOR, « ChatGPT is a bullshit generator. But it can still be amazingly useful », *AI snake oil*, 6 décembre 2022, <<https://aisnakeoil.substack.com/p/chatgpt-is-a-bullshit-generator-but>>, consulté le 2 septembre 2023.

3. *Ibid.*

4. C'est nous qui traduisons : « What makes a great song great is not its close resemblance to a recognizable work. Writing a good song is not mimicry, or replication, or pastiche, it is the opposite. It is an act of self-murder that destroys all one has strived to produce in the past », Nick CAVE, sans titre, janvier 2023, <<https://www.theredhandfiles.com/chat-gpt-what-do-you-think/>>, consultée le 2 septembre 2023.

posture, il dit aussi à notre sens quelque chose de fondamental à la création artistique, qui concerne plus spécifiquement la notion de style, ici en question : si le style est avant tout une question de singularité, dans quelle mesure un procédé informatique, basé uniquement sur l'imitation peut-il jamais prétendre à cette singularité ? Comme le dit à nouveau Nick Cave : les œuvres artistiques « se fondent sur la lutte humaine interne, complexe de la création et, pour autant que je sache, les algorithmes n'ont ni sensations ni sentiments¹ ».

Or il nous semble que, quelle que soit la sophistication à laquelle peuvent être amenés les systèmes d'AA, et plus généralement d'intelligence artificielle, aucun de ces deux aspects – sensation et sentiments – ne sont visés dans leurs développements. Même si le domaine de l'informatique affective (ou « *affective computing* » en anglais) transpose le type de traitement mis en œuvre dans l'AA vers le domaine des émotions humaines et des sentiments humains, elle ne prétend pas à plus que de l'imitation. De même, des systèmes de captation de plus en plus complexes permettent de mesurer certains aspects du monde physique, mais ils ne sont en rien comparables à l'intrication complexe dans laquelle ont lieu les sensations humaines. Pour le dire plus simplement, comme le fait Ben Medlock, spécialiste des systèmes de reconnaissance de geste par AA : « Le corps est le chaînon manquant pour des machines véritablement intelligentes² ».

Cette position rejoint celle de Varela, Thompson et Rasch, notamment dans *The Embodied Mind*, qu'ils modélisent avec le concept d'*énaction* : à l'opposé du cognitivisme, qui cherche à réduire la réalité à un ensemble de règles explicites, désincarnées, et en prolongement critique de l'émergentisme, qui traite la réalité comme une combinaison probabiliste, ces auteurs affirment l'import-

tance du corps et de l'expérience vécue. Ils se placent ainsi dans le prolongement de la phénoménologie et dans son hybridation avec les sciences cognitives. C'est dans cette même perspective qu'il nous semble fertile de se situer, en ce qui concerne le sujet que nous abordons ici : les usages de l'AA dans la création artistique. Ainsi, quelle que soit l'efficacité des systèmes de génération esthétique par AA, il nous semble que l'action d'un (ou plusieurs corps) est toujours nécessaire afin que le résultat soit artistiquement prégnant. D'un point de vue opérationnel, la question n'est alors plus si ce que produisent ces systèmes est en soi bon ou pas artistiquement, mais : comment s'en saisir artistiquement et y impliquer le corps et la créativité humaines ?

Bien évidemment les réponses à cette question sont innombrables, et dépendent *in fine* de chaque pratique singulière, mais nous pouvons néanmoins ébaucher quelques pistes hypothétiques. D'une manière générale, cette hypothèse implique qu'il est nécessaire de ménager des dispositifs de régulation pour les humains, afin qu'ils puissent adapter au mieux leurs utilisations de ces systèmes à leurs usages créatifs singuliers. Ces dispositifs de régulation peuvent se situer principalement aux deux stades principaux du processus d'AA, comme nous l'avons vu précédemment : au moment de la constitution des espaces latents, par l'établissement de taxonomies les plus pertinentes, d'une part, et d'autre part au moment de leur restitution, afin d'infléchir ou corriger – notamment de manière itérative, au fil de plusieurs essais et expérimentations – les résultats produits par les systèmes de génération par AA. Dans les deux cas, il nous semble important de pouvoir ménager pour les utilisateur·rices autant que possible de points de contrôle sur les paramètres des systèmes de classification et de génération.

Au stade de l'apprentissage, donc de l'établissement des espaces latents à partir de corpus de données, il faudrait idéalement que l'usagè·e puisse décider de la façon dont les choix sont faits dans les systèmes de classification et de recommandations, c'est-à-dire qu'elle ou il puisse déterminer avec autant de flexibilité et d'extension que possible les critères et orientations des algorithmes de classification, c'est-à-dire les paramètres des taxonomies, comme on l'a vu plus

1. *Ibid.*, c'est nous qui traduisons : « they are predicated upon the complex, internal human struggle of creation and, well, as far as I know, algorithms don't feel » – nous avons cherché dans notre traduction de la fin de la citation à rendre le double sens anglais de *feel* en utilisant les notions de sensation et de sentiment.

2. Ben MEDLOCK, « The body is the missing link for truly intelligent machines », *aeon.co*, mars 2017, <<https://aeon.co/ideas/the-body-is-the-missing-link-for-truly-intelligent-machines>>, consultée le 2 septembre 2023.

haut. En effet, l'immense majorité des systèmes existants se fondent sur des espaces latents qui ont été constitués de la même manière pour tous les utilisateur·rices des systèmes génératifs. Un des problèmes importants que pose cependant cette hypothèse, au-delà de l'aspect technique, est le coût écologique de cet entraînement¹. Les systèmes actuels d'apprentissage ne sont pas conçus de manière à pouvoir permettre ce genre d'usages, mais peut-être serait-il possible à l'avenir de laisser lors de l'apprentissage des marges de manœuvre permettant l'inflexion après-coup des paramètres sur lesquels aura été constitué un espace latent ?

Au stade de la génération des artefacts, il s'agit également, afin d'augmenter l'expressivité de ces dispositifs, de pouvoir donner autant que possible de contrôle sur les résultats, et ce en réduisant autant que possible la durée de la boucle entre perception et rétroaction de la part de l'utilisateur·rice. En effet, dans la majeure partie des dispositifs existants à la date de rédaction de cet article, seule une requête initiale est disponible, qui fournit un résultat donné, et il est la plupart du temps impossible d'ajuster ce résultat après-coup sans relancer l'ensemble de la génération, pour parfois obtenir un résultat totalement différent. En quelque sorte (et cela n'est pas étranger au domaine d'origine des auteur·rices : la musique) le développement d'« instruments » dédiés à l'interaction avec ces systèmes nous semble être une piste prometteuse, qui permettrait d'aller vers une approche plus éactive de l'interaction avec ces systèmes. De telles interfaces éactives² permettraient ainsi de mettre en œuvre des boucles de rétroaction plus serrées entre la production, la perception et l'évaluation des résultats produits,

afin d'en infléchir de manière bien plus fine et expressive. Pour l'instant, les systèmes de génération existants ne sont cependant pas assez rapides pour que cela soit réellement réalisable, mais étant donné le dynamisme du domaine de l'AA, il y a fort à parier que le niveau de performance nécessaire à ce que ceci devienne possible sera atteint sous peu, d'ici quelques années au plus.

Le point commun entre ces propositions est qu'elles mettent l'accent sur l'importance de l'heuristique dans le processus de création, du fonctionnement par tâtonnements, qui est également le principe sur lequel est fondé l'apprentissage automatique : le système apprend de manière empirique, et progresse en fonction des ajustements et retours qu'il reçoit, que ce soit par les développeur·ses ou par les utilisateur·rices. Par ailleurs, cette perspective favorise le potentiel, déjà implicitement offert par le fonctionnement de ces systèmes génératifs de « faire le système à sa main », autrement dit d'affiner le type de résultats qu'ils produisent pour qu'ils puissent correspondre au mieux au style de son utilisateur·rice, de manière évolutive et dynamique.

Enfin, pour en revenir à l'état actuel des systèmes génératifs par AA, avec toutes les limitations qu'ils présentent, il reste possible de s'en servir d'une manière créative, aussi « factice » que puisse paraître leur production en tant que telle. En effet, ce problème n'est pas en soi différent de celui qui se pose dès le début de l'usage de procédés génératifs – comme par exemple avec l'usage par le compositeur Iannis Xenakis des procédés stochastiques : *in fine* il reste toujours aux artistes la décision poétique et artistique finale d'utiliser tel ou tel résultat généré, les façons de les agencer, et la manière de les « bricoler » au fur et à mesure de leur génération. Cette problématique est en effet comparable à l'émergence de l'usage des samples en musique (le « *sampling* ») dans les années 1980 (ou même, avant cela, dès l'émergence de la musique électroacoustique) : si ceux-ci ne sont pas des éléments originaux en eux-mêmes, ils peuvent devenir un matériau à remodeler et participer en cela à des créations véritablement singulières. Ce n'est donc pas un hasard si on retrouve aujourd'hui avec les produits des systèmes génératifs par AA des débats – en particulier sur la question de l'auctorialité – quasiment

1. Comme cela est évoqué dans Katyanna QUACH, « AI me to the Moon... Carbon footprint for training GPT-3 same as driving to our natural satellite and back », *The Register*, novembre 2020, <https://www.theregister.com/2020/11/04/gpt3_carbon_footprint_estimate/>, consultée le 2 septembre 2023, avec un certain humour (noir), l'empreinte carbone de l'entraînement du système sur lequel se base *ChatGPT* est équivalent à un aller-retour terre-lune en voiture thermique.

2. Cette approche a notamment été initiée dans le domaine artistique au sein du réseau européen d'excellence pour les interfaces éactives, cf. notamment Claude CADOZ, « Enactive Interfaces ? » dans *The Virtual Workshop of the 7th Enactive Conference*, 2004.

identiques à ceux qui avaient cours au début de l'utilisation du *sampling*. Cela est avant tout dû à notre sens à la nouveauté du procédé, et en particulier à son aspect spectaculaire, qui devrait selon toute vraisemblance s'évaporer pour laisser la place à des questionnements esthétiques et éthiques, voire politiques, plus fertiles.

Conclusion

Si l'on revient aux définitions de l'intelligence artificielle et en particulier des techniques d'apprentissage que nous citons au début de cet article, le caractère principal des résultats que produisent les systèmes génératifs que nous avons examinés est qu'il peut être « comparable » à des résultats humains. Comme nous l'avons vu par la suite, cela est possible car l'apprentissage automatique procède par imitation – et en particulier par l'imitation de *styles*. Cependant, quelque soit leur vraisemblance et leur comparabilité à des artefacts produits par des humains, les résultats de cette imitation peuvent difficilement être esthétiquement satisfaisants en eux-mêmes, de par leur manque de singularité consubstantielle. Afin de pouvoir en faire un usage artistique, et pour reprendre les termes de Deleuze, Guattari et Massumi, il s'agit alors d'utiliser ces éléments comme matériaux et de leur donner de la consistance, afin que les éléments du résultat « tiennent ensemble ».

Face aux débats qui agitent les professionnels et l'opinion publique sur le rôle à venir de ces systèmes, et sur les risques potentiels qu'ils pourraient faire courir, pour beaucoup motivés par la peur des fantasmes sur la « singularité », ou autres scénarios de prise de pouvoir par les machines, il nous semble important voire urgent de replacer le débat sur les moyens d'agir sur ces procédés, de prendre ou reprendre le pouvoir sur ce qu'elles produisent et sur la façon dont elles le font – en effet il semble bien que le pouvoir soit déjà pris : non pas par les machines mais par les intérêts des multinationales qui les développent. Face à cette

prise de pouvoir et à leurs effets sociaux et esthétiques, il nous paraît important de remettre dans les mains des utilisateur·rices (artistes, chercheur·euses ou simples usagers) les moyens d'agir sur la conception des systèmes d'AA et en particulier sur les modalités d'interaction et de personnalisation qu'ils leur permettent. C'est à cette fin que nous avons tenté de donner quelques éléments de compréhension du fonctionnement de ces systèmes et des enjeux épistémologiques et esthétiques qu'ils soulèvent.

Pia BALTAZAR et KIVANÇ TATAR