



HAL
open science

Résister aux boîtes noires. Design et intelligence artificielle

Anthony Masure

► **To cite this version:**

Anthony Masure. Résister aux boîtes noires. Design et intelligence artificielle. Cités, 2019, L'intelligence artificielle : enjeux éthiques et politiques, N°80 (4), pp.31. <10.3917/cite.080.0031>. <hal-02917049>

HAL Id: hal-02917049

<https://hal.science/hal-02917049v1>

Submitted on 26 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

RÉSISTER AUX BOÎTES NOIRES. DESIGN ET INTELLIGENCES ARTIFICIELLES

Anthony Masure

Presses Universitaires de France | « Cités »

2019/4 N° 80 | pages 31 à 46

ISSN 1299-5495

ISBN 9782130821151

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-cites-2019-4-page-31.htm>

Distribution électronique Cairn.info pour Presses Universitaires de France.

© Presses Universitaires de France. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

*Résister aux boîtes noires. Design et intelligences artificielles*¹

ANTHONY MASURE

À en croire Henry A. Kissinger, ancien secrétaire d'État du président américain Richard Nixon, « l'époque des Lumières se termine : philosophiquement, intellectuellement – à tous points de vue – la société humaine n'est pas préparée à l'essor de l'intelligence artificielle² ». Selon lui, la logique rationnelle ayant façonné le monde actuel serait désormais bouleversée par une nouvelle révolution technologique dont nous n'aurions pas encore tiré toutes les conséquences : « son aboutissement pourrait être un monde reposant sur des machines alimentées par des données et des algorithmes et non pas sur des normes éthiques ou philosophiques³ ».

L'intelligence artificielle (IA) est régulièrement agitée comme étendard d'un progrès à la fois fascinant et révoltant, et ce alors même que son concept est déjà ancien. En effet, dès les années 1950, le mathématicien Alan Turing et les équipes du National Physical Laboratory (NPL) rendent public l'Automatic Computing Engine (ACE), une des premières machines programmables. Poussant la vieille distinction philosophique entre le corps et l'esprit (*hardware/software*⁴) à son paroxysme, Turing en

1. Cette contribution s'appuie en partie sur une conférence éponyme donnée par l'auteur en septembre 2018 à l'Institut supérieur des arts multimédia de la Manouba (Tunis) à l'occasion du colloque « MCRIDD ». Voir : <<http://www.anthonymasure.com/conferences/2018-09-isamm-design-ia-tunis>>. Site consulté le 11 septembre 2019.

2. Henry A. Kissinger, « How the Enlightenment Ends. Philosophically, intellectually – in every way – human society is unprepared for the rise of artificial intelligence », *The Atlantic*, juin 2018, [En ligne], <<https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>>. Site consulté le 11 septembre 2019.

3. *Ibid.*

4. On doit cette distinction au mathématicien John von Neumann. Voir Boris Saulnier, « Compte-rendu de l'ouvrage de John von Neumann *L'Ordinateur et le cerveau* [1955–1956] »,

vient à considérer explicitement la possibilité d'un cerveau électronique. Dans son article « Les ordinateurs et l'intelligence ¹ », publié la même année, Turing se demande à quelles conditions une entité calculante pourrait être douée d'intelligence et, le cas échéant, comment le reconnaître du point de vue humain. La capacité d'une machine à faire preuve d'intelligence est envisagée par Turing sous l'ordre de la simulation, et ce peu importe que personne ne soit en mesure de comprendre le fonctionnement interne de la machine :

Nous voulons [...] accepter la possibilité [...] qu'une équipe d'ingénieurs puisse construire une machine qui fonctionne, mais dont les modalités de fonctionnement ne peuvent être décrites de manière satisfaisante par ses concepteurs, parce qu'ils ont appliqué une méthode en grande partie expérimentale ².

Cette prise de distance avec la nécessité de rendre intelligibles les systèmes techniques fait écho au champ de la cybernétique ³, dont les principes généraux ont déterminé de nombreux systèmes contemporains de calcul, d'interfaces, et d'interaction qui façonnent encore en grande partie l'époque contemporaine. La cybernétique de Wiener reprend l'idée du mathématicien Von Neumann qu'une machine (qu'un ordinateur) puisse être comparable au cerveau humain via la métaphore de la boîte noire. Mais si l'on associe couramment le concept de « boîte noire » à la cybernétique, celui-ci vient plus précisément des théories comportementalistes. À la suite des expériences sur les « réflexes conditionnels » de Ivan Pavlov ⁴, les psychologues John Broadus Watson et Burrhus Frederic Skinner développent le « behaviorisme », une méthode permettant d'étudier les relations statistiques entre l'environnement et le comportement en faisant fi du psychisme. Selon Watson, « la psychologie devrait se défaire des notions de conscience ⁵ ». Les processus inobservables s'effectuent dans une boîte noire – la fameuse *black box* –, qui ne constitue pas un objet de recherche

mars 2003, [En ligne], <http://boris.saulnier.free.fr/DOCS/200303_Saulnier_OrdreCerveau.pdf>. Site consulté le 11 septembre 2019.

1. Alan M. Turing, « Les ordinateurs et l'intelligence » [« Computing machinery and intelligence », 1950], trad. fr. Julien Basch et Patrice Blanchard dans *La Machine de Turing*, édition dirigée par Jean-Yves Girard, Paris, Seuil, coll. « Sciences », 1995.

2. Alan M. Turing, « Les ordinateurs et l'intelligence », *op. cit.*, p. 139.

3. On date généralement la naissance de la cybernétique de la publication en 1948 de l'ouvrage *La Cybernétique. Information et régulation dans le vivant et la machine* du mathématicien Norbert Wiener.

4. Ivan Petrovitch Pavlov, *Conditioned reflexes*, Londres, Routledge and Kegan Paul, 1927.

5. John Broadus Watson, « Psychology as the Behaviorist Views it », *Psychological Review*, n° 20, 2013, p. 158-177.

car seules comptent les observations des « comportements manifestes » – par exemple : un stimulus en réaction à un bruit. De même, dans la vision comportementaliste de l'apprentissage, l'apprenant est semblable à une boîte noire car on ne sait pas – et il n'y a pas besoin de savoir – ce qui passe à l'intérieur. L'individu est le seul résultat de son environnement : il suffit d'analyser ses entrées (*inputs*) et sorties (*outputs*). Cette opacité intrinsèque à la compréhension voire au déclenchement d'un comportement « provoqué » renvoie à la différence structurelle séparant les IA « symboliques » des « connexionnistes », qu'il faut à présent préciser. Ce tournant, qui constitue l'objet de cet article, marque une rupture profonde au sein des IA et rend nécessaire un réexamen de leurs conséquences sur la condition humaine, à penser dans sa pluralité.

Le terme de *deep learning* est forgé en 2006 par l'ingénieur Geoff Hinton¹. Techniquement, cette méthode « connexionniste » désigne un type d'intelligence artificielle dérivé du *machine learning* (apprentissage automatique) où la machine a pour objectif d'« apprendre par elle-même » – contrairement à l'approche plus ancienne de la logique « symbolique » (celles des premières formes d'IA²) où elle se contente d'exécuter des règles prédéterminées. Le *deep learning* se base sur un réseau de « couches » de neurones artificiels s'inspirant du cerveau humain. Ces derniers permettent de découper une tâche complexe en sous-catégories, l'idéal étant d'automatiser l'ensemble du processus pour viser une méthode « non supervisée ». Comme le notent les chercheurs Dominique Cardon, Jean-Philippe Cointet et Antoine Mazières³, c'est bien l'explosion des données caractéristiques de la massification des usages numériques et du *big data* des années 2010 qui impose pour de bon l'approche neuronale. Celle-ci se révèle très performante pour traiter les domaines aussi généralistes que le traitement du signal, de la voix, de la parole ou du texte, mais surtout pour faire face à de nouveaux défis « comme la détection de *spams*, les techniques de filtrage collaboratif utilisées pour la recommandation, la prédiction de stock, la recherche d'information ou l'analyse des réseaux sociaux⁴ ».

1. Geoffrey E. Hinton, Simon Osindero, Yee-Whye Teh, *Neural Computation*, vol. XVIII, n° 7, Cambridge, MIT Press, juillet 2006, p. 1527-1554, [En ligne, site consulté le 11 septembre 2019], <<https://10.1162/neco.2006.18.7.1527>>.

2. Pour une chronologie détaillée de la querelle entre les approches symboliques et connexionnistes, nous renvoyons le lecteur à Dominique Cardon, Jean-Philippe Cointet, Antoine Mazières, « La revanche des neurones », *Réseaux*, n° 211, novembre 2018, [En ligne, site consulté le 11 septembre 2019], <<https://neurovenge.antonomase.fr>>.

3. *Ibid.*

4. *Ibid.*

Il est essentiel de noter ici que le caractère auto-évolutif des techniques du *deep learning* le rend structurellement opaque, y compris pour les développeurs informatiques : son efficacité est conditionnée par le recours au paradigme de la « boîte noire ». Cette opacité engendre de nombreuses craintes et fantasmes¹, liées à une perte voire à une automatisation du sens moral. Si l'on entend sous le terme de « morale » la vieille acception de « ce qui suit la lumière de la raison² », il est alors possible de suggérer que les technologies du *deep learning* opèrent une substitution de la raison par la rationalité. L'opacité de cette médiation technique rejoue les craintes déjà énoncées par la philosophe Hannah Arendt à propos de « l'idéal moderne qui est de réduire [l'entendement humain] en symboles mathématiques³ ». Dans cette absence de raison, et par extension, de morale, « les êtres humains ne sont plus [...] que des animaux [*animale rationale*] capables de “calculer les conséquences”⁴ ». Or l'effet « boîte noire » du *deep learning* va même plus loin, puisque le « calcul des conséquences » évoqué par Arendt est lui aussi délégué aux machines. Les lois des nombres, celles des symboles mathématiques, pourraient-elles, dès lors, remplacer la capacité humaine à définir collectivement des « règles de conduite admises inconditionnellement et considérées comme applicables⁵ » ? Autrement dit : un monde mû par des techniques automatiques (non supervisées par des êtres humains) est-il souhaitable ?

Ce type de question n'est pas réservé aux domaines de l'ingénierie et de la philosophie morale. En effet, historiquement, les champs de la technique, de l'esthétique, et par extension du design, auront affronté les affres d'environnements de vie soumis à des logiques de rendement et d'optimisation des comportements. En effet, l'apparition du design s'inscrit dans une opposition à la baisse de qualité des productions de la « seule » ingénierie industrielle. En qualifiant ces productions d'« ersatz⁶ » au tournant du XIX^e siècle, l'artiste et écrivain William Morris préfigure le design comme

1. Yann Moulrier Boutang, Ariel Kyrou, « Mama IA Mamamouchi. Le dynamitage de 7 idées reçues pour en finir avec le Bourgeois numérique, ce Monsieur Jourdain de l'intelligence artificielle », *Multitudes*, n° 72, octobre 2018, p. 7-15.

2. Alain Rey (dir.), *Dictionnaire historique de la langue française*, Paris, Le Robert, 2011.

3. Hannah Arendt, *Condition de l'homme moderne* (1958), trad. fr. Georges Fradier, Paris, Pocket, coll. « Évolution », 2001, p. 335.

4. *Ibid.*, p. 356.

5. Alain Rey (dir.), *op. cit.*

6. William Morris, « L'âge de l'ersatz », conférence donnée le 18 novembre 1894 au New Islington Hall, Manchester. Reproduit dans *L'Âge de l'ersatz et autres textes contre la civilisation moderne*, trad. fr. Olivier Barancy, Paris, Encyclopédie des Nuisances, 1996.

l'utopie politique d'une réconciliation entre les conditions de production et la démocratisation des progrès techniques. Or l'idéal fallacieux dénoncé par Morris d'une production industrielle découplée d'enjeux artistiques et critiques prend une nouvelle dimension à l'ère des IA du *deep learning*. Ayant pour ambition d'annexer l'ensemble des activités humaines, celles-ci revendiquent depuis une dizaine d'années des capacités d'imagination et de création traditionnellement associées à l'art et au design. Plus précisément, le design dit « génératif » pourrait profondément modifier les pratiques de design. Grâce à l'IA (nous promet-on), des services comme TheGrid.io¹ (2015) ou Wix ADI² (2017) pourraient créer automatiquement des sites Web, tandis que le programme Dreamcatcher du logiciel Autodesk Netfabb³ (2018) pourrait générer un panel de formes optimales répondant à des contraintes données. Bien que ces « ersatz » ne soient pas très intéressants esthétiquement (car rapidement stéréotypés) et socialement (car déconnectés d'une expertise des usages), ceux-ci se révèlent économiquement puissants car ils court-circuitent le contre-pouvoir du design. Si l'on entend sous ce nom une intelligence des formes, le design peut se comprendre comme une résistance aux forces capitalistes puisque ces dernières, pour se développer, ont pour idéal une chaîne de production où le formel ne fait plus l'objet d'une expérience, d'un faire.

Le paradigme des boîtes noires du *deep learning* interroge donc directement les pratiques de design, à la croisée de la production d'objets (d'un travail du « faire ») et de leur inscription sociale. D'où ce paradoxe : alors qu'il n'aura cessé de se débattre avec les logiques de marché, le design se voit désormais profondément modifié par des intelligences artificielles en quête de rendement efficace. Afin d'examiner les enjeux et contre-pouvoirs des boîtes noires des intelligences artificielles, il nous faut donc opérer un double mouvement : étudier comment le design se voit annexé par ces nouvelles sortes d'IA « opaques » et, dans le même temps, montrer en quoi le design pourrait mettre à jour leurs dimensions impensées. Que devient le design dans un monde où la création pourrait s'automatiser ? Que peuvent faire les designers face à la multiplication des « boîtes noires » ?

1. *TheGrid.io*, 2011, [En ligne, site consulté le 11 septembre 2019], <<https://thegrid.io>>.

2. « Introducing Wix Artificial Design Intelligence, the first-ever AI that designs a stunning website for you », *Wix.com*, [En ligne, site consulté le 11 septembre 2019], <<https://www.wix.com/about/adi-get-access>>.

3. Marine Protais, « Autodesk lance (enfin) son logiciel de design génératif », *L'Usine nouvelle*, juin 2017, [En ligne], <<https://www.usinenouvelle.com/editorial/autodesk-lance-enfin-son-logiciel-de-design-generatif.N557153>>.

Sur quelles valeurs fonder le développement des IA, alors que celles-ci font vaciller les « anciennes » catégories de pensée ?

Pour traiter ces questions, nous verrons en quoi les spécificités du *deep learning*, au début des années 1990, renforcent le caractère inintelligible des technologies numériques. Nous prolongerons ces réflexions en nous appuyant sur les travaux du théoricien des médias Vilém Flusser qui, dès les années 1970, analyse les enjeux du développement d'écritures et d'images produites sans intervention humaine. Il nous faudra, pour finir, analyser les conséquences de ces programmes supposément « autonomes » sur les pratiques de design.

BLACK BOX SOCIETY, L'ÈRE DU SOUPÇON

Bien qu'elle se révèle très puissante d'un point de vue technique, l'approche du *deep learning* comprend pourtant plusieurs problèmes majeurs. Tout d'abord, le classement des données de sortie (injectées en même temps que les entrées) peut comprendre de nombreux biais, ne serait-ce que par les présupposés qu'il induit. Il en va ainsi, par exemple, quand des chercheurs se proposent d'identifier automatiquement l'orientation sexuelle d'individus à partir de l'analyse de leurs visages¹ : qu'en est-il de la pertinence de l'étiquetage des jeux de données ?

Les capacités de calcul et l'agrégation de larges jeux de données n'étant pas à la portée de tous, un autre problème réside dans la privatisation de ces développements techniques. Les chercheurs en *deep learning* les plus reconnus travaillant pour les GAFAM (Google, Amazon, Facebook, Apple, Microsoft), il n'est guère étonnant de constater que ces grands groupes technologiques dominent aujourd'hui ce secteur. Facebook, sous la houlette de Yann Le Cun², développe principalement des traitements d'analyse d'images et de conversations, intégrables directement dans l'« architecture toxique³ » (Olivier Ertzscheid) de sa plateforme publici-

1. Michal Kosinski, Yilun Wang, « Deep neural networks are more accurate than humans at detecting sexual orientation from facial images », *Journal of Personality and Social Psychology*, vol. CXIV, n° 2, février 2018, p. 246-257, [En ligne, site consulté le 11 septembre 2019], <https://www.gsb.stanford.edu/sites/gsb/files/publication-pdf/wang_kosinski.pdf>.

2. Communément reconnu comme l'inventeur de l'« apprentissage profond », l'ingénieur Yann Le Cun rejoint Facebook en 2013, la même année où Geoffrey Hinton, son directeur de post-doctorat, rejoint Google. D'autres exemples pourraient compléter cette liste.

3. Olivier Ertzscheid, « Trump, Google, l'idiot utile, les architectures techniques toxiques et le meilleur des algorithmes possibles », *Affordance.info*, 17 décembre 2018, [En ligne, site consulté le

taire aux deux milliards d'utilisateurs. De par leur caractère programmable et programmant, les algorithmes numériques opèrent des prises cognitives d'une autre nature que les flux abrutissants des mass médias télévisuels :

Lorsque ces algorithmes opèrent dans des espaces qui sont ceux d'architectures techniques toxiques (parce qu'elles amplifient certains biais cognitifs, parce qu'elle valorisent la viralité au détriment de la véracité ou de vérifiabilité, parce qu'elles n'ont pour objectif que la spéculation sur nos données et sur les espaces discursifs qui les rendent exploitables, etc.), alors le principe [du bien commun] cesse d'être opérant car gangrené par une logique spéculative principalement incarnée par la rentabilité publicitaire¹.

Au-delà de la publicité, les promesses de rentabilité sont multiples. IBM ne cesse de vanter les mérites de sa plateforme Watson (2011), censée entre autres améliorer les traitements contre le cancer². Google, via Tensor Flow³ (2015), propose, quant à lui, une plateforme *open source* pouvant être utilisée en local ou à distance dans le *cloud* : on peut donc supposer que les résultats de la communauté serviront directement à l'améliorer.

Enfin, un troisième problème du *deep learning*, peut-être le plus important, réside dans son caractère opaque : ce sont des boîtes noires dont personne – y compris les programmeurs – ne comprend exactement comment fonctionnent leurs opérations internes liées aux interactions et rétroactions avec les données.

L'encodage progressif de toutes les expériences humaines en « nourriture » destinée à alimenter des *black boxes* entraîne donc une méfiance généralisée envers les algorithmes, c'est-à-dire envers les règles logiques qui sous-tendent le fonctionnement des programmes numériques. Aussi, plutôt que d'accuser les machines de nous manipuler, il s'agirait plutôt de mettre à jour leurs logiques délibérément masquées conçues par des êtres humains. Dans son essai *Black Box Society. Les algorithmes secrets qui contrôlent l'économie et l'information*⁴ (2015), le professeur de droit Frank

11 septembre 2019], <https://www.affordance.info/mon_weblog/2018/12/meilleur-algorithmes-possibles-trump-idiot.html>.

1. Olivier Ertzscheid, « Trump, Google, l'idiot utile, les architectures techniques toxiques et le meilleur des algorithmes possibles », *Affordance.info*, 17 décembre 2018, [En ligne], <https://www.affordance.info/mon_weblog/2018/12/meilleur-algorithmes-possibles-trump-idiot.html>.

2. Casey Ross, Ike Swetlitz, « IBM's Watson supercomputer recommended "unsafe and incorrect" cancer treatments, internal documents show », *StatNews.com*, 25 juillet 2018, [En ligne], <<https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>>.

3. « Google Tensor Flow », [En ligne], <<https://www.tensorflow.org>>.

4. Frank Pasquale, *Black Box Society. Les algorithmes secrets qui contrôlent l'économie et l'information*, trad. fr. Florence Devesa, Phil Adams et Lucia Di Bisceglie, Paris, Fyp, 2015.

Pasquale montre que, dans la lignée des dispositifs foucauldien, un certain nombre d'algorithmes révèle les nombreuses connivences existant entre les sphères publiques et privées. Les exemples ne manquent pas : un moteur de recherche archivant la moindre trace d'activité en ligne (Google), un média social attribuant un score de « fiabilité » à ses utilisateurs (Facebook), un gouvernement notant sa population (score de crédit social dans certaines villes de Chine), une police cherchant à prédire les comportements à risque (PredPol : Predict Prevent Crime), un juge s'appuyant sur des imbrications statistiques pour quantifier les risques de récidive (The Justice Department's National Institute of Corrections ¹), des systèmes de reconnaissance faciale visant à déterminer le genre (Face++) d'une personne, etc. À lire Pasquale, il saute aux yeux que les boîtes noires engendrent une société délétère créant « des pouvoirs invisibles », « une concurrence déloyale », des discriminations, etc. Il donne ainsi l'exemple du roman transhumaniste et dystopique *Super triste histoire d'amour* ² de l'écrivain Gary Shteyngart (2010) :

Incapables de s'opposer aux géants [de la technologie], les personnages [...] mesurent leur personnalité et leur « sex appeal » avec des applications sur smartphones. Leurs notes de solvabilité sont affichées publiquement dans des magasins. [...] Ils ne se préoccupent pas beaucoup de la signification de ces scores, ni de la façon dont ils sont calculés. Ils veulent juste qu'ils soient élevés. Le niveau de classement de la boîte noire est source d'identité, la dernière réserve de valeur « objective » dans un monde où l'instabilité et la faible capacité d'attention sapent les valeurs les plus complexes du soi ³.

Face à ce faisceau de menaces sournoises contre la défense des libertés individuelles et des minorités, mettre à jour les « secrets » de ces organismes de contrôle devient alors le mot d'ordre d'une société civile en proie aux pires manipulations. Pasquale, en conclusion, plaide pour le développement d'« alternatives publiques » dans la technologie et la finance pour se diriger « vers une société intelligible ». Mais est-ce vraiment possible, dès lors que la spécificité technique des IA du *deep learning* est de produire des programmes évolutifs et imprévisibles ?

1. Julia Angwin, Jeff Larson, Surya Mattu et Lauren Kirchner, « Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica.org*, mai 2016, [En ligne, site consulté le 11 septembre 2019], <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>.

2. Gary Shteyngart, *Super triste histoire d'amour* [*Super Sad True Love Story*, 2010], trad. fr. Stéphane Roques, Paris, L'Olivier, 2012.

3. Frank Pasquale, *Black Box Society*, *op. cit.*, p. 278.

Cette capacité des technologies numériques à auto-programmer l'être humain en « fonctionnaire », entendu au sens de pure entité fonctionnelle dénuée de sens moral, a été entrevue dès le début des années 1970, soit donc à l'aube de l'informatique personnelle, par le théoricien des médias Vilém Flusser. Chercheur aux multiples facettes, Vilém Flusser est notamment célèbre pour ses écrits autour des appareils photographiques et des caméras vidéo. Pour Flusser, la photographie fait partie des « techno-images » produites par les « appareils » électroniques. Les photographies et les vidéos ne « signifient » pas le monde extérieur mais sont constituées des textes invisibles des programmes :

La compétence de l'appareil doit être supérieure à celle de ses fonctionnaires. Aucun appareil photo correctement programmé ne peut être entièrement percé à jour par un photographe [...]. C'est une *black box*. [...] Voilà ce qui caractérise le fonctionnement de tout appareil : le fonctionnaire est maître de l'appareil grâce au contrôle qu'il exerce sur ses faces extérieures (sur l'*input* et sur l'*output*), et l'appareil est maître du fonctionnaire du fait de l'opacité de son intérieur¹.

En faisant du « fonctionnaire » un concept amoral, et non pas une simple désignation des personnes employées par les gouvernements, Flusser montre comment les technologies numériques mettent en crise les critères et catégories de pensée de l'idéal moderne. Selon Flusser, la seconde guerre mondiale représente l'aboutissement du « programme » (projet) occidental, à savoir une compréhension de la technique consistant à traiter n'importe quel élément (y compris les êtres humains) comme des objets manipulables, sans sens moral :

Une telle ontologie programmatique a suscité l'invention d'ordinateurs et d'instruments intelligents. Elle mène à la transformation de la société en système cybernétique composé d'appareils et de fonctionnaires. Les hommes sont programmés pour fonctionner comme pièces du jeu symbolique. Ils sont chiffrés et numérotés. Ils deviennent computables dans des statistiques et des cartons perforés. Ils sont programmés d'une manière telle qu'ils acceptent volontiers leur programmation. [...] Bien sûr, une telle société post-industrielle n'est pas encore arrivée à son stade de réalisation parfaite. Mais nous avons déjà ses modèles : Eichmann comme modèle du fonctionnaire, Kissinger comme modèle du programmeur, Auschwitz comme modèle de la société entière².

1. Vilém Flusser, *Petite philosophie de la photographie* (1985), trad. fr. Jean Mouchard, Oberhausbergen, Circé, 1998, p. 36. Le titre original, en portugais du Brésil, signifie « Philosophie de la boîte noire » (*Filosofia da Caixa Preta*).

2. Vilém Flusser, « IV. Notre Travail », in *Post-histoire* [version française inédite, 1982], propos liminaire de Catherine Geel, introduction de Anthony Masure, postface de Yves Citton, Paris, T&P Work UNit, 2019.

Ce basculement dans une cybernétique avilissante marque, selon Flusser, l'entrée dans la « post-histoire ». Cette « objectivation de l'homme par l'homme ¹ » est rendue possible par la soumission des « fonctionnaires » à l'« appareil », c'est-à-dire à un programme dont ils ne peuvent qu'exécuter les règles. On pourrait alors penser que les programmeurs, parce qu'ils écrivent les programmes, pourraient être en mesure de les expliquer et de les contrôler. Or, dit Flusser en substance – même pour le secrétaire d'État Henry A. Kissinger ² – une telle tâche est vouée à l'échec car elle revient à appliquer aux programmes des catégories de pensée obsolètes. Les programmeurs n'échappent en rien au pouvoir de l'appareil :

Dans la véritable société post-industrielle la bureaucratie sera automatisée, invisible et efficace. [...] En apparence, ce seront les programmeurs des boîtes noires qui gouverneront. La technocratie. En réalité, ce seront les appareils eux-mêmes qui gouverneront. Les programmeurs seront eux-mêmes programmés pour programmer les appareils, et ils seront programmés par d'autres appareils ³.

Selon Flusser, en l'absence d'un point de vue critique et d'un travail artistique, nous courons le risque de devenir analphabètes – un risque que pointait déjà Walter Benjamin dans sa lecture philosophique de l'histoire de la photographie ⁴. Il s'agit alors de « ne pas être dévoré ⁵ par la boîte » :

La boîte, qu'elle soit appareil photo ou média, tend automatiquement à nous dévorer. Les photographes essaient de duper les boîtes pour leur faire produire de l'information. C'est la lutte entre la liberté humaine et ses propres dispositifs. Chaque photographie témoigne isolément de cette lutte ⁶.

Si l'on replace ces analyses dans le cadre des intelligences artificielles, il est intéressant de constater que les jeux d'ajustement entre les *inputs* ou *outputs* de ces boîtes noires témoignent de la lutte formulée par Flusser : cherchant coûte que coûte à leur faire « cracher » des informations, nous risquons de n'être que les fonctionnaires d'un jeu dont les règles nous échappent. Cette idée d'assujettissement de l'être humain – son assignation

1. Vilém Flusser, « I. Notre sol », *ibid.*

2. Henry A. Kissinger, « How the Enlightenment Ends », *op. cit.*

3. Vilém Flusser, « IV. Notre Travail », *op. cit.*

4. Walter Benjamin, « Petite histoire de la photographie » (1931), *Études photographiques*, n° 1, tirage à part, trad. fr. André Gunthert, novembre 1996.

5. Le vocabulaire de Flusser fait écho à cet article emblématique : Marc Andreessen, « Why Software Is Eating the World », *The Wall Street Journal*, 20 août 2011, [En ligne, site consulté le 11 septembre 2019], <<https://a16z.com/2011/08/20/why-software-is-eating-the-world>>.

6. Vilém Flusser, « Comment ne pas être dévoré par la boîte », 14^e Rencontres internationales de la photographie, Arles, 9 juillet 1983. Document inédit consulté à la Vilém Flusser Archive, UdK Berlin.

à de la pure utilité – nous emmène au plus près des enjeux croisés du design et des intelligences artificielles. Face au formatage des expériences humaines dans ce qu’elles ont de plus singulier, quelles marges de manœuvre peut-on inventer ? Des contre-pouvoirs sont-ils encore envisageables ?

FACE AU *DEEP LEARNING*, LES CONTRE-POUVOIRS DU DESIGN

Responsabiliser l’injonction à l’efficacité

Le premier élément de réponse que nous proposons consiste à examiner comment le design dit « de services » s’est très bien accommodé de techniques permettant d’invisibiliser ceux que le sociologue Antonio Casilli appelle les « tâcherons du clic ¹ » (les personnes constituant et ajustant manuellement les jeux de données) et, de façon plus générale, des rapports de domination. Les exemples d’application « immoraux » des IA ne manquent pas ² : profilage des goûts et génération de contenus personnalisés (Netflix, 2017), analyse de « parcours utilisateurs » au sein de contextes marchands (Adobe), détection de passagers ivres (Uber, 2018), détection du genre d’une personne aux *check-in* des aéroports (Detroit, États-Unis, 2018), « robots recruteurs » (Corée du Sud, 2018), etc.

Dans chacun de ces exemples, le *deep learning* est utilisé à des fins coercitives pour les libertés individuelles. De façon plus générale, le fait que les intelligences artificielles soient majoritairement issues de grandes firmes néolibérales jette un sérieux doute sur leur capacité à œuvrer pour le bien commun ³. Enfin, le caractère structurellement opaque du *deep learning* interroge directement la notion de responsabilité, et donc de neutralité : qui faudra-t-il blâmer quand un programme « auto-apprenant » tuera un patient hospitalisé pour lui avoir formulé un mauvais diagnostic ⁴ suite à

1. Antonio A. Casilli, *En attendant les robots. Enquête sur le travail du clic*, Paris, Seuil, coll. « La Couleur des idées », 2019.

2. Par souci d’économie, nous n’indiquons pas les références précises de ces exemples car ces derniers peuvent facilement être retrouvés en ligne.

3. Ben Tarnoff et Moira Weigel, « Why Silicon Valley can’t fix itself », *TheGuardian.com*, 3 mai 2018, [En ligne, site consulté le 11 septembre 2019], <<https://www.theguardian.com/news/2018/may/03/why-silicon-valley-cant-fix-itself-tech-humanism>>.

4. Aaron M. Bornstein, « Is Artificial Intelligence Permanently Inscrutable? Despite new biology-like tools, some insist interpretation is impossible », *Nautilus*, 1^{er} septembre 2016, [En ligne, site consulté le 11 septembre 2019], <<http://nautil.us/issue/40/Learning/is-artificial-intelligence-permanently-inscrutable>>.

l'application de motifs (*patterns*) simplistes ? La question est alors de savoir si, sous couvert d'efficacité, le design doit se lier à une conduite technique qui travaille contre nous en faisant de l'opacité un prérequis à l'optimisation fonctionnelle.

Face à ces nouvelles menaces, l'appel à l'intelligibilité de Frank Pasquale est toujours d'actualité. Certains ingénieurs pensent ainsi qu'il serait possible de créer des IA « explicables¹ », or cette recherche de transparence nécessite de produire une surcouche technique pour fonctionner². Dans le même registre, les demandes d'« éthique³ » peinent à s'appuyer sur des critères partagés, les régulations et les valeurs se heurtant aux frontières des États-nations. Comment comparer des IA chinoises, américaines, arabes ou européennes ? Faut-il œuvrer à « décoloniser » leurs *patterns* ? Analysée par Vilém Flusser, cette crise de critères moraux devenus inopérants au sein des environnements numériques donne l'impression que nous ne sommes toujours pas prêts à affronter ces enjeux, quand bien même, comme nous l'avons vu en introduction, le réel pouvoir des IA est largement surévalué.

Désautomatiser le design

Le deuxième élément d'éclairage propose d'interroger le risque d'une automatisation des processus de design. Comme les jeux de données nécessaires au fonctionnement du *deep learning* s'appuient sur des contenus déjà formatés par un travail algorithmique (facilement indexables donc issus du Web, etc.), les productions obtenues ressemblent souvent aux *datasets* de départ. Il saute rapidement aux yeux que la « sémiotique simplissime⁴ » des exemples que nous donnions en introduction (TheGrid.io et Wix ADI) n'est guère plus intéressante que les *templates* (modèles de mise en

1. « Explainable Artificial Intelligence », *Wikipedia*, [En ligne], <https://en.wikipedia.org/wiki/Explainable_Artificial_Intelligence>.

2. Loup Cellard et Anthony Masure, « Le design de la transparence : une rhétorique au cœur des interfaces numériques », *Multitudes*, n° 73, dossier « Tyrannies de la transparence », hiver 2019.

3. « The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems », *Accessnow.org*, 16 mai 2018, [En ligne, site consulté le 11 septembre 2019], <<https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>>.

4. Jean Lassègue, « Des grilles et des rubans. Les machines formelles d'Alan Turing », Paris, B42/Fork, *Back Office*, n° 2, mai 2018, « Penser, classer, représenter », [En ligne, site consulté le 11 septembre 2019], <<http://www.revue-backoffice.com/numeros/02-penser-classer-representer/jean-lassegue-grilles-rubans>>.

page) dominant le design d'interface contemporain. Mais chercher à analyser ce que devient le design dans un monde où la création pourrait s'automatiser revient à considérer que le design serait *in fine* une source d'invention permanente. Sans les IA, le design n'échappe pourtant en rien aux effets de mode (*trends*) et aux processus de copie. En ce sens, les méthodes du *deep learning* ne feraient finalement que pousser plus avant le design vers une puissance de production de marchandises et de *patterns* automatisés dans lequel il se serait déjà trop largement engagé.

Il serait toutefois réducteur de ne voir dans la copie qu'un processus sans intérêt, alors qu'une bonne partie de l'histoire du design (et de l'art) montre au contraire que celle-ci peut être source d'invention, y compris dans son caractère absurde. Le projet *Balenciaga AI* (2018) de l'artiste Robbie Barrat propose ainsi une nouvelle collection générée à partir des anciennes productions de la maison Balenciaga, mettant par-là (involontairement ?) en évidence le caractère répétitif de l'existant¹. De façon plus radicale, le projet *Made in Machinale*² (2018) travaille les porosités entre le design nordique et la culture industrielle chinoise de Shenzhen en mettant en tension le rôle du designer face à l'apprentissage machine. Cette recherche empirique explore la combinaison de l'approche nordique (réponse à des besoins) et de la logique chinoise (adéquation au marché et aux composants disponibles) pour générer un nouveau processus de création. Concrètement, les composants, matériaux et marchandises déjà disponibles en ligne sur le marché chinois (Alibaba.com, etc.) sont transmis à un réseau de neurones qui a été formé aux idéaux du design nordique. Ce métissage algorithmique génère un grand nombre de propositions étranges, qu'un curateur humain peut ensuite trier, faire produire et commercialiser – bouclant ainsi les boucles des réseaux de neurones. En hybridant le *deep learning*, *Made in Machinale* interroge directement l'histoire et les fondamentaux du design. Que se passe-t-il lorsque le produit n'est pas créé par un besoin fonctionnel mais par un caprice du marché ? Qui décide quelles fonctionnalités sont prioritaires, nécessaires, ou même souhaitables ? Ces marchandises existent-elles simplement parce qu'il est possible de les produire ?

1. Katharine Schwab, « This AI designs Balenciaga better than Balenciaga. It's really into asymmetry and color-blocking », *FastCompany.com*, 22 octobre 2018, [En ligne, site consulté le 11 septembre 2019], <<https://www.fastcompany.com/90223486/this-ai-designs-balenciaga-better-than-balenciaga>>.

2. Simone Rebaudengo, Sami Niemelä (dir.), « Made in Machina/e », 2018, [En ligne, site consulté le 11 septembre 2019], <<https://madeinmachina.com>>.

D'autres expériences de ce type seraient nécessaires pour étudier plus finement les limites de ce qu'on croit tenir pour acquis – et pour humain – dans le design. On retrouve ici la prémonition de Vilém Flusser quant à la seule échappatoire à un monde en proie aux pires manipulations sournoises : pour peu qu'il ne se laisse pas absorber par les circonvolutions technologiques, seul l'art (et par extension le design) saurait faire « cracher » aux boîtes noires des sorties (*outputs*) imprévus.

Déjouer l'esthétique inauthentique

Le troisième élément de compréhension des mutations du design au contact du *deep learning* consiste à se pencher sur les esthétiques qu'il produit. Malgré les injonctions à l'innovation, il est frappant de constater que ces IA soient majoritairement mobilisées pour reproduire des démarches antérieures. Il faut donc s'interroger sur la pertinence de réduire la création à de simples prouesses techniques : produire un énième Rembrandt, Mondrian, paysage, visage, etc., comme si l'apparition de toute nouvelle technique devait relever de l'académisme, à savoir l'imitation de règles, de traditions ou d'anciennes techniques¹. Il en va ainsi, par exemple, de l'apparition de la photographie qui emprunte les vieux habits de la peinture avant d'être « découverte », dans un second temps, par des artistes parvenant à la faire exister comme art. En ce sens, de nombreuses démonstrations techniques réalisées sous le nom d'IA échoueraient à faire art, si l'on entend sous ce nom une capacité à dérouter (conduire) une technique dans des directions qui ne seraient pas immédiatement attendues.

Pour échapper à ces impasses, et dans l'intervalle d'une « découverte » des IA (à supposer que celle-ci soit possible), il est tout de même possible de jouer avec elles pour avérer leur caractère absurde – une démarche que n'aurait pas renié Vilém Flusser. Conçu par John Fass et Alistair McClymont, le projet *Of Machines Learning to See Lemon*² (2018) met en évidence les incohérences et chausse-trappes de la catégorisation

1. Pierre-Damien Huyghe, « Le devenir authentique des techniques », conférence au Centre national de la recherche technologique, Rennes, 2004, [En ligne, site consulté le 11 septembre 2019], <<http://pierredamienhuyghe.fr/pagesaudio/DevenirAuthentiqueTechniques.html>>.

2. John Fass, Alistair McClymont, « Of Machines Learning to See Lemon », *AlistairMcClymont.com*, 2018, [En ligne, site consulté le 11 septembre 2019], <<http://www.alistairmcclymont.com/artwork/of-machines-learning-to-see-lemon>>.

algorithmique du *deep learning* en présentant une série d'objets de diverses natures, disposés sous forme de grille dans un meuble de rangement aux multiples cases. Leur coexistence et leur positionnement dans les boîtes ne peuvent qu'intriguer le regardeur, mis face à une curation « non humaine » :

Ce processus de classification invisible vise habituellement à produire des décisions automatisées, qui peuvent avoir de lourdes conséquences sur les libertés individuelles et collectives. Les possibles avantages de l'apprentissage machine [*machine learning*] sont nombreux, mais nous courons le risque de développer des technologies d'une telle complexité que notre capacité à les façonner pour servir le bien commun devient très limitée¹.

D'une façon approuvante, les designers graphiques du studio Chevalvert ont travaillé, lors d'un workshop (atelier de travail) intitulé « Machine Jacking² » (2017), à détourner des IA (Siri, *chatbots*, OCR, etc.) de ce pour quoi elles sont programmées. En se mettant « dans la peau » d'une IA, il s'agit ici de confronter des « cadrages subjectifs » aux règles logiques. Il faut ainsi déconstruire deux mythes à propos des intelligences artificielles dites créatives : d'une part que l'artiste (ou le designer) aurait le contrôle total sur sa production, et d'autre part que la machine serait totalement autonome. De façon plus intéressante, « les artistes génératifs contrôlent habilement à la fois l'ampleur et l'emplacement du hasard introduit dans leurs productions³ ».

CONCLUSION – FAIRE DÉRIVER LE GOUVERNER

L'étude du concept de boîte noire – via ses racines cybernétiques jusqu'à ses ramifications contemporaines dans le *deep learning* appliqué à l'IA – nous a permis d'éclairer les risques d'une absence de recul critique sur ces technologies. Pour échapper aux discours alarmistes tels que celui de Henry A. Kissinger⁴, il nous faut affronter l'absurdité d'un monde technique faisant vaciller les catégories traditionnelles de la pensée (humain/

1. *Ibid.*, notre traduction.

2. « Design graphique et intelligence artificielle : vers un design algorithmique ? », *Stereolux.org*, Nantes, octobre 2017, [En ligne, site consulté le 11 septembre 2019], <<https://www.stereolux.org/blog/design-graphique-et-intelligence-artificielle-vers-un-design-algorithmique>>.

3. Jason Bailey, « Why Love Generative Art? », notre traduction, juillet 2018, *Artnome*, [En ligne, site consulté le 11 septembre 2019], <<https://www.artnome.com/news/2018/8/8/why-love-generative-art>>.

4. Henry A. Kissinger, « How the Enlightenment Ends », *op. cit.*

non humain, nature/culture, inventé/généré, etc.). En révélant la part machinique des êtres humains, les intelligences artificielles du *deep learning* invitent non seulement à réinventer les pratiques de création, mais aussi à dérouter, par l'art et le design le pouvoir totalitaire du capitalisme cognitif. Le design pourrait alors être envisagé, aux antipodes des ambitions avilissantes d'une automatisation aveugle, comme une capacité à faire dériver tout gouvernail¹.

Nos trois propositions de contre-pouvoirs aux menaces des intelligences artificielles – responsabiliser l'injonction à l'efficacité ; désautomatiser le design ; déjouer l'esthétique inauthentique – visent à contrecarrer l'idéologie cognitiviste à l'œuvre derrière les boîtes noires. D'autres directions de travail pourraient bien sûr être déployées : nous ne pouvons qu'inciter les designers à les activer pour que ce formatage des cadres de vie et des expériences humaines n'ait pas pour fatalité de se lier de façon irréversible au design².

1. *kubernêtês*, « l'art du gouvernail ».
2. Remerciements à Ariel Kyrou, Saul Pandelakis et Alexandre Saint-Jevin.