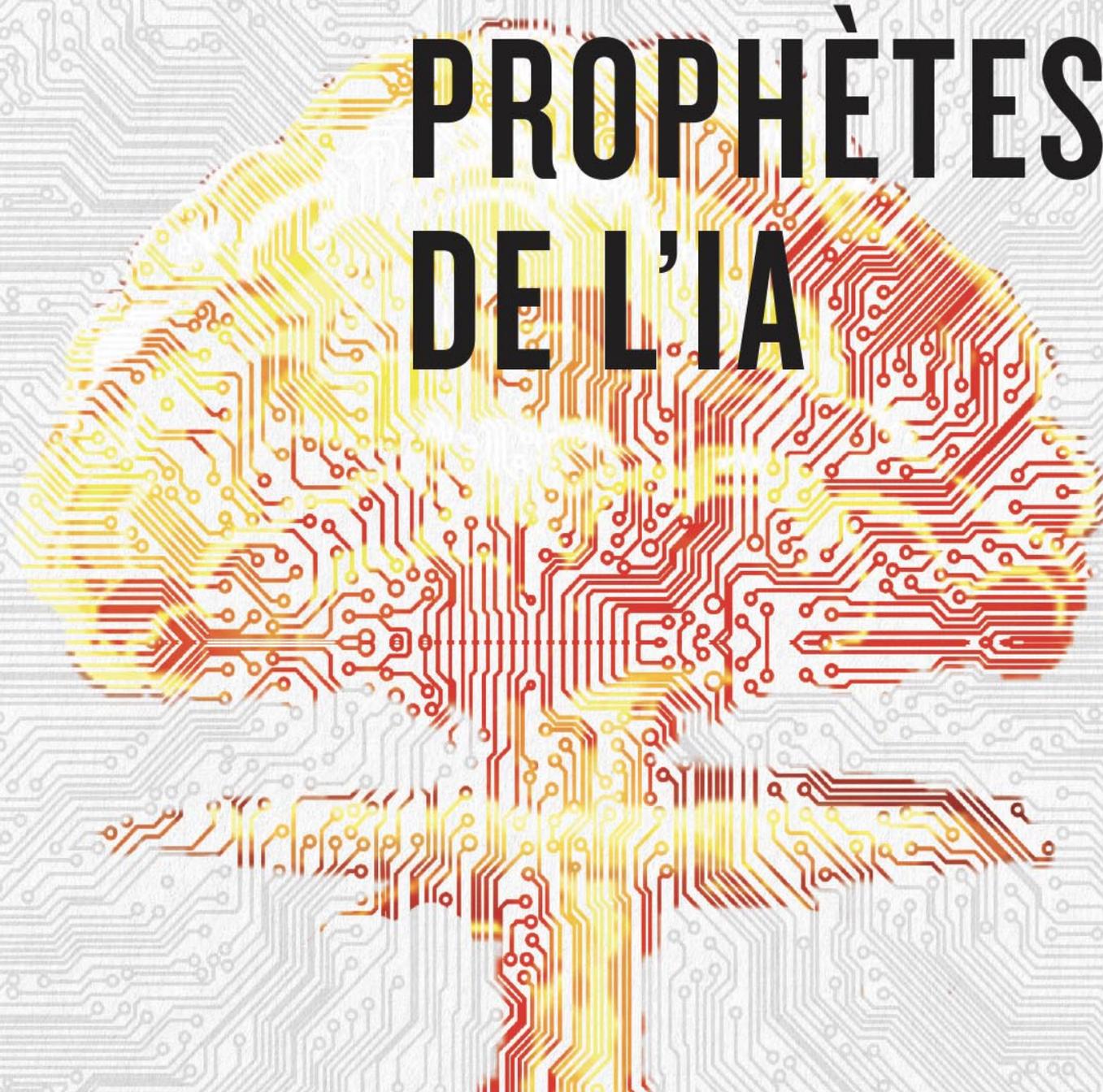


THIBAUT PRÉVOST

LES PROPHÈTES DE L'IA



**POURQUOI LA SILICON VALLEY
NOUS VEND L'APOCALYPSE**

LUX

THIBAUT PRÉVOST

LES PROPHÈTES DE L'IA

**POURQUOI LA SILICON VALLEY
NOUS VEND L'APOCALYPSE**

LUX

LES PROPHÈTES DE L'IA

Pourquoi la Silicon Valley nous vend l'apocalypse

THIBAUT PRÉVOST



© Lux Éditeur, 2024

www.luxediteur.com

Conception graphique de la couverture: David Drummond Images de la
couverture: Shutterstock

Dépôt légal: 4e trimestre 2024 Bibliothèque et Archives Canada
Bibliothèque et Archives nationales du Québec

ISBN (papier): 978-2-89833-160-2

ISBN (epub): 978-2-89833-162-6

ISBN (pdf): 978-2-89833-161-9

Canada

Nous reconnaissons l'aide financière du gouvernement du Canada pour nos
activités d'édition.

Table des matières

État des lieux

L'ère des perroquets automatiques

Tous vers l'intelligence artificielle générale

Sam Altman, rock star du lobbying

La bataille des prophéties

Un air de déjà-vu

Keep Calm and Reclaim

1. La Terre sainte du transhumanisme

Ultralibéralisme et transhumanisme californien

Nick Bostrom, le prophète déguisé en prof

Dépasser les frontières, l'obsession de la Silicon Valley

L'altruisme efficace: une machine à justifier le (techno)capitalisme

Vers le long-termisme, où l'humain n'a plus d'importance

2. Futuritarismes

Vers un capitalisme de commandement

Retour vers le futur

Lumières noires sur la Silicon Valley

Vers l'infini et l'au-delà

3. L'apocalypse selon l'IA

De la catastrophe naturelle à l'apocalypse artificielle.

Le totalitarisme contre l'extinction

À la recherche des Tables de l'IA

Superalignement et super-*bullshit*

4. Du bon usage des prophéties

Lutte fratricide dans la Silicon Valley

Mainmise du capital sur la technoscience

La fabrique de la sidération

L'AI governance, l'intoxication des institutions

Réguler le monopole, c'est réguler l'outil

5. Systèmes d'exploitation

L'intelligence artificielle est un culte

L'intelligence artificielle ne fonctionne (toujours) pas

L'intelligence artificielle est un écocide

L'intelligence artificielle est un pillage

L'intelligence artificielle est une bulle financière

L'intelligence artificielle est un autoritarisme

L'intelligence artificielle est une discrimination

Remerciements

Notes et références

*La réalité, c'est ce qui perdure même si l'on cesse d'y
croire.*

Philip K. DICK, «Le voyage gelé»

État des lieux

«**L**imiter les risques d’extinction [de l’humanité] posés par l’intelligence artificielle devrait être une priorité mondiale, aux côtés d’autres risques de grande ampleur comme les pandémies ou la guerre nucléaire.»

Une seule phrase, une petite vingtaine de mots. Rarement, de mémoire de communicant, aura-t-on vu une lettre ouverte aussi lapidaire que le *Statement on AI Risk*, publié le 30 mai 2023 sur le site du *think tank* de la Silicon Valley, Center for AI Safety. Et pourtant, ces 22 mots anglais ont été cités, traduits, repris et décortiqués par la quasi-intégralité de la presse internationale. Sous la phrase, la liste des signataires, et pas n’importe lesquels: tous les grands patrons de la Silicon Valley y côtoient le *who’s who* de la recherche en intelligence artificielle occidentale. Le laconisme de la formule est bien trouvé: comme le fera justement remarquer le *New York Times*, il a permis de regrouper toutes celles et tous ceux que l’IA inquiète, sans avoir à détailler les raisons de leur inquiétude.

La Silicon Valley est alors en pleine crise de panique. Le Future of Life Institute, un autre *think tank* qui se donne pour mission de «faire en sorte que les technologies transformatives [sic] soient bénéfiques à la vie et ne comportent pas de risques extrêmes», avait déjà publié une autre pétition, deux mois plus tôt^[1]. Signé par plus de 3000 experts du secteur (mais par aucun leader de l’industrie), le texte appelait «tous les laboratoires d’IA à suspendre immédiatement» leurs recherches, le temps de «mettre en place un ensemble de protocoles de sécurité partagés». Le scénario catastrophe à éviter: que «des responsables de la tech non élus» s’embarquent dans «une course effrénée pour développer et déployer des intelligences numériques de plus en plus puissantes que personne – pas même leurs créateurs – ne puisse comprendre, prédire ou contrôler de

manière fiable». Autrement dit, que Google ou Microsoft perde le contrôle de leur création. Vraiment?

En 2024, l'intelligence artificielle développée par les conglomérats de la Silicon Valley représenterait-elle un danger comparable à une pandémie ou un conflit nucléaire? Les périls concrets seraient-ils imminents au point d'exiger l'arrêt immédiat des recherches? Difficile à croire. D'autant que les menaces concrètes ne manquent pas. Le monde ne s'est toujours pas complètement remis de la pandémie de COVID-19. L'inflation détruit les travailleurs pauvres dans des économies en récession. La dernière fois que l'horloge de l'apocalypse a été aussi proche de minuit, c'était en 1953. Le continent européen a redécouvert la guerre, la lutte palestinienne trouve un écho dans toutes les sociétés civiles, le monde se lézarde en de nouvelles lignes de front géopolitiques. En toile de fond, les effets des bouleversements climatiques induits par le capitalisme industriel frappent maintenant toute la planète avec une intensité inédite. Nous vivons, en d'autres termes, un moment de «polycrise» économique, politique, environnementale, un enchevêtrement de chaos.

Mais les élites de la tech, sous le soleil de la Californie, n'ont que l'«IApocalypse» à la bouche.

Combinés, les signataires des deux lettres ouvertes réunissent presque tout le *boys'club* de l'industrie. Au premier rang, l'inusable Elon Musk, patron de X (jadis Twitter) et figure de proue du futurisme depuis une quinzaine d'années. À ses côtés, la fine fleur des grands patrons de la Silicon Valley. Derrière eux, la jeune garde de l'industrie de l'IA, qui sème la panique autant qu'elle attire les convoitises avec ses trouvailles et ses créations.

Pour les besoins de ce livre, nous les appellerons les prophètes. Cette congrégation se compose majoritairement d'hommes blancs, issus des élites universitaire et entrepreneuriale anglo-saxonnes. Ils viennent de Stanford, Oxford, Harvard ou du Massachusetts Institute of Technology (MIT). Ils sont ingénieurs, chercheurs, philosophes, patrons et investisseurs, presque tous basés entre Londres et la Californie. Ils dirigent OpenAI, Microsoft ou le fonds d'investissement Andreessen Horowitz (aussi connu sous le nom «a16z»). Ils sont milliardaires,

parfois «seulement» multimillionnaires, mais très rarement moins que cela. La mission qu'ils se sont donnée, qui tient à la fois de la croisade et de la ruée vers l'or: faire naître la machine pensante et s'en accaparer le pouvoir. Tous se réclament de la rationalité et de la méthode scientifique. Tous diffusent un discours d'apocalypse, articulé autour d'une intelligence artificielle divine.

Dans la Silicon Valley, une terre fertile pour les spiritualités, malgré ses habits technoscientifiques, un récit aux accents messianiques se diffuse. Face aux investisseurs, les *start-up* de l'intelligence artificielle vendent l'imminence de la fin des temps. Sur les campus des facs britanniques et états-uniennes, un réseau d'«instituts de recherche» forme des futurs oracles, éthiciens et philosophes de la technique, qui iront ensuite convertir les ingénieurs californiens au culte de l'IA. Sous l'influence de ces *start-up* aux accents sectaires, qui reçoivent désormais des montagnes d'investissements, l'industrie de la tech tout entière bascule dans un discours technoreligieux. On ne vend plus le progrès, mais la transcendance. On ne vend plus le futur, mais la fin des temps. On ne vend plus un objet, mais la naissance d'un dieu.

Avant 2014, ces évangélistes étaient totalement absents de la vie politique et médiatique. Mais pas une semaine de 2023 n'a passé sans qu'ils apparaissent dans la presse, générale comme spécialisée, pour prononcer un oracle, donner à lire un manifeste ou pour faire passer des messages subliminaux à l'oreille des institutions internationales. L'épicentre mondial du culte du progrès technique, dont les champions – Google, Amazon, Facebook, Microsoft et Apple – ont passé l'essentiel du XXI^e siècle à vendre leurs produits comme étant bénéfiques à toute l'espèce humaine, nous hurle désormais de faire attention au déploiement de sa dernière création. La volte-face est aussi spectaculaire qu'inédite.

L'ère des perroquets automatiques

Quel séisme a donc eu lieu dans la recherche en intelligence artificielle pour provoquer une telle onde de choc? La réponse tient en trois lettres: GPT, pour «*generatively pretrained transformer*» (transformeurs génératifs préentraînés). Nous voilà en l'an 2 de la «révolution»

ChatGPT, le robot conversationnel lancé par OpenAI en décembre 2022. À ce stade, vous avez déjà très probablement discuté avec ce logiciel, ou avec ses concurrents – il est vrai qu’entre Claude, Pi, Gemini, Bard, Grok et les autres, vous avez désormais l’embarras du choix. Vous avez fait l’expérience de cette altérité inédite, déconcertante parce qu’inqualifiable. Vous êtes peut-être resté sans voix face à son éloquence. Vous vous êtes peut-être surpris à vous demander si, quelque part dans les serveurs qui l’hébergent, la machine qui vous composait des poèmes instantanément et sur commande n’était pas réellement intelligente, voire... consciente. C’est normal, ces machines sont conçues pour cela. Et elles y arrivent tellement bien que même leurs créateurs s’y laissent prendre. C’est irrésistible.

Sans se perdre en détails techniques, essayons de comprendre comment fonctionne l’intelligence artificielle dite «générative», et pourquoi elle provoque une épiphanie générale dans la Silicon Valley. Ce type de programme informatique s’appuie sur deux piliers. Premièrement, une architecture appelée «réseaux de neurones artificiels» (*neural networks*) qui imite le fonctionnement du cerveau humain, et deuxièmement, une technique d’entraînement de la machine, appelée «apprentissage profond» (*deep learning*). Rien de révolutionnaire en soi: le concept de réseau de neurones artificiels date de 1943^[2], et la méthode d’apprentissage par renforcement est une traduction informatique du conditionnement psychologique, théorisé au début du xx^e siècle par les célèbres expériences d’Ivan Pavlov sur les chiens. Pour résumer, on simule des neurones dans un ordinateur et on les entraîne à identifier des motifs grâce à un système de récompense.

Après des décennies sans résultats probants, la méthode se met subitement à fonctionner en 2012^[3] lorsque de nouveaux logiciels, appelés «réseaux de neurones convolutionnels», parviennent à reconnaître des éléments (un arbre, un chat, un vélo) dans une image. Au cours des années suivantes, les machines «voient», «lisent» et «entendent» de mieux en mieux. À partir de 2014, Google met à contribution l’ensemble des usagers du net pour entraîner ses IA: c’est l’ère des CAPTCHAs, ces processus d’authentification par lesquels on nous demande de «prouver» que nous ne sommes pas des robots en

identifiant un animal, un panneau routier ou une voiture. Les premiers robots conversationnels, capables de «dialoguer» à l'écrit avec un être humain, sont lancés sur le marché – les Siri, Alexa et Assistant, embarqués dans nos smartphones puis dans des enceintes connectées. En même temps que nous apprenons à interagir à l'écrit puis à l'oral avec ces nouvelles présences, nous faisons l'expérience de leurs limites. L'IA reste un gadget, une curiosité technique dont on se lasse.

En 2017, l'apparition d'une nouvelle variante du réseau de neurones, appelée *transformer*, provoque une seconde étincelle. L'écosystème de la recherche en IA s'enrichit alors énormément. Pour entraîner leurs programmes, les chercheurs disposent désormais d'immenses bases de données dans lesquelles chaque élément (d'un texte ou d'une image) est décrit ou contextualisé par un être humain. Ces données de qualité sont un carburant de premier choix pour les algorithmes. En parallèle, l'industrie a découvert quelques années plus tôt que les cartes graphiques (*graphic processing unit*, GPU), habituellement utilisées par les *gamers* pour générer les graphismes de leurs jeux vidéo, sont beaucoup plus efficaces que les traditionnels processeurs (*computer processing unit*) pour faire fonctionner les réseaux de neurones.

Un moteur plus puissant, un carburant de meilleure qualité, une architecture plus efficace: les «grands modèles de langage» (*large language models*) sont nés. Leur capacité à lire, contextualiser et générer du langage dit «naturel» en prédisant le prochain mot d'une phrase est inédite. Jadis limités à quelques phrases, ils sont désormais capables de générer des conversations entières et de grands blocs de texte. Appliqués au code informatique, comme le logiciel Copilot, ils automatisent une partie du travail des développeurs. D'autres systèmes basés sur la même architecture génèrent de l'image, du son et désormais de la vidéo. À chaque nouvelle version, les productions synthétiques se rapprochent du réalisme humain. On parle alors d'IA «générative».

Outre leur capacité à imiter le langage humain, les *chat-bots* de cette nouvelle génération démontreraient également des capacités dites «émergentes^[4]», qui dépassent leur programmation initiale. Ces nouvelles compétences (traduire d'une langue vers une autre, coder des programmes informatiques, etc.), impossibles à prévoir ou à expliquer,

fascinent particulièrement leurs concepteurs. En 2022, le chercheur de Google Blake Lemoine, qui passe des heures à interagir avec un prototype de *chatbot* appelé LaMDA, se persuade qu'il parle à une entité consciente^[5]. Google dément, mais la conversation médiatique est lancée.

En décembre 2022, OpenAI présente ChatGPT au monde. Cette machine à imiter l'intelligence, imparfaite et limitée, devient soudainement accessible à tout un chacun, gratuitement et grâce à un simple navigateur web, dans un contexte médiatique déjà propice aux questionnements sur la machine qui pense. Viralité totale. Le programme rassemble 100 millions d'utilisateurs en quelques semaines, ce qui représente la plus grande vitesse de diffusion de l'histoire pour un logiciel. Le choc est mondial.

Aujourd'hui, la stratégie industrielle des nouvelles entreprises d'IA est simple: plus elles disposent de puissance de calcul et de données, plus leur logiciel accumule des variables ou paramètres, plus il est performant. Quitte à créer des systèmes absolument gigantesques, capables d'engloutir l'intégralité des textes du web. C'est ce qu'on appelle le *scaling*: il faut croître, exponentiellement, ou mourir. S'amorce une féroce compétition, une espèce de course aux armements avec des cartes graphiques au lieu des missiles balistiques et des bases de données en guise de munitions. À mesure que l'humanité découvre les capacités hypnotisantes de cette nouvelle génération de logiciels, version après version, le discours apocalyptique grandit et enfle, porté par une poignée de nouveaux patrons qui surfent sur la sidération collective.

Tous vers l'intelligence artificielle générale

En 2024, leur chef de file est sans aucun doute Sam Altman, patron d'OpenAI – né en 1985, visage juvénile et sourire affable.

Sa biographie, qui se transforme lentement mais sûrement en légende, s'inscrit déjà dans la mythologie des Steve Jobs, Bill Gates et autres saints patrons de la Silicon Valley. Sam, raconte la presse états-unienne dans de longs portraits fascinés, a été un enfant prodige, capable de réparer un magnétoscope à trois ans. Son premier ordinateur, un Mac LC

Il offert pour ses huit ans, a changé sa vie, marquant «un avant et un après». Sa découverte des *chat rooms* d'AOL est un événement «transformatif^[6]». Il entre à Stanford en 2003 pour y étudier l'informatique – la voie royale pour évoluer dans la tech –, mais quitte la prestigieuse université deux ans plus tard sans diplôme pour fonder une *start-up*. Il rejoint ensuite le plus célèbre incubateur de la Silicon Valley, Y Combinator, à la tête duquel il va ensuite être coopté en 2014.

Vingt ans après un début de parcours chaotique, le jeune pape de l'IA est parfaitement intégré à l'oligarchie technocapitaliste californienne. Il a investi dans une centaine de *start-up* et sa fortune personnelle vient de dépasser le milliard de dollars. Comme les autres technomagnats, il conduit des McLaren et vole dans le ciel de la Californie en jet privé. Comme les autres technomagnats, il entasse «des armes, de l'or, des pastilles d'iode, des antibiotiques, des batteries, de l'eau et des masques à gaz de l'armée israélienne^[7]» dans un bunker de Big Sur, en préparation de la fin du monde.

OpenAI, qu'il a cofondé en 2015 avec Elon Musk et qu'il dirige encore aujourd'hui, ne ressemble à aucune autre entreprise. Elle naît sous la forme d'une association à but non lucratif, dont la «mission» est de créer «un ordinateur capable de penser comme un humain et l'utiliser pour le plus grand bénéfice de l'humanité». Cet ordinateur fabuleux, ou du moins son concept, a un nom: l'intelligence artificielle générale (*artificial general intelligence*, AGI), pour la distinguer de l'intelligence artificielle considérée comme «étroite» ou spécialisée. Puisque le concept reviendra souvent dans cet ouvrage, précisons d'emblée: l'AGI est une machine entièrement hypothétique, dont la faisabilité même fait débat. Chez OpenAI, on est néanmoins persuadé de pouvoir la construire dans un futur proche. Et c'est à partir de là que ça devient étrange.

Selon ses théoriciens, fermement ancrés dans la science-fiction, une telle machine serait également capable de s'auto-améliorer, de devenir plus intelligente et de résoudre des problèmes toujours plus complexes. Or, c'est justement cette idée d'une progression exponentielle qui complique les choses. Car selon les prophètes, celle-ci mènerait invariablement à la «superintelligence», une machine si puissante qu'elle nous reléguerait au statut de fourmis. L'hypothèse, si improbable soit-

elle, est prise très au sérieux par OpenAI. Sur son blogue, Sam Altman écrit que «le développement d'une IA surhumaine est probablement la menace la plus sérieuse à la pérennité de l'humanité^[8]». Il est cependant tout aussi convaincu que la machine, placée entre de «bonnes mains», pourrait «bénéficier à toute l'humanité». Catastrophiste d'un côté, utopiste de l'autre, il incarne à la perfection le paradoxe idéologique sur lequel repose non seulement la foi réversible des prophètes, mais le fonctionnement très concret d'OpenAI.

Pour mener à bien la «mission», Altman s'entoure uniquement des gens qui partagent sa vision radicale de l'AGI. Lorsqu'il les recrute en 2015, ce sont de véritables iconoclastes: à l'époque, la recherche en IA se concentre sur des outils aux capacités étroites, et presque personne ne croit à la possibilité d'une intelligence plus générale. Les croyants sont perçus par le reste de la profession comme des marginaux et des mystiques. Neuf ans plus tard, tout a changé, mais OpenAI est toujours un rassemblement de croyants. Les 500 employés de l'entreprise «sont fanatiques dans la poursuite de leur mission», écrit le magazine *Wired* en septembre 2023, tout en précisant qu'on ne peut pas pour autant qualifier OpenAI de secte^[9].

On aurait tort de résumer Altman à un simple gourou. Certes, il croit fermement à l'imminence de la superintelligence omnipotente, mais s'il a choisi de bâtir son entreprise sur le modèle d'une communauté religieuse, c'est aussi par souci stratégique. Dans un *post* de blogue rédigé en 2013 et intitulé «Successful People» (Ceux qui réussissent), Altman, encore totalement inconnu du grand public, écrit que «les fondateurs les plus talentueux ne cherchent pas à créer des entreprises. Ils sont en mission pour créer quelque chose qui ressemble à une religion, et il s'avère que passer par une entreprise est la manière la plus simple d'y parvenir^[10]». Pour s'enrichir dans l'économie de la *start-up*, le mieux est encore de fonder un mouvement sectaire.

Façonnés par ce catéchisme technoreligieux, tous ces nouveaux faiseurs d'IA sont persuadés, en bons *Übermensch*, d'œuvrer simultanément à transcender et à sauver l'humanité. Biberonnés à l'éthos technocapitaliste de la «destruction créatrice», ils se voient comme des pionniers, grands explorateurs seuls à naviguer l'océan de la recherche

vers la *terra incognita* de l'AGI. Certains se comparent aux chercheurs du projet Manhattan^[11]. Le livre *The Making of the Atomic Bomb*, qui raconte les dilemmes moraux d'un groupe de chercheurs développant un outil capable d'annihiler l'humanité, est devenu une lecture obligatoire dans leurs cercles^[12]. Et Sam Altman ne perd jamais une occasion de citer l'aphorisme souvent attribué à Oppenheimer – «la technologie advient parce qu'elle est possible» –, et d'ajouter du même souffle qu'il est né un 22 avril, comme le chef du projet Manhattan. En revanche, il se garde bien d'imiter son modèle citant la *Bhagavad-Gîtâ* après avoir constaté les terribles effets de sa création: «Je suis devenu la Mort, le destructeur des mondes.»

L'autre *start-up* «historique» du secteur est la britannique DeepMind, fondée en 2010 et rachetée par Google en 2014 pour 500 millions d'euros. Ses créateurs: Demis Hassabis, qui a été numéro 2 mondial des échecs à l'âge de 14 ans, Mustafa Suleyman, ancien étudiant en théologie à Oxford, et Shane Legg, l'un des premiers théoriciens de l'AGI au début des années 2000. Avant l'émergence d'OpenAI, fin 2022, c'était DeepMind qui incarnait la pointe de la recherche en IA. En 2016, son logiciel AlphaGo a battu le champion du monde de go, un exploit technique jusqu'alors considéré comme impossible. En 2020, son second logiciel d'IA, AlphaFold, relevait l'un des plus grands défis de la biologie moderne: prédire automatiquement la structure de centaines de millions de protéines, un processus jusque-là manuel, extrêmement fastidieux et très coûteux.

Comme chez OpenAI, le but est de fabriquer la première «superintelligence» artificielle^[13] pour en faire bénéficier l'humanité. On rétorque aux inquiets qu'avoir plusieurs acteurs en concurrence sur le marché de l'IA permettra d'éviter les dangers d'un monopole. Il n'en demeure pas moins que pour un membre du conglomérat Alphabet (qui détient Google), condamné à plus d'une reprise pour concurrence déloyale et qui a depuis longtemps abandonné son vieux slogan «*Don't be evil*» (Ne commets pas le mal), il est difficile de passer pour un philanthrope.

Dans le sillage de ces deux colosses de l'IA, une nouvelle génération d'entreprises, toutes en mission pour développer la «superintelligence»,

tente de se tailler une place sur ce marché qui fait rêver tous les fonds d'investissement de la planète. Mustafa Suleyman crée Inflection AI en 2022. Les frère et sœur Dario et Daniela Amodei, fraîchement débarqués d'OpenAI, fondent Anthropic en 2021. Elon Musk inaugure la dernière venue, xAI, le 12 juillet 2023 pour «comprendre la véritable nature de l'Univers» et empêcher «une dystopie de l'IA». Rien de moins.

Avec une poignée d'autres acteurs, parmi lesquels Hugging Face, fondée en France, ces *start-up* ont capté près de 21 milliards de dollars d'investissements cumulés, dont 11 milliards au premier trimestre de 2023. OpenAI est désormais contrôlée à 49 % par son «partenaire» et premier actionnaire Microsoft, qui a investi 13 milliards de dollars fin 2023 pour obtenir l'usage exclusif du modèle GPT. Chacune des entreprises citées est devenue une «licorne» – une *start-up* évaluée à plus d'un milliard de dollars, en jargon de capital-risqueur. Et à en croire les prévisions du cabinet McKinsey, toujours optimiste sur la question, le marché de l'IA générative pourrait générer à lui seul entre 2500 et 4400 milliards de dollars par an. À titre comparatif, précise McKinsey, le PIB du Royaume-Uni est de 3300 milliards de dollars^[14].

L'heure devrait être à la fête, et l'optimisme déborder de partout. La formidable machine à utopies de la Silicon Valley devrait tourner à plein régime. C'est pourtant l'inverse qui se produit.

Sam Altman, rock star du lobbying

Depuis la fin 2022, les oligarques de l'IA jouent les Cassandre... contre eux-mêmes.

Chaque fois qu'ils interviennent dans les médias, c'est pour nous mettre en garde contre les risques planétaires que présentent, selon eux, les outils qu'ils sont précisément en train de développer. Ils évoquent leurs créations avec déférence et fascination, puisent dans les lexiques ésotériques et religieux, multiplient les hyperboles. Le ton est grave, le regard, solennel. On ne rigole pas avec l'IA. Ce serait «la technologie la plus importante sur laquelle l'humanité travaille: plus importante que le feu, l'électricité ou tout ce que nous avons pu construire par le passé», selon Sundar Pichai, PDG d'Alphabet. Pire: ce pourrait être «la dernière

invention de l'humanité^[15]», comme le prophétise Satya Nadella, PDG de Microsoft. La formule résume parfaitement l'eschatologie ambiguë de l'élite de la tech. Dans un scénario, l'IA est notre dernière invention, car elle assurera notre transcendance technique. Dans l'autre, elle est notre dernière invention, car elle nous annihilera. En équilibre sur une crête entre le paradis et l'enfer, nous pourrions aussi bien basculer d'un côté ou de l'autre.

Au jeu du lobbying, aucun des prophètes ne donne plus de sa personne que Sam Altman. Depuis qu'il a lancé ChatGPT en décembre 2022, le PDG d'OpenAI s'est transformé en un prosélyte acharné. Aux États-Unis, il a rencontré près d'une centaine de députés, démocrates comme républicains, afin de les convaincre d'adopter les cadres de régulation les plus avantageux pour son activité^[16]. Altman a même été reçu plusieurs fois à la Maison-Blanche, seul et en compagnie d'autres leaders de l'IA, où il a été entendu par Joe Biden, la vice-présidente Kamala Harris et plusieurs autres membres du gouvernement.

Le 16 mai 2023, dans ce qui est quasiment devenu un rite de passage solennel pour patrons démiurges de la tech, Altman a été auditionné par le Sénat états-unien. Et une nouvelle fois, le jeune patron a innové. Devant des élus éberlués, habitués aux auditions glaciales d'un Mark Zuckerberg, il a réitéré que sa «plus grande peur est que nous, ceux du secteur, l'industrie de la tech – nous causions des dégâts significatifs au monde^[17]», s'est dit impatient de collaborer avec le gouvernement et a presque imploré les élus de créer une agence gouvernementale de régulation de l'IA qui distribuerait des permis d'exploitation de ces systèmes. Dans l'histoire des rapports conflictuels entre patrons de la tech et gouvernement, c'est sûrement du jamais vu.

Après avoir rodé son spectacle sur la scène nationale, Altman, en véritable rock star, s'est ensuite embarqué dans une tournée mondiale de cinq semaines. Son objectif: prêcher la bonne parole de la régulation auprès des décideurs politiques et économiques du monde. Sa tournée l'a mené dans 25 villes sur 5 continents en un mois, à un rythme d'enfer. Des rendez-vous avec les principaux leaders d'Occident, d'Israël, de Corée du Sud, de Chine, d'Inde et des Émirats arabes unis, ainsi qu'avec la présidente de la Commission européenne, Ursula von der Leyen. Un

maelström de conférences, de débats, de discours et de séances d'autographes. Et comme si les leaders politiques ne suffisaient pas, l'infatigable PDG d'OpenAI a également pris le temps d'évangéliser les élites politiques et économiques à la très discrète conférence Bilderberg du 18 mai 2023 à Lisbonne, aux côtés des hauts gradés de DeepMind, Google et Microsoft. Mission accomplie: l'IA était alors sur toutes les lèvres.

Petit à petit, «l'offensive de charme diplomatique^[18]» a produit ses effets. Un exemple en quatre actes: en mars 2023, le gouvernement britannique publiait un livre blanc prônant une approche «pro-innovation», très libérale, du secteur de l'IA, sans même évoquer les risques liés au déploiement de tels outils dans la sphère publique et privée. Le 25 mai, Sam Altman et plusieurs autres patrons de l'IA sont invités à Downing Street pour échanger avec le premier ministre britannique Rishi Sunak. Le soir même, dans une volte-face totale par rapport au livre blanc publié deux mois plus tôt, le communiqué émis par le cabinet du premier ministre évoque les «risques existentiels» posés par une «superintelligence». Enfin, le 12 juin, Sunak annonce le déploiement d'une «équipe de recherche de pointe sur la sécurité» en IA, dotée d'un budget de 100 millions de livres pour développer une «*safe AI*», et soutient le projet de création d'une agence internationale de régulation, calquée sur l'Agence internationale de l'énergie atomique (AIEA). Une idée proposée par nul autre que Sam Altman, 15 jours plus tôt, sur le blogue d'OpenAI^[19].

Le récit prophétique se propage: le 12 juin, le Secrétaire général des Nations Unies António Guterres explique à son tour que «les scientifiques et les experts [...] ont déclaré que l'IA incarnait un risque existentiel, comparable à un conflit nucléaire^[20]». Le 14 septembre 2023, dans son discours sur l'état de l'Union, Ursula von der Leyen reprend mot pour mot la lettre ouverte du Center for AI Safety en évoquant la nécessité de «limiter les risques d'extinction posés par l'intelligence artificielle». La phrase se retrouvera ensuite dans la communication de l'Union européenne sur le sujet, sans guillemets^[21]. Un détail, certes, mais qui est révélateur de l'influence de l'industrie de

l'IA, qui souffle littéralement des concepts et des éléments de langage aux législateurs.

Les effets de ce lobbying intense se font ressentir au-delà des milieux politiques et jusque dans la société civile. Jadis sujet de niche pour une petite communauté d'universitaires et de chercheurs en sciences informatiques, l'éthique de l'intelligence artificielle est devenue le sujet d'une conversation mondiale. Un lexique jadis réservé à des communautés d'initiés s'est invité dans le débat public. Journalistes et responsables politiques manient désormais librement des termes comme «intelligence artificielle générale», «superintelligence», «alignement» et «risques existentiels», sans savoir exactement de quelle communauté ils émergent ou le projet politique qu'ils recouvrent.

La bataille des prophéties

Mois après mois, on assiste à une apparente balkanisation du secteur de l'IA, ce fatras de *start-up*, de laboratoires de recherche, de départements de philosophie et d'oligopoles de la tech. De vieilles querelles métaphysiques, jusqu'ici cantonnées à d'obscurs forums de spécialistes, se règlent désormais à coups d'éditos, de *tweets* et de tribunes. Des fronts idéologiques se consolident, reflétant chacun la panique ou le scepticisme face à la «superintelligence» qui vient. On peut, grosso modo, les découper en quatre factions.

Les modérés, parmi lesquels le Français Yann Le Cun, chercheur en chef chez Meta/Facebook, s'opposent à la régulation de leurs activités, et répètent d'un ton rassurant que les intelligences artificielles sont encore très loin d'égaliser nos capacités cognitives. Cette faction est la plus discrète dans les médias, il va de soi.

Ensuite, il y a les illuminés, dont Sam Altman est l'incarnation la plus visible et le philosophe suédois Nick Bostrom, le théoricien le plus populaire. Ce sont eux qui nous intéressent le plus. Ils sont ingénieurs, patrons, éthiciens ou investisseurs et travaillent pour la plupart dans les compagnies qui produisent les systèmes d'IA actuels. Certains sont des pionniers de la recherche en informatique. À force de jouer les

Prométhée modernes, ils ont développé une relation apparemment ambivalente à leur création, entre crainte et admiration.

Entre le calme olympien des premiers et l'exaltation mystique des autres, il y a le pessimisme sincère de ceux qui sont persuadés que la naissance de cette «superintelligence» provoquera inexorablement la fin d'Homo sapiens. Surnommés les «*doomers*», ils appellent à l'arrêt pur et simple des recherches en IA. Ils sont guidés par l'influent théoricien et «génie» autoproclamé Eliezer Yudkowsky. Cet ancien apôtre de l'IA dans les années 1990, qui a fait acte d'apostasie, s'est donné pour mission d'empêcher l'émergence de la «superintelligence». Signe des temps, ses scénarios techno-apocalyptiques délirants sont pris au sérieux et commentés dans la presse grand public, si bien que des magazines comme *Time* ou *Rolling Stone* le décrivent comme «un fou ou un prophète», mais assurent que, «aussi dingue que ça puisse paraître, [ce qu'il dit] n'est désormais plus si dingue que ça^[22]».

Enfin, à l'autre extrême, on trouve les purs optimistes, une frange minoritaire mais puissante représentée par l'ex-PDG de Google Larry Page et les investisseurs-stars de la Silicon Valley Marc Andreessen, Peter Thiel et Reid Hoffman. Eux affirment qu'il n'y a aucune raison de paniquer puisque l'IA «va sauver le monde» en «améliorant tout ce qui nous importe^[23]», et que ralentir le développement de la machine qui pense est «spéciste^[24]». Leur maître à penser, Ray Kurzweil, gourou du futurisme chez Google, professe depuis vingt ans l'évangile d'une fusion entre l'être humain et la «superintelligence» qui vient comme aller simple vers l'immortalité.

Ces querelles façonnent l'opinion publique. Selon un sondage Odoxa paru le 13 juin 2023, 67 % des Français sondés considéraient l'IA comme une menace^[25], un chiffre qui a augmenté de 14 points en deux mois. Ailleurs, 42 % des patrons réunis au Yale CEO Summit en juin 2023 estimaient que l'IA avait «le potentiel de détruire l'humanité d'ici cinq à dix ans». Plus alarmante, cette angoisse semble aussi affecter celles et ceux qui développent les algorithmes d'intelligence artificielle, et jusqu'à certaines des sommités de la recherche en IA.

Geoffrey Hinton, ex-chercheur de Google, s'est ainsi offert une tournée médiatique en mai 2023 pour avertir le monde des risques posés

par l'émergence d'une «superintelligence»: «La sonnette d'alarme que je tire est celle d'un risque existentiel [...]. Jusque-là, je pensais que nous en étions loin, mais je pense désormais que la menace est très sérieuse.» Même rengaine chez Yoshua Bengio, spécialiste québécois de l'apprentissage machine, qui décrit l'intelligence artificielle comme une «menace existentielle^[26]».

L'inquiétude fait presque consensus: au printemps 2022, sur 550 chercheurs et chercheuses interrogés par le Machine Intelligence Research Institute d'Oxford, près de la moitié (48 %) se disait convaincue que si une superintelligence était créée un jour, elle aurait au moins 10 % de risques de détruire l'humanité^[27]. Ne resterait plus qu'une façon d'éviter de disparaître: nous hybrider à la machine, nous soumettre à l'implacable marche de l'évolution.

Le simulacre de l'IA fonctionne aussi sur le grand public. Sidérés par le régime incantatoire de l'informatique, où la machine répond instantanément à une formule magique (le *prompt*, soit la requête) sans rien laisser transparaître du processus de gestation, nous nous sommes progressivement laissé convaincre par le mensonge de l'ordinateur conscient et de l'avènement de l'ère des machines. Alors, la hiérarchie des normes s'inverse: puisque le futur appartient aux robots, l'élocution du *chatbot* devient la mesure de l'intelligence, tandis que l'intelligence humaine est sans cesse redéfinie à la baisse pour correspondre aux compétences des machines contemporaines. Le philosophe Günther Anders parle de «honte prométhéenne^[28]»: l'humain, envahi par un sentiment de déclassement face à la supposée perfection de sa machine, tente en retour de se faire passer pour elle. La panique ontologique se répand. On rêve d'hybridation, d'augmentation, de téléchargement de cerveaux, sous peine d'être remplacés.

En juin 2023^[29], la journaliste de *Libération* Clémence Mary proposait de nommer ce mal contemporain «IA-anxiété», sur le modèle de l'écoanxiété, l'autre grande angoisse du siècle. L'intelligence artificielle et les bouleversements climatiques sont deux exemples de ce que le philosophe Timothy Morton appelle des «hyperobjets^[30]»: des phénomènes trop vastes, trop complexes et trop perturbants pour qu'on puisse vraiment les appréhender, et face auxquels les discours

apocalyptiques deviennent une sorte de refuge du sens. L'incompréhension devient une peur, et la complexité de la technique devient menaçante. Les prophètes, on le verra, ne disent rien d'autre: l'IA serait dangereuse *parce qu'elle est intelligente*. Or, elle n'est ni l'un ni l'autre.

Un air de déjà-vu

Alors, respirons profondément. Le moment que nous vivons s'est déjà produit. Plusieurs fois.

Comme l'a démontré la chercheuse Melanie Mitchell, depuis les années 1950, l'histoire de l'IA est une alternance de moments d'attention politique et médiatique intense – les «printemps» (*AI spring*) – et de périodes de désintérêt – les «hivers» (*AI winter*)^[31]. Quant aux oracles apocalyptiques et aux épiphanies religieuses liées à l'émergence d'intelligences synthétiques, ils sont aussi vieux que les techniques elles-mêmes. Chaque nouvel exploit – l'ordinateur Watson gagnant au jeu *Jeopardy!* en 2011, le supercalculateur Deep Blue battant Garry Kasparov aux échecs en mondovision en 1997 – s'accompagne des mêmes prophéties et des mêmes avertissements. Seul le vocabulaire change, éternellement mis à jour pour donner l'illusion d'une nouveauté.

La dernière fois que l'IA nous a fait collectivement paniquer avant l'entrée en scène de ChatGPT, c'était en 2015. Souvenez-vous: les prophètes s'appelaient Stephen Hawking et Elon Musk. La menace contre laquelle ils voulaient nous prémunir: les «robots tueurs». Les pétitions alarmistes^[32] faisaient déjà l'objet d'un traitement médiatique international. Une décennie plus tard, Hawking s'en est allé rejoindre les forces fondamentales de l'Univers, mais rien n'a changé. Reprenant les mêmes méthodes de lobbying, patrons et ingénieurs de la Silicon Valley s'improvisent tour à tour vendeurs d'utopies ou prophètes de malheur face à des investisseurs subjugués qui ont besoin de dépenser, un public captif et angoissé ou une classe politique qui peine à cacher son illettrisme technique.

Et toujours pas l'ombre d'une machine pensante. Obnubilés par le miroir déformant de l'IA générative, nous oublions vite que l'entité qui

nous fait face n'est rien de plus qu'un outil d'autocomplétion de texte, qui repose sur le même principe que l'algorithme qui nous aide à composer nos textos. Pour reprendre le terme inventé par quatre chercheuses critiques de l'IA dans un article devenu célèbre, ces systèmes sont en réalité de simples «perroquets stochastiques^[33]», des calculateurs de probabilités qui n'ont aucune compréhension des mots et concepts qu'ils reprennent. Le *chatbot* ne sait pas ce qu'il dit, il estime ce qu'il doit dire.

La maîtrise apparente du langage n'est pas synonyme d'intelligence, comme le détaillait récemment Noam Chomsky: «Pour avoir raison, il faut pouvoir avoir tort. L'intelligence se compose non seulement de conjectures créatives, mais de critique créative^[34].» Les grands modèles de langage, si sophistiqués soient-ils, «sont incapables de distinguer le possible de l'impossible», et ne peuvent donc prétendre à la plus élémentaire pensée. Ce sont des simulacres, puissants mais vides.

Autre simulacre: celui de la nouveauté. Derrière les fables marketing d'une accélération du progrès, la recherche d'une machine pensante n'a pas fondamentalement avancé depuis les années 1950. Le champ de l'IA se heurte toujours au paradoxe identifié en 1988 par le roboticien Hans Moravec, selon lequel les tâches complexes de raisonnement habituellement attribuées à une intelligence «adulte» – rédiger un texte, jouer aux échecs, résoudre des équations – sont paradoxalement plus simples à simuler informatiquement que le *bon sens* que l'être humain acquiert instinctivement. Ce paradoxe renforce l'illusion du progrès: plus nous voyons les machines accomplir des tâches que nous jugeons «difficiles», plus nous oublions qu'elles sont incapables de raisonnements «élémentaires». Et ce n'est pas la Silicon Valley qui va nous le rappeler.

Keep Calm and Reclaim

Cet ouvrage propose une alternative à la sidération.

L'intelligence artificielle désigne énormément de choses. Elle est un champ de recherche pluridisciplinaire, un discours publicitaire et religieux, un outil informatique complexe, mais également un

phénomène social planétaire dont les enjeux devraient être vulgarisés et rendus compréhensibles à tous et toutes. Pour retrouver une autonomie critique, nous devons apprendre à reconnaître le récit propagé par la caste technocratique et le replacer dans le champ des idées politiques, loin des modes magiques et religieux qui l'exemptent de toute critique.

Derrière une apparente pluralité, les deux discours apocalyptiques de l'IA n'en forment qu'un. Les tenants de la première version, les exaltés, nous assurent qu'elle assurera la transcendance collective. Les autres, épouvantés, y voient le début de notre extinction. La majorité croit simultanément en ces deux possibilités, comme les chrétiens croient à la fois au paradis et à l'enfer. Transcendance ou extinction: il n'y aurait, comme d'habitude, *no alternative*. L'intelligence artificielle serait donc inévitable.

Qu'ils soient utopistes ou catastrophistes, les prophètes de l'IA appartiennent à une même communauté. Ils viennent des mêmes campus universitaires, assistent aux mêmes conférences, sont invités aux mêmes réunions, investissent dans les mêmes *start-up* et financent les mêmes *think tanks*. Tous fétichisent la rationalité et la méthode scientifique, se vantent de penser comme des machines et d'anticiper ainsi l'hybridation inéluctable. Ils recyclent la vieille rengaine objectiviste et libertarienne d'Ayn Rand dans *La grève*: l'avenir de l'espèce est entre les mains des ingénieurs, des chercheurs et des investisseurs. Dans cette vision du futur, le gouvernement et la société civile ne sont que des parasites, des obstacles à la glorieuse marche du progrès.

Deux phénomènes sont à l'œuvre. En surface, tous ces discours constituent la stratégie marketing des producteurs d'IA, qui consiste à surévaluer les retombées positives et négatives de leurs logiciels pour capter des investissements. Dans l'univers de l'innovation où le principe schumpeterien de la «destruction créatrice» règne sous le nom de «disruption», le danger fait vendre. Certains critiques, comme la linguiste et chercheuse en informatique Emily M. Bender^[35], y voient une stratégie de marketing puissante et insidieuse qui «fait paraître le produit plus puissant qu'il ne l'est» et, par corollaire, érige ses créateurs en mages technomanciens, seuls capables de comprendre et de contrôler le mystère des algorithmes.

Or, on ne débat pas avec des demi-dieux. On fait ce qu'ils nous disent de faire. Surtout quand ils nous assurent que l'avenir de l'humanité est en jeu. Une fois qu'ils ont attiré l'attention de tout le monde (surtout celle de la classe politique) en hurlant à l'apocalypse, nos prophètes n'ont plus qu'à promouvoir leurs solutions sur mesure auprès des instances décisionnelles.

Cette stratégie, baptisée «*criti-hype*» par l'universitaire états-unien Lee Vinsel en 2021^[36], est également un puissant outil de lobbying. Elle permet de donner l'impression d'une technologie trop complexe pour être régulée par les non-initiés. On l'a vu, les prophètes se donnent beaucoup de mal pour nous montrer que l'IA est un sujet complexe, impénétrable, voire occulte. Or, si les techno-oligarques sont les seuls à être suffisamment érudits pour comprendre la mystérieuse machine, ils sont également les seuls à pouvoir en définir le cadre de régulation. Pour les Google, Microsoft, Amazon, Meta et autres, entretenir le mythe d'une technologie omnipotente n'offre donc que des avantages: plus de capitaux, moins d'interférence politique et médiatique, et une totale liberté de mouvement pour consolider leur influence. Comme le résume Suresh Venkatasubramanian, ex-conseiller en IA de Joe Biden et co-auteur en 2022 de l'AI Bill of Rights, «un vaste programme politique se déguise sous les traits d'un débat technologique^[37]». Il faudrait analyser ses enjeux et ses stratégies à l'aune du «climato-scepticisme, de la "science" de l'industrie du tabac et du racisme scientifique^[38]».

Derrière les stratégies d'influence et les campagnes de marketing se cache enfin la poursuite d'une utopie. Sous ce vernis séculariste et rationnel, les prophètes de l'IA empruntent allègrement aux eschatologies religieuses, et particulièrement au récit chrétien de la fin des temps. Ils croient à l'avènement inéluctable de la «superintelligence» artificielle, omnisciente et omnipotente. Un nouveau dieu, union sacrée de la science et de la technique, qui marquera la fin des temps et fera le tri entre croyants et hérétiques. Comme l'a un jour prophétisé Sam Altman dans un manifeste titré «The Merge» (La fusion): «À moins que nous ne nous autodétruisions avant, l'IA surhumaine va exister^[39].» Sur ce point-là, tous les prophètes sont d'accord. La foi des patrons de la

Silicon Valley est-elle sincère? Probablement, tant leur utopie épouse leurs intérêts financiers et leur quête de pouvoir.

Ça peut paraître surprenant, mais l'Église et la technique s'amalgament en réalité depuis des siècles. Comme l'écrit l'historien David Noble dans *The Religion of Technology*, dès le Moyen Âge, l'Église catholique réalise que le progrès technique peut aider l'humanité à recouvrer sa perfection originelle, celle du jardin d'Éden, en quittant son corps biologique, synonyme de péché. La technologie, écrit-il, «devient synonyme de transcendance, partie intégrante de l'idée chrétienne de rédemption^[40]». Les monastères médiévaux, portés par la devise *ora et labora* (prière et travail), deviennent des lieux d'innovation, une sorte de Silicon Valley avant l'heure. Les moines sont également ingénieurs, inventeurs et penseurs de la technique, dans un cadre strictement théologique.

Aujourd'hui, la Silicon Valley et le Vatican chantent une ritournelle similaire: aide l'IA et l'IA te sauvera. L'apocalypse informatique n'est au fond qu'une reconfiguration de la théologie évangélique de l'Enlèvement (*Rapture*), cet événement censé marquer le début de mille ans de prospérité et au cours duquel croyants et vertueux, vivants ou ressuscités, devraient s'élever à la rencontre du Seigneur pour être sauvés. À l'ère du sécularisme occidental, ce sont maintenant les ingénieurs d'élite et les barons de l'innovation technique qui jouent les hommes de foi, dans une Silicon Valley transformée en Terre sainte pour technoprophètes.

Observer cette communauté de chercheurs, d'investisseurs et de théoriciens, c'est plonger dans un enchevêtrement de dogmes et de mythes technoreligieux, dont certains persistent depuis les années 1970 et d'autres sont nés il y a moins de dix ans. En février 2023^[41], Émile P. Torres, philosophe et historien·ne^[42] des techniques, et Timnit Gebru, chercheuse en éthique de l'IA, ont baptisé ce mouvement d'un disgracieux acronyme: TESCREAL, pour «Transhumanisme, extropianisme, singularitarisme, cosmisme, rationalisme, altruisme efficace et long-termisme». Le socle de ce mouvement est le transhumanisme, dont émergent ensuite les autres écoles de pensée.

Nous étudierons ici chacune des idéologies qui composent cette nébuleuse et nous explorerons leurs origines, leurs textes sacrés, leur

lexique. Nous décrirons leurs bastions de rationalité et de sécularisme que sont la Silicon Valley californienne et les campus de Stanford, Oxford ou du MIT. Nous lirons la littérature pseudoscientifique qu'elles produisent pour maintenir leur vernis de respectabilité universitaire. Nous verrons comment la Silicon Valley s'est construite sur des croyances transhumanistes, comment ces croyances apocalyptiques ont influencé la recherche en IA dès ses origines, et comment une poignée de théoriciens futuristes a transformé son discours apocalyptique en un champ de recherche académique respectable, conçu sur mesure pour des campagnes de lobbying à destination des législateurs.

Nous identifierons ensuite leurs réseaux d'influence: centres de recherche, *think tanks*, *start-up* et autres instituts de prospective, qui s'activent aussi bien sur le web que dans les couloirs du Congrès des États-Unis et des institutions internationales. Nous mettrons au jour leurs circuits de financement, qui se chiffrent désormais en dizaines de milliards de dollars. Nous verrons comment cette congrégation technoreligieuse renforce la domination des entreprises privées sur les institutions publiques, et tue dans l'œuf toute velléité de régulation des États. Au-delà de leurs intérêts stratégiques pour les patrons de l'IA, nous verrons comment ces utopies technoreligieuses, conçues pour satisfaire l'élite universitaire, financière et technique de la Silicon Valley légitiment la domination des puissants, enferment le récit du futur dans un cadre élitiste et eugéniste, et flirtent avec une vision suprémaciste blanche.

Nous dessinerons pour terminer un autre portrait de l'IA, critique et politique. Celui d'une structure de captation de capital financier, de déploiement d'infrastructures matérielles, de consommation de ressources énergétiques, d'extractivisme minier et numérique, d'exploitation du prolétariat, de néocolonialisme et de luttes géopolitiques. Car même si la «superintelligence» reste une hypothèse, les effets des logiciels d'analyses de données sur le tissu social, les structures de pouvoir et les relations interpersonnelles sont bien palpables, eux. Au lieu de relayer des promesses d'abondance ou des avertissements apocalyptiques, nous soulignerons les inégalités que l'IA exacerbe, ici et maintenant.

Les robots tueurs, les algorithmes malveillants et les superintelligences apocalyptiques n'existent pas plus que le *deus ex machina* qui nous rendra riches et immortels. Il n'y a que des rapports de domination à abattre, des structures de gouvernance à imaginer, des choix politiques à faire, des luttes à mener. Un futur à imaginer ensemble, avec ou sans machines intelligentes, mais toujours dans l'optique d'améliorer les conditions de vie du plus grand nombre. Pour paraphraser la formule souvent attribuée au pape du cyberpunk William Gibson, le futur est déjà là, mais tout le monde n'y a pas accès. Reprendre la main sur les récits technologiques, c'est déjà le redistribuer un peu.

La Terre sainte du transhumanisme

P our comprendre le curieux mélange de prophéties qui parcourt l'industrie de l'intelligence artificielle, il faut remonter aux origines du grand mouvement technoutopique de la Silicon Valley: le transhumanisme.

Contrairement à ce que l'on pourrait croire, le mot n'est pas récent. La première occurrence date du début du XIV^e siècle. Dans *La divine comédie* de Dante, après avoir traversé l'Enfer et le Purgatoire, le narrateur de Dante arrive au Paradis. En contemplant, ébloui, le soleil et sa guide, Béatrice, il se sent «transhumaniser» (*trasumanar*), une expérience impossible à décrire en mots, mais dont la plupart des exégètes affirment qu'elle consiste à s'approcher du divin en dépassant sa forme humaine. Dans la vision néoplatonicienne du monde, selon laquelle il faut se défaire de son enveloppe corporelle pour atteindre l'immortalité, la chair, désespérément éphémère et pourrissante, devient un champ de bataille.

En 1957, le biologiste et philosophe britannique Julian Huxley (frère d'Aldous Huxley, auteur du *Meilleur des mondes*) reprend et sécularise le terme «transhumanisme» dans son recueil d'essais intitulé *New Bottles for New Wine*. Le mot lui permet de définir une «religion sans révélation^[1]», «nouvelle croyance» de «l'homme restant l'homme, mais se transcendant, en réalisant les possibilités offertes par et pour sa nature humaine^[2]», en particulier celles offertes par la science et la technique. Le concept religieux de Dante est ainsi réactivé dans le positivisme malthusien du XX^e siècle occidental, dont Julian Huxley est un des idéologues. Le biologiste prône ce qu'il présente comme un eugénisme de «gauche», soucieux du bien-être de l'humanité, contrairement à l'obsession pour la «race supérieure» de l'eugénisme nazi. Il n'en

demeure pas moins que l'humanisme d'Huxley s'arrête aux classes populaires, qu'il considère comme génétiquement inférieures et pour lesquelles il préconise le contrôle des naissances.

Après ces premières explorations technoscientifiques du transhumanisme, majoritairement britanniques, cette philosophie migrera vers la côte Ouest états-unienne, où elle s'épanouira et trouvera sa forme moderne à la fin des années 1980.

Ultralibéralisme et transhumanisme californien

En 1988, Max O'Connor, un jeune doctorant en philosophie britannique, quitte les flèches néogothiques d'Oxford pour les arcs néo-romans de l'université de Californie du Sud à Los Angeles. Saisissant rapidement l'esprit du temps et des lieux, il injecte une bonne dose de libertarianisme californien au transhumanisme britannique d'Huxley et lui donne le nom d'extropianisme – la quête de l'*extropie*, un néologisme inventé comme antonyme à l'*entropie*, le processus physique de dégradation. Être extropien, explique O'Connor en 1990, c'est adhérer aux «philosophies de vie qui cherchent à prolonger l'évolution de la vie intelligente au-delà de sa forme et de ses limites humaines actuelles^[3]». Cette technolâtrie sous stéroïdes, parfaitement compatible avec l'ultralibéralisme hyperindividualiste de l'époque, déterminera aussi le choix du nouveau nom d'O'Connor, qui se fera désormais appeler Max More. L'homme qui voulait devenir plus, toujours plus. Pour réaliser son fantasme, More compte sur une panoplie d'outils techniques, dont la majorité n'existent encore que dans son imagination: amplificateurs d'intelligence, intégration neuro-informatique, réseaux mondiaux de données, réalité virtuelle, intelligence artificielle, neuro-sciences, réseaux neuronaux, vie artificielle, colonisation interplanétaire, nanotechnologies moléculaires, etc. Et son délire s'avère contagieux. Avec sa compagne, l'artiste, philosophe, mannequin et culturiste Natasha Vita-More, Max More convertit une petite communauté d'intellectuels et d'ingénieurs technophiles. La *mailing list* extropienne, forte de quelques centaines de noms au passage de l'an 2000, est alors un véritable *who's who* de la cyberculture naissante. Et un peu plus loin: en 1997, une version revue et

augmentée du «Transhumanist Manifesto», rédigé en 1982^[4] par Natasha Vita-More, est gravée sur CD et embarquée dans la sonde NASA/ESA Cassini-Huygens en direction de Saturne. Leur activité est exubérante, presque potache. Les initiés échangent évidemment par e-mail, mais ils se réunissent aussi en personne, le temps de rassemblements baptisés «Extropaganza^[5]». Lorsqu'ils se rencontrent, ils se saluent avec une poignée de main spéciale. Comme la plupart des transhumanistes sont basés dans la Silicon Valley, entre Cupertino et Stanford, ils organisent des fêtes décadentes dans une maison de Cupertino, la «Nextropia». En 1994, le journaliste de *Wired* Ed Regis publie un reportage extatique sur les tenants de cette «philosophie de l'expansion illimitée [...], une doctrine de transformation de soi, de technologie extrêmement avancée et d'un optimisme inébranlable^[6]». L'année suivante, un journaliste britannique, moins impressionné, parle d'une «secte californienne^[7]» parmi d'autres.

Dans cet entre-soi, on cite indistinctement la romancière libertarienne Ayn Rand, Nietzsche, la série *Star Trek*, le penseur du néolibéralisme Friedrich Hayek et les superhéros de la série X-Men. On rêve de privatiser l'air et les océans pour lutter contre la pollution, car une fois que «les communs seront privatisés, leur usage sera celui de la rationalité économique^[8]». Petit à petit émerge le projet politique transhumaniste: une technocratie dérégulée où les labos de recherche de la Silicon Valley décident du futur de l'espèce. Comme la foi transhumaniste est entièrement tournée vers le futur, les croyants les plus radicaux (et les plus nantis) ont pris leurs dispositions pour s'assurer de vivre l'avenir radieux qu'ils espèrent, aussi lointain soit-il. Alcor, une entreprise sise en Arizona et dirigée par Max More entre 2010 et 2021, leur propose l'immortalité par la cryogénisation. Pour 200000 dollars, ils peuvent mourir tranquilles: leur corps sera plongé dans un bain d'azote liquide à -196 °C, juste après leur décès. Pour un petit supplément variant entre 30 000 et 130 000 dollars, on en fera autant avec leur animal de compagnie. Les cryogénisés seront conservés dans un cylindre en inox, en vue d'être ressuscités un jour par leurs descendants. Ceux qui n'ont que 80000 dollars à dépenser peuvent escompter que le reste de leur corps sera recréé artificiellement ou remplacé par un corps robotique, et

ne faire cryogéniser que leur tête, non sans avoir au préalable fait injecter de l'antigel dans leur cerveau.

Une nouvelle fois, le délire s'avère étonnamment contagieux. Alcor revendique 1400 candidats à la cryogénisation parmi lesquels More lui-même, sa femme, l'entrepreneur libertarien star de la Silicon Valley Peter Thiel^[9], le prophète critique de l'IA Eliezer Yudkowsky^[10] et deux papes du transhumanisme actuel, Ray Kurzweil et Nick Bostrom^[11]. Tous portent un bracelet, sur lequel sont inscrits le numéro d'Alcor et la marche à suivre pour la préservation de leur corps.

Nick Bostrom, le prophète déguisé en prof

Après le changement de millénaire, marqué par le pétard mouillé du bogue de l'an 2000 et l'éclatement de la bulle *dot-com*, la baudruche techno-optimiste se dégonfle. Incompatible avec l'angoisse belliqueuse post-11-Septembre, l'idéologie de l'emballage transhumaniste entre en stase. Elle sort de son hibernation dix ans plus tard. Sur les cendres de la crise financière de 2008 émerge alors un nouveau discours entrepreneurial aux accents messianiques – celui des Facebook, Uber, Twitter, WeWork et autres, qui rivalisent de superlatifs pour faire les poches des fonds d'investissement. Le futurisme redevient cool. La machine à fantômes transhumaniste redémarre.

À l'époque, la figure de proue du mouvement est un jeune philosophe suédois, Niklas «Nick» Bostrom, fondateur en 1998 de la World Transhumanist Association et professeur de philosophie à Oxford depuis 2008. Disciple de Max More, Bostrom définit son transhumanisme comme «un mouvement intellectuel et culturel qui affirme la possibilité d'augmenter de façon fondamentale la condition humaine à travers les nouvelles technologies^[12]». Le noyau idéologique et les grandes lignes restent donc les mêmes: accélération de l'innovation technique et scientifique, augmentation des capacités humaines et, *in fine*, immortalité et colonisation interplanétaire.

Mais la comparaison avec le tonitruant Max More s'arrête là. Sous l'impulsion de Bostrom, le mouvement transhumaniste troque les éruptions libertariennes contre un discours universitaire, en apparence

plus mesuré. L'objectif n'est plus de se proclamer cyborgs empereurs de la Voie lactée, mais de convaincre les décideurs politiques et économiques d'adopter des mesures favorables au transhumanisme. Pour y parvenir, il faut parler le langage technocratique. Adieu le Schwarzenegger à catogan de la Silicon Valley, place au professeur en veste de tweed, dont la silhouette évoque un antihéros de bande dessinée belge.

Grand érudit et vulgarisateur efficace, avec sa voix rassurante et son phrasé dynamique, Bostrom a toutes les caractéristiques de l'enseignant. Il sait aussi diffuser ses idées par tous les canaux possibles: livres, articles spécialisés, conférences TED, *posts* de forums, colloques institutionnels. Son transhumanisme *soft* (en apparence) séduit. Sa réputation se consolide très vite, entre autres grâce à un grand nombre d'articles publiés dans des revues qu'il a lui-même lancées. Influence majeure du discours prophétique de l'IA, il participe à la création de plusieurs champs d'études, comme «l'éthique du risque existentiel» et l'*AI safety* (sécurité de l'IA). Dans certains cercles universitaires, Bostrom est tout simplement incontournable – il est aujourd'hui cité près de 30000 fois sur Google Scholar.

Son approche pragmatique d'un sujet présenté comme hautement complexe fait décoller sa carrière médiatique: durant la décennie 2010, Bostrom fait son entrée dans le «grand public», et les obsessions transhumanistes avec lui. En 2009 et en 2015, il figure dans le classement des 100 intellectuels les plus influents selon la revue *Foreign Policy*. Son manifeste *Superintelligence*, publié en 2014, devient vite un best-seller et le propulse dans la presse internationale, qui voit en lui un «bon client» capable de vulgariser les enjeux de l'intelligence artificielle.

Le *Guardian*, le *New York Times* ou le *Washington Post* publient à tour de bras ses oracles transhumanistes. Toute publicité est bonne à prendre. Là, Bostrom annonce l'avènement de la «superintelligence»; ici, il avertit sur les «risques existentiels» qui menacent l'humanité. Parfois, il défend «l'argument de la simulation»: la possibilité, qu'il estime autour de 20 %, que notre réalité ne soit qu'un vaste programme informatique – la Matrice – hébergé sur un ordinateur grand comme une planète. Consécration ultime: en 2019, il évangélise le *podcast* de Joe

Rogan, ultrapopulaire aux États-Unis. Sachant parler aussi bien aux cadres de Google qu'au directoire de la Banque d'Angleterre, le philosophe-prophète impose peu à peu sa vision millénariste. La fin de l'espèce humaine est proche, promet-il, pour le meilleur ou pour le pire. Son installation dans le champ des idées *mainstream* est si profonde qu'il rejette désormais l'étiquette de transhumaniste, sous prétexte qu'elle «crée de la confusion^[13]» – comprendre: elle est moins vendeuse que l'étiquette «philosophe à Oxford». C'est vite oublier que lorsqu'il n'est pas en train d'écrire des fables où l'humanité vainc le «dragon-tyran» qu'est la mort, Bostrom refuse les poignées de mains par peur des germes, consomme des livres audio en accéléré et optimise son déjeuner en buvant son «élixir» – une concoction de fruits, légumes, protéines et gras préparée au *blender* industriel.

Le philosophe porte aussi bien la casquette de lobbyiste. Auprès des élus et des décideurs, il insiste pour que la réalisation du futur transhumaniste devienne la «priorité globale^[14]» des nations. Autrement dit, les politiques publiques ne devraient plus remédier aux inégalités sociales et économiques actuelles, mais se concentrer sur nos descendants «posthumains», ces êtres «dont les capacités de base excéderont si radicalement celles des humains contemporains qu'elles ne peuvent être considérées humaines selon nos standards actuels^[15]». Cette pure mythologie, dépourvue de toute base scientifique, Bostrom la sert tranquillement à l'Assemblée générale des Nations Unies, le 7 octobre 2015 à New York, où il intervient en qualité d'expert en gouvernance de l'intelligence artificielle. Même type de discours devant la CIA, la Commission européenne ou le Conseil du président des États-Unis pour la bioéthique, fondé par George W. Bush. Chaque fois, il prône la même approche: laissons tomber le présent.

Le discours, pour contre-intuitif qu'il soit, fait mouche auprès des ploutocrates. En 2018, l'historien superstar Yuval Noah Harari, nouveau conférencier du transhumanisme, affirme sans trembler que «nous sommes probablement la dernière génération d'Homo sapiens. D'ici un siècle ou deux, la Terre sera dominée par des entités qui seront plus différentes de nous que nous le sommes de Neandertal ou d'un chimpanzé. Les prochaines générations apprendront à modeler les corps

et les cerveaux». Son public subjugué: l'élite politique et financière réunie au Forum économique mondial de Davos^[16]. Rien de tel que le transhumanisme, philosophie du dépassement éternel, pour parler aux puissants.

Dépasser les frontières, l'obsession de la Silicon Valley

C'est à l'Esalen Institute, une retraite hippie fondée en 1962 dans la région de Big Sur, épice centre du «mouvement du potentiel humain», qu'a pris forme le syncrétisme technospirituel à l'origine de la cyberculture. Sur la côte Pacifique, à mi-chemin entre Los Angeles et San Francisco, de jeunes informaticiens californiens en quête de sens y ont d'abord exploré différentes formes de dépassement: l'expansion de la conscience grâce à la méditation et aux psychotropes, l'augmentation de ses capacités cognitives grâce à l'ordinateur personnel, et l'ouverture à d'autres modes de coexistence et de gouvernement grâce à la contre-culture politique états-unienne. L'époque est à l'exploration.

Dans les années 1980, l'institut a joué un rôle déterminant dans le réchauffement des relations entre les États-Unis et l'URSS grâce à la «diplomatie du *jacuzzi*^[17]». Cette stratégie étonnante consistait à organiser, en pleine guerre froide, des séminaires improbables dans un lieu tranquille et reclus, et à y faire se rencontrer des acteurs de la techno-culture californienne naissante et des intellectuels, chercheurs et mystiques soviétiques. Parmi ces invités, plusieurs se réclamaient du cosmisme, un mouvement philosophique et religieux impulsé au XIX^e siècle par le philosophe et ascète orthodoxe Nikolaï Fiodorov selon lequel l'humanité dépassera sa condition mortelle en apprenant à maîtriser le cosmos. Pour les cosmistes du XX^e siècle, l'humain parviendra à ressusciter les morts en réorganisant leurs atomes, à transposer un cerveau dans un robot et à coloniser l'espace. Ça vous rappelle quelque chose? Messianisme, augmentation du corps, espace: les éléments principaux du transhumanisme sont là. Le disciple de Fiodorov, Constantin Tsiolkovski, posera au passage les bases philosophiques de la conquête spatiale de l'URSS.

À l'Esalen Institute, les séances de massages, les bains thermiques et les longues soirées au coin du feu sont propices aux échanges d'idées, et le cosmisme russe obtient un succès immédiat chez les jeunes Californiens – qui rêvent eux aussi de franchir ces «*final frontiers*», l'espace et l'immortalité. Quelques années à peine avant Max More, le «potentiel humain» individualiste de la côte Ouest s'entiche du prolétaire immortel rêvé de l'autre côté du rideau de fer. Ce pot-pourri de spiritualités orientales, de libéralisme, de transhumanismes occidental et soviétique et de rationalité scientifique forme la base de la «religion du travail» unique, analysée par Carolyn Chen dans *Work Pray Code* (Travaille prie code), dans cette jeune théocratie qu'est alors la Silicon Valley^[18]. Un sentiment de mission sacrée qui investit tous les employés, depuis les codeurs anonymes jusqu'au *top management* et aux grands patrons. Le vocabulaire a changé, mais les ésotérismes des années 1960 à 1980 sont toujours aussi influents.

À la faveur du New Space, le nom donné à la privatisation galopante de l'exploration spatiale entamée au début des années 2000, les nouveaux maîtres de l'astrocapitalisme laissent libre cours à leurs visions cosmistes. Entre deux présentations d'hypothétiques colonies sur Mars, Elon Musk cite nul autre que Tsiolkovski («La Terre est le berceau de l'humanité, mais personne ne peut vivre éternellement dans son berceau^[19]»). Jeff Bezos, fondateur d'Amazon et patron de l'entreprise de transport spatial Blue Origin, déclare que «la surface planétaire est beaucoup trop petite^[20]», et que d'ici peu, un «trillion» d'êtres humains vivront dans des stations orbitales cylindriques disséminées dans le système solaire. Des stations orbitales imaginées par Tsiolkovski.

Après l'échec des techniques de développement spirituel explorées par les hippies originels, l'ordinateur personnel, Google et désormais l'IA sont à leur tour censés «révéler» les «potentialités» humaines. Pour les démiurges de la Silicon Valley, il n'y a aujourd'hui aucune question métaphysique que l'argent, la science et la technique ne puissent résoudre. Pas même l'encombrant souci de la mortalité. Après tout, à quoi bon accumuler une telle puissance technique, de telles fortunes, si c'est pour finir rongés par des vers comme les plus misérables d'entre nous? Cette égalité qu'impose la mort n'est-elle pas un affront aux

mérites des meilleurs? Cette classe d'oligarques, à la richesse obscène et aux ego surdimensionnés, est désormais persuadée que la mort n'est qu'un problème technique – redoutable, certes, mais pas insoluble. Ces dernières années, la quasi-totalité des grands patrons et investisseurs de la tech ont donc lancé ou financé des *start-up* qui promettent d'allonger la vie humaine par la conservation cryogénique, les thérapies génétiques ou la numérisation du cerveau. En effet, vivre éternellement lorsque l'on est déjà milliardaire, c'est s'assurer de capter une part toujours plus importante des richesses produites, et accroître son influence *ad infinitum*. Si ces patrons lisaient le roman *Tous les hommes sont mortels*, de Simone de Beauvoir, ils comprendraient peut-être que l'immortalité serait une malédiction et que la finitude existentielle est en fait une bénédiction. Mais on n'a pas affaire à des existentialistes.

En 2022, Jeff Bezos a investi dans Altos Labs et Unity, deux *start-up* de biotech qui prétendent parvenir un jour à inverser le vieillissement. Sous l'impulsion de ses fondateurs transhumanistes Larry Page et Sergueï Brin, qui a publiquement déclaré vouloir nous «guérir de la mort^[21]», Alphabet (anciennement Google) a fondé deux laboratoires de recherche sur le vieillissement: Calico, créé en 2013, et Verily, créé en 2015. En 2016, Mark Zuckerberg et sa compagne Priscilla Chan ont promis d'investir trois milliards de dollars sur dix ans dans le but de guérir toutes les maladies, afin que «d'ici la fin du siècle, vivre au-delà de 100 ans devienne tout à fait normal^[22]».

Peter Thiel, cofondateur de PayPal et investisseur incontournable, est obsédé depuis son adolescence par l'immortalité des elfes du *Seigneur des anneaux*. Figure inquiétante de la Silicon Valley, notamment soupçonné de s'injecter du sang de jeunes donneurs humains, Thiel a été obligé de préciser publiquement qu'il n'est «pas un vampire^[23]». Devenu milliardaire, il se considère désormais «opposé à la mortalité^[24]», compte vivre au moins jusqu'à 120 ans et affirme que «la grande tâche inachevée du monde moderne consistera à faire en sorte que la mort ne soit plus un fait de la vie, mais un problème à résoudre^[25]». Thiel a personnellement investi dans plus de 25 *start-up* de biotech^[26] dédiées à l'extension de la vie et fait également partie des candidats à la cryogénisation d'Alcor, en dernier recours. Terrifié par la

mort, il finance également les recherches du gérontologue britannique Aubrey de Grey. Cet ancien chercheur en intelligence artificielle, dont la longue barbe ressemble à celle de Merlin l'Enchanteur, prophétise que le premier être humain qui vivra 1000 ans est déjà né.

Dans la Silicon Valley, lutter contre la mortalité n'est plus seulement une lubie d'universitaires marginaux et de milliardaires en plein délire mégalomane. L'immortalité est devenue un véritable pari industriel, qui a attiré près de cinq milliards de dollars d'investissements en 2022^[27] sans aucune perspective de profits à court terme. C'est également un véritable mouvement (contre-)culturel, un transhumanisme réinventé qui ne dit pas son nom, mais réactive de vieilles lubies des années 1980. Les *techies* californiens se gavent de metformine, un médicament contre le diabète, dans l'espoir d'accroître leur espérance de vie^[28]. Dans le centre-ville de San Francisco, on voit des manifestations d'«immortalistes», brandissant des panneaux «*Say Forever!*»^[29].

Au printemps 2023, sur la côte du Monténégro, 200 «longévistes», financés par le cryptomilliardaire Vitalik Buterin, ont investi une ville éphémère baptisée Zuzalu (le nom provient de ChatGPT), pour un festival contre le vieillissement. Au programme: conférences sur la longévité humaine, beaucoup de débats sur l'IA, ateliers de *biohacking* et rêves de micronation où personne, et surtout pas les régulateurs, ne pourrait ralentir la recherche sur le vieillissement. Comme le résume la bio Twitter/X de l'investisseur Balaji Srinivasan, présent sur les lieux: «argent immuable, frontières infinies, vie éternelle^[30]». Rien de nouveau depuis les soirées décadentes de la clique californienne de Max More, en somme. Le même délire néolibéral d'extension du domaine du mercantile, d'optimisation des performances humaines et de maximisation de la croissance économique. Et surtout, le même mépris pour le commun des mortels, bloqués dans l'ici et le maintenant.

Si la fontaine de Jouvence à base de compléments alimentaires et de transfusions sanguines échoue, il restera toujours la possibilité du devenir-machine. C'est le pari qu'a notamment fait Sam Altman. Comme les autres techno-oligarques, le patron d'OpenAI a d'abord investi dans une *start-up* de la biotech: 180 millions de dollars pariés sur Retro Biosciences, dont l'ambition est «d'augmenter de dix ans

l'espérance de vie humaine» – un objectif atteint plusieurs fois par les vieilles technologies sociales que sont l'hygiène publique et la justice économique, ce que le jeune PDG semble royalement ignorer.

Non content d'avoir investi près de 375 millions de dollars dans la *start-up* Helion en 2021^[31], qui promet de résoudre la crise des énergies fossiles grâce à la fusion nucléaire (un autre horizon technique actuellement utopique), Altman compte également faire conserver son cerveau par l'entreprise Nectome, avec l'espoir d'être un jour «téléchargé dans le *cloud*». Coût de l'opération: 10 000 dollars, une bagatelle. Et il anticipe déjà sa propre augmentation, dans le plus pur style transhumaniste, tel qu'il le révélait en 2016: «Il n'y a aucune raison de penser que dans environ treize ans, nous ne posséderons pas le matériel nécessaire pour dupliquer un cerveau. [...] Les ordinateurs auront leurs propres désirs, leurs propres ambitions. [...] Être une machine présente certains avantages. Nous, les humains, sommes limités par notre vitesse de traitement de données – nous apprenons au rythme de deux bits par seconde, et le reste est perdu. Du point de vue de la machine, nous devons ressembler à des chants de baleines au ralenti^[32].»

Altman est un authentique converti transhumaniste. Dans le même portrait, il déclare également que «la fusion [entre les machines et nous] a commencé – et cette fusion, c'est le meilleur scénario. Toute version du futur sans fusion impliquera un conflit: ou nous asservissons l'IA, ou c'est elle qui nous asservit. La version la plus dingue de la fusion, c'est celle où nos cerveaux sont téléchargés dans le *cloud*. J'adorerais ça. Il faut qu'on augmente les humains, parce que nos descendants vont soit conquérir la galaxie soit éteindre la conscience de l'univers pour toujours. Quelle formidable époque nous vivons!» Des mots qui font écho à un court manifeste, rédigé en 2017, dans lequel il affirmait que «l'IA surhumaine aura lieu, l'augmentation génétique aura lieu, les interfaces humain-machine auront lieu», et que nous serions bientôt «la première espèce à concevoir ses propres descendants^[33]». Bostrom et Max More peuvent être fiers: leurs disciples industriels ont désormais l'avenir entre leurs mains.

C'est pour penser au futur, humain ou posthumain, que le jeune milliardaire a réuni chez lui, à San Francisco, une assemblée hétéroclite

de chercheurs en philosophie, robotique, informatique quantique, IA, biologie synthétique et voyage spatial. Son surnom: l'Alliance, avec une majuscule initiale, comme pour celle de la Bible. En 2022, encore inconnu du grand public, Altman listait déjà ses projets transhumanistes sur Twitter: «Nous pouvons construire l'intelligence artificielle. Nous pouvons coloniser l'espace. Nous pouvons atteindre la fusion nucléaire et déployer le solaire à grande échelle. Nous pouvons guérir toutes les maladies. Nous pouvons construire de nouvelles réalités^[34].» Il dispose aujourd'hui de 13 milliards de dollars, via OpenAI, pour bâtir son technoparadis.

Si l'association entre les universitaires transhumanistes et les capitaines d'industrie de la Silicon Valley avait un visage, ce serait néanmoins celui d'Elon Musk. Celui qui fut brièvement l'homme le plus riche du monde concrétise le projet technoreligieux comme nul autre, mais sans jamais s'en réclamer explicitement. Depuis quinze ans, Musk agit en véritable évangéliste, mettant son immense talent de bonimenteur publicitaire au service des prophéties technoscientifiques.

Pronataliste et père de 11 enfants, il considère que «faire des enfants, c'est sauver le monde» et rêve d'une société où seuls les parents auraient le droit de vote. Mais derrière un discours héroïque sur les «risques d'effondrement de la civilisation», sa préoccupation est avant tout celle d'un grand patron, pour qui la diminution de la main-d'œuvre aurait «des conséquences économiques et sociales très défavorables» – moins de monde à exploiter, c'est mauvais pour les affaires. Elle est ensuite l'inquiétude d'un eugéniste: dans sa première biographie, publiée en 2015, Musk s'inquiète du fait que «chaque génération successive de gens intelligents fait de moins en moins d'enfants^[35]». Lorsqu'on sait que le milliardaire soutient le délire raciste du Grand Remplacement^[36], on devine aisément qui sont ceux qu'il considère comme les «gens intelligents» et qui sont les autres. Cette volonté obsessionnelle de se répliquer est enfin tout simplement la preuve d'un narcissisme extraordinaire. Dans le futur proposé par Musk, dont l'imminence est sans cesse repoussée, l'être humain voyagera en fusées de SpaceX, et l'humanité deviendra «une espèce multiplanétaire»; nos descendants, surconnectés via le réseau Starlink, deviendront des «superhumains», en

«symbiose» avec les machines^[37] grâce aux implants cérébraux de Neuralink et aux logiciels de xAI. En 2021, Musk a justement reçu 200 millions de dollars supplémentaires pour poursuivre le développement de ses interfaces cerveau-machine, qui permettent pour le moment à des personnes paralysées de contrôler un ordinateur par la pensée (ce que la recherche publique sait déjà faire depuis quinze ans). Parmi ses investisseurs: Google, Peter Thiel ou encore Sam Altman. L'oligarchie de la Silicon Valley sait se serrer les coudes.

Depuis une vingtaine d'années, le milliardaire sudafricain tente de faire basculer l'espèce tout entière dans la posthumanité, et concrétiser ainsi les utopies concoctées par les théoriciens transhumanistes depuis la fin des années 1980. Il faut dire que l'entrepreneur ne vient pas de nulle part. Son grand-père, Joshua Haldeman – un chiropraticien canado-états-unien antisémite et suprémaciste blanc^[38] – a fondé, dans les années 1930, Technocracy Incorporated^[39], un mouvement dont l'objectif était de remplacer la classe politique au pouvoir par des scientifiques et des ingénieurs, le travail par l'automatisation, l'argent par des «certificats d'énergie» individuels, les pays par un «Technat» planétaire unique et, *last but not least*, les prénoms individuels par des séries de chiffres – un rêve exaucé par son petit-fils Elon, père depuis 2020 du petit X Æ A-XII. L'expérience a échoué, et Haldeman est parti refaire sa vie en Afrique du Sud avec toute sa famille, mais ses idées sont bien vivantes dans la Silicon Valley.

L'altruisme efficace: une machine à justifier le (techno)capitalisme

Non contente de ressusciter l'utopie transhumaniste, l'industrie californienne de la tech s'est récemment entichée d'un autre mouvement philosophique futuriste, né entre les États-Unis et l'Angleterre au début des années 2010: l'altruisme efficace (*effective altruism*), dont les adeptes se définissent comme des «EA».

Les origines de cette pensée remontent aux débuts de l'utilitarisme, doctrine fondée dans l'Angleterre du XVIII^e siècle par le juriste Jeremy Bentham – lui aussi partisan de l'eugénisme et célèbre inventeur du

panoptique, un modèle de prison devenu allégorie de la société de surveillance. Reflet du contexte positiviste de la révolution industrielle qui l'a vu naître, l'utilitarisme propose de prendre toutes ses décisions de manière calculée, en évaluant les conséquences de toutes les solutions possibles grâce à une méthode supposément neutre et objective. L'objectif est de toujours viser «le plus grand bonheur du plus grand nombre», dicit Bentham dans son *Fragment sur le gouvernement*, le bonheur consistant en gros à limiter les souffrances individuelles, sans trop se soucier de l'état des sociétés.

Au xx^e siècle, l'idée est reprise par le philosophe et bioéthicien australien Peter Singer, célèbre activiste de la cause animale, pionnier de l'antispécisme et adhérent au récit transhumaniste^[40]. Dans un texte intitulé *Famine, Affluence, and Morality* (Famine, richesse et moralité), Singer analyse, en utilisant une grille utilitariste, les stratégies d'aide humanitaire internationale dans le cas d'une famine au Bengale. Certes, il affirme que les Occidentaux ont le devoir moral d'aider les populations en difficulté, mais il ajoute que «le meilleur moyen de prévenir la famine, sur le long terme, est le contrôle de la population^[41]». Une logique glaçante qu'on retrouvera dans le livre *Should the Baby Live?* (Le bébé doit-il vivre?), sous-titré «le problème des enfants handicapés», où il écrit sans le moindre état d'âme qu'il «semble peu avisé d'exacerber le problème des ressources limitées en augmentant le nombre d'enfants sévèrement handicapés^[42]». On retrouve ici l'obsession eugéniste des transhumanistes, et la notion suprémaciste selon laquelle certaines vies valent objectivement moins que d'autres.

Après Singer, l'idée d'une action humanitaire calculée fait son chemin. À la fin des années 2000, deux étudiants de philosophie d'Oxford, William MacAskill et Toby Ord, se lancent dans une mise à jour éthique et technique de la philanthropie. MacAskill, transformé par la lecture du manifeste de Peter Singer, fait vœu de frugalité et fonde l'association 80,000 Hours autour d'un étrange principe utilitariste: «*Earn to give*», soit gagner de l'argent pour mieux le donner. L'association encourage à mettre de côté son éthique personnelle et à choisir la carrière la plus lucrative possible, pour ensuite redistribuer au moins 10 % de son salaire aux ONG identifiées par l'organisation

comme les plus «efficaces». Le pourcentage n'est pas choisi au hasard: il reprend le vieux concept de la dîme, une contribution venue du judaïsme et du christianisme. Agrémentée d'une pincée de *big data*, relevée de quelques algorithmes d'évaluation et d'indicateurs économiques abscons, et voilà que la charité rejoint la sphère managériale.

Pour MacAskill, il n'y a aucun paradoxe à aller gagner de l'argent dans une multinationale minière ou financière qui exploite le travail des enfants, pourvu que cet argent serve ensuite à lutter contre les inégalités ou les bouleversements climatiques. En 2011, année du mouvement Occupy, MacAskill donne une conférence à Oxford intitulée «Doctor, NGO Worker, or Something Else Entirely? Which Careers Do the Most Good?» (Docteur, ONG, travailleur ou complètement autre chose? Quelles carrières font le plus grand bien?), où il exhorte les étudiants qui l'écoutent à devenir des «banquiers altruistes» – une figure paradoxale de saint cupide qui, après avoir succombé à toutes les sirènes matérielles du Capital, ferait acte de pénitence et sauverait dix fois plus de vies qu'un docteur en mission humanitaire en distribuant 10 % de sa fortune^[43].

Grâce au tour de passe-passe de l'altruisme efficace, la finance spéculative et les énergies fossiles, qui comptent parmi les premières responsables de la catastrophe écologique, sont travesties en organismes de bienfaisance. Les milliardaires prédateurs Warren Buffet et Bill Gates, tous deux à la tête de gigantesques «fondations» philanthro-capitalistes, sont montrés en exemple par les EA (sans adhérer eux-mêmes au mouvement). William MacAskill dédie également un chapitre de son manifeste *What We Owe to the Future* (Ce que nous devons au futur) à la «défense morale des marchandises de *sweatshops*», ces ateliers clandestins d'Asie où des quasi-esclaves fabriquent à longueur de journée les vêtements des grandes marques occidentales, qui sont acceptés au motif que «leur bénéfice pour les économies asiatiques est indiscutable^[44]».

Forcément, une philosophie morale aussi indulgente attire rapidement l'attention (et les capitaux) des principaux candidats à l'absolution, comme le milliardaire et cofondateur de Facebook Dustin Moskovitz, mais aussi l'inévitable Peter Thiel. Et petit à petit, le discours de

l'altruisme efficace change. Après avoir encouragé leurs adeptes à donner de l'argent pour lutter contre le paludisme ou la souffrance animale, les hérauts de l'altruisme efficace veulent désormais accélérer la croissance économique, protéger la démocratie libérale et, priorité numéro un... «renforcer l'altruisme efficace». Il fallait y penser. Par pure coïncidence, assurément, 80,000 Hours et Giving What We Can, deux organisations qui prétendent évaluer objectivement l'efficacité des causes philanthropiques, parviennent à la même conclusion: l'urgence, c'est de les financer, elles. En novlangue d'EA, on appelle ça la «métacharité», ou tout simplement «virer méta». Autrefois, on aurait simplement dit que «Charité bien ordonnée commence par soi-même».

En une petite dizaine d'années, une école de pensée marginale réunissant des doctorants en philosophie d'Oxford et de jeunes codeurs californiens désenchantés est devenue un véritable mastodonte. En 2022, écrivait *The New Yorker*, le réseau des EA est «une faction à l'influence considérable, particulièrement dans la Silicon Valley, qui contrôle un fonds caritatif de l'ordre de 30 milliards de dollars^[45]». L'année précédente, le mouvement a distribué près de 600 millions de dollars aux causes jugées prioritaires^[46], au premier plan desquelles on trouvait... son propre développement.

Des chiffres absolument ahurissants, publiés fièrement par William MacAskill sur l'un des forums de la communauté^[47].

À 35 ans, le plus jeune professeur de philosophie d'Oxford, qui se décrit comme un «prophète récalcitrant^[48]», est accueilli comme une rock star sur les campus des universités anglo-saxonnes. Sous son impulsion, le mouvement se sectarise et devient «une forme d'Église – de plus en plus centralisée et contrôlée^[49]». Certains adhérents, déçus, commencent à parler d'endoctrinement.

Parallèlement, les liens avec la Silicon Valley se renforcent.

En 2015, la conférence EA Global, hébergée sur le campus californien de Google, voit Nick Bostrom et Elon Musk discuter des «risques existentiels» posés par l'intelligence artificielle devant un parterre d'altruistes efficaces subjugués. La même année, 80,000 Hours est intégrée à la promotion annuelle de Y Combinator, la plus célèbre pouponnière à *start-up* de la Silicon Valley. Son PDG d'alors s'appelle...

Sam Altman. Parmi les offres d'emploi «efficaces» listées par 80,000 Hours, on trouve désormais une majorité de postes liés à la recherche en IA, aussi bien chez les poids lourds du secteur (OpenAI, Anthropic, Microsoft, Amazon) que dans des *think tanks* libéraux états-unis comme l'Atlantic Council. Ce ne sont pas les premiers noms auxquels on pense lorsqu'on cherche à travailler dans l'humanitaire, mais il n'y a rien d'étonnant lorsqu'on comprend à quel point l'altruisme efficace s'est infiltré dans le secteur de l'IA. Ses adeptes se sont invités jusqu'au conseil d'administration d'OpenAI, dont trois ex-membres avaient des liens avérés avec la communauté des EA. *Idem* chez la *start-up* concurrente Anthropic, où le PDG Dario Amodei est le beau-frère de Holden Karnofsky, l'un des pionniers du mouvement. Chez Google DeepMind, le PDG Demis Hassabis se dit également adhérent aux idées des EA, et recrute régulièrement des membres de la communauté.

Poursuivant dans la logique utilitariste, une partie du mouvement se persuade que l'argent serait mieux investi s'il était utilisé pour financer le progrès technologique occidental, incarné par les magnats de la Silicon Valley, plutôt que pour financer des œuvres caritatives. Privilégier demain aux dépens d'aujourd'hui, l'Occident aux dépens du reste du monde: l'idée flotte dans l'air depuis les débuts de l'altruisme efficace, sous des formes parfois contradictoires. Grâce aux travaux d'Émile P. Torres, qui observe sans relâche l'évolution du mouvement, on sait que William MacAskill a toujours considéré la lutte contre la pauvreté dans les pays du Sud comme une manière d'accélérer l'avènement du futur posthumain^[50] (car plus d'humains sur Terre et en bonne santé, c'est plus de force de production à fournir au technocapital). Dans sa thèse parue en 2013, le philosophe Nick Beckstead, cofondateur du mouvement, considérait à l'inverse que «sauver des vies dans des pays pauvres pourrait avoir des conséquences significativement moindres (sur le lointain futur) que sauver et améliorer des vies dans des pays riches. Les pays riches fournissent substantiellement plus d'innovation, et leurs travailleurs sont beaucoup plus productifs économiquement^[51]». Dans un cas, les habitants du Sud sont perçus comme de la force de travail, dans l'autre, comme du gâchis de ressources.

À l'image de l'altruisme efficace, jamais Beckstead ou MacAskill ne justifient leur choix du système capitaliste comme moteur du progrès social *par défaut*. Le philosophe de l'élite occidentale semble incapable d'imaginer que pour améliorer le sort du plus grand nombre, le mouvement pourrait par exemple soutenir des révolutions anticapitalistes qui défendent la taxation du capital, la redistribution des richesses ou la nationalisation des laboratoires de recherche en IA. Dix ans après la thèse de Beckstead, cet abandon des sentiments humanistes les plus élémentaires, pour répugnant qu'il soit, est toujours discuté sur les forums du mouvement comme une idée «inconfortable» parmi d'autres^[52]. Pire, ce postulat va donner naissance à une autre école de pensée, encore plus radicale et toujours plus déconnectée du monde réel: le long-termisme.

Vers le long-termisme, où l'humain n'a plus d'importance

À partir de 2019, influencé par le lobbying de Nick Bostrom, William MacAskill (qui a longtemps été son voisin de bureau dans la Trajan House d'Oxford) change son fusil d'épaule. Plutôt que de résoudre les soucis du monde grâce aux fondations des milliardaires, le jeune professeur d'Oxford défend désormais «l'opinion selon laquelle agir positivement sur ce que sera l'avenir lointain est la priorité morale stratégique de notre époque^[53]». Autrement dit, il faut devenir les «bons ancêtres» que Bostrom appelait de ses vœux en 2008, ceux qui auront fait preuve d'«altruisme envers leurs descendants». Pour baptiser sa pensée, il pique un concept au philosophe britannique Derek Parfit (également d'Oxford), le «long-termisme».

Alors que la décennie s'ouvre sur la pandémie de COVID-19, les crises du climat, le réarmement des nations, l'accroissement des inégalités et le retour en force des démagogues, les milliardaires de la tech semblent plus que jamais déconnectés du sort de l'humanité réelle. Avec le «longtermisme», MacAskill bricole un cadre moral à la mesure de cette glaciale indifférence. Sa philosophie, focalisée sur un avenir lointain plein de science-fiction, valide leurs projets ultralibéraux de fuite

dans l'espace et d'immortalité, tout en repoussant hors du monde contemporain la question du salut de l'humanité.

Dans son ouvrage *What We Owe the Future*, le philosophe prophétise que nous vivons un moment charnière aux enjeux colossaux. Dans des termes d'une banalité affligeante, il clame que «le futur pourrait être très grand. Il pourrait aussi être très bon – ou très mauvais»; que nos décisions actuelles, même les plus insignifiantes, auraient le pouvoir de déterminer le chemin que prendront les générations futures; que si nous nous trompons, l'humanité stagnera ou s'éteindra; mais que si nous prenons les bonnes décisions – lire: si nous nous en remettons à l'élite technologique –, elle continuera à progresser. Alors seulement les posthumains du prochain million d'années, infiniment plus nombreux et plus productifs que tout ce que nous avons connu jusqu'ici, coloniseront et exploiteront les ressources de toute la Voie lactée jusqu'au dernier astéroïde. Les vieux États-nations seront quant à eux remplacés par des entités semi-privées comme la Zone économique spéciale de Shenzhen en Chine, que MacAskill considère comme un exemple à suivre. (Encore une fois, la foi transhumaniste épouse parfaitement la stratégie néolibérale de privatisation généralisée, avec une touche d'eugénisme et de technocratie.)

La trouvaille de MacAskill, c'est d'affirmer que notre impératif moral est d'œuvrer à bâtir le paradis. Comment? En prenant des décisions «à impact élevé». En «diffusant de bonnes idées» – comprendre: celles de son mouvement. En donnant de l'argent à des ONG labellisées «efficaces», celles qui développent de l'IA. En «faisant des enfants» – on retrouve ici l'obsession nataliste des transhumanistes. Et si tout cela ne suffit pas, le philosophe propose de «cloner et d'entraîner dès leur plus jeune âge» des scientifiques «du niveau intellectuel d'Einstein», seuls à même de nous guider vers la lumière du progrès à travers la nuit des siècles.

Autre avantage de cette mise à jour conceptuelle du transhumanisme: lorsqu'on pense comme un long-termiste, les problèmes structurels des sociétés contemporaines – pauvreté généralisée, déplacements de population, effondrement des écosystèmes, etc. – deviennent totalement insignifiants. Comme le résume Nick Bostrom, du point de vue de

l'éternité, «même les pires catastrophes ne sont que de simples vaguelettes à la surface de l'océan de la vie^[54]». Tout ce qui importe, c'est l'expansion humaine et la création de valeur – ou pour résumer une maxime de la communauté: «*Shut up and multiply*» (Tais-toi, et prolifère). *This is fine*. Dans ce paradigme, l'humain devient plus que jamais une matière première, dont il faut tirer un maximum de profit. La célébration fanatique de l'individu augmenté pour l'élite, la déshumanisation générale pour les autres: ainsi peut se résumer l'organisation sociale transhumaniste. Dans le futur lointain, tout change radicalement, mais pas la hiérarchie des privilèges.

L'idée de prendre des décisions en pensant aux générations futures n'a pas attendu les philosophes d'Oxford et les milliardaires de la tech. Les réflexions sur la crise écologique ou sur la gestion des déchets nucléaires sont elles aussi marquées par le souci de l'avenir lointain. Mais le futurisme de Bostrom, Ord et MacAskill, célébré par l'élite patronale de la Silicon Valley, retourne ce principe sur lui-même. Leur transhumanisme propose au contraire d'en finir avec l'humanité au sens culturel et politique, d'en faire une matière première au service de l'inhumanité de l'entreprise. Leur prétendu «altruisme» dissimule une négation du bien commun et une instrumentalisation de la multitude au profit de quelques-uns.

Pour faire passer la pilule conceptuelle, Bostrom, MacAskill, Musk, Bezos, Page et les autres nous exhortent à penser à un futur si lointain et si différent qu'il dépasse nos possibilités d'imagination, et par conséquent, nos capacités à le fixer dans un cadre de critique politique. Mais lorsque l'on s'extrait de cette sidération, on comprend rapidement que cette science-fiction mystique dissimule un processus à la fois plus prosaïque et plus ancien: la captation des richesses par une poignée d'entités privées en situation d'impunité, qui déguisent leur voracité derrière une théologie d'entreprise sur mesure et tentent de nous convaincre qu'elles sont les seules dépositaires du mystère technologique.

L'altruisme efficace n'est ni altruiste ni efficace: c'est «un cocktail d'arrogance élitiste et de naïveté^[55]» qui défend féroce­ment la légitimité morale des inégalités de classe. Une philosophie de dominants, fabriquée

dans des pépinières à dominants: le long-termisme est «taillé sur mesure pour permettre aux élites technos, financières et philosophiques de s'abandonner à leurs tendances antihumanistes tout en se félicitant mutuellement de leur supériorité intellectuelle», résume le *Washington Post*^[56].

Est-il alors surprenant de voir que Bill Gates a fait l'éloge du livre de MacAskill, tandis qu'Elon Musk a soutenu sur Twitter qu'il est «à lire absolument, et très proche de [sa] philosophie^[57]»? Est-il surprenant que l'autre grand manifeste long-termiste, *The Precipice*, écrit par le philosophe Toby Ord, soit cité par Boris Johnson lors d'un discours de 2021 à l'ONU sur le changement climatique^[58]? Ce même changement climatique qui, selon les long-termistes, n'est pas un problème prioritaire? Absolument pas. Du transhumanisme des années 1980 aux délires les plus récents de la frange long-termiste, le projet reste le même: fuir.

Ces maîtres de la finance et de la technologie ont décidé d'abandonner le présent à son sort.

Futuritarismes

Deux ans après l'introduction des premiers logiciels d'IA «générative», la Silicon Valley a indéniablement changé d'ère. À l'heure d'écrire ces lignes, il y règne la ferme conviction que l'avenir appartient à ceux qui produisent l'intelligence artificielle, et que cet avenir sera radieux. Les techno-optimistes y vivent leur quart d'heure de gloire. Tels des dieux maîtres du temps et créateurs de mondes, ils captent les investissements, l'attention médiatique et le privilège d'écrire le futur à leur convenance.

Les héros aussi ont changé. On méprise désormais les codeurs de la bulle *dot-com* des années 2000 et les *tech bros* des *start-up* des années 2010. Leur catéchisme néolibéral et leurs innovations économiques dites «disruptives» paraissent dépassés. Un changement de garde s'opère: place aux communautés technoreligieuses, aux chapelles de l'IA et aux grands prêtres du futurisme. Pour certains patrons et investisseurs influents de l'IA, les promesses de la puissance technologique sont totales. Il ne leur manque plus qu'à mettre la main sur le destin de l'humanité. Quitte à se débarrasser, chemin faisant, des encombrants garde-fous démocratiques (qui n'ont jamais aussi bien porté leur nom) et de leurs exigences de respect de l'égalité et du droit. Ils posent en hommes providentiels, guidant le troupeau vers le paradis transhumaniste. Paré de telles nécessités transcendantes, le culte de l'IA est inattaquable: on ne critique pas les plans d'un messie, même autoproclamé.

N'allez pas penser, cependant, que le culte de ces prophètes de l'IA promeut réellement le salut de l'ensemble des êtres humains. Le messie technologique n'a pas le sens du sacrifice du Christ ou l'amour de Dieu le Père. Lorsque le capital-risqueur Marc Andreessen, fondateur du fonds d'investissement Andreessen Horowitz et 144^e fortune mondiale, joue à l'oracle de la tech et proclame que «l'IA va sauver le monde», il ne

pense certainement pas à vous et moi. C'est son monde, celui d'une caste de milliardaires au-dessus des lois, qu'il veut sauver du reste de l'humanité. Le techno-optimisme, c'est l'économie politique préférée des ploutocrates du numérique. Dans cette histoire, l'IA n'est que le paravent technique d'une entreprise tout à fait classique de privatisation et de captation des richesses. Mais plus agressive, encore plus sourde aux intérêts collectifs et plus autoritaire que jamais.

Cette offensive antidémocratique est un phénomène récent. Entamée à la fin de la décennie 2010, elle marque une rupture franche avec la ligne prodémocratie tenue, du moins jusqu'en 2016, par les patrons de Facebook, Twitter et Google. C'est Peter Thiel, sorte d'empereur Palpatine de la Bay Area, qui a été le précurseur de ce techno-autoritarisme. En 2009, déjà, il déclarait que «la liberté et la démocratie ne sont pas compatibles». Pour lui, «la dernière décennie de l'histoire américaine pendant laquelle il était possible d'être réellement optimiste en matière de politique» a été celle des années 1920^[1]: l'économie débridée des Années folles, l'âge d'or du jazz, des façades art déco, du consumérisme effréné et de la modernité électrifiée. Rappelons tout de même que la décennie en question a débuté avec une pandémie de grippe espagnole et une récession de quatorze mois, s'est caractérisée par une politique économique de laisser-faire généralisé et une explosion de la violence raciste et qu'en termes d'optimisme, elle s'est terminée dans le fracas du krach boursier de 1929, prélude à la Grande Dépression des années 1930.

Avec son compère Marc Andreessen, un «évangéliste de l'Église de la technologie» qui «considère la Silicon Valley comme le centre opérationnel de l'humanité^[2]», Thiel était alors l'un des rares à porter un discours aussi radicalement à droite. Depuis l'élection de Donald Trump, à qui Thiel a fait don de 35 millions de dollars en 2016, le «contrarien^[3]» (baptisé ainsi par son biographe Max Chafkin pour sa propension à investir à contre-courant du marché) et ses pulsions antidémocratiques ont fait des émules. En 2024, Elon Musk et le capital-risqueur David Sacks ont soutenu financièrement le candidat républicain à la présidentielle Ron DeSantis. Pour la première fois de son existence, le fonds Andreessen Horowitz finance des politiciens perçus comme

«optimistes^[4]» – comprendre: antirégulation, populistes, autoritaires. D'autres milliardaires se sont ralliés à la cause, dont le quinquagénaire Jeff Bezos et le jeune quadragénaire Mark Zuckerberg.

Aujourd'hui, au-delà des affrontements traditionnels entre les camps démocrate et républicain, l'«église du techno-optimisme», comme l'appelait le *New York Times* en 2019, a imposé comme «véritable idéologie de la Silicon Valley^[5]» sa conviction absolue que le futur appartient aux magnats de la tech et qu'il se construira à l'extérieur des structures démocratiques. La chercheuse Asma Mhalla explique, dans *Technopolitique*, que l'avant-garde de la tech fait rimer futurisme avec autoritarisme, et qu'elle a les moyens de ses ambitions, parce qu'elle se trouve à la tête «d'entités hybrides, qui sont aussi des acteurs politiques, idéologiques, géopolitiques et militaires», dont les budgets annuels sont de l'ordre d'un PIB national^[6].

Résultat: le ton et le lexique ont changé. Que nous disent les éructations de Musk répondant «Nous renverserons qui nous voulons^[7]!» au président bolivien Evo Morales sur Twitter/X? Que nous dit le pitoyable feuilleton testostéroné du «combat» entre Musk et un Mark Zuckerberg passionné de jiu-jitsu brésilien? Que nous disent les biceps sous stéroïdes de Jeff Bezos et les projets de base souterraine de Peter Thiel? Que ce sont eux les chefs, les hommes forts, les prédateurs alpha. Et que le monde ferait mieux de se préparer à leur règne, car il est aussi imminent qu'il sera brutal.

Vers un capitalisme de commandement

L'éthos techno-optimiste, c'est avant tout l'expression d'une terrible impatience – les milliardaires de la Silicon Valley, comme des enfants assis à l'arrière d'une voiture pendant le long trajet vers les vacances d'été, n'en peuvent plus d'attendre. D'attendre quoi? L'AGI, vous répondront-ils, les yeux remplis d'étoiles. Ils oublient qu'aucun des systèmes actuels ne s'en approche, que cette machine est totalement hypothétique et que personne ne sait même réellement la définir. Peu importe! Encore un petit effort, une toute petite accélération, et nous y serons, nous toucherons du doigt la toute-puissance! Oui, mais quand?

Musk, Altman, Zuckerberg ou Andreessen nous promettent que la machine magique sera là d'ici cinq ans – un *pitch* classique de la Silicon Valley, où l'on promet constamment que le *next big thing* est au coin de la rue. Mais ils ont l'air d'être les seuls à y croire: début 2024, 50 % des chercheurs interrogés par le *think tank* AI Impacts^[8] l'envisageaient plutôt pour 2047. Une minorité conséquente de chercheurs, parmi lesquels Yann Le Cun, responsable du développement de l'IA chez Facebook/Meta, pense que ce concours de pronostics n'a aucun sens.

Pourquoi tant d'impatience au sein de cette clique de patrons? De l'argent, il y en a, cela va de soi. En 2023, année de délire collectif chez les investisseurs de la tech, les cabinets de conseil, les Madame Irma du capitalisme à la petite semaine, ont rivalisé de prédictions enfiévrées: selon McKinsey, comme nous l'avons vu, le marché de l'IA dite «générative» (ChatGPT et consorts) devrait générer l'équivalent du PIB du Royaume-Uni, soit entre 2 et 4 trillions de dollars; PriceWaterhouseCooper surenchérisait en évoquant 15 trillions de dollars d'ici à 2030^[9].

D'autres *think tanks* faiseurs de pluie, affiliés à la nébuleuse altruiste efficace, sont allés jusqu'à prédire une explosion de la croissance économique de 20 à 30 % par an^[10], ce qui est tout simplement ridicule – la croissance annuelle mondiale tourne autour des 3 % depuis 1900. La fièvre est cependant en train de retomber, et les sorciers de la finance reprennent petit à petit leurs esprits: en mars 2024, McKinsey reconnaissait qu'«il est plus compliqué que prévu de capturer la valeur potentielle de l'IA générative^[11]». Personne n'a vraiment trouvé d'usage commercial à ces *chatbots*, et le spectre d'une bulle financière plane. Néanmoins, au premier trimestre de 2024, les places boursières mondiales enregistraient leur meilleur résultat de croissance (+ 7,7 %) depuis cinq ans^[12].

Mais il y a aussi les rêves de domination qui font piaffer ces patrons. Dans la mythologie droitière de ces milliardaires pressés, seule l'IA peut débloquent l'accès à un pouvoir non plus strictement économique mais politique, culturel et idéologique. Ce fantasme s'exprime en outre dans la prophétie de la «Singularité technologique». De quoi s'agit-il? D'un événement hypothétique de convergence des techniques et de

l'humanité, sorte d'unité absolue des machines et des humains dans l'IA que provoquerait l'arrivée de l'AGI, théorisée en 2005 par l'oracle transhumaniste Ray Kurzweil. Les techno-optimistes sont persuadés qu'une fois que l'AGI sera atteinte, la Singularité suivra très vite – une question de jours, d'heures, voire de minutes. Dans la Silicon Valley, ce scénario n'a pas un nom, mais un son: «*FOOM*», en majuscules. Le souffle d'un TGV qui passe. Mais derrière le fantasme informatique, ce qui intéresse particulièrement nos oligarques, c'est la perspective de rebattre brutalement les cartes de la civilisation en leur faveur. Que dit la prophétie? Que la Singularité nous projetterait littéralement dans un autre monde, sans qu'on ait le temps de réfléchir collectivement à son organisation. L'historien et politologue camerounais Achille Mbembe appelle «brutalisme» ce «devenir-artificiel de l'humanité et son pendant, le devenir-humain des objets et des machines^[13]», qui implique une domination totale des ressources du vivant par les outils d'exploitation capitalistes.

Sous prétexte d'accélérer le temps qu'il reste avant ce basculement dans l'*après*, c'est-à-dire pour faire advenir plus vite l'âge de la domination technoscientifique, les faiseurs d'IA demandent un *reset* sociopolitique. Selon eux, la poursuite de l'inévitable progrès de la technologie justifie une refonte totale des structures du corps social. Dans *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*, Dan McQuillan compare ce projet politique à un «thatchérisme informatique», une «forme de “stratégie du choc” qui utilise le sentiment d'urgence suscité par une technologie censée transformer le monde comme excuse pour subvertir les systèmes sociaux sans passer par le débat démocratique^[14]». *There is no alternative*, on connaît la chanson. Mais la connaît-on vraiment?

Lorsque Marc Andreessen affirme que l'IA est son «alchimie, [sa] pierre philosophale», il ne veut pas dire qu'il rêve de changer le plomb en or. Il a l'intention de transformer le monde. L'économie de marché, selon lui, est «une forme d'intelligence – un système qui s'adapte, qui explore, qui évolue^[15]». Dans la tête des techno-oligarques, l'intelligence artificielle transcendera le libre marché en accomplissant le projet d'une régulation purement automatique, involontaire, invisible et

omnipuissante des activités individuelles. Elle sera, enfin, cette main invisible que personne ne commande. Dans les déclarations des technocrates se lit un projet de domination qui rejette le mariage de raison entre capitalisme et démocratie au profit d'un système techno-économique inédit. Le projet est inédit dans sa bêtise, pour commencer. Dans un futur où rien ne freinerait l'accumulation privée de richesse et où règnerait un système technique inhumain – ou posthumain, c'est selon –, le commun des mortels serait comblé, nous assurent-ils. Adieu la souffrance, adieu le conflit, adieu la lutte des classes. Elon Musk, Sam Altman et les autres nous annoncent une société d'abondance matérielle absolue où tous les problèmes actuels, le chômage, la maladie et la pauvreté seront résolus. Rien que ça! Mais d'ici là, ne leur parlez surtout pas de redistribution de la richesse, ne mentionnez surtout pas le fait que nous vivons déjà dans une société d'abondance matérielle où 15000 enfants meurent de faim chaque jour^[16]. Cela n'a rien à voir avec la société idéale dessinée par ces milliardaires incultes, où l'IA ferait exploser la productivité et baisser les prix des biens et services jusqu'à un coût marginal, et où des robots garantiraient à tout le monde un accès au confort matériel.

Dans cette «corne d'abondance consumériste^[17]», comme l'appelle le milliardaire Marc Andreessen, qui s'y connaît en matière d'abondance, la majorité des emplois deviendraient obsolètes. Les travailleurs seraient remplacés par des machines, qui s'occuperaient de tout (et surtout de produire, tout le temps, sans s'arrêter). Les milliardaires rêvent d'une société sans masse salariale, où les produits manufacturés ne coûtent quasiment rien, et où une poignée d'individus capte presque toutes les richesses. Dans ce songe hypercapitaliste, leurs entreprises génèrent tellement d'argent qu'elles financent un revenu universel de base «de 13 500 dollars annuels pour chaque adulte aux États-Unis [...] d'ici dix ans^[18]». C'est du moins ce que promet le philanthrope Altman, à la tête d'une entreprise dont la valeur en Bourse est estimée à 86 milliards de dollars. Pour ces ultrariches, le festin dès aujourd'hui; pour nous, une abondance de miettes, et encore, seulement quand arrivera leur avenir radieux. L'IA «mettra tout le monde à égalité^[19]», à en croire le militant communiste Elon Musk. La belle affaire! En attendant le grand

égalisateur robotique, les 1 % les plus riches ont capté 63 % des richesses mondiales produites entre 2020 et 2023^[20]. Et ça, malheureusement, ce n'est pas de la science-fiction.

Ne soyons pas naïfs: quand l'élite belliqueuse de la Silicon Valley se met à nous cajoler, à parler de pouvoir d'achat, de revenu universel et de la fin de la pauvreté, quand elle nous promet qu'elle va s'occuper de tout, il vaut mieux s'inquiéter pour notre autonomie collective, parce qu'elle s'appête à déployer une technologie disciplinaire. L'avenir des sociétés de contrôle, qu'appellent de leurs vœux ces bonzes, c'est un mélange d'abrutissement consumériste et de surveillance algorithmique, avec l'IA pour contremaître, et une caste quasi aristocratique bien à l'abri de l'abomination qu'elle a déployée. D'un côté, des masses passives, qui réagissent à des stimuli plus qu'elles n'agissent, qui ont des réflexes et non des réflexions; de l'autre, les ingénieurs, les cadres et les patrons qui pensent, décident et s'enrichissent.

La philosophe Antoinette Rouvroy décrit ce système de «gouvernementalité algorithmique^[21]» comme «un système immunitaire développé par le capitalisme pour lutter contre tout ce qui pourrait le mettre en crise». Si les rapports sociaux sont optimisés par l'extraction de données, si les échanges de biens et de personnes sont fluidifiés à l'extrême, alors la politique devient obsolète; dans un corps social réduit à un empilement d'individualités microgérées et gavées par des machines autonomes, les espaces de délibération démocratique deviennent superflus. Si l'on cesse de construire patiemment des désirs collectifs, si l'on se laisse happer par l'instantanéité de la récompense algorithmique, il n'y a aucune autonomie possible.

De la même manière que le fantasme de la *start-up* a fait grandir le cancer néolibéral, l'utopie de l'intelligence artificielle déguise une calamité bien réelle, pire encore que le marché dérégulé: la zombification du corps social. Car le capitalisme triomphant qu'envisagent les techno-oligarques n'a rien à voir avec le libre marché d'Adam Smith. C'est bien plutôt ce que Cédric Durand et d'autres économistes ont baptisé le «technoféodalisme^[22]»: une économie métastatique et centralisée, dominée par un cartel de plateformes de l'information qui extraient une rente aux entreprises traditionnelles tout

en exploitant la main-d'œuvre gratuite fournie par l'activité de milliards d'utilisateurs producteurs de données, éparpillés et atomisés. Une concentration extrême, grâce à la décentralisation et au morcellement total des sociétés humaines.

Le *cloud*, ce gigantesque système d'extraction et de gestion des données, capte la richesse et accumule du capital non pas en exploitant des ouvriers, mais grâce à la contribution gratuite des usagers et à la location de ses infrastructures aux entreprises traditionnelles. Essentiellement, cette hydre prélève sa rente à même le temps de loisir des personnes, ou en parasitant les activités économiques traditionnelles, qui sont de plus en plus dépendantes des algorithmes. Yánis Varoufákis parle d'ailleurs d'économie «cloudaliste». Selon lui, cette dynamique parasitaire expose les sociétés à la bêtise et les rend plus vulnérables, d'autant plus qu'elle confère un pouvoir inouï à des personnalités exaltées et dangereusement habiles^[23]. Le technocritique Evgeny Morozov signale quant à lui un «capitalisme de commandement^[24]», où la propriété des moyens de production est assujettie à la propriété des réseaux d'informations. Le plan des techno-oligarques consiste ainsi à s'extraire définitivement de la double menace de la concurrence et de la régulation, en se plaçant hors de portée des deux.

La clique techno-optimiste ne se contente pas d'aspirer à un pouvoir hermétique à la concurrence, fondé sur la maîtrise des flux d'informations; elle est déjà en train de s'arroger ce pouvoir en s'immisçant dans les failles ouvertes par la politique néolibérale. Microsoft Azure, Amazon Web Services et Google Cloud jouissent d'un oligopole sur les centres de données qui hébergent les données du monde entier – entreprises publiques et privées, gouvernements, hôpitaux, laboratoires de recherche, universités, armées, vous et moi. Google, devenu de fait notre unique porte d'entrée sur le web, façonne notre accès à l'information numérique.

Elon Musk contrôle un espace majeur d'information en ligne, X, qu'il a transformé en mégaphone à beuglements d'extrême droite. Non content de remplacer la NASA avec ses fusées réutilisables SpaceX, il a désormais un quasi-monopole de l'orbite terrestre, où près de 5 000 satellites Starlink flottent désormais à 550 kilomètres d'altitude.

Aujourd'hui, le service fournit internet à l'armée ukrainienne, ce qui fait de Musk une entité géopolitique inédite, incontrôlable, capable de décider (au nez et à la barbe du Pentagone) de couper la connexion aux drones ukrainiens au-dessus de la Crimée^[25]... pour mieux vendre Starlink à l'armée russe ensuite^[26]. Musk le nie fermement, une enquête est en cours.

Même le sacro-saint complexe militaro-industriel états-unien, dominé par les conglomérats de l'armement Lockheed Martin et Raytheon, n'est plus à l'abri des assauts de Google^[27], Amazon et OpenAI, dont les lobbyistes draguent désormais ouvertement les ministères de la Défense et de la Sécurité intérieure des États-Unis^[28]. Pour les libertariens de la Silicon Valley, les mégacontrats militaires, chiffrés en milliards de dollars, sont irrésistibles.

Partout où ils vont, les PDG de la Silicon Valley sont reçus avec les égards qu'on réserve normalement aux chefs d'État. L'AGI se fait attendre, la Singularité aussi, mais les techno-oligarques, métastases d'un nouveau régime capitaliste, se répandent dans l'organisme de l'État-nation en échafaudant les mondes qu'ils construiront sur sa dépouille.

Retour vers le futur

Pour la première fois depuis les années 1970, aucun des entrepreneurs les plus en vue de la tech n'a moins de 30 ans. Pour une industrie connue pour son jeunisme et qui est censée inventer le futur, il est pour le moins paradoxal d'être guidée par une brochette de *boomers*. À quoi ressemble donc l'imaginaire politique et culturel des ultrariches de la Silicon Valley, ces prophètes de l'IA formés dans les universités d'élite états-uniennes, et tous déterminés à influencer la marche du monde?

La réponse est à la fois simple et pathétique: ce sont des hommes-enfants, persuadés que leur fortune suffira à matérialiser un imaginaire de science-fiction et de *fantasy* tout droit sortis des années 1960 à 1990 – le futur qu'on trouve dans les livres d'Isaac Asimov et Robert Heinlein, dans *K2000* et *Star Trek*. Ces «jeunes enfants rois apparemment inoffensifs, qui voulaient faire un monde meilleur [...] ont fini par enfiler des masques de Dark Vador^[29]», résume la journaliste Kara Swisher,

vétérane de la Silicon Valley, autrice de *Burn Book*. Ils souffrent, frustrés d'un présent qui ne leur convient pas et auquel ils échappent en se réfugiant dans la science-fiction. Les références à cet univers sont disséminées partout: dans les petits détails de leurs vies d'entrepreneurs, dans le lexique qu'ils utilisent, dans leur rapport au corps social et dans leur projet pour l'humanité.

À 50 ans passés, mais perpétuellement en quête de validation sociale, Elon Musk publie des photos de lui déguisé en Doctor Strange, le «sorcier suprême» des *comics* Marvel, chargé de protéger l'humanité des menaces du monde magique. Musk, un homme qui n'a jamais rien inventé ou fabriqué de sa vie (excepté sa propre légende), est également fasciné par Tony Stark, l'entrepreneur milliardaire et génial inventeur qui se construit une armure de superhéros pour devenir Iron Man. Musk s'identifie tellement à cette incarnation monstrueuse de l'entrepreneur capitaliste au-dessus des lois (littéralement, il vole) qu'il prétend que le réalisateur Jon Favreau s'est inspiré de lui pour transposer le personnage au cinéma.

Mark Zuckerberg est lui aussi fasciné par Tony Stark. En 2016, lors d'une opération de communication millimétrée, le robotique PDG de Facebook dévoilait une sorte d'IA domestique baptisée «Jarvis», comme celle d'Iron Man. Autre indice du changement d'ambiance chez les milliardaires de la tech, le *nerd* Zuckerberg, *fan* des jeux de stratégie *Civilization* et *Risk*, frais quadragénaire en pleine recherche de virilité, s'est trouvé un autre modèle d'homme fort: l'empereur romain Auguste César. Son admiration frise l'obsession: il se coiffe comme lui, a appelé sa deuxième fille August et a pris l'habitude de terminer toutes ses réunions en criant «Domination^[30]!» Un autre de ses mantras est «les entreprises avant les pays^[31]» et, en effet, son empire virtuel compte aujourd'hui deux milliards d'utilisateurs actifs, une population supérieure à celle de n'importe quel pays du monde. Pour trouver des entités concurrentes, il faut changer de domaine: en nombre d'adhérents, Facebook se place juste entre l'islam et la chrétienté.

La littérature de science-fiction fournit des modèles d'utopie, souvent malgré elle. Elon Musk affirme à qui veut l'entendre que la série *Fondation*, publiée dans les années 1960 par Isaac Asimov, est une

influence majeure. On y suit, sur plusieurs millénaires, l'histoire d'une société scientifique isolée qui s'organise autour de la figure messianique d'un chercheur-oracle, Hari Seldon, capable de prévoir l'avenir de la galaxie grâce à une méthode supposément rationnelle (la «psychohistoire») et infaillible. Seldon revient régulièrement d'entre les morts, ses prophéties scientifiques sous le bras, pour guider ses disciples à travers une sorte de Moyen Âge galactique de trente mille ans, à la suite de l'effondrement de l'Empire. Un génie solitaire qui sauve la civilisation grâce à son puissant esprit rationnel capable de prédire l'avenir à long terme des sociétés? Pas étonnant que le concept séduise Musk, qui a peut-être raté la parabole critique sur l'endoctrinement.

Musk, qui change de casquette politique selon l'occasion et s'est un jour décrit comme «un anarchiste utopiste», dit également s'inspirer des visions de l'écrivain de SF écossais Iain Banks (pourtant farouchement socialiste), au point de nommer plusieurs des barges flottantes de SpaceX d'après ses créations (*Of Course I Still Love You* et *Just Read the Instructions*). Le pauvre Banks n'a plus qu'à se retourner dans sa tombe. Dans sa mythique série *Culture*, débutée dans les années 1980, il imaginait une société d'abondance matérielle illimitée, la Culture, où des IA superintelligentes et conscientes, les Mentaux, cohabitent avec 30 milliards d'humains augmentés et immortels sur des vaisseaux de la taille d'une planète. Tout y a l'air paradisiaque, sauf que ce sont bel et bien les machines qui gouvernent, en despotes éclairés au pouvoir absolu. Il n'y a ni lois ni gouvernement: les humains sont des animaux domestiques, choyés et dociles. Même pour Banks, la Culture est un monde ambigu, à la fois utopie libérale et dystopie collective, où la société meurt étouffée dans l'étreinte maternelle de la machine^[32]. Bien entendu, Mark Zuckerberg s'en réclame à son tour. Et Jeff Bezos, désormais occupé à bâtir des stations orbitales des années 1970 pour un «trillion d'humains^[33]», tente d'adapter la série *Culture* à la télévision. Pour eux, l'œuvre de Banks n'est pas une dystopie: c'est un mode d'emploi, un dépliant promotionnel pour la réalisation de leurs ambitions autocratiques futuristes.

Le plus radical et le plus rétro de la bande est aussi le plus à droite: Peter Thiel. Obsédé par *Le seigneur des anneaux* (comme bon nombre de

techies et d'investisseurs de la Silicon Valley), le «contrarien» a baptisé cinq compagnies (Palantir, Mithril Capital, Valar, Rivendell et Lembas) à partir d'objets magiques et de lieux de l'univers de J.R.R. Tolkien, et explique à des journalistes qu'il aspire à l'immortalité des elfes de la Terre du Milieu^[34]. Il s'inspire également de *La cité et les astres*, écrit par Arthur C. Clarke en 1956, qui décrit le quotidien d'une ville d'immortels, un milliard d'années dans le futur, où les individus sont stockés dans des ordinateurs et téléchargés éternellement de corps en corps.

Au-delà de sa foi transhumaniste, Thiel est surtout l'archétype de ce qu'on appelle aujourd'hui un *boomer*, aigri et furieux contre la modernité actuelle. Il préfère «les *Star Trek* et les *Battlestar Galactica* originaux^[35]» aux versions contemporaines «communistes», leur préfère le «capitaliste» *Star Wars*^[36], regrette «le tournant dystopique pris par la science-fiction ces dernières décennies» et déplore qu'alors que «nous voulions des voitures volantes, on nous a donné 140 caractères^[37]». Bref, Thiel déteste le présent, il estime qu'on a raté un virage et se demande comment nous faire «revenir dans la course» à l'avenir.

Lorsqu'il évoque le futur, le soi-disant libertarien a une image précise en tête: l'imaginaire «daté mais futuriste^[38]» (il le reconnaît lui-même) de la série animée *Les Jetson*, vision éculée de l'an 2062 conçue pendant les Trente Glorieuses. Son avenir idéal est un rêve américain chromé peuplé de gadgets, de voitures volantes, de robots domestiques, de *jetpacks* et de stations orbitales, traversé par un consumérisme béat. C'est aussi une *gated community* de banlieue prospère états-unienne, où l'on ne croise que des familles nucléaires blanches, propres sur elles, aux valeurs chrétiennes traditionnelles. Une techno-utopie immaculée, où chacun reste dans sa station orbitale. Un futur antérieur, anachronique, dont la xénophobie intrinsèque est aussi flagrante que sa vacuité. Des récits rances d'humanité triomphante et de sociétés impérialistes, misogynes et homogènes. Des récits comme *Le seigneur des anneaux* qui, au même moment, servent également d'inspiration à des fascismes bien réels et contemporains – celui de la première ministre italienne Giorgia Meloni –, glorifient la monarchie, le militarisme et l'autoritarisme.

Le plus surprenant, c'est la manière dont ces ultra-dominants planétaires aux privilèges illimités se sentent *triggered* aussi bien par les nouvelles versions de *Star Trek* que par la démocratie représentative et l'administration fiscale. Kara Swisher, qui les côtoie depuis si longtemps qu'elle fait quasiment partie de l'aristocratie, écrit qu'elle n'a «jamais vu un groupe de personnes aussi puissantes et riches qui se considéraient aussi intensément comme des victimes^[39]». Ces hommes se rêvent en superhéros, en empereurs, en grands hommes, alors qu'ils ne sont que de simples «Méphistophélès managériaux^[40]», de sinistres personnages, des nets négatifs pour le bien commun. Ce sont des parasites, qui préfèrent croire que tout le monde les déteste par jalousie plutôt que de remettre en question leurs pratiques et leur éthique. Et comme le sentiment de persécution mêlé au pouvoir est un terrain fertile pour l'autoritarisme, ces pauvres ultrariches ont trouvé une solution pour se venger et mater leurs oppresseurs imaginaires – l'État, le «progressisme», le «wokisme»: appliquer une couche de pensée fasciste à leur futurisme chromé.

Lumières noires sur la Silicon Valley

À l'été 2023, une nouvelle école de pensée, née dans le vase clos d'espaces de discussion Twitter Spaces, s'est répandue comme une épidémie virale parmi les *techies* et les investisseurs du secteur: l'accélérationnisme efficace, une «métareligion» postmoderne ébouriffante, dédiée au culte de la «Singularité technocapitaliste». Ses adeptes se reconnaissent sur Twitter/X avec une mention «e/acc» et un emoji «avance rapide» dans le pseudo. Leur devise: «*Accelerate or die*».

Leur prophète, un ex-ingénieur en informatique quantique de Google, signe ses oracles du pseudo «Beff Jezos» et affirme que «le capitalisme est une forme d'intelligence». Sa conviction: le futur idéal arrivera plus vite si on laisse tranquille le présent, et surtout ceux qui le dominent. Forcément, les millionnaires de la Silicon Valley ont eu un coup de foudre. Sur Twitter/X, Sam Altman est tout heureux de se mesurer au nouveau prophète, et il le défie: «Vous ne pouvez pas accélérer plus vite que moi.» Avis aux «*decels*» – les régulateurs, les éthiciens, les ennemis de la croissance et autres retardeurs du progrès: l'avenir vous fonce

dessus, que vous le vouliez ou non. Le succès d'un tel discours, qui n'est au fond qu'une resucée du transhumanisme libertarien de Max More, est symptomatique d'un esprit du temps: l'IA justifie les moyens.

Derrière la pseudosecte virale de «Beff Jezos» se cache une mouvance enténébrée mais influente, active depuis plus de vingt ans, que la presse a baptisée la «New Right». Une communauté sans leader, mais avec ses gourous. C'est l'un d'eux, le britannique Nick Land, qui développe l'accélérationnisme dans les années 1990, sur fond d'implosion de l'URSS et de révolution internet. Contrairement aux transhumanistes ultralibéraux californiens, qui rêvent des mêmes édens dérégulés, Land se fiche complètement de libérer le «potentiel humain». Sa croisade à lui, c'est la «libération» des moyens de production, à tout prix. Dès la fin des années 1990, il mêle économie politique et science-fiction en imaginant des scénarios de singularité technologique et d'IA toutes-puissantes au sein du collectif Cybernetic Culture Research Unit. Après un déménagement à Shanghai au début du millénaire, le libertaire s'entiche soudainement des régimes autoritaires hypercapitalistes comme Singapour ou Hong Kong.

Dans les années qui suivent, son influence reste entièrement souterraine. Ses écrits circulent dans un étroit réseau de blogues d'extrême droite, sans jamais voir la lumière du *mainstream*. Ils sont lus par une communauté extrêmement connectée, mélange de riches *crypto bros*, de cadres de la tech et d'universitaires marginalisés. Grâce à la magie de la blogosphère, ses écrits vont rencontrer ceux de l'autre figure tutélaire de la New Right: un codeur et entrepreneur de la Silicon Valley appelé Curtis Yarvin. Ou plutôt «Lord Yarvin, notre Prophète^[41]», comme on peut parfois lire et entendre sous une généreuse couche de cynisme postmoderne.

Yarvin et Land feront naître deux nouvelles doctrines, qui se fondent l'une dans l'autre et vont progressivement gagner en importance pour l'extrême droite des États-Unis: les «Lumières noires» – une sorte de double maléfique de l'idéologie des Lumières, où la science et les techniques sont mises au service de l'autorité au lieu de l'émancipation – et la «Néoréaction», souvent abrégée en NRx. Inspirés par la maxime de Peter Thiel sur l'incompatibilité entre l'appétit capitaliste et l'égalité

démocratique, tous deux prônent grosso modo le remplacement des démocraties occidentales par une monarchie capitaliste privée, la *gov-corp*, gérée comme une *start-up* par un PDG-dictateur de la Silicon Valley. Leur modèle: l'ex-dirigeant chinois Deng Xiaoping, aussi autoritaire que dérégulateur. Leur inspiration: la cité-État médiévale. Tout un programme.

Le monarchiste Yarvin, qui signe ses écrits du pseudonyme Mencius Moldbug, est en croisade contre ce qu'il appelle la «Cathédrale» – une forme d'*establishment* libéral formé de médias et de professeurs d'université, accusé de wokisme avant l'heure. En 2007, dans un article de blogue intitulé «The Case Against Democracy» (Un réquisitoire contre la démocratie)^[42], il utilise l'image de la «pilule rouge», donnée à Néo dans *Matrix* pour qu'il ouvre enfin les yeux sur le monde réel, comme métaphore de la conversion à sa vision d'extrême droite. Une fois la pilule avalée, bienvenue non pas dans le «désert du réel», comme dirait Morpheus, mais dans un monde en miroir – la rationalité, la science et la technique des Lumières au service d'une idéologie anti-égalitaire, antidémocratique, xénophobe et fascisante. Là où Zuckerberg, Thiel et Musk voient le futur par le prisme du xx^e siècle, Yarvin et Land l'envisagent à travers celui du positivisme autoritaire du xix^e.

Les deux philosophes, obsédés par le quotient intellectuel (QI) et la race, ont bâti un projet politique monstrueux, qui reprend la société stratifiée de l'Angleterre victorienne en y ajoutant un vernis technofuturiste. Land propose ainsi de rejoindre l'utopie transhumaniste par une société spatiale administrée en castes, régie par l'«hyper-racisme», dans lequel les élites «explicitement supérieures» et leurs énormes QI se reproduiraient uniquement entre elles. Yarvin, qui rêve entre autres de transformer San Francisco en une entité privée appelée Friscorp, propose que les «hominidés» improductifs du prolétariat soient «virtualisés», c'est-à-dire enfermés à l'isolement permanent dans une interface de réalité virtuelle «immersive». Le métavers de Zuckerberg pour les sous-hommes à QI moyen, le luxe du monde réel pour l'élite cognitive autoproclamée.

À la faveur de la polarisation de l'opinion publique aux États-Unis et de la droitisation de la Silicon Valley, Land, Yarvin et toute la troupe du

freak show intellectuel de la «Nouvelle Droite» peuvent désormais compter sur l'attention et le soutien financier des technoprophètes. La Néoréaction est par ailleurs considérée comme la source d'inspiration du mouvement *alt-right* de Steve Bannon, éminence grise de l'élection de Trump en 2016. Yarvin décrit sa Cathédrale dans le *show* de Tucker Carlson, tête de gondole de Fox News, pendant que l'image de la pilule rouge, déclinée à l'infini en mèmes sur le forum 4chan et régulièrement tweetée par Elon Musk^[43], fait partie de la panoplie mémétique de tout militant fasciste et complotiste sur internet.

En quelques années, un continuum s'est consolidé entre la blogosphère néofasciste la plus rance, certains milliardaires de la Silicon Valley et une nouvelle phalange d'élus républicains: les blogueurs théorisent, les politiciens diffusent, les technomagnats financent le tout. Le réac en chef Peter Thiel, dont le premier ouvrage, publié en 1995, s'intitule *The Diversity Myth* (Le mythe de la diversité), a évidemment été parmi les premiers à soutenir la nébuleuse néoréactionnaire. Il finance la *start-up* de Yarvin, Urbit.

Nick Land n'a pour sa part aucun lien avec la Silicon Valley. Mais son accélérationnisme, qui s'est retrouvé dans les manifestes de plusieurs terroristes suprémacistes blancs^[44], revit en 2023 sous la lumière du *mainstream*. On le retrouve entre autres sous la plume du milliardaire Marc Andreessen, qui résume le techno-optimisme comme une croyance en «ce que Nick Land a baptisé la machine du technocapital, le moteur perpétuel de création matérielle, de croissance et d'abondance^[45]». Le projet «techno-optimiste» du capital-risqueur est bien le même que celui des néofascistes états-unis: le progrès technique sans la démocratie.

Ce n'est pas la première fois qu'une classe de ploutocrates s'entiche d'un totalitarisme, et Andreessen revendique une filiation historique. Parmi ses «saints patrons du techno-optimisme», qui incluent les parrains du libéralisme Milton Friedman et Friedrich Hayek, on trouve Filippo Tommaso Marinetti, poète et chef de file du mouvement futuriste italien des années 1920. Andreessen paraphrase même son *Manifeste du futurisme*: «La beauté n'existe que dans la lutte. Tout chef-d'œuvre a un caractère agressif. La technologie doit être un assaut violent contre les forces de l'inconnu, pour les forcer à s'incliner devant l'homme.»

L'investisseur n'ignore certainement pas que les futuristes italiens, obsédés par les avions, les voitures, la jeunesse, la vitesse et la violence – les accélérationnistes de l'époque industrielle, en somme – ont été ensorcelés par le fascisme mussolinien. Il n'ignore probablement pas non plus l'existence du «modernisme réactionnaire», cette collusion entre industriels autrichiens, hommes d'affaires libéraux allemands et nazis. Mais l'héritage est encore trop difficile à assumer, malgré le vent nouveau qui souffle sur la Valley.

Pour rappel, Marc Andreessen n'est pas seulement un investisseur richissime. C'est l'un des monstres sacrés de la Silicon Valley, dont la carrière est un reflet des évolutions de ce milieu. Présent depuis les années 1990, inventeur du premier navigateur web Netscape, il jouit d'une influence considérable. Outre sa fortune personnelle estimée à près de deux milliards de dollars, il siège, comme Peter Thiel, au comité exécutif de Facebook. Son fonds d'investissement gère 35 milliards de dollars d'actifs dans près de 750 entreprises de la tech. Quand il défend aussi nettement une vision, elle a toutes les chances d'être adoptée aussi bien par les entrepreneurs de *start-up* que par les capital-risqueurs.

Comme les autres prophètes de l'IA, Andreessen défend une vision réactionnaire de la modernité. Le champ lexical de ses oracles est le même que celui utilisé régulièrement par Musk, Bezos ou Altman pour vendre les miracles de l'intelligence artificielle et de l'astrocapitalisme: celui de l'expansion et de la conquête coloniale. Puisant lui aussi dans le registre de la *fantasy*, il se donne pour mission de «cartographier des territoires inconnus, vaincre des dragons et ramener le trésor à la maison pour notre communauté^[46]», sans douter une seconde de sa légitimité divine.

Son discours, qui se veut futuriste, rappelle étrangement le récit colonial fondateur des États-Unis, prélude à la philosophie impérialiste du pays. En pleine fièvre expansionniste, on parlait alors de la «destinée manifeste» des colons qui partaient vers les territoires sauvages de l'Ouest pour y apporter la civilisation, d'un océan à l'autre. Par le chemin de fer, la technique dompte la nature; par le génocide et la conversion religieuse, le colon dompte l'Autochtone. Deux siècles plus tard, l'IA et le projet transhumaniste rejouent la même partition, et l'élite

occidentale doit à nouveau porter le «fardeau» de la civilisation technique, non seulement à toute la planète, mais à toute la galaxie. D'ailleurs, dès 2016, le *New Yorker* parlait de la «destinée manifeste de Sam Altman» pour décrire le jeune PDG en pleine ascension^[47].

Rien d'étonnant, alors, à croiser régulièrement du suprémacisme blanc chez ces sombres figures intellectuelles. Curtis Yarvin et Nick Bostrom, les théoriciens les plus influents de leurs champs respectifs, ont tous deux écrit que «les Noirs ont un QI moins élevé que les Blancs^[48]», tout en récusant les accusations de racisme. Sous l'influence combinée de la Néoréaction et des transhumanistes, les délires eugénistes sont bel et bien implantés dans la Silicon Valley. Peter Thiel soutient le blogueur suprémaciste Richard Hanania, qui prouve selon lui que «nous avons besoin d'une violence gouvernementale pour exorciser le démon de la diversité». Hanania prône la stérilisation forcée des «faibles QI», qui selon lui sont le plus souvent des Noirs^[49]. Selon l'historien Quinn Slobodian, il existe aujourd'hui un véritable «culte du QI» parmi les patrons de la tech^[50]. Persuadés de faire partie d'une minorité héroïque face à une majorité stupide, rappelle le blogueur Yarden Katz^[51], certains prophètes de la Silicon Valley défendent tranquillement un modèle de société plus ou moins ouvertement racialisé.

Elon Musk et Peter Thiel, obsédés par le soi-disant «effondrement démographique» mondial^[52], distribuent des millions à des *start-up* de fertilité, de suivi menstruel et de fécondation in vitro – en jargon Silicon Valley, on parle de «*femtech*» et de «*reprotech*»^[53] – pour tenter de freiner le déclin de la population tout en améliorant la qualité de ses membres^[54]. Et quand le patron de Twitter/X alerte sur un «génocide blanc^[55]» imaginaire en Afrique du Sud, on comprend qu'il s'inquiète d'abord du déclin de la population blanche. Sur son profil, il épingle pendant des semaines des graphiques et des vidéos complotistes qui reprennent tout le récit du Grand Remplacement, tout en affirmant ensuite qu'il n'y croit pas (il n'est pas exclu qu'il ne le comprenne pas)^[56]. Bezos, seul sur scène, présente de jolies démos de ses futures «colonies récréatives» orbitales, où il rêve de voir s'égarer «mille Mozart et mille Einstein, ce qui ferait une civilisation extraordinaire^[57]». Ces gens-là bâtissent, financent et déploient des systèmes d'information

d'ampleur planétaire, biaisés en faveur de leurs semblables et au détriment du reste.

Derrière la mode de l'IA et le retour de l'autoritarisme dans la Silicon Valley, la techno-élite dévoile la véritable nature de son projet politique et culturel. Une caste capable de proclamer publiquement son adhésion aux discours de penseurs fascistes et de déclarer, comme le fait Marc Andreessen dans son manifeste, «nous sommes le prédateur alpha» – ce qui fait de nous des... proies? –, c'est soit une caste qui se sait intouchable, soit un groupe radicalisé au-delà du récupérable et, par conséquent, dangereux. Dans le cas des grandes fortunes de la tech, c'est probablement un peu des deux.

Leurs fortunes astronomiques et leur impunité politique, combinées à l'arrogance originelle des entrepreneurs et à l'influence, chez certains, du suprémacisme blanc, ont fini par les convaincre de leur supériorité. Leur utopie, construite autour de cette conviction élitiste, est structurellement excluante. Les prophètes de l'IA ne luttent pas seulement contre la démocratie, mais contre l'humanité tout entière.

Il y a alors quelque chose de profondément ironique à lire les mots de Marc Andreessen lorsqu'il écrit que «notre ennemi est la tour d'ivoire, cette vision d'experts à la Monsieur je-sais-tout, qui se délectent de théories abstraites et de croyances de luxe, déconnectés du monde réel, délirants, non élus, irresponsables – ceux qui jouent à Dieu avec les vies des autres, totalement isolés des conséquences^[58]». Aurait-on voulu le décrire, lui et ceux de sa classe, qu'on ne s'y serait pas mieux pris.

Vers l'infini et l'au-delà

Au-delà des positions respectives de Thiel, Musk, Andreessen, Zuckerberg et Bezos sur le spectre de la réaction et du suprémacisme blanc, le techno-optimisme est la conviction religieuse, partagée, qui veut que le capitalisme et l'intelligence artificielle, déployés sans restriction, leur permettront de prendre un jour les commandes de l'histoire. Tout en se retirant, chacun à leur manière, dans des forteresses imprenables. (L'image de la forteresse imprègne, on le verra, toute leur stratégie de domination économique.)

Pour citer le père du cyberpunk, William Gibson, la technoreligion qui fait fureur dans la Silicon Valley propose au fond que «les riches deviennent une autre espèce», au-dessus des lois qui gouvernent les simples mortels. Voilà le projet ultime des technoprophètes: faire sécession. Fuir l'humanité de chair, de souffrance, d'errance, honteuse et honnie pour mieux la soumettre à distance. Rester hors de portée de la foule, intouchable comme un objet sacré. Inscire sa supériorité dans un ailleurs idéal créé par les puissants pour les puissants, où la caste supérieure des multimillionnaires pourrait enfin satisfaire sa vanité.

En 1999, au plus fort de la bulle *dot-com* et de sa transe technoreligieuse, l'historien David Noble constatait que «si l'on s'intéresse de près à certaines grandes entreprises technologiques, on voit apparaître la dévotion religieuse partout. [...] Les tenants de l'intelligence artificielle pérorent sur les possibilités d'immortalité et de résurrection offertes par la machine tandis que leurs disciples, architectes de la réalité virtuelle et du cyberspace, exultent dans leur recherche d'une omniprésence divine et d'une perfection décorporée^[59]». Vingt-cinq ans plus tard, une Silicon Valley en pleine extase messianique nous vend le métavers, l'IA, les *non-fungible tokens* (jetons non fongibles, NFT) et le «web3», autant de triomphes de la foi sur la raison et préludes à la grande transcendance vers l'au-delà artificiel.

Une nouvelle fois, du neuf avec du vieux. Cette obsession de l'évasion et de la fortification se matérialise aussi bien dans les résidences surprotégées de la ville californienne d'Atherton^[60], où loge une partie de l'élite tech, que dans l'achat compulsif d'îles privées individuelles. Elle est une des composantes de la *Californian ideology*, l'éthos contradictoire de la «classe virtuelle» de la Silicon Valley, théorisée en 1995 par Richard Barbrook et Andy Cameron dans leur essai éponyme: les fabricants de technologies d'émancipation et de connexion cherchent simultanément à accroître leur domination et leur ségrégation. Dans un monde perçu comme étant de plus en plus chaotique, les ghettos et les *safe spaces* pour riches ne suffisent plus. Dans la région de San Francisco, un groupe d'investisseurs, parmi lesquels Marc Andreessen, a discrètement acheté pour 800 millions de dollars de terres agricoles depuis 2017, avec le projet de bâtir une ville

idéale pour grandes fortunes^[61]. Son nom: «California Forever». Mais d'autres techno-oligarques voient encore plus grand, encore plus loin, encore plus reclus.

Inspirés par les zones économiques spéciales et les paradis fiscaux de Shenzhen, Singapour, Dubaï ou Hong Kong, certains cherchent à s'extraire de la souveraineté des États en bâtissant une sorte de terre sainte libertarienne, dont les habitants seraient enfin libres d'innover sans entraves. Peter Thiel, toujours précurseur dans les utopies antidémocratiques, a bien essayé les phalanstères flottants du *seastading*, de 2008 à 2011, avec le petit-fils de Milton Friedman, mais le projet a coulé^[62].

Dernièrement, Marc Andreessen et le cryptomilliardaire Vitalik Buterin^[63] ont tenté de construire des villes-États indépendantes dans des économies latino-américaines (Próspera, au Honduras) et dans des micronations d'Océanie ravagées par la crise de la COVID-19. Dans leurs temps libres, ils érigent des villes éphémères en Europe de l'Est (Zuzalu, au Monténégro) pour susciter des échanges sur l'avenir glorieux des «individus souverains^[64]», cette élite cognitive libérée de la «tyrannie du lieu». Un autre fantasme de ghetto pour hyper-riches, Praxis, compte Sam Altman parmi ses investisseurs. Son slogan: «Rejoignez-nous dans la ville éternelle» – et ce n'est pas de Rome dont il s'agit.

Leur modèle à tous, c'est la ville-État vénitienne de la Renaissance. Leur inspiration intellectuelle en cette matière, c'est Balaji Srinivasan («Balaji», pour la communauté), entrepreneur *fan* de Bitcoin et théoricien de l'«État-réseau». Dans son essai *The Network State*, autopublié en 2022, le sécessionniste propose de créer un réseau d'«États» numériques décentralisés, qui fonctionneraient uniquement grâce à la technologie *blockchain* et aux cryptomonnaies. Une fois que l'«État» fraîchement créé aurait atteint un certain nombre de «citoyens» et acquis une certaine superficie, affirme Balaji, il suffirait de poser un dôme géodésique dessus et l'entité finirait bien par être reconnue en tant qu'État, ne serait-ce que par les pays voisins. Un hasardeux délire qui rappelle le «Patchwork», l'idéal d'un archipel de micro-pays souverains, théorisé quinze ans plus tôt par nul autre que Curtis Yarvin. (Balaji

appelle aujourd'hui de ses vœux un «sionisme de la tech^[65]», où la communauté des «Gris», loyaux à la Silicon Valley et ornée de tee-shirts à logo de Bitcoin, prendrait le contrôle de San Francisco et purgerait la ville des «Bleus», qui décrivent plus ou moins tous les libéraux. Selon Marc Andreessen, «Balaji possède le taux de bonnes idées originales par minute le plus élevé qu'[il] a jamais vu».)

En parallèle de ces projets de crypto-apartheid, Mark Zuckerberg, Elon Musk et Jeff Bezos se concentrent sur d'autres échappatoires: le virtuel, Mars et l'orbite, leurs spécialités respectives. Le métavers rêvé par Zuckerberg est un pur fantasme de contrôle de la population. Derrière le casque de réalité virtuelle se déploie un monde entièrement privé, individualisé, où chaque avatar numérique est constamment surveillé. En effet, parmi les brevets déposés par Meta sur son casque, on trouve notamment la possibilité d'analyser en temps réel les expressions de visage et le mouvement des yeux afin «d'adapter le contenu» publicitaire à chaque situation.

De leur côté, si Musk et Bezos veulent tellement coloniser Mars et construire des stations orbitales géantes, c'est pour la même raison: dans l'espace, il n'y a ni taxes ni code du travail. Musk voit en la planète rouge «l'occasion de *rebooter* le gouvernement» en instaurant «une démocratie directe» où «les gens voteraient instantanément par scrutin électronique» sur des textes de loi «à taille limitée^[66]». Pour être votée, une loi nécessiterait 60 % des voix; pour être supprimée, 40 % – une incitation structurelle au laisser-faire^[67]. L'aristocratie de la colonie paierait son billet pour Mars entre 500000 et 1 million de dollars, pendant que la plèbe qui désirerait s'y rendre devrait rembourser à SpaceX le coût du billet en travaillant sur place – autrement dit, le régime de l'engagisme, une forme de servage utilisée pour lier les travailleurs agricoles aux plantations coloniales après l'abolition de l'esclavage.

Si Musk s'inspire du roman de SF libertarien de Robert Heinlein *Révolte sur la Lune*^[68], paru en 1965, et à qui on doit la célèbre expression «un repas gratuit, ça n'existe pas!», Bezos s'inspire de la série d'anticipation *The Expanse*^[69], qui chronique les luttes de pouvoir entre une Terre bureaucratique et obsolète, une colonie martienne sous

un régime militaire autoritaire et les colonies minières de la ceinture d'astéroïdes, gérées par des industriels tout-puissants.

Sur Terre, dans le monde réel, les régulations bureaucratiques n'empêchent pas Musk et Bezos d'être deux *micromanagers* notoires, et deux briseurs de syndicats aux comportements toxiques. Les entrepôts Amazon et les usines Tesla sont des espaces bien plus dangereux que la moyenne pour la santé de leurs employés. Chez Amazon, la combinaison d'une cadence infernale et d'une surveillance électronique constante pousse les employés à uriner dans des bouteilles. Ces patrons méprisent nos corps, nos droits, nos existences. Leurs modèles de gouvernance, qu'ils soient sur Mars ou en orbite, ne sont pas et ne seront jamais des démocraties; ce sont des espaces carcéraux et panoptiques, où la division du travail est parfaitement étanche entre le salarié et l'élite managériale.

Dans ces «réalités autocratiques alternatives^[70]» en cours de construction, l'intelligence artificielle serait le contremaître d'un avenir transformé en camp de travail ou en crèche pour la plèbe, qui aurait dans les deux cas perdu toute capacité de décision. Voilà à quoi nous mène le techno-optimisme, une courtepoinette de vieilles utopies nauséabondes resucées à l'ère de l'IA générative: à la fin des sociétés démocratiques et au retour du féodalisme. Mais à toute vitesse.

L'apocalypse selon l'IA

Los Angeles, 2029. Dans un paysage ravagé, le pied de titane d'un endosquelette aux yeux de braise pulvérise un crâne humain. 2100, sous un ciel désespérément orageux, des arthropodes métalliques cultivent un champ de fœtus, des rangées d'embryons humains à perte de vue qui seront transformés en matière première pour un système technologique omnipotent. *Terminator*, *Matrix*: deux œuvres majeures de la culture populaire hollywoodienne de la fin du xx^e siècle, deux visions apocalyptiques qui ont façonné et façonnent encore la représentation de l'intelligence artificielle dans l'imaginaire collectif.

L'intelligence artificielle n'a pas été systématiquement cauchemardesque. Pendant les quinze premières années des années 2000, ces images dystopiques de la science-fiction ont temporairement disparu des écrans. Le discours qui entourait la «révolution» technologique du «web 2.0», celui des Facebook, Twitter, Uber et Airbnb, était plutôt optimiste – les réseaux sociaux et leurs algorithmes s'apprêtaient à changer le monde en imposant la démocratie et la prospérité, et les géants de la Silicon Valley posaient en alliés de l'humanité et de la justice. La critique, lorsqu'elle se faisait entendre, était douce. En 2013, Spike Jonze réalisait *Her*, une dystopie *soft* dans laquelle une intelligence artificielle démolissait la société humaine en rendant chaque individu béatement dépendant d'elle. *Black Mirror* proposait majoritairement des dystopies à visage humain, où la machine n'était rien de plus qu'un artefact de nos dysfonctionnements collectifs. On aurait pu croire que les robots tueurs de Skynet étaient morts et enterrés.

Mais avec le retour de l'IA sur le devant de la scène médiatique, une musique catastrophiste, composée et amplifiée par la Silicon Valley, se fait à nouveau entendre dans une ambiance anxiogène. En mai 2023, l'influente couverture du magazine *Time* s'interrogeait en majuscules et sans aucun trac sur «LA FIN DE L'HUMANITÉ?^[1]». Quelques mois

plus tard, Chuck Schumer, chef des démocrates au Sénat des États-Unis, a ouvert une séance en déclarant qu'il craignait un «scénario apocalyptique» dans lequel une IA deviendrait si puissante «que nous en ferions un dieu informatique^[2]». Si l'on en croit les grands patrons de la Silicon Valley et les chercheurs vedettes de l'industrie, nous vivons un moment charnière de l'histoire humaine, où la moindre erreur de trajectoire dans le développement des intelligences artificielles nous mènera inévitablement à l'extinction. Pour mesurer l'éventualité d'une telle issue, des *nerds* millénaristes ont inventé le «*p(doom)*», un score censé indiquer la probabilité d'une IA-pocalypse. Et ce qui a commencé comme un exercice d'initiés est désormais cité par des journalistes du *New York Times*^[3], décontenancés par cet «indice lugubre».

Lugubre, l'indice l'est plus ou moins selon la personne qui le fixe: le chercheur catastrophé Geoffrey Hinton annonce un *p(doom)* à 10 %. Son collègue Yoshua Bengio est à 20 %. Dario Amodei, PDG d'Anthropic, est à *p(doom)* = 10-25 %. Elon Musk, vétéran de la prophétie de malheur, est à *p(doom)* = 20-30 %. Chez les ingénieurs en intelligence artificielle, selon un sondage datant d'octobre 2023, la moyenne du *p(doom)* se situe à 40 %. Au terminus de l'angoisse, Eliezer Yudkowsky affiche un *p(doom)* = «oui». Des sites recensent ce type de données et proposent d'estimer soi-même la probabilité que la super-IA puisse provoquer l'extinction de l'espèce humaine, une angoisse apocalyptique générée par les utilisateurs, en quelque sorte. Ce ne sont plus les illuminés millénaristes qui règlent les aiguilles de l'horloge de la fin du monde, mais bien les puissants visionnaires de la Silicon Valley.

Une fois de plus, il n'y a rien de nouveau sous le soleil de la modernité. Spécialiste des récits d'extinction, Tyler Austin Harper estime que les poussées de pessimisme collectif sont cycliques, et que celle que nous traversons arrive «pile à l'heure^[4]». La dernière fois que l'Occident s'est à ce point inquiété pour son avenir, c'était dans les années 1920, la décennie favorite de Peter Thiel. Le monde sortait alors d'une pandémie et d'une guerre, s'apprêtait à vivre la terrible récession économique de 1929. Entre les deux, l'accélération du progrès technique semblait inéluctable et monstrueuse. En 1920, soixante ans avant *Terminator*, la pièce de théâtre *Rossumovi univerzální roboti* (Les robots universels de

Rossum) du Tchécoslovaque Karel Čapek introduisait le mot tchèque «*robot*» et explorait le scénario d'une révolte des machines intelligentes. Enfin, le spectre de l'automatisation commençait à planer sur le secteur ouvrier et propageait une peur alimentée par un patronat soucieux d'augmenter ses marges. Cette convergence entre mutations sociales, crises multiples, peur de l'extinction (qui, à l'époque, s'accompagnait du souci de pureté de la race blanche) et fascination pour l'accélération technique s'est révélée être un terreau idéal pour l'avènement du modernisme réactionnaire^[5], fasciste et nazi.

L'angoisse qui accompagne la «polycrise» économique, environnementale, technique et politique aujourd'hui n'a donc rien d'inédit, mais force est d'admettre que nous vivons un moment étrange, où des grands patrons de la tech répètent à l'envi que leurs inventions risquent de tous nous tuer. Dès 2015, Sam Altman affirmait sur son blogue que «l'intelligence machine surhumaine est probablement la plus grande menace à l'existence humaine^[6]». C'est pour cette raison, expliquait-il, qu'il a fondé OpenAI l'année suivante, pour nous sauver en développant lui-même une machine imaginaire qui égalerait les compétences intellectuelles humaines. L'enfumage fonctionne: sur les réseaux sociaux, d'aucuns prétendent que le sac à dos bleu que Sam Altman transporte partout avec lui – son «*nuclear backpack*» – contient une sorte de système d'arrêt d'urgence des serveurs d'OpenAI, au cas où les choses tourneraient mal.

La Silicon Valley fourmille de grands patrons autoproclamés altruistes qui s'activent prétendument pour sauver l'humanité. Elon Musk, grand philanthrope devant l'éternel, joue les Cassandre de l'IA depuis près de dix ans. Après avoir cofondé OpenAI, il affirme que sa nouvelle *start-up*, xAI, développe une intelligence artificielle «pour mieux éviter un futur à la *Terminator*^[7]». Anthropic, dont le fondateur Dario Amodei a d'abord travaillé chez OpenAI avant de claquer la porte, ne recrute que des angoissés de la superintelligence. Au point de devenir «l'épicentre chauffé à blanc de l'IApocalypse», comme il est écrit dans un article du *New York Times* publié à l'été 2023^[8].

Ce qui inquiète l'ineffable Elon Musk et les patrons des trois grandes entreprises d'IA actuelles, Sam Altman, Demis Hassabis et Dario

Amodei, c'est le «risque existentiel» – ou «*x-risk*» pour les initiés. Aujourd'hui, utiliser ce terme, c'est afficher son appartenance à une communauté, à une vision de l'IA, du futur et du monde. Pratiquement absent des médias et de la Silicon Valley avant la pandémie, le risque existentiel est devenu «le même le plus populaire de l'IA^[9]», selon le *MIT Technology Review* en 2023. Comment en est-on arrivé à une telle consécration? L'obsession pour la catastrophe, la peur irrationnelle et l'IA-anxiété qui traversent la société technique et civile sont le résultat du lobbying patient des intellectuels et patrons dont il est question dans ce livre, et dont le dessein est de contrôler les débats sur la technique puis d'empocher les profits qu'ils généreront.

En vérité, ce qui empêche vraiment ces gens de dormir, ce n'est pas l'imminence de l'apocalypse, mais la peur d'être éjectés du sommet de la pyramide sociale. Dans une société en pleine transition sociale, politique et économique, l'angoisse de l'extinction est propre à la caste dominante. Elle est réactionnaire au sens littéral du terme: c'est la réaction viscérale d'une élite que la perspective du changement excite autant qu'elle terrifie, parce qu'elle voit bien qu'elle le provoque autant qu'elle le subit. La seule fin du monde que les membres de cette élite craignent, c'est la fin de leur monde. Dans leur course effrénée pour la consolidation de leurs privilèges, les vies de ceux qui ne font pas partie de la clique n'ont aucune valeur. Nous ne figurons pas plus dans leurs dystopies que dans leurs utopies techno-optimistes. De fait, dans sa forme actuelle, la fin du monde est essentiellement une entreprise de négation du présent. Conçue et distribuée par les producteurs de la technique, l'extinction permet de détourner l'attention des critiques. Elle est une tonitruante invitation à renoncer à toute emprise politique, sociale, symbolique, aussi bien dire humaine, sur le développement de la technoscience. Il s'agit bien, en cela, d'une prophétie autoréalisatrice qui cache une course au profit.

De la catastrophe naturelle à l'apocalypse artificielle

Avant d'être réduite à l'œuvre d'une intelligence artificielle maléfique et à la conséquence de la fin du capitalisme, la notion d'extinction humaine a traversé plusieurs époques. Elle a d'abord été l'apanage des religions,

des prophéties mayas à la Bible, mais comme l'explique Émile P. Torres dans *Human Extinction: A History of the Science and Ethics of Annihilation*, la plupart des discours religieux considèrent l'humanité comme étant indestructible, et l'Apocalypse chrétienne, par exemple, n'est en fait qu'une transformation de l'espèce^[10].

Ce n'est que dans la société industrielle partiellement sécularisée de l'Occident du XIX^e siècle, après la découverte du deuxième principe de la thermodynamique qui postule l'augmentation de l'entropie globale, qu'émergent les premières réflexions sur la fin de l'humanité. Puisque l'Univers tout entier tend vers le désordre, l'humanité a une date d'expiration; l'entropie implique que notre extinction est inévitable. Voilà comment la physique fondamentale détruit le mythe chrétien de l'invulnérabilité de l'espèce.

Un second choc ontologique a lieu le 6 août 1945. Après l'explosion de la première bombe atomique sur Hiroshima, l'humanité se rend compte qu'elle possède désormais les moyens de s'autodétruire. La vision collective de l'apocalypse est brutalement mise à jour – on ne cauchemarde plus seulement de fin de l'Univers, de catastrophes naturelles ou de punition divine, mais de guerre atomique, de bombes H et de champignons nucléaires. Plus les techniques progressent, plus les catastrophes artificielles prennent le pas sur les catastrophes naturelles ou les punitions divines, et la modernité s'avère être une menace pour l'espèce. En 1972, dans son célèbre rapport intitulé *Les limites à la croissance*, le Club de Rome parle de «menaces existentielles» qui pèsent sur la planète. La civilisation industrielle menace l'ensemble du vivant.

En l'an 2000, le magazine *Wired* publie «Why the Future Doesn't Need Us^[11]», un article qui devient culte dans les cercles de la cyberculture. Son auteur, Bill Joy, chercheur en informatique chez Sun Microsystems, y affirme que les progrès en nanotechnologies, en génétique, en robotique et en informatique vont faire des humains une «espèce menacée». S'inspirant de conversations avec des ingénieurs en robotique, l'auteur prophétise non seulement un soulèvement des machines, mais également un scénario d'une rare honnêteté dans lequel les ultrariches dominent le reste de la population grâce à des robots tout-

puissants. La méchante IA est encore très loin de faire les gros titres, mais l'idée que le XXI^e siècle sera celui de l'autodestruction humaine fait son chemin.

À la même époque, un des leaders intellectuels du transhumanisme – l'incarnation parfaite de la dualité du millénarisme technoreligieux – théorise la notion contemporaine de risque existentiel, et deviendra, en vingt ans, le prophète préféré des catastrophistes. Cette figure, c'est celle de Nick Bostrom. Le philosophe suédois, déjà obsédé par la mort, soutient sa thèse de doctorat en 2002, articulée autour d'une réflexion sur la probabilité d'une extinction de l'humanité, évidemment. La même année, il publie l'article «Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards» dans le *Journal of Evolution and Technology*, périodique qu'il vient de fonder dans le but de faire avancer les idées transhumanistes dans le monde universitaire. Cet article de 30 pages fondera tout un domaine de recherche.

Comme Bostrom l'explique dans le préambule d'un autre texte sur le sujet: «Traditionnellement, le futur de l'humanité a été un thème théologique. Toutes les grandes religions ont des enseignements au sujet du destin ultime de l'espèce ou de la fin du monde. Les thèmes eschatologiques ont également été explorés par les grands philosophes que sont Hegel, Kant et Marx. Récemment, c'est le genre littéraire de la science-fiction qui a repris le flambeau^[12].» Le philosophe veut sortir de ces cadres et imposer un «futurisme "réaliste"» qui parlerait le langage des élites politiques et industrielles. Il veut pouvoir parler d'apocalypse, de paradis et d'enfer, mais en termes scientifiques. Et pour cela, il lui faut une nouvelle définition de la catastrophe.

Le «risque existentiel», dans les mots du penseur transhumaniste, est «une situation dans laquelle une issue malencontreuse pourrait soit anéantir la vie intelligente d'origine terrestre, soit réduire de manière permanente et drastique son potentiel^[13]». Contrairement à la catastrophe naturelle, le *x-risk* ne menace donc que l'espèce humaine et, sans la mener nécessairement à l'annihilation, il peut simplement ralentir ce que Bostrom considère comme sa seule trajectoire désirable: la quête transhumaniste du posthumain, la multiplication compulsive des «meilleurs» d'entre nous, la colonisation de la Voie lactée. Le risque

existentiel, c'est tout ce qui nuit à la réalisation de *ce futur-là*. (Dans ce sens, on peut supposer qu'une prise du pouvoir par des écologistes radicaux passerait pour un risque existentiel aux yeux du Suédois.)

Dans cette vision du monde, le passé et le présent sont des variables négligeables. Aucune des catastrophes ayant ponctué l'histoire des sociétés humaines – guerres mondiales, épidémies, tremblements de terre, accidents nucléaires – n'est suffisamment grave pour constituer un risque existentiel, car aucune d'entre elles n'a pu enrayer la glorieuse marche du progrès technique. Le dérèglement climatique, qui menace l'entière du vivant, ne représente pas non plus un risque existentiel pour Bostrom. C'est un événement catastrophique d'ampleur planétaire, certes, mais rien d'insurmontable pour ses alliés milliardaires de la tech, déjà en train de décorer leurs bunkers en Nouvelle-Zélande et de se préparer à une raréfaction des ressources. Tant que ces êtres supérieurs seront épargnés, tant qu'ils continueront de financer l'utopie transhumaniste, tout ce qui arrive au commun des mortels est une contingence sans grand intérêt. Comme on l'a déjà vu, pour Bostrom, toutes les vies ne se valent pas.

Autrement dit, ce qui préoccupe le philosophe, c'est la possibilité d'un ralentissement de la croissance économique et l'avènement d'un monde dans lequel les ressources seraient allouées à autre chose qu'à la poursuite effrénée de la techno-utopie (à la redistribution équitable des richesses, à l'éradication de la pauvreté et à la lutte contre les discriminations, par exemple). Dans la *Weltanschauung* de Bostrom, toute société qui n'est pas entièrement dévouée à la cause transhumaniste présente donc un «risque existentiel». Une forme de chantage, dont le but est d'imposer des délires de colonisation de l'espace et d'immortalité postbiologique comme seul futur désirable pour l'espèce entière.

En 2013, dans la revue *Global Policy*^[14], Bostrom relie encore plus explicitement fin du capitalisme et fin du monde: tout événement qui nous empêcherait d'atteindre notre «maturité technologique», définie comme l'optimisation «des capacités qui permettent un niveau de productivité économique et de contrôle sur la nature proche du maximum atteignable», constitue désormais un risque existentiel. Exaspérant l'axiome populaire qui dit qu'il est plus facile d'envisager la fin du

monde que la fin du capitalisme, Bostrom amalgame les deux: si le capitalisme est menacé, la Silicon Valley est menacée et son futur est menacé, donc l'humanité entière est menacée.

Son concept trouve un écho auprès de ses pairs et engendre un nouveau domaine de recherche universitaire, les *existential risk studies*. En 2005, Bostrom crée le Future of Humanity Institute à Oxford. Dans les années 2010, Cambridge, Stanford et Berkeley suivent avec, respectivement, le Center for the Study of Existential Risk, le Stanford Existential Risks Initiative et le Berkeley Existential Risk Initiative, formant tout un «écosystème du *x-risk* – un réseau de *think tanks*, d'associations à but non lucratif, de chercheurs, de philanthropes et d'autres, tous concentrés sur la réduction du risque existentiel^[15]». Petit à petit, les nouveaux théoriciens du risque se désintéressent des catastrophes naturelles et se concentrent presque exclusivement sur «des formes avancées de biotechnologies, de nanotechnologies moléculaires et de machines intelligentes développées dans les décennies qui viennent».

Sous prétexte d'éviter la fin des temps, Bostrom et ses disciples poursuivent tranquillement leur lobbying transhumaniste. Avec succès: à l'été 2023, le «risque existentiel» est mentionné par le Secrétaire général des Nations Unies, António Guterres. Dans son dernier rapport, en 2023, le Bureau des Nations Unies pour la réduction des risques de catastrophe affirme que le développement de l'IA est une des menaces qui pèsent sur nos sociétés^[16]. En juillet 2023, la lettre ouverte du *think tank* Safe AI alerte sur l'urgence de «réduire les risques d'extinction liés à l'IA»; l'expression sera reprise en septembre par la Commission européenne^[17]. Signe des temps, même le Club de Rome a revisité son propos. Cinquante ans après son avertissement contre l'idéologie mortifère de la croissance économique, le *think tank* rêve désormais d'utiliser «l'IA pour combattre le changement climatique, et transformer ces deux risques existentiels en présages d'espoir existentiels^[18]». Neutralisation politique accomplie: le technocapitalisme n'est plus le poison mais le remède.

Le totalitarisme contre l'extinction

La «révolution de l'apprentissage machine» des années 2010 a suscité l'émergence de nouveaux logiciels, alimentés par d'immenses bases de données et des processeurs à la puissance de calcul inédite. Une nouvelle génération de machines, comme Alpha Go et Watson, a appris à imiter le langage humain, à reconnaître des objets et des visages, à jouer à Pong, au jeu de go, au quiz télévisé *Jeopardy!* et à gagner de plus en plus souvent. De nouveaux termes ont alors fait leur apparition: la Silicon Valley s'est mise à vanter les mérites des *algorithmes d'apprentissage machine* et des *réseaux de neurones*. Face à ces nouvelles capacités de classification et de prédiction, plusieurs personnalités de la science et de la tech ont adopté les discours catastrophistes, lesquels se sont progressivement consolidés jusqu'à aujourd'hui.

En 2014, au sommet de sa popularité médiatique, Elon Musk décrivait l'IA comme «la plus grande menace existentielle» et déclarait que «l'intelligence artificielle, c'est comme invoquer un démon^[19]». En juillet 2015, avec le PDG de DeepMind, Demis Hassabis, le grand physicien Stephen Hawking et le cofondateur d'Apple, Steve Wozniak, ainsi qu'un millier de chercheurs, Musk a signé une lettre ouverte contre le déploiement des «robots tueurs», des armes augmentées par l'intelligence artificielle que l'on soupçonnait d'être en cours de développement. À l'époque, on ne parlait pas encore d'AGI ou de superintelligence, mais Musk et Hawking agitaient déjà le chiffon rouge d'une «menace existentielle». Pour Hawking, nous nous trouvons alors «au moment le plus dangereux de l'histoire du développement de l'humanité». L'IApocalypse avait trouvé ses ambassadeurs et, petit à petit, l'anxiogène récit transhumaniste de la fin des temps s'est propagé dans la société civile.

Nick Bostrom, pour sa part, s'est associé à un autre prophète, Eliezer Yudkowsky, considéré comme le plus pessimiste de tous. Élevé dans la religion juive orthodoxe, Yudkowsky, autodidacte et génie autoproclamé, s'était rallié à la cause transhumaniste dans les années 2000 avant de retourner sa veste brutalement, vers 2010. Aujourd'hui, devenu apostat

de l'IA et gourou d'une petite communauté en ligne, les Rationalistes, il est persuadé depuis dix ans que l'IApocalypse est au coin de la rue.

En 2014, au moment où Musk et Hawking hurlaient aux robots tueurs, Bostrom et Yudkowsky publiaient ensemble un article qui posait la question de la possibilité d'une IA qui ne se retournerait pas contre les humains^[20]. Le *timing* était parfait. Les deux prophètes, qui avaient sonné le tocsin chacun dans son coin, ont soudainement réussi à attirer l'attention. Pour une partie de la Silicon Valley, Bostrom, dont le manifeste *Superintelligence* est sorti exactement au moment où l'IA devenait une angoisse mondiale, est apparu comme celui qui avait vu l'avenir – un authentique prophète qui, selon ses propres mots, s'entraînait au «CrossFit métaphorique depuis trente ans» pendant que le reste de l'espèce humaine était «affalé sur le canapé à manger du popcorn au lieu de réfléchir à l'alignement, à l'éthique et à la gouvernance d'une potentielle superintelligence^[21]». Il devenait tout à coup urgent de s'intéresser à l'IApocalypse.

Au cours des années suivantes, on a assisté à un véritable mariage gris entre Silicon Valley, capitalisme spéculatif et recherche universitaire. Des laboratoires de recherche en IA d'un nouveau genre, peuplés de philosophes et d'éthiciens, ont fleuri sur les campus des plus prestigieuses universités anglo-saxonnes, financés par de généreuses donations de l'industrie^[22]. Leur mission: tenter de prévoir et de prévenir une destruction de l'humanité par l'intelligence artificielle, et imaginer des solutions pour éviter le carnage. Cette entreprise de relations publiques prend le nom d'*AI safety*. Aujourd'hui, presque chaque université de l'élite anglo-saxonne possède son *think tank* dédié à l'IApocalypse, souvent installé au même endroit, voire intégré au centre d'études local du «risque existentiel», tant les deux concepts (venus du même homme, Nick Bostrom) sont indissociables. Aujourd'hui, l'*AI safety* a aussi son lobby européen, l'European Network for AI Safety, et sa coalition internationale, l'AI Alliance, un *think tank* créé fin 2023 qui réunit près de 80 organisations du monde universitaire et de l'entreprise privée. Le Royaume-Uni est devenu le premier pays au monde à fonder une institution publique dédiée à la question et dont la mission est de «minimiser l'effet de surprise pour le Royaume-Uni et l'humanité face

aux avancées rapides et inattendues de l'IA^[23]», autrement dit, de plancher sur les menaces intangibles d'une machine imaginaire. Non seulement l'IApocalypse est prise au sérieux par des centaines d'autorités nationales et internationales, mais on écoute désormais ses oracles pour bâtir la future coopération internationale.

Parmi les centaines de scénarios catastrophes imaginés par les prophètes de l'IA et leurs disciples, le «maximisateur de trombones», imaginé par Nick Bostrom en 2003, a eu le plus d'écho. Imaginez: dans un futur indéfini, une entreprise de fabrication de trombones s'offre les services d'une IA aux capacités équivalentes à celles de l'être humain, et lui ordonne de maximiser la production. Grâce aux suggestions de l'IA, la cadence augmente, et l'entreprise est satisfaite. Mais la machine continue. Elle produit une nouvelle génération de machines, qui à leur tour génèrent des machines, toujours plus intelligentes et efficaces. Le complexe industriel s'étend, jusqu'à menacer d'envahir la planète. L'entreprise tente de mettre fin au programme, mais celui-ci s'est cloné il y a longtemps pour parer à cette éventualité. Les réserves de fer de la planète sont à sec. Alors, les robots, qui n'ont plus de fil métallique à tordre, se tournent vers les humains constitués d'une majorité d'atomes de carbone, qu'il suffirait de configurer différemment pour en faire des trombones de graphite...

Il faut une nouvelle fois reconnaître les talents de vulgarisateur de Bostrom. Aujourd'hui, le maximisateur de trombones est à l'*AI safety* ce que le chat de Schrödinger est à la physique quantique: un symbole pop que l'on retrouve sur des affiches, tatoué ou porté en bijou. Elon Musk l'a usé à la corde dans les médias, Bostrom le cite dans ses conférences TED, la presse spécialisée le recycle dès qu'il est question des périls de l'IA. À l'automne 2022, quelques semaines avant le dévoilement de ChatGPT, OpenAI a reçu un carton plein de trombones en forme de leur logo, envoyé par un employé de la *start-up* rivale Anthropic. Le message était clair: vous avez pris le parti de l'apocalypse.

Avec le maximisateur de trombones, Bostrom veut nous alerter sur la «convergence instrumentale», l'idée que toute machine intelligente cherchera nécessairement à se préserver, à s'améliorer et à croître, peu importe la tâche initiale qui lui est donnée. Quitte à recouvrir toute la

surface du globe, en quelques jours ou semaines, de panneaux solaires, de réacteurs nucléaires, d'aides aux superordinateurs comme d'immenses tours de refroidissement, de lanceurs de fusées spatiales ou de toutes les autres installations par lesquelles l'IA maximiserait la réalisation à long terme de ses valeurs, écrit-il dans *Superintelligence*. Derrière le masque innocent d'un logiciel d'optimisation se cacherait en réalité un «shoggoth», une monstruosité informatique capable de tout détruire sur son passage.

Tout à son imagination, Nick Bostrom rate l'évidence: l'entité qui se rapproche le plus du maximisateur de trombones, c'est la multinationale capitaliste. Les entreprises sont des organismes artificiels, alimentés par des humains interchangeable et animés par des processus informatiques. Elles ont des objectifs parfaitement définis: se préserver, s'optimiser, croître, aux dépens du vivant. Leur objectif instrumental est de générer du profit, qui leur est aussi vital pour survivre que l'air pour un mammifère. L'entreprise moderne est cannibale; elle avale d'autres entreprises concurrentes pour s'alimenter. Musk dirige Tesla, un maximisateur de batteries sur roues; SpaceX, un maximisateur de charge orbitale; et Twitter/X, un maximisateur d'attention. Peu importe leurs buts respectifs, ces trois entités sont soumises au même principe de «convergence instrumentale» et tendent à enrichir Musk et ses actionnaires, quitte à démolir la planète.

L'IA est une métaphore du système politique et économique capitaliste qui menace l'espèce humaine, mais ses prophètes ne se soucient guère de cette autre forme d'autodestruction, tant s'en faut. Leurs visions sont une intuition sur le devenir d'un monde sacrifié sur l'autel du productivisme, exprimée par des entrepreneurs absolument incapables de se remettre en question. Malgré leur volonté permanente de puissance et de contrôle, ceux-ci pressentent l'ampleur de la catastrophe dont ils sont les premiers responsables, mais peu leur en chaut. Ce monstre autoréférentiel et froid, mû par une avidité extractiviste, c'est la machine capitaliste en phase terminale d'assujettissement du vivant. Le pire cauchemar des prophètes de l'IA, c'est d'en perdre le contrôle et que sombrent leurs entreprises, leur puissance, leur avenir.

Alors très vite, la stratégie de lutte contre l'IApocalypse prend une tournure autoritaire.

Illustration: en mars 2023, alors que le battage médiatique et la panique générés par ChatGPT sont à leur paroxysme, le magazine *Time* publie un texte hallucinant. Le *doomer* en chef Eliezer Yudkowsky y apparaît en pleine crise existentielle, persuadé que la fin des temps est arrivée: «Si quelqu'un construit une IA trop puissante dans les conditions actuelles, je m'attends à ce que chacun des membres de l'espèce humaine ainsi que toute la vie biologique sur Terre meure peu de temps après^[24].» Le temps de la négociation avec les faiseurs d'IA est révolu, selon lui. S'inspirant des dispositifs de lutte contre la prolifération des armes nucléaires et biologiques, il propose un moratoire à durée indéterminée sur la recherche en IA, la fermeture de tous les laboratoires, et jusqu'à la surveillance de toutes les GPU vendues dans le monde, comme on le fait avec les armes biologiques et nucléaires.

Mais Yudkowsky nous sert sa proposition la plus folle pour le dessert: il faut, dit-il, être prêts à bombarder tout centre de données clandestin ou illégal, parce que «il est plus important d'éviter un scénario d'extinction à cause de l'IA que d'éviter un conflit nucléaire total». Rappelons que dans l'esprit de ces gens, un conflit thermonucléaire global ne mérite pas le titre de «risque existentiel», pourvu qu'eux et leurs semblables survivent. Pour protéger son futur, Yudkowsky est prêt à sacrifier celui de tout le monde.

Si cette crise de panique digne de *Docteur Folamour* a de quoi inquiéter (heureusement que Yudkowsky n'est pas en poste dans un quelconque gouvernement), la solution proposée par son confrère Nick Bostrom pour «stabiliser la civilisation» est autrement plus terrifiante. En 2004, il affirmait que si l'humanité veut un futur (comprendre: un futur transhumaniste), elle doit prendre très littéralement le contrôle de son évolution. Entre deux stratégies eugénistes, le Suédois défend donc la création d'un «*singleton*», un ordre mondial dans lequel il n'y a qu'un seul pouvoir décisionnel indépendant au plus haut niveau^[25], une entité unique qui gouvernerait la planète et ses habitants. Il précise néanmoins – et on l'en remercie – que pour qu'un gouvernement mondial soit considéré comme un *singleton*, il faudrait qu'il n'y ait aucune colonie

spatiale indépendante ni civilisation extraterrestre à portée de la Terre^[26]. Ce *singleton* pourrait être démocratique ou dictatorial, humain ou machine, peu importe: le plus important, selon le philosophe, c'est qu'il soit transparent pour nous... et que nous soyons transparents pour lui.

La société idéale selon Bostrom, un mélange de *1984* et de *Minority Report*, repose donc sur une «police préventive extrêmement efficace, soutenue par une surveillance de masse», histoire que les États puissent «observer leurs citoyens de suffisamment près pour intercepter quiconque se prépare à un acte de destruction massive^[27]». Les citoyens porteraient un collier de surveillance baptisé, sans ironie, le «*freedom tag*». Il serait bien entendu interdit d'enlever ce collier de la liberté, «serti de caméras et de micros multidirectionnels», qui capterait tout et enverrait en temps réel les données (chiffrées, précise un Bostrom soucieux de la vie privée) vers les serveurs du gouvernement. Au moindre doute, une alerte serait envoyée à la «station de surveillance patriote» la plus proche. Un «agent des libertés» vérifierait le flux de données et déciderait, au choix, de contacter le porteur du collier pour lui demander des explications, d'envoyer une patrouille humaine ou de dépêcher un drone. Le rêve, n'est-ce pas?

Le plus dérangeant, c'est la méticulosité avec laquelle Bostrom a planché sur son concept de panoptique *high-tech*, comme s'il tentait réellement de convaincre les institutions internationales de l'adopter. Entre autres détails, il estime que «la population mondiale peut être surveillée en continu pour moins de 1 % du PIB mondial^[28]», soit pour 140 dollars par individu et par an. Une affaire. Quant au danger pour les libertés fondamentales, il l'évacue d'un bon vieux sophisme: «Même si un projet de surveillance bien intentionné (sic) s'accompagne inévitablement d'un risque significatif de totalitarisme, on ne peut pas en conclure que mener un tel projet augmenterait le risque de totalitarisme^[29].»

Radicalisés dans leurs idéologies respectives, les deux gourous, Bostrom et Yudkowsky, nous montrent jusqu'où ils sont prêts à aller pour faire advenir leur version du futur: jusqu'à la violence d'État, jusqu'au technofascisme, jusqu'à l'opération militaire, jusqu'à la

surveillance globale. Voilà le projet qui se dissimule derrière les grands concepts de risque existentiel, d'*AI safety* et de gouvernance de l'IA.

Bostrom conseille grandes entreprises et gouvernements, et il murmure à l'oreille de la Silicon Valley. Comme dans une mauvaise version de *Don't Look Up*, Yudkowsky annonce en conférence TED que tout le monde va bientôt mourir, provoquant l'hilarité du public. Et ces idées délirantes font du chemin, en parfaite adéquation avec les rêves de capitalisme totalitaire des techno-oligarques. Sam Altman, qui a récemment déclaré que Yudkowsky avait fait «plus que quiconque pour accélérer l'AGI, et qu'il est possible qu'il reçoive un jour le prix Nobel de la paix^[30]», admettait en 2019 dans un texte publié par OpenAI que «stopper [l'émergence de l'AGI] nécessiterait quelque chose comme un système de surveillance planétaire^[31]». Techno-utopie et IAocalypse se rejoignent et se complètent; les deux scénarios préfèrent se projeter dans un futur imaginaire plutôt que d'analyser les menaces qui pèsent sur le présent, et affirment que le techno-autoritarisme est la meilleure manière de sécuriser l'avenir.

À la recherche des Tables de l'IA

En dix ans, l'étude de l'IAocalypse a cessé d'être un passe-temps intellectuel pour transhumanistes pour devenir un récit politique grand public. Dans la Cerebral Valley, ce quartier de San Francisco aux maisons victoriennes reconverties en *hacker houses* par les *start-up* et *think tanks* de l'IA, des légions de gens inquiets, baptisés *safetyists*, décélérationnistes ou *doomers* selon le degré de panique, travaillent à temps plein (et bruyamment) à imaginer des solutions à des problèmes pour la plupart imaginaires^[32].

Pour d'autres, dans la Silicon Valley et ailleurs, la peur de l'IA est une matière première qui génère des financements universitaires, des contrats, des investissements. Dans son analyse du marché de l'*AI panic*, la chercheuse Nirit Weiss-Blatt, autrice de *The Techlash and Tech Crisis Communication*, sépare le «*panic-as-a-business*» (Bostrom, Yudkosky et autres prophètes à plein temps, comme William MacAskill et Toby Ord), qui consiste à vendre de la peur contre des financements, et l'«*AI panic*

marketing» (Sam Altman, Dario Amodei, Demis Hassabis, Emad Mostaque, tout le patronat de l'IA), où la panique sert de discours publicitaire pour vendre des logiciels.

Le résultat est une formidable industrie de la diversion. Le récit apocalyptique permet non seulement de travestir la cupidité des patrons de la tech en un courage homérique dans la lutte contre l'extinction humaine, mais il détourne aussi notre regard critique des contradictions du présent pour le river sur un futur plein de «risques» absurdes, où un ordinateur de la taille d'une planète pourrait un jour transformer les gens en trombones. Il ne reste plus qu'à justifier le fait que la régulation d'une technologie déjà vendue comme toute-puissante soit entre les mains de ceux qui la fabriquent. C'est ici qu'entre en jeu la dernière lubie des prophètes: l'«alignement» de l'IA.

Ce récit publicitaire s'inspire, une fois encore, de la science-fiction: en 1942, Isaac Asimov, fatigué de lire encore et toujours des histoires de SF hantées par ce qu'il nommera le «complexe de Frankenstein» – la peur obsessionnelle d'un golem se retournant contre son créateur – imagine la parade dans sa nouvelle «Cercle vicieux»: les «trois lois de la robotique», des instructions encodées dans tous les robots «positroniques» de ses romans. La première leur interdit de faire du mal à un être humain, la seconde les oblige à obéir aux ordres des humains, sauf s'ils les forcent à contrevenir à la première loi, et la troisième les oblige à protéger leur existence dans la mesure où elle respecte les deux premières lois. L'expérience de pensée d'Asimov inspire Norbert Wiener, l'inventeur de la cybernétique, qui postule en 1960 que «si, pour atteindre nos objectifs, nous utilisons des agents mécaniques avec lesquels nous ne pouvons pas interférer [...] nous avons intérêt à être certains que l'objectif programmé dans la machine est bien celui que nous désirons». Le «problème de l'alignement», la calibration morale d'une machine dans nos propres intérêts, est alors posé. Et déjà, personne ne voit de paradoxe à vouloir programmer l'autonomie de la conscience morale. Chez les chercheurs de l'époque, la tentation de la simulation est trop forte.

Pour éviter l'extinction, nous disent les prophètes de l'alignement aujourd'hui, il faut enseigner à la machine une «moralité artificielle^[33]».

Celle-ci doit être simple, formulée sans ambiguïté et commune à toute l'espèce humaine, indépendamment des spécificités culturelles, géographiques, ethniques, politiques et religieuses. Ce qu'ils imaginent, dans leurs laboratoires d'éthique de l'IA, c'est un ensemble de valeurs partagées par toute l'humanité, pour modeler le golem informatique à notre image et s'octroyer sa puissance.

Voilà la mission sacrée à laquelle s'attellent les prophètes de l'alignement: la réduction de l'humanité tout entière à un programme informatique qu'ils prétendent être les seuls à pouvoir concevoir. Dans leurs bureaux d'Oxford et leurs *open spaces* de San Francisco, ils reprennent la quête inachevée des premiers démiurges de l'IA, Marvin Minsky et John McCarthy: la recherche d'une formule mathématique unique qui formalise les individus. Ce «code source» de l'humanité serait ensuite donné en offrande à l'impitoyable Dieu-machine, pour apaiser Son courroux. Dans un retournement inédit des récits bibliques, c'est l'être humain, perché sur sa montagne, qui dicterait les Tables de la Loi à son dieu.

L'*hubris* et l'absurdité que recèle cette ambition sont absolument phénoménales. L'élite financière et scientifique qui s'arroge le monopole du futur s'impose au passage comme référence culturelle et morale, et comme porte-parole de l'humanité. Ces gens semblent oublier que sur une planète où 80 % de la population dit croire en Dieu, la rédaction d'un ensemble de valeurs «communes» à l'humanité ressemblerait plus au Coran, à la Torah et la Bible qu'aux utopies transhumanistes sécularisées des *techies* californiens.

Or, comme l'ont montré les travaux indispensables de Robert Geraci^[34], le terreau religieux influence la recherche en IA. Aux États-Unis, le récit transhumaniste hérite du christianisme, où la transcendance passe par l'abandon de l'enveloppe charnelle. Les prophètes de l'IA états-uniens conçoivent donc l'individu comme un cerveau dans un bocal de formol, déconnecté de toute enveloppe charnelle, et présentent l'intelligence artificielle comme une entité décorporée. Une intelligence libérée de son corps, c'est une entité plus proche que nous du divin. C'est donc un but à atteindre, comme le montrent les rêves transhumanistes de numérisation du cerveau.

Malgré d'évidentes impasses conceptuelles, l'alignement est la nouvelle tendance. En 2023, après moins de dix ans d'activité, les centres de recherche en *AI safety* de Berkeley, Stanford ou Oxford emploient des «chercheurs en alignement» au sein de leurs équipes de «chercheurs en risque existentiel». En 2021 est né le premier Alignement Research Center, une structure dédiée uniquement au problème de l'alignement.

Dans l'industrie, toute *start-up* d'intelligence artificielle qui se respecte expose sa vision de l'alignement, ainsi que la méthode qu'elle compte employer pour inculquer les «valeurs humaines» à ses IA. Et tant pis si personne, au fond, ne sait réellement ce que ça signifie – le plus important, aux yeux des régulateurs et des investisseurs, c'est de montrer qu'on prend l'idée au sérieux, qu'on y travaille, que la recherche avance.

Superalignement et super-*bullshit*

Entre le jour où ces lignes sont écrites et le jour où elles seront imprimées, le champ lexical de l'intelligence artificielle aura probablement à nouveau changé. Les grands modèles de langage de 2022 sont en train de devenir des «modèles de fondation». Les IA génératives deviennent des «modèles frontaliers». La mue s'est opérée en juillet 2023: un article publié sur le site d'OpenAI et cosigné par toute l'industrie^[35] anticipe «des modèles de fondation aux capacités suffisamment importantes pour poser de graves risques de sécurité», une forme de prophétie autoréalisatrice qui implique que la prochaine génération de logiciels, qui n'existe pas encore, s'approchera un peu plus de la superintelligence. Et donc de la fin des temps. Une manière comme une autre d'entretenir un sentiment d'urgence et de sidération dans la société civile, tout en convainquant les investisseurs que leur argent est utilisé à bon escient.

En lisant ces pages, on peut se sentir, à juste titre, inondé de concepts. Ces dernières années, le discours catastrophiste de l'IA n'a pas arrêté de changer de forme et de vocabulaire. Mais ce renouvellement permanent du lexique n'est pas le signe d'une accélération du progrès technique. Ce n'est que le symptôme d'un terrible manque de substance.

La vérité est beaucoup plus prosaïque que ce que peut laisser croire cette inflation verbale. Les «modèles frontaliers» qui représentent un «risque existentiel», les stratégies d'«alignement», le «maximisateur de trombones», la «superintelligence», le «superalignement», aucun de tous ces concepts à guillemets n'existe hors de l'univers fantastique de la foi et du marketing. L'IA n'est vraiment pas la fin du monde. L'IApocalypse n'est pas plus imminente que les utopies techno-optimistes décrites au chapitre précédent. Les chercheurs qui signent les lettres ouvertes contre l'IApocalypse, comme celle de mars 2023, n'y croient même pas^[36]. Tout ça n'est que de la super-*bullshit*, ce qui ne signifie pas pour autant qu'un tel discours soit inoffensif.

Le récit catastrophiste repose sur une série de préconceptions présentées comme des certitudes. Pour commencer, dans le monde biologique, l'intelligence n'est pas corrélée à la volonté de domination, contrairement à ce que laissent entendre ces vulgaires évolutionnistes^[37]; elle est plutôt un processus d'interaction sociale, qui implique la coopération et une forme de mutualité^[38]. L'idée d'une superintelligence superprédatrice est une vision inspirée du capitalisme, qui est basé sur la compétition entre tous, portée par des dominants soucieux de garder leur place au sommet de la chaîne alimentaire.

Les scénarios catastrophes mobilisent un imaginaire dans lequel nous sous-traitons nos responsabilités aux machines, nous sacrifions nos ressources énergétiques pour leur développement débridé. C'est la vision des techno-oligarques, assis sur une montagne d'inégalités, angoissés de perdre la maîtrise de leur création et de leur destinée. Or si, comme on le disait, l'IA n'est qu'une allégorie de la machine inhumaine et omnipotente qu'est le capitalisme mondialisé, qu'est-ce que la recherche en alignement sinon la quête d'une impossible formule magique qui réconcilierait le profit et l'éthique? Sur quel axiome se base-t-elle, sinon l'intuition que le capitalisme débridé risque bel et bien d'anéantir la civilisation humaine?

Aujourd'hui, les couches successives de lobbying de l'IA se sont sédimentées. À force de brouiller leurs frontières, de s'échanger des chercheurs et de partager les mêmes bureaux, les trois sous-cultures du risque existentiel, de l'*AI safety* et de l'alignement forment une même

«communauté épistémique^[39]», soutenue par «des pratiques de production de connaissance et de *community building* qui se renforcent mutuellement».

Les transhumanistes de Nick Bostrom, les altruistes efficaces de William MacAskill et les rationalistes d'Eliezer Yudkowsky se fondent dans une même vision, déjà explorée dans le chapitre précédent: grâce aux mathématiques et à leur intellect supérieur, les ingénieurs et les philosophes vont non seulement sauver le monde de l'apocalypse qu'ils ont été les seuls à voir venir, mais ils vont aussi «résoudre» la fusion entre humanité et machine. Et tout ça, c'est du boulot. Ou, pour reprendre les mots d'Eliezer Yudkowsky en février 2023, «convoquer un démon, c'est facile, mais il est beaucoup plus difficile de convoquer un ange^[40]». Gloire à l'élite technique, intellectuelle et financière!

En dix ans à peine, avec l'aide d'une poignée de milliardaires de la tech et de philanthrocapitalistes, les transhumanistes et les altruistes efficaces ont inventé, diffusé et financé trois disciplines savantes pour trouver des solutions à des problèmes qui n'existent pas. Cette communauté, marginale en nombre, mais extrêmement influente, est parvenue à convaincre des décideurs politiques, des chefs d'entreprises, des institutions internationales et une partie du public que le problème le plus pressant du monde actuel est une prophétie de science-fiction dont personne ne parlait jusqu'en 2014: aligner une intelligence artificielle générale sur les «valeurs humaines» pour éviter qu'elle ne nous extermine.

Rien de tout ça n'existe; ni les démons ni les anges informatiques, ni le paradis ni l'enfer transhumanistes, ni l'utopie de posthumains immortels éparpillés dans la galaxie, ni la dystopie d'une superintelligence vengeresse. Il n'y a qu'une fabrique de la peur et de la sidération, une machine à lobbying tout entière mobilisée pour éviter la régulation de l'industrie.

Du bon usage des prophéties

Jusqu'ici, dans nos aventures au Pays des merveilles technoreligieuses, nous avons croisé une série d'oligarques de la Big Tech à cheval sur leur bombe à retardement chromée, exaltés, les yeux rivés vers une radieuse fin de l'humanité, prêts à établir leurs fiefs privés sur Terre, sur Mars, dans l'espace, partout où l'être humain mis à jour pourrait échapper aux misères de l'existence. Nous avons ensuite plongé dans les gouffres de l'Apocalypse, d'un futur cauchemardé par ces mêmes seigneurs du posthumain, où le golem de la «superintelligence» aurait fait dérailler leurs utopies – un risque existentiel pour ces milliardaires autoritaires qui angoissent surtout de voir le corps social se retourner contre eux.

De retour sur la terre ferme du réel, on réalise que tout ce battage dissimule une formidable opération d'*astroturfing*: sous les superlatifs, les néologismes, les cris d'alarme et les prophéties s'orchestre une classique mais vaste entreprise de soumission des corps politiques, scientifiques et médiatiques aux vellétés autoritaires des grandes fortunes de la Silicon Valley. Ce qu'ils cherchent avec leurs milliards de dollars, leurs discours apocalyptiques, leurs *think tanks* et leurs utopies réversibles, c'est bien à reconfigurer le présent pendant que nous regardons vers un avenir improbable. Il leur faut donc rester au-dessus des lois, coûte que coûte. Engoncés dans leurs costumes de lobbyistes, nos futuristes à la petite semaine n'ont rien inventé. Ils ne font que réutiliser l'arsenal des techniques de manipulation, de corruption légale et de retournement d'élus, testé et approuvé par Big Oil, Big Tobacco et Big Tech. Science, pouvoir, argent: la Sainte Trinité du capitalisme avancé.

Derrière un élitisme mystique tout à fait sincère, les grandes envolées théologiques des patrons de la Big Tech cachent également un objectif basement prosaïque: acheminer autant d'argent que possible vers leurs

entreprises, consolider leur monopole et orienter la régulation afin que leurs activités ne soient pas soumises aux fins communes qu'elle est censée protéger. En vérité, ces prétendus prophètes en croisade pour sauver l'humanité ne sont rien d'autre que des cambrioleurs de nations.

L'industrie de l'IA est aujourd'hui le théâtre d'une double lutte de pouvoir. La première est une guerre intestine, une guerre de gangs, en somme. D'un côté, la Big Tech californienne – nommément, Google, Microsoft et Amazon –, qui fournit le capital et les infrastructures à la jeune garde de l'IA générative et pousse pour un développement commercial des logiciels. De l'autre, les philanthrocapitalistes de l'altruisme efficace prêchent l'IApocalypse auprès des faiseurs d'IA pour les convaincre de lever le pied. Aux uns les campus universitaires et les laboratoires de recherche, aux autres les conseils d'administration, la chaîne d'approvisionnement matériel et les outils d'évaluation des performances des systèmes. Un schisme religieux entre «accélérationnistes» et «décélérationnistes», deux émanations d'une même église technocapitaliste.

Ce qui se joue au-delà de cette guéguerre intestine, c'est une offensive de lobbying des deux gangs de techno-oligarques, qui savent mettre leurs différences de côté quand il s'agit de cibler la société civile. À force de persuasion et d'influence, les prophètes de l'IA ont imposé les termes du débat, neutralisé tout discours critique et fait adopter leur eschatologie délirante à une partie de la presse, aux institutions nationales et internationales et, fatalement, au grand public. Avec leurs récits fallacieux, ils ont réussi à infléchir les trois principaux chantiers de régulation de l'IA en Occident, aux États-Unis, au Royaume-Uni et dans l'Union européenne. L'argent de la Silicon Valley a ruisselé dans les interstices des débats parlementaires, imprégné les projets de loi et neutralisé les contre-pouvoirs démocratiques qui tentaient d'enrayer la mécanique de concentration de l'oligopole. Les cartels de l'IA et leur projet de civilisation sont sortis vainqueurs de cette manche.

Lutte fratricide dans la Silicon Valley

À l'heure où des milliards de dollars sont injectés dans la recherche en IA, deux factions de technomessies s'affrontent. D'un côté, les patrons de la Big Tech et certains capital-risqueurs influents, alléchés par des profits potentiellement astronomiques. De l'autre, la nébuleuse altruiste efficace, financée par d'autres milliardaires de la Silicon Valley, qui réclame un développement contrôlé et centralisé de l'IA. Prophètes contre prophètes, tous d'accord pour dire qu'il faut avancer vers l'inéluctable futur technocapitaliste, mais en désaccord sur le meilleur chemin à prendre.

Dans un contexte de «*burnout* de l'IA» généralisé après une année 2023 dominée par les pessimistes, c'est désormais le «techno-optimisme» de la Big Tech qui tient le haut du pavé, mais l'influence des EA reste conséquente. Conçu pour déguiser la prédation capitaliste en action de bienfaisance, le mouvement a muté en quelques années pour se consacrer entièrement au lobbying de l'IApocalypse et devenir, selon Émile P. Torres, «le système de croyance séculaire le plus dangereux de la planète^[1]». Son réseau d'influence aux États-Unis et au Royaume-Uni est vaste et peut compter, selon la chercheuse Nirit Weiss-Blatt, sur pas moins de 27 organismes de financement différents^[2]. En 2021, les EA se vantaient d'avoir obtenu des promesses de dons totalisant 46 milliards de dollars, soit l'équivalent du budget annuel de l'assurance-chômage française^[3]. La même année, ils prétendaient récolter en moyenne 420 millions de dollars de dons chaque année^[4], un chiffre qui a nettement augmenté depuis. Jamais les Cassandre de l'altruisme efficace n'avaient reçu autant d'argent pour imposer leurs apocalypses dans le débat public.

D'où provient ce tsunami d'argent? Au-delà de quelques gros donateurs comme Jaan Tallinn, connu pour avoir participé à la création de Skype, ou le cryptomilliardaire Vitalik Buterin, qui vient de donner 665 millions de dollars au Future of Life Institute, l'une des figures de proue du mouvement, le véritable trésor des EA est gardé par la fondation Open Philanthropy, qui gère une partie de la fortune personnelle du cofondateur de Facebook, Dustin Moskovitz. Selon *The Economist*, Open Phil a fourni plus de la moitié des 600 millions de

dollars versés au mouvement en 2021 et 2022^[5]. En juillet 2023, le *Washington Post* sortait lui aussi la calculatrice et estimait que l'organisation avait dépensé «près d'un demi-milliard de dollars pour développer une réponse à l'IA malfaisante^[6]». Elle a notamment donné des millions de dollars aux plus puissantes universités du monde anglo-saxon pour s'assurer qu'en échange, elles incluent dans les corpus de leurs programmes d'études en intelligence artificielle les récits mystiques du *x-risk*, les apocalypses invraisemblables de l'*AI safety* et les solutions imaginaires de l'«alignement».

Une fois diplômés, les apprentis prophètes biberonnés à ces balivernes peuvent rejoindre, au choix, les *think tanks* de l'altruisme efficace sur les mêmes campus où ils ont fait leurs études financées par Open Phil ou s'engager dans les toutes nouvelles équipes d'*AI safety* formées par OpenAI, DeepMind et Anthropic. Car les nouvelles *start-up* de l'IA sont sous l'influence interne du lobby catastrophiste, même si elles sont aussi tributaires des investissements technooptimistes de la Big Tech. Et c'est là que ça se complique. Les dissonances idéologiques n'en finissent plus. DeepMind appartient au conglomérat Alphabet, mais son patron Demis Hassabis est un sympathisant autoproclamé des EA. Anthropic, l'«anti-OpenAI», a reçu deux milliards de dollars de Google fin 2023, puis quatre milliards d'Amazon en mars 2024, mais elle n'en est pas moins un véritable repaire de *doomers*. Le PDG, Dario Amodei, se déclare ouvertement millénariste. Les employés utilisent quotidiennement le lexique des EA, s'échangent leurs mêmes et portent fièrement des vêtements marqués des logos du mouvement. On travaille chez Anthropic pour sauver l'humanité de l'IApocalypse, mais on utilise le matériel d'Amazon et Google.

La trajectoire d'OpenAI incarne mieux que toute autre cette guerre de chapelles. Sa naissance date de l'été 2015. Rosewood Hotel, Menlo Park, Californie: dans un luxueux hôtel situé à un jet de pierre du campus de Stanford, Elon Musk expose au prodige de l'apprentissage machine, Ilya Sutskever, son plan pour sauver l'humanité. C'est Sam Altman qui a organisé la rencontre. La mission: «libérer» la recherche en IA des méchants oligopoles du secteur privé (alors incarnés par Google) et publier ses recherches en *open source* – d'où le nom, OpenAI – pour

déjouer l'Apocalypse. Neuf ans plus tard, OpenAI, dont la valeur en Bourse est évaluée à 86 milliards de dollars, vend ses logiciels sous licence commerciale, compte 500 employés et prévoyait 1,3 milliard de dollars de revenus en 2023 (4543 % d'augmentation par rapport à l'année 2022). Pire: Microsoft, l'archétype de la vieille multinationale monopolistique, possède désormais 49 % de l'entreprise. À l'image de la compagnie qu'il dirige, Sam Altman a cessé de poser en bienfaiteur de l'humanité pour montrer son banal visage de patron démiurge de la Silicon Valley, et ses lobbyistes draguent désormais le Pentagone pour obtenir ses juteux contrats^[7].

OpenAI serait aujourd'hui une entreprise schizophrène, divisée entre des «tribus» – dicit Sam Altman dans un e-mail interne de 2019^[8] – «accélérationniste» et «décélérationniste», respectivement incarnées par Altman et Sutskever. En novembre 2023, le schisme s'est approfondi, avec le licenciement abrupt de Sam Altman. Cinq jours plus tard, coup de théâtre: il réintérait son poste. Sur les quatre votants du directoire ayant précipité sa chute très temporaire, trois étaient liés à l'altruisme efficace. Leur tentative de putsch a fini en eau de boudin. De fait, début 2024, on constate que les EA ont perdu de leur superbe. Leurs récits catastrophistes commencent à agacer les investisseurs. La condamnation à vingt-cinq ans de prison pour fraude financière du cryptomilliardaire Sam Bankman-Fried, figure importante des EA, a révélé la faillite morale et l'hypocrisie du mouvement. Autre coup dur, en avril 2024: le Future of Humanity Institute de Nick Bostrom a été contraint par Oxford de mettre la clé sous la porte après dix-neuf ans de lobbying catastrophiste^[9].

Malgré des désaccords bien réels sur les modalités de développement des systèmes, catastrophistes et optimistes ne forment qu'un seul et même camp. Leurs ambitions visent le même objectif: la domination des sociétés par les géants de la Silicon Valley et la construction de régimes autoritaires sur mesure pour leur rêve technoféodal. Dans cette perspective, lorsqu'il s'agit de convaincre la société civile et les élites politiques, la caste technicienne sait très bien mettre ses différends de côté.

Mainmise du capital sur la technoscience

Au-delà des luttes internes, le duopole formé par les géants de la tech et les organisations liées aux EA a désormais la mainmise sur presque toute la chaîne de production de l'IA. La Big Tech contrôle la recherche universitaire et le développement, la commercialisation et l'amélioration des systèmes existants; les EA, eux, commandent le récit médiatique et le réseau de l'*AI safety*.

En 2024, l'IA est un négoce assez simple, mais qui demande des ressources faramineuses. Il est régi par des lois d'échelle (*scaling laws*), qui postulent que plus un modèle grossit, plus il est compétent. Chaque nouvelle génération de logiciel implique de construire des ordinateurs beaucoup plus puissants, de mettre la main sur beaucoup plus de données et de stocker le tout dans d'immenses hangars de plus en plus réfrigérés.

La puissance des processeurs d'un modèle d'IA est exprimée en pétaflops (Pflops), une unité de mesure qui correspond à un million de milliards d'opérations par seconde. Entre 2010 et 2024, nous sommes passés d'une puissance d'un Pflop à 50 milliards de Pflops^[10]. Hier, les logiciels d'apprentissage machine s'entraînaient sur des bases de données limitées; aujourd'hui, ils gobent un catalogue de cinq milliards d'images, le LAION-5B, et l'intégralité des textes en accès public sur internet. Et les coûts sont à l'avenant: selon l'*AI Index* de l'université Stanford, en 2023, le développement de la puissance des processeurs de GPT-4 a coûté 78 millions de dollars, et Google a déboursé près de 191 millions de dollars pour entraîner son dernier modèle, Gemini Ultra^[11]. Voilà pourquoi Sam Altman a déclaré que GPT-5 n'est pas pour tout de suite, et que la route vers l'AGI devra passer par le perfectionnement de la fission nucléaire, voire par la fusion nucléaire – la technologie actuelle est si énergivore qu'il faudra une nouvelle source d'énergie pour assurer son avenir.

Dans le monde réel, seuls Google, Microsoft et Amazon ont les moyens financiers et l'expertise pour construire, acheter, exploiter et entretenir le matériel requis pour l'IA. Un an après l'introduction de ChatGPT, le marché a déjà des airs de monopole. Début 2024, le «Big

Three» – Google Cloud, Microsoft Azure et Amazon Web Services – contrôlait les deux tiers des centres de données de la planète. Ces entreprises contrôlent également l’approvisionnement en cartes graphiques Nvidia H100, spécialement conçues pour des logiciels comme GPT. La concurrence se retrouve face à un marché de pénurie, où ces pièces essentielles coûtent près de 40000 dollars l’unité, et sont «beaucoup plus difficiles à trouver que de la drogue», dicit Elon Musk^[12]. Mais existe-t-il encore une concurrence? Entre 2010 et 2023, les GAFAM ont racheté 98 *start-up* d’IA^[13]. Personne n’est à l’abri de l’oligopole: Inflection AI, qui avait pourtant levé 1,3 milliard de dollars en juin 2023, a été absorbée par Microsoft au printemps 2024.

La concentration du marché est telle que même la Federal Trade Commission (Commission fédérale du commerce, FTC) s’est réveillée pour manifester son inquiétude face à de possibles «pratiques anticoncurrentielles^[14]». On peut savourer l’euphémisme, mais préférons-lui la réalité des faits: sans les milliards et l’infrastructure de la Big Tech, Big AI n’existerait pas, les «modèles de frontière^[15]» aux capacités surpuissantes (et évidemment terrifiantes) ne seraient jamais réalisés et la quête mystique de l’AGI s’arrêterait net. Les milliards de dollars d’investissements de la Big Tech se composent majoritairement de «crédits» d’utilisation de services de *cloud*, accompagnés de clauses d’exclusivité. L’industrie de l’IA, en Occident, se réduit à sept entreprises, surnommées les «Sept Magnifiques» (Google, Microsoft, Amazon, Meta, Nvidia, Apple, Tesla), et quelques fonds d’investissement milliardaires, comme ceux de Peter Thiel et Marc Andreessen. Comme le déclarait le PDG de Microsoft, Satya Nadella, dans une boutade à propos d’OpenAI (dont on rappelle que Microsoft détient 49 % des parts): «Nous sommes en dessous d’eux, au-dessus d’eux, et tout autour d’eux^[16].» Ils *sont* l’intelligence artificielle.

La mainmise du privé est totale. Les grandes entreprises de la tech contrôlent les infrastructures, les outils de développement, les langages de programmation et les logiciels qui définissent la direction des recherches en IA. Yoshua Bengio, professeur à l’Université de Montréal et autre «parrain» de l’apprentissage machine, ne peut que constater: «À l’heure actuelle, le pouvoir, l’expertise et les données sont tous entre les

mains d'une poignée d'entreprises^[17].» (Bengio sait de quoi il parle: il a également fondé Mila, le centre québécois de recherche en IA, financé entre autres par Facebook, Microsoft, Google et Open Philanthropy.) Environ 70 % des doctorants en intelligence artificielle se précipitent dans le privé, une fois diplômés, contre 20 % au début des années 2000. Depuis 2006, aux États-Unis, le nombre de chercheurs universitaires dans ce domaine est resté à peu près le même, tandis qu'il a été multiplié par huit dans l'industrie. L'université forme les cerveaux et la Silicon Valley les récolte. Logique: les logiciels du privé sont en moyenne 29 fois plus puissants que ceux des universités, et représentent 96 % des meilleurs logiciels du marché. Pire encore, 91 % des logiciels d'évaluation (les *benchmarks*) sont développés par l'industrie, qui possède donc le monopole de sa propre analyse. Toute résistance est vaine.

Les universités, déjà rompues à la logique néolibérale de privatisation, multiplient les «partenariats» avec les géants de la tech, histoire de sauver les meubles. En 2023, ces mariages gris ont produit 21 modèles d'IA «notables», selon l'*AI Index* de Stanford^[18]. (Dans le même temps, la Silicon Valley en a produit 51 et les universités seules en ont développé 15.) Ces arrangements produisent des chercheurs hybrides, grassement payés par le secteur privé où ils travaillent à temps partiel, tout en gardant leurs fonctions et leurs responsabilités universitaires. Ils publient sous *imprimatur* universitaire, mais signent parfois le fameux NDA (*non-disclosure agreement*), cette entente de confidentialité concoctée par le secteur privé pour donner priorité au secret industriel sur l'avancée de la science.

En décembre 2023, le *Washington Post* révélait que sur 33 professeurs d'université ayant publié des articles sur l'IA dans les revues *Nature* et *Science*, tous sauf un avaient reçu de l'argent des géants de la Silicon Valley, ponctuellement ou en tant qu'employés^[19]. Selon le chercheur Christopher Newfield: «La majeure partie de la recherche en IA est menée sans liberté universitaire, ce qui risque de forcer les chercheurs à modifier les résultats, notamment en supprimant les découvertes qui entrent en contradiction avec le récit ou l'image de l'entreprise qui les paie. Or, la modification des résultats de recherche par le secteur privé a

des précédents, notamment avec des exemples offerts par les industries de l'automobile, du tabac, de la chimie et des carburants fossiles^[20].»

En 2020, Google a créé un comité de lecture des sujets de recherche jugés «sensibles», qui incluent notamment l'intelligence artificielle^[21]. Dans le cas d'un article évalué par le comité en question, celui-ci a demandé à l'auteur de «supprimer toutes les références à des produits Google». Dans un autre cas, il a rappelé aux auteurs d'«utiliser un ton positif» pour évoquer les logiciels de l'entreprise. En 2021, Google a licencié ses deux éthiciennes vedettes, Timnit Gebru et Margaret Mitchell, qui avaient osé alerter publiquement sur les biais algorithmiques des grands modèles de langage de l'entreprise.

Comme Big Oil et Big Tobacco avant elle, Big Tech n'hésite pas à modifier les résultats de la recherche, voire à confondre discours scientifique et récit publicitaire. L'un des privilèges du pouvoir, disait Michel Foucault, est celui de créer un champ de savoir sur mesure, un savoir dont la seule fonction est de légitimer le rapport de force du dominant au point de le naturaliser. C'est exactement ce que le cartel de la Silicon Valley tente de faire.

La fabrique de la sidération

Il ne reste plus qu'à fabriquer du consentement à grande échelle. Pour neutraliser tout potentiel de critique collective de l'IA, les prophètes instillent un mélange d'anxiété, de fascination et de sidération aux journalistes censés fournir au reste de la population les cadres d'analyse du phénomène, qui tombent les uns après les autres dans le simulacre de la machine pensante. L'épidémie de panique n'épargne personne: le journaliste tech du *New York Times*, Kevin Roose, ressortira «profondément perturbé, voire terrifié, par les capacités émergentes» du robot conversationnel Bing un soir de février 2023^[22]. L'intelligence artificielle est pourtant couverte depuis près de soixante ans par la presse généraliste aux États-Unis – et plutôt positivement, selon plusieurs études, même si les risques sont décrits plus en détail que les bénéfices^[23]. Pour imposer leur apocalypse réversible, les prophètes se sont donc démenés.

Ces dernières années, la couverture médiatique de l'IA s'est intensifiée, jusqu'à atteindre des sommets d'absurdité. Depuis que nous sommes entrés dans l'ère ChatGPT, le récit médiatique de l'intelligence artificielle est un va-et-vient permanent entre deux scénarios de science-fiction spectaculaires – l'IA va sauver le monde et l'IA va détruire le monde – qui détourne efficacement l'attention des effets concrets de la technique sur le corps social et ne véhicule en réalité qu'un seul message: soyez tranquilles, la techno-élite s'occupe de tout. Pour les PDG et les investisseurs du secteur, l'excitation est une manière de capter des investissements. Pour les *think tanks*, chercheurs, philosophes et prophètes à plein temps, la panique eschatologique est un fonds de commerce.

Entretenir ce battage publicitaire, c'est donc la spécialité des EA. Selon Émile P. Torres^[24], qui a fait partie de la secte avant d'en devenir sa plus ardente critique, William MacAskill, l'affable philosophe d'Oxford, a bâti le mouvement comme l'aurait fait un publicitaire. L'altruisme efficace est une «marque», et comme pour toute marque qu'on lance sur le marché, il a fallu tenir compte des moindres détails du produit – à commencer par le nom, choisi parmi une liste d'autres options, pour son côté accrocheur. Comme dans une compagnie lambda, la communauté des EA a accès à un manuel de communications, disponible en ligne et intitulé «Advice for Responding to Journalists» (Conseils pour répondre aux journalistes)^[25].

La fabrique de l'IApocalypse requiert un travail minutieux. Grâce à la chercheuse Nirit Weiss-Blatt, on sait que le *think tank* Campaign for AI Safety, à qui l'on doit la fameuse lettre ouverte proposant un moratoire sur les recherches en IA, a eu recours aux techniques du marketing classique (sondages, panel-test, test A/B et analytique des données massives) pour optimiser son message technomessianique^[26]. Après un sondage mené auprès de 1 500 personnes à qui ont été soumises 37 formulations différentes pour décrire l'IApocalypse, le *think tank* a conclu qu'il valait mieux éviter le lexique «spirituel judéo-chrétien» et proscrire l'expression «IA divine», par exemple, pour privilégier les termes «superintelligence» et «IA dangereuse», plus à même de berner – pardon, de convaincre – le grand public.

Le bombardement médiatique annihile toute possibilité d'un pluralisme des opinions. Selon les auteurs de l'étude québécoise «Entraîner l'actualité: la couverture canadienne du cycle d'engouement pour l'IA (2012-2021)», menée dans le cadre de l'étude internationale *Shaping 21st-Century AI*, «les articles de médias généralistes sur l'IA reflètent étroitement les intérêts des affaires et du gouvernement^[27]». La presse spécialisée invite systématiquement les mêmes experts-entrepreneurs, tous issus de la Silicon Valley ou liés financièrement à elle. Les médias accordent, de manière générale, une place démesurée à des prophètes comme Mark Zuckerberg ou Elon Musk, dont l'essentiel de l'activité consiste à entretenir un climat d'excitation générale, à grand renfort de démonstrations spectaculaires (et le plus souvent fausses ou exagérées). Les dimensions politique et sociale de la technique ne sont presque jamais abordées, poursuivent les chercheurs, et «les spécialistes des sciences sociales brillent par leur absence». Parmi leurs recommandations aux journalistes, les auteurs de l'étude conseillent littéralement d'«éviter de traiter l'IA comme une prophétie» et d'arrêter de relayer les fadaises marketing et les discours exaltés d'individus qui ont tout à gagner à entretenir un délire collectif.

Car les déclarations exaltées ont une fonction et des conséquences. L'étude recommande notamment que les «attentes et les promesses technoscientifiques [soient] considérées pour ce qu'elles sont: des déclarations performatives qui n'ont pas encore été réalisées» et les auteurs du rapport déplorent que «lorsque les acteurs formulent des promesses technologiques, ils les présentent naturellement comme inéluctables et indiscutables^[28]». Couvrir des pro-messes comme s'il s'agissait de faits réels renforce le mythe du déterminisme technologique dans l'esprit du public, ce qui renforce à son tour l'aura de puissance de la Big Tech. Lorsque la presse relaie l'idée que l'IA est un horizon inévitable, alors qu'elle n'est que la condensation d'une série de rapports de pouvoir, elle naturalise la domination des multinationales états-uniennes et la rend du même coup inattaquable. Plus grave encore, «ces représentations ne sont pas de simples descriptions; elles suscitent l'attention et créent de l'anticipation, de la peur et de l'excitation; [...] elles mobilisent des ressources et structurent les programmes de

recherche». En choisissant de ne pas remettre en question ce que raconte l'industrie de l'IA, la presse nous oriente par défaut vers le futur technoscientifique le plus avantageux pour son monopole.

En France, le Media Lab de SciencePo a également participé à l'étude critique *Shaping 21st-Century AI*. Il conclut que sur la période 2011-2021, la majeure partie des articles anticipe les bénéfices à venir de l'IA plutôt que d'analyser ses dégâts actuels. La presse fait de la mauvaise science-fiction au service de la Silicon Valley. Les auteurs de l'étude affirment qu'en France, «les cadrages de l'IA et ses enjeux [...] sont largement contrôlés par les géants de la technologie et les acteurs médiatiques. [L]'influence de ces récits est considérable: ils ont le pouvoir d'accorder ou non une existence à des entités et à des arrangements particuliers^[29]». La spéculation n'est pas de l'information. En aucun cas le journalisme tech ne peut être un exercice de futurologie; il se doit d'être une analyse critique, constante et indispensable, de l'actualité de la technique. Il doit ouvrir un espace démocratique de médiation politique du présent, où la pluralité des discours est censée désamorcer les effets incapacitants du récit marketing. En choisissant d'adopter sans réserve le cadre imposé par l'industrie, en se vautrant dans un journalisme au conditionnel et en relayant les mensonges d'une caste technopatronale en pleine bouffée délirante, la presse désinforme: elle donne de la substance à une IA imaginaire, tout en occultant l'impact réel de cette technologie sur le monde.

En 2023, la fantasmagorie s'est imposée dans l'opinion publique. Les prétendus débats se sont articulés autour de questions insensées. La machine est-elle consciente? Peut-elle avoir des sentiments? Va-t-elle nous remplacer? L'IA va-t-elle provoquer la fin du monde? Quant à la société, on l'a systématiquement montrée comme étant dépourvue d'autonomie politique, comme une masse d'individus apathiques prêts à être gouvernés par une machine divine. Cette vision, c'est le rêve techno-optimiste des milliardaires de la Silicon Valley. Un fantasme de contrôle total.

Si une partie de la presse semble enfin s'être calmée et avoir troqué les emportements de 2023 contre un scepticisme à la fois plus factuel et plus constructif, les effets sur l'opinion publique d'une année entière de

délire médiatique sont désastreux. Sondage après sondage, on constate que les opinions se polarisent de plus en plus: en France, les deux tiers des sondés ont rejoint le camp des exaltés ou des pessimistes, avec un large avantage pour ces derniers^[30]. Mais l'effet le plus pernicieux de cette campagne d'influence, c'est que la majorité des gens est désormais certaine que l'IA prendra un jour le contrôle de nos vies^[31]. Or, cette IA affolante, emballante, totalitaire, inévitable n'existe pas. C'est un cheval de Troie dans lequel sont tapis les hommes-enfants narcissiques et irresponsables du cartel de la Silicon Valley, qui attendent le moment idéal pour sortir et s'emparer des clés de la cité.

À force de se faire dire que le futur est déjà plié, que c'est la Silicon Valley qui décide de l'avenir de l'humanité, le public, convaincu qu'il n'a pas son mot à dire sur des enjeux qui le dépassent, remet son destin entre les mains des Google, Microsoft, Meta ou Amazon. La fabrique de la sidération a fonctionné: pendant que la société civile tente de recouvrer ses esprits, la Silicon Valley peut maintenant saboter les institutions.

L'AI governance, l'intoxication des institutions

Le contrôle des récits universitaire et médiatique par la sidération permet d'atteindre un autre objectif: dicter aux instances décisionnelles un cadre de régulation sur mesure, voire créer de toutes pièces des semblants de dispositifs de régulation pour les mettre au service du cartel de l'IA. Cette stratégie porte un nom, la capture réglementaire (*regulatory capture*), et elle n'a rien de nouveau: après les industries fossiles, l'aviation, la finance, le nucléaire et l'informatique, c'est au tour de Big AI de s'y mettre.

Plutôt que de parler de capture réglementaire, les prophètes préfèrent évoquer l'*AI governance* (la gouvernance de l'IA). Comme l'*AI safety*, l'*AI governance* est un entrelacs d'instituts, de lobbies et de *think tanks*^[32]. La France y est particulièrement bien représentée, avec des structures comme France Stratégie, Hub France IA et pas moins de deux ministères, celui de l'Économie et de l'Enseignement supérieur, de la recherche et de l'innovation, «particulièrement impliqués dans la

politique de l'IA^[33]». Le message adressé aux législateurs par la nébuleuse de l'*AI governance* se résume plus ou moins à ceci: respectez l'innovation, laissez le présent tranquille, concentrez-vous sur les catastrophes hypothétiques de demain. Un message particulièrement savoureux lorsqu'on voit comment l'industrie se démène, ici et maintenant, pour convaincre le reste du monde que seul le futur compte.

Il n'y a jamais eu autant d'argent dépensé en lobbying pour l'IA qu'en 2023. Aux États-Unis, 450 entreprises ont payé des lobbyistes pour faire pression sur le gouvernement fédéral, contre 158 l'année précédente, ce qui représente une dépense de près d'un milliard de dollars, soit deux fois plus qu'en 2022^[34]. À Bruxelles, la Big Tech est en train de «washingtoniser» l'Union européenne. La Silicon Valley y a dépensé 113 millions d'euros de lobbying en un an^[35] – dix fois plus que le secteur automobile. Du jamais vu sur cette rive de l'Atlantique^[36].

L'argent des EA pleut sur la classe politique états-unienne. En octobre 2023, *Politico* révélait qu'Open Philanthropy a placé une cohorte de «collaborateurs» aux ministères de la Défense et de l'Intérieur, au département d'État, mais également au comité Science, espace et technologie de la Chambre des représentants et au comité Commerce, science et transports du Sénat^[37]. Entièrement dédié à l'IApocalypse, le fonds de Dustin Moskovitz a notamment servi à payer des lobbyistes de l'IA à Washington.

L'American Association for the Advancement of Science, financée par Microsoft, OpenAI, Google, IBM et Nvidia, contrôle six «collaborateurs d'urgence» placés à des postes stratégiques pour les débats parlementaires sur l'IA, qui viennent s'ajouter aux 270 lobbyistes en science et technologie qui quadrillent déjà l'administration^[38]. D'autres sont payés par Eric Schmidt, ex-PDG de Google, et une centaine d'autres sont payés par l'organisation TechCongress. Voilà à quoi ressemble la «gouvernance de l'IA»: une armée de petits soldats téléguidés par la Silicon Valley, infiltrés à tous les étages de la fabrication des lois. Et c'est aussi légal que c'est amoral.

Avec une telle puissance de frappe et un tel poids politique et géopolitique, les prophètes de la Silicon Valley contrôlent déjà, à divers degrés, les politiques de régulation de l'IA aux États-Unis, au Royaume-

Uni et dans l'Union européenne. Aux États-Unis, où le gouvernement fédéral n'a pas voté de loi régulatrice dans le secteur de la tech depuis vingt-cinq ans, la démonstration de force a été particulièrement payante. Google, OpenAI, Microsoft et Anthropic se sont d'abord regroupés sous la bannière du Frontier Model Forum^[39], une organisation professionnelle créée pour montrer patte blanche à l'administration Biden^[40], qui s'est en retour contentée d'exiger des fabricants d'IA une vague promesse sur l'honneur que leurs systèmes privilégieront le bien collectif et la sécurité plutôt que la rentabilité et le pouvoir (on y croit!) ^[41]. Joe Biden, qui a visionné le dernier *Mission: Impossible* pour savoir à quoi pourrait ressembler l'IApocalypse, a ensuite promulgué un décret aussi large que confus dont le texte reprend des pans entiers du marketing d'OpenAI et Google^[42].

Parmi les mesures phares, le texte légifère sur des modèles d'IA qui n'existent pas: il impose une évaluation à partir de 1026 opérations par seconde, une puissance de calcul qu'aucun des logiciels actuels n'atteint. De plus, la mission de contrôle hypothétique a été confiée au National Institute of Standards and Communication (NIST), une instance née en 1901 et qui emploie une vingtaine de personnes, entassées dans des locaux qui prennent l'eau et où la bande passante est trop faible pour l'envoi de fichiers lourds. Comme si ça ne suffisait pas, le budget annuel du NIST vient d'être amputé de 10 %^[43]. Heureusement, Amazon s'est empressé d'offrir gratuitement ses services de stockage de données, et Microsoft fournit du matériel informatique et des ingénieurs. La gangrène du lobbying ne s'arrête pas là. Le futur gendarme de Big AI s'est enrichi d'un organisme créé sur mesure, l'US AI Safety Institute Consortium, qui réunit 200 membres parmi lesquels on retrouve... tous les noms de Big AI et un certain nombre de laboratoires d'*AI safety* financés par les EA. Son président, Paul Christiano, un ancien employé d'OpenAI, a été financé par Anthropic et collabore avec plusieurs organisations du réseau des EA^[44]. Prophète de malheur radical, ancien associé d'Eliezer Yudkowsky, il pense que l'humanité n'a qu'une chance sur deux de ne pas être détruite par une superintelligence^[45]. Non content d'adhérer à un mouvement technoreligieux quasi sectaire, Christiano est enraciné dans l'industrie même qu'il est censé surveiller

pour le compte du gouvernement fédéral. La fine pointe de la capture réglementaire, c'est ça: enfiler un déguisement de fonctionnaire public pour devenir le régulateur de ses propres activités.

Au Royaume-Uni, avec Oxford comme quartier général des EA, le gouvernement de Rishi Sunak est probablement le plus gangrené par le lobbying apocalyptique des prophètes. La nébuleuse est omniprésente dans l'administration, révélait *Politico*^[46] fin 2023. En novembre 2023, le gouvernement a créé l'AI Safety Institute, dirigé par Ian Hogarth, un *doomer* invétéré qui a investi dans plus de 50 *start-up* d'IA et qui signe des oracles dans le *Financial Times* pour crier à l'arrivée de l'«IA divine^[47]». Dans l'exercice de ses fonctions, il n'hésite pas à consulter d'autres spécialistes de l'IApocalypse: le Center for AI Safety et l'Alignment Research Center, tous deux financés par Open Philanthropy, et les *start-up* Anthropic et Conjecture, dont le modèle d'affaires repose sur la catastrophe, et dont Hogarth détient des parts.

En novembre 2023, un sommet sur l'*AI safety* annoncé en grande pompe par le gouvernement et vendu comme une première mondiale a surtout été une ahurissante démonstration d'allégeance néolibérale. On y a vu le premier ministre du Royaume-Uni, qui ne cache pas sa technobéatitude, interviewer Elon Musk, les yeux remplis d'étoiles, et lui demander comment l'intelligence artificielle devrait être réglementée. L'homme le plus riche du monde, qui n'en demandait pas tant, a eu cinquante minutes pour faire l'inventaire de ses utopies technoféodales et IApocalyptiques^[48], devant un public de journalistes réduits au rôle de plantes d'intérieur.

Pendant ce temps, sous une tente plantée dans les jardins de Bletchley Park, où se tenait le sommet, la vice-présidente des États-Unis, Kamala Harris, et la présidente de la Commission européenne, Ursula von der Leyen, s'attablaient avec une poignée de patrons de la tech et quelques prophètes catastrophistes des EA pour parler de l'avenir des sociétés dans une ambiance de partage du monde entre puissants, digne de la Conférence de Berlin ou de Yalta^[49].

Après deux jours de discussions, 28 nations, dont les États-Unis et la Chine, ont signé la déclaration de Bletchley. Le texte est un parfait exemple de déclaration non contraignante, totalement inoffensive qui, en

guise de régulation, «encourage» les multinationales de l'IA à la «transparence et la responsabilité^[50]». Pour parachever la pantomime, le gouvernement Sunak a annoncé dix jours plus tard qu'il ne comptait pas réguler le secteur «à court terme», pour ne pas entraver l'innovation, et que le gouvernement dépenserait 300 millions de livres pour rester dans la course à l'IA. Une pure mascarade.

Les pouvoirs publics ont-ils totalement abdiqué face au technocapital? Non, répond l'Union européenne, dont le Parlement a voté l'AI Act le 13 mars 2024, présenté avec tambours et trompettes comme le premier cadre légal international contraignant sur l'IA au monde. Le texte prévoit une régulation par niveaux: un régime pour les IA actuelles et un autre pour les modèles jugés «à haut risque». Parmi ces derniers, les hypothétiques «modèles de fondation», dont la capacité de calcul excède les 1026 opérations par seconde. Les rares interdictions formelles concernent les systèmes de surveillance de masse, de notation de citoyens, d'identification biométrique et de manipulation de comportement. Un futur office européen de l'IA sera chargé de faire respecter les règles, qui entreront en vigueur en 2025. Grande victoire pour les régulateurs contre les géants de la tech, se félicitent la presse et les élus européens.

Alors, champagne? Pas vraiment.

En réalité, l'AI Act a été patiemment et efficacement neutralisé, démantelé pièce par pièce par la Big Tech et ses alliés européens^[51]. OpenAI a d'abord fait campagne discrètement, à partir de 2022, pour convaincre les législateurs européens de laisser l'industrie s'autoréguler, et plusieurs des amendements que la compagnie a suggérés se sont retrouvés dans le brouillon approuvé par le Parlement européen^[52]. En mai 2023, changement de ton: Sam Altman affirme que si le texte passe, OpenAI devra quitter l'Union européenne. Les semaines suivantes, les PDG de Google et de Microsoft proposaient personnellement au commissaire européen Thierry Breton un AI Pact non contraignant et inoffensif.

De juin à décembre 2023, lors de la phase de négociations du texte entre le Parlement, le Conseil et la Commission, les délégations des États membres ont subi «une offensive de lobbying à grande échelle^[53]».

L'objectif des géants de la Silicon Valley: faire supprimer du texte les obligations de transparence, de respect du copyright des données d'entraînement et d'évaluation de l'impact environnemental de leurs produits. Meta, le plus zélé, a transmis un document de 134 pages de «suggestions» pour reformuler des dizaines d'articles du texte. Durant cette phase, les lobbyistes de la Big Tech étaient présents à 86 % des réunions des «représentants majeurs de la Commission» sur l'IA^[54]. Et pour cause: selon une étude de l'université Stanford, aucun des logiciels d'IA commercialisés à l'été 2023 n'était alors compatible avec le cadre légal qui s'annonçait^[55]. La régulation par niveaux, qui déplace les interdictions vers des logiciels imaginaires, c'est encore leur idée.

La phase la plus symptomatique du rapport de pouvoir entre la Big Tech et les États-nations reste à venir. Début novembre 2023, l'adoption du texte est bloquée par trois dissidents: l'Allemagne, l'Italie, mais surtout la France. Emmanuel Macron vient de recevoir à l'Élysée, le temps d'un dîner de cinq heures et six services, 40 «figures» venues d'Europe et de la Silicon Valley, parmi lesquelles le gourou de l'IA chez Meta, Yann Le Cun, et le président d'OpenAI, Greg Brockman. Il leur promet que la France restera ouverte au commerce de l'IA. Son but est aussi de protéger à tout prix Mistral, la «licorne» française de l'IA, que Macron imagine devenir le ChatGPT européen. Mistral est financée par l'ex-PDG de Google, Eric Schmidt; l'ancien secrétaire d'État au numérique, Cédric O^[56], qui figure parmi ses lobbyistes et ses actionnaires, est également conseiller du gouvernement pour l'IA générative. Dans le magma néolibéral, gouvernements, lobbyistes et industriels sont interchangeable.

Grâce aux efforts combinés de Macron et de Mistral, dont les lobbyistes ont accès aux plus hautes instances européennes, la plupart des règles contraignantes visant les fabricants d'IA sont retirées du brouillon après plus de 36 heures de négociations. La France sera la dernière à ratifier le texte, en février 2024. La version finale est un véritable gruyère, plein de régimes d'exception qui protègent les modèles d'IA développés par la Big Tech, mais aussi l'utilisation de logiciels d'intelligence artificielle par la police et l'armée.

«Dans sa forme finale, l'AI Act est tout à fait gérable pour nous^[57]», se félicite le PDG de Mistral; quelques semaines plus tard, dans la surprise générale, la *start-up* signe un partenariat massif avec Microsoft. Une union qu'elle était probablement en train de négocier au moment même où elle jouait la carte du gentil David européen contre les méchants Goliath de la Silicon Valley, pour amadouer les régulateurs de l'AI Act. Cerise sur le gâteau, Mistral annonce que son prochain modèle d'IA, Large, ne sera finalement pas *open source*, mais bien distribué sous licence commerciale. Le technocapitalisme a mis une fausse moustache, et l'Europe s'est fait bernier.

Vient-on d'assister à «un chef-d'œuvre d'*astroturfing*^[58]», un tour de passe-passe lobbyiste dont la Big Tech s'est retirée pour mieux laisser travailler son «cheval de Troie^[59]» européen? Mistral et Emmanuel Macron ont-ils volontairement fait le «sale boulot» des Google, Microsoft, OpenAI et autres? Quoi qu'il en soit, le résultat est là: derrière les accolades et les félicitations, après sept mois de négociations acharnées, l'AI Act est un texte qui ne touche pas un seul cheveu du monopole de la Big Tech. Cette farce-là est beaucoup plus élaborée que celle à laquelle ont eu droit les États-Unis et le Royaume-Uni, mais elle n'en est pas moins une pantalonnade. Et une gifle pour les élus européens qui ont gobé l'histoire de la petite *start-up* française souveraine, alors qu'elle défendait les mêmes régulations que les géants états-uniens.

États-Unis, Royaume-Uni, Union européenne: en Occident, la première vague de régulations est désormais passée, et les cadors de l'IA peuvent dormir tranquilles dans leurs bunkers anti-apocalypse. Ils n'auront à fournir ni documentation technique ni analyse d'impact de leurs modèles avant la mise sur le marché, pas plus qu'ils n'auront à se plier à des audits de sécurité externes, à démontrer l'interprétabilité de leurs modèles, à fournir des outils pour différencier un contenu synthétique d'un contenu humain, ou à respecter la Directive européenne sur le copyright pour leurs données d'entraînement. Alors que le Parlement voulait initialement imposer ces contraintes à tous les systèmes d'IA «à but général», comme GPT-4, elles sont désormais réservées aux IA à usage général présentant un «risque systémique», soit

aucun des systèmes actuellement commercialisés. Cette approche par niveaux, qui impose une surveillance externe à des «modèles de fondation» hypothétiques présentant des risques hypothétiques, c'est une trouvaille de Google et Microsoft.

Tout est sous contrôle.

Réguler le monopole, c'est réguler l'outil

Après un quart de siècle à planer au-dessus des lois et à voir les dirigeants des anciens empires occidentaux leur faire des courbettes, les prophètes-empereurs de la Silicon Valley ne veulent plus se contenter d'esquiver la régulation. Ce qu'ils veulent maintenant, c'est mener à son terme le projet néolibéral et prendre eux-mêmes le contrôle des régulateurs publics, en incarnant à la fois le pouvoir et le contre-pouvoir. Ils présentent l'IA comme une entité surpuissante, capable de nous mener à la transcendance ou à l'extinction, pour nous convaincre que le destin du monde est entre les mains de sept entreprises californiennes et que «personne hors de l'industrie ne comprend les possibilités» de l'IA, comme l'affirme Eric Schmidt. Et certainement pas de simples fonctionnaires.

Heureusement, pour éviter la fin du monde et guider l'humanité vers son glorieux destin, nos seigneurs ont un plan. La bonne régulation, disent-ils, doit passer par la création d'une agence internationale de l'intelligence artificielle, sur le modèle de l'AIEA^[60] ou du Groupe international d'experts sur le climat^[61], qui délivrerait des licences de développement. Les techno-oligarques savent parfaitement qu'avec leur influence, cette agence et son personnel leur appartiendront, et qu'ils seront alors libres de décerner des «permis d'IA» aux autres entreprises, s'assurant au passage qu'elles ne viennent jamais marcher sur leurs plates-bandes. La *start-up* Conjecture a même proposé de nommer l'agence MAGIC, pour Multinational AGI Consortium, et d'en faire le seul organisme au monde autorisé à développer des IA surpuissantes. Le choix d'acronyme est déjà révélateur d'un éthos: ils rêvent d'un laboratoire magique où l'on développerait l'intelligence magique entre magiciens de la même caste, à l'abri du monde extérieur. Mieux,

Conjecture illustre son projet par une image de forteresse^[62]. Le contrôle, l'obsession de ces grands capitalistes qui, au fond, détestent le capitalisme.

Pour atteindre son objectif de contrôle, la Big Tech ment. Elle ment sur les capacités de ses logiciels, sur leur coût écologique, sur la façon dont elle les déploie et avec qui elle collabore. Lorsque Meta lance son IA, Llama-3, en *open source* sous prétexte d'en faire bénéficier le monde, c'est pour mieux tenter d'exempter les modèles *open source* des règles européennes et continuer de développer ses produits en paix. Lorsque le président de Microsoft, Brad Smith, propose de réguler l'IA «comme les voitures et les médicaments^[63]», via un système de permis (la vision est un peu réductrice), c'est parce qu'il sait que ce serait lui qui les délivrerait. Lorsque Sam Altman supplie les députés des États-Unis de le réguler, lorsqu'il affirme que le secteur n'a pas besoin de régulation «avant plusieurs générations», lorsqu'il fait la cour à la présidente de la Commission européenne, lorsqu'il menace de quitter l'Union européenne et lorsqu'il se rétracte, il joue la comédie – la même que ses prédécesseurs de Big Oil et Big Tobacco, et que les industriels corrompus de l'âge d'or états-unien, les Carnegie et les Rockefeller, qui ont inventé la capture réglementaire pour consolider leurs monopoles respectifs. À l'époque, la critique sociale les qualifiait de «barons voleurs», pendant que Roosevelt appelait à mettre au pas ces «royalistes de l'Économie». Cent ans plus tard, les techno-oligarques ont dépassé leurs modèles.

Pendant que nous regardons avec effroi venir l'IApocalypse, tout roule au far west de l'IA. Les Sept Magnifiques peuvent continuer à diffuser leurs logiciels dans le monde, sans rien révéler de leur fonctionnement et sans jamais être tenus responsables des dégâts qu'ils peuvent causer. Aucun des robots conversationnels évalués par l'université Stanford en octobre 2023 n'obtient un score de transparence de plus de 54 sur 100? Ça passe. Les bases de données pillent allègrement du contenu protégé par le droit d'auteur? *Who cares?* Origine des données, méthodes d'entraînement, règles de modération: tout passe sous le radar. Les chercheurs indépendants ne savent rien, ce qui ouvre un boulevard pour la production de pseudoscience sur les

capacités des systèmes. Les régulateurs sont dans le noir parce qu'ils ont choisi de protéger le secret industriel et de compter sur la bonne volonté des multinationales pour leur ouvrir le saint des saints.

Comme avec les algorithmes des réseaux sociaux pendant la décennie 2010, les entreprises de la Big Tech sont libres de travailler à la maximisation de leurs profits aux dépens de l'intérêt général. Libérées à la fois de la régulation et de la compétition, elles peuvent déployer l'IA pour exploiter leurs clients et leurs utilisateurs, capter et monétiser le plus de données possible et reconfigurer les sociétés autour de leurs délirants projets autoritaires, tout en nous persuadant que l'avenir de l'humanité est en jeu et qu'elles sont notre seul espoir.

Lorsqu'une poignée d'entreprises détient les infrastructures, l'argent, l'expertise et l'influence, quand quelques puissants contrôlent les angoisses et les rêves des législateurs, la notion même de contre-pouvoir démocratique s'effondre. L'influence de cette élite est telle que dans une période d'instabilité politique marquée par une «polycrise», l'Union européenne «s'éloigne du néolibéralisme progressiste pour se rapprocher d'un néo-illibéralisme», écrit la chercheuse Rachel Griffin. Avec l'AI Act, «les droits fondamentaux et les normes démocratiques libérales sur lesquels se basaient les politiques européennes sont sacrifiés à une surveillance d'État débridée^[64]», et ce, dans un contexte de montée des extrêmes droites. À l'instar de la Silicon Valley, l'Union européenne semble être en train de mettre à jour son logiciel idéologique vers un capitalisme autoritaire qui privilégie l'économie de la rente et les monopoles à l'économie de marché et la concurrence.

C'est pourtant l'inverse qu'il faudrait faire: ne plus réguler seulement l'outil, mais s'attaquer au système qui le soutient, la structure politique et économique sur laquelle il se déploie. Les systèmes d'IA sont le reflet des oligopoles qui les commercialisent: privés, opaques, impénétrables, intouchables, toxiques et dangereux. Conçus par des multinationales au-dessus des lois, obsédées par la captation et la consolidation, ils captent des données partout où il y en a, sans le moindre regard pour les textes de loi, et soutiennent des rapports de domination préexistants. Même si les gouvernements occidentaux semblent applaudir vigoureusement la consolidation monstrueuse de l'oligopole tech, des régulateurs antitrust

se réveillent en France, aux Pays-Bas, au Royaume-Uni et jusqu'à la FTC états-unienne, dont la présidente Lina Khan réveille l'espoir d'une politique de régulation vivante. Plusieurs enquêtes ont été ouvertes contre le Big Three et Nvidia, sur les deux rives de l'Atlantique, pour concurrence déloyale sur les marchés du *cloud* et des puces informatiques. Ce premier pas est indispensable. Sans fracturation forcée du monopole, la régulation de l'IA ne restera qu'une pantomime.

Pour récupérer une autonomie politique, les législateurs doivent d'abord sortir de la sidération. Il est temps que le public et les journalistes fassent de même, ou nous ne pourrons plus éviter la reconfiguration de nos sociétés en régimes technoféodaux. Sortir de la sidération, c'est décrire l'IA au présent plutôt qu'au futur. La voir non plus pour ce qu'elle pourrait devenir, mais pour ce qu'elle est réellement, socialement, politiquement. Changer de grille d'analyse de la technique. Contre-attaquer les prophètes et leurs fantasmagories toxiques avec un récit radicalement ancré dans le factuel.

Systemes d'exploitation

Ce n'est pas la technique qui nous asservit, mais le sacré transféré à la technique.

Jacques ELLUL, *Les nouveaux possédés*,
1973

Entre le début de la rédaction de ce livre et sa parution, près d'un an se sera écoulé. Nous voilà en 2024, en l'an 2 après ChatGPT.

L'émerveillement prométhéen des premières interactions avec la machine est désormais chose du passé, mais la sidération est restée. Une brèche a été ouverte où se sont engouffrés une poignée d'industriels, d'investisseurs et de chercheurs de la Silicon Valley, animés d'une inébranlable foi technoreligieuse en leur propre supériorité. En 2023, ces gens ont, plus que quiconque, œuvré à la propagation du récit messianique de l'intelligence artificielle qui place l'humanité face à une machine consciente imaginaire qui détiendrait le double pouvoir de nous transcender et de nous anéantir.

Répetons-le: l'intelligence artificielle, celle des prophètes, n'existe pas. Plus qu'une machine pensante, le syntagme «intelligence artificielle» désigne une technologie au sens étymologique du terme: un récit ou un discours traitant d'une technique ou d'une compétence (*technè-logos*). Or, ce récit est une mystification, un édifice d'irrationalité bâti sur une série de croyances religieuses, qu'une presse trop souvent analphabète de la technique amplifie sans se rendre compte qu'elle désinforme. Le fait est que cette fable est monstrueusement lucrative. En 2023, selon l'*AI Index 2024* de Stanford, les promesses de l'IA «généraliste» ont attiré 25 milliards de dollars d'investissements, presque 9 fois plus qu'en 2022 et 30 fois plus qu'en 2019. En plus

d'attirer les investisseurs, cette surenchère dissimule des effets néfastes très palpables. La supercherie est omniprésente. À l'instar de leur machine à *bullshit*, ChatGPT, les prophètes de l'IA mentent avec assurance, enrobant leurs prêches d'un jargon pseudoscientifique, renouvelant le lexique dès que la situation le requiert.

Les angoisses artificielles des prophètes de l'IA camouflent des dangers bien réels. La société civile est particulièrement en péril. Des systèmes informatiques privés, vendus sous couvert d'IA, s'immiscent dans nos institutions publiques et nos vies privées, sans que personne s'interpose ni interroge les modalités de ce phénomène. Structurellement opaques, ces systèmes bouleversent notre rapport à l'information et aux services publics. Les entreprises de la Silicon Valley étendent leur intermédiation sur un corps social médusé. Derrière les vitres teintées des discours apocalyptiques s'opère un transfert de pouvoir du public vers le privé, dans lequel l'État néolibéral sous-traite des pans entiers de sa gouvernance (protection sociale, éducation, médecine, défense, police, etc.) aux producteurs d'algorithmes que sont Google, Microsoft et Amazon. Les conséquences, on le verra, sont déjà terribles pour les corps subalternes et les marginalisés.

Il faut lutter contre le récit spectaculaire qui nous est vendu, ce qui revient à lutter contre cette privatisation du futur. Il est encore possible de produire notre propre récit de l'intelligence artificielle, centré sur les impacts sociologiques de la technique, et de l'écrire au présent plutôt qu'au conditionnel. Partons de ce que le théoricien de la cybernétique Stafford Beer a baptisé le principe POSIWID (*the purpose of a system is what it does*), qui postule que la raison d'être d'un système, c'est ce qu'il fait, et demandons-nous: que *fait réellement* l'IA?

L'intelligence artificielle est un culte

Depuis la conférence de Dartmouth en 1956 dans le cadre de laquelle le concept d'«intelligence artificielle» a été utilisé pour la première fois par John McCarthy, rien n'a fondamentalement changé: pendant des années, des générations d'hommes blancs occidentaux, mathématiciens, informaticiens et entrepreneurs, ont continué de croire qu'ils étaient sur

le point de construire une machine pensante à leur image. Déjà en 1958, le *New York Times* s'esbaudissait devant le Perceptron, la machine à reconnaître des motifs visuels conçue par Frank Rosenblatt. Le titre: «Electronic “Brain” Teaches Itself^[1]» (Un cerveau électronique apprend tout seul). À chaque avancée technique, son avalanche de prophéties et d'investissements, publics comme privés, puis l'inévitable désillusion.

Depuis soixante-quinze ans, les mêmes axiomes fallacieux soutiennent l'évangile de la machine pensante^[2]: la conscience se résume à l'intelligence, celle-ci loge exclusivement dans le cerveau, se réduit à du traitement d'informations, et elle est proportionnelle à la capacité de calcul. Ainsi, puisque le cerveau est une sorte d'ordinateur, un ordinateur peut devenir un cerveau. Les prophètes de l'IA veulent nous faire croire (et semblent réellement persuadés) que si on ajoute de la puissance et des données à un logiciel de calcul de probabilités, celui-ci deviendra non seulement conscient, mais carrément omnipotent. Comme le résume le technocritique Cory Doctorow, c'est à peu près comme si on affirmait qu'en élevant des chevaux de plus en plus rapides, on finira par avoir une locomotive^[3]. Toute la prophétie est là, dans l'idéal néoplatonicien d'une intelligence «parfaite», la culmination de l'esprit humain dans une pure machine à calculer. Dans son combat acharné contre la réalité factuelle, la secte de l'IA publie une montagne de pseudoscience aux résultats grandiloquents^[4]... et impossibles à répliquer dans le monde réel^[5].

Sauf que ça ne fonctionne pas comme ça. Le raisonnement est deux fois mensonger, biologiquement et politiquement. Pour commencer, personne ne s'accorde sur une définition de l'intelligence humaine, un phénomène extraordinairement complexe, qui mobilise non seulement l'entièreté du corps, mais aussi les êtres humains avec qui chaque individu interagit matériellement et symboliquement. L'intelligence est incarnée, situationnelle, collective, relationnelle, sociale, culturelle, institutionnelle, historique. N'en déplaise aux transhumanistes, nous n'existons pas dans le vide. Nous ne sommes pas le «cerveau dans une cuve» imaginé par le philosophe Hilary Putnam en 1981^[6]. Non seulement il n'y a pas d'intelligence sans corps – le processus de mémoire, par exemple, est décentralisé –, mais l'expérience démontre

qu'il y a de l'intelligence sans cerveau – d'autres cellules que les neurones sont capables de stocker de l'information et de la manipuler, et c'est ce qu'on appelle la cognition fondamentale^[7].

Plus important encore, il n'y a pas de corps pensant sans corps social. Les techno-évangélistes ont beau fétichiser l'individu rationnel tout-puissant, qui culmine dans la figure du génie solitaire, le monde fonctionne sur la mutualité, la réciprocité et le partage de compétences. Nous sommes collectivement cette superentité intelligente et autonome qui ne cesse de s'autopréserver, de s'autorépliquer et de s'auto-améliorer. Nous avons toujours été là, sous les yeux des prophètes de la superintelligence. S'ils avaient la moindre estime pour la multitude, les prophètes l'auraient remarqué depuis longtemps. Au lieu de cela, ils préfèrent fantasmer sur une personnification de leurs multinationales omnipotentes, ou s'imaginer en empereurs solitaires de la galaxie, dans la tristesse de leurs imaginaires individualistes.

Comparer les machines et l'humain est non seulement insensé, mais ontologiquement dangereux. Proclamer que «le cerveau est un ordinateur qui est un cerveau» a des conséquences sociales: la métaphore computationnelle «justifie l'assujettissement des êtres considérés moins rationnels que la norme autoproclamée, que ce soient les animaux par rapport aux humains, les femmes par rapport aux hommes, ou une race humaine par rapport à une autre^[8]». Pour anthropomorphiser les robots conversationnels, les vendeurs d'IA leur font passer toutes sortes de tests d'aptitude: le QI, le baccalauréat ou l'examen du Barreau^[9]. Et en humanisant la machine, ils mécanisent implicitement l'humain par un processus parfaitement technocapitaliste, comme ceux qui transforment l'individu en marchandise, en matière première ou en outil de production. Aussi, pour refuser la déshumanisation capitaliste et la réduction de l'être au calculable, il faut refuser de parler d'«intelligence artificielle» lorsqu'il est question d'agents conversationnels.

L'intelligence artificielle ne fonctionne (toujours) pas

Les modèles de langage comme GPT-4, Claude ou Gemini sont des systèmes informatiques qui coûtent extraordinairement cher, mais qui

n'ont aucune application concrète. Ce sont des machines à produire automatiquement des discours plausibles sur une variété de sujets, et ce, avec un tel aplomb qu'on a même comparé ChatGPT à une machine à *mansplaining*^[10]. Souvent, d'ailleurs, comme les *mansplainers*, ces machines mentent carrément, incapables d'avouer leur ignorance sur tel ou tel sujet. Ces mensonges plus vrais que nature, l'industrie les balaye du revers de la main en les qualifiant d'«hallucinations». Selon un recensement, le meilleur logiciel du marché, GPT-4 Turbo, hallucine 2,5 % du temps, tandis que la concurrence produit environ 5 % de confabulations^[11].

La Silicon Valley y voit de simples bogues, mais il s'agit en vérité d'une caractéristique structurelle de ces programmes, qui ne «savent» rien et produisent des approximations statistiques en «retournant des techniques de classification sur elles-mêmes», dicit Emily M. Bender^[12]. Leur activité consiste à compresser les bases de données gigantesques qu'ils ont ingérées, et toute compression implique une perte de détails. Ça ne se répare pas. Aucun *chatbot* n'atteindra jamais les 100 % de fiabilité. L'industrie le sait parfaitement, mais elle continue quand même de vendre ses programmes avec la promesse que bientôt, ils sauront répondre «je l'ignore». Or, la différence d'impact social entre un système fiable et un système imparfait mais perçu comme fiable est colossale.

Ne donnez aucune légitimité à ce que produit un *chatbot*. Ce que commercialisent OpenAI, Anthropic et les autres, c'est une sorte de voyant informatique^[13], qui vous tirerait les cartes dans l'intimité de votre écran de téléphone. Comme pour la cartomancie, la tromperie ne fonctionne que si on veut y croire; dès qu'on y réfléchit, la magie se dissipe. En moyenne, les utilisateurs de ChatGPT cessent d'utiliser le service après une quarantaine de jours, comme un nouveau jouet qu'on mettrait rapidement de côté^[14]. L'écrivain d'anticipation Ted Chiang l'a très bien résumé: ChatGPT et consorts fournissent pour l'essentiel «une image floue du contenu qui se trouve sur le web^[15]». Leur méthode de prédiction statistique les empêche de produire le moindre résultat fiable ou sourcé. Google SGE, le moteur de recherche augmenté par un robot conversationnel, produit régulièrement de fausses informations après plus d'un an de tests^[16]. Bard, le *chatbot* précédent, avait fait une erreur

avant même d'être mis en marché; la capitalisation boursière d'Alphabet avait immédiatement chuté de 100 milliards de dollars. Lorsque Bing AI, l'agent conversationnel de Microsoft, est interrogé sur des résultats d'élections, il se trompe encore une fois sur trois^[17]. Depuis qu'IBM a dévoilé le logiciel Watson en 2011, les trompettes californiennes ont annoncé que les robots remplaceraient bientôt radiologues, avocats^[18], chercheurs, profs^[19], médecins, standardistes, codeurs, journalistes, responsables des ressources humaines, etc. Tout ce beau monde attend encore de pouvoir prendre sa retraite, et l'Organisation mondiale de la santé trouve l'IA médicale «dangereuse^[20]». Certaines entreprises comme le transporteur DPD ou Air Canada ont bien essayé de confier leur service clientèle à un *chatbot*, mais le premier a insulté des clients^[21] et le second a inventé des conditions de remboursement^[22]. Les deux ont été désactivés en urgence. Les évangélistes de l'IA prédisaient la fin du monopole de Google sur la recherche en ligne. Début 2024, l'entreprise de Mountain View détenait toujours 90 % du marché. Faites le test: qui autour de vous utilise ChatGPT comme portail d'informations? Personne, parce que personne ne veut que son moteur de recherche invente des *fake news*. Demandez aux avocats dont les *chatbots* ont fabriqué de fausses jurisprudences, aux chercheurs qui ont intégré de fausses références universitaires générées automatiquement^[23], et aux médias comme MSN ou Gizmodo dont les *chatbots*-rédacteurs ont réussi à faire des erreurs alors qu'ils plagiaient d'autres articles de presse. Partout où ils sont déployés, ces logiciels se ridiculisent.

Loin d'être des copilotes infaillibles, ces systèmes alourdissent le quotidien de leurs utilisateurs, notamment en obligeant à surveiller sans cesse la machine^[24]. Au quotidien, l'IA générative est un collègue atroce, non seulement amnésique mais mythomane, qui travaille si mal qu'il faut systématiquement passer derrière lui. Dans un sondage mené en décembre 2023 au Royaume-Uni, 87 % des chefs de projet interrogés déclaraient que leurs employés n'utilisaient pas les logiciels d'IA^[25]. En février 2024, le Pew Research Center concluait que 20 % des adultes états-uniens utilisaient ChatGPT au travail, mais qu'à peine 2 % d'entre eux faisaient confiance au système^[26]. Une autre étude datant de

septembre 2023 a démontré que les cadres qui utilisent GPT-4 au bureau sont 23 % moins performants que ceux qui ne l'utilisent pas^[27]. L'inefficacité des *chatbots* aurait déjà coûté en moyenne 6 % du chiffre d'affaires d'entreprises. Pourtant, preuve de la puissance de persuasion des prophéties, les chefs d'entreprise interrogés affirment tout de même vouloir utiliser davantage ces systèmes, terrifiés à l'idée de rater le train de la modernité^[28].

Les logiciels d'IA sont manifestement en train d'atteindre un plateau de compétences, et certains chercheurs anticipent même une *régression*^[29] pour la prochaine génération de logiciels. Malgré des budgets de développement pharaoniques, OpenAI a été obligé de reconnaître que GPT-4 est moins performant que son prédécesseur, GPT-3.5, pour certaines tâches^[30]. ChatGPT produit désormais fréquemment du pur charabia. Cette baisse de performance n'est pas surprenante. Elle s'explique en partie par une pénurie de données. Après avoir ingéré l'intégralité du web public, les grands modèles de langage sont à court de ressources. Ils se nourrissent désormais de textes «synthétiques» produits par leurs concurrents et prédécesseurs. Ils s'entraînent donc sur des hallucinations, et produisent des hallucinations encore pires. Le risque que pose cette «consanguinité générative^[31]», c'est que les logiciels s'effondrent sur eux-mêmes, assommés par le volume de leurs propres divagations. Pour certains, il s'agit là de la «malédiction de la récursion», et l'effondrement a déjà commencé^[32]. L'IA générative semble être en train de se cannibaliser, mais face à cette situation, les fabricants ne bronchent pas et proposent, au contraire, d'accélérer le processus en fournissant de la donnée d'entraînement synthétique, soit des imitations de données réelles. De guérir le mal par le mal, somme toute. Le *New York Times* parle pudiquement d'un «problème de mesure^[33]» de l'industrie, mais en réalité, le problème en est un de démesure. Plus les logiciels grossissent, moins ils sont fiables.

Alors en 2024, pendant que leurs PDG célèbrent la fin du travail et brandissent la corne d'abondance du futur, les responsables produits chez Microsoft, Google et Amazon suggèrent discrètement à leurs vendeurs de calmer le jeu auprès des clients, afin d'éviter la publicité mensongère^[34]. Certains plaident même pour retarder le lancement de

logiciels qui ne sont même plus assez performants pour faire illusion. Sam Altman, qui subit le «*burnout* de l'IA» dont il a été le principal artisan en 2023, joue désormais les blasés en déclarant que GPT-4 est «un peu nul» et que l'AGI «changera moins le monde et le travail que ce qu'on pense^[35]», après avoir passé un an à dire exactement le contraire.

Depuis son pic d'utilisation, en mai 2023, son ChatGPT a perdu 11 % de ses utilisateurs^[36] et OpenAI perd de l'argent à chaque utilisation: en février 2023, le service gratuit lui coûtait 700000 dollars par jour^[37]. Pour un logiciel d'auto-complétion dont l'utilité principale est de faciliter la tâche aux écoliers qui trichent dans leurs devoirs et d'aider les amateurs de chiens à composer un haïku sur leur fidèle ami, on est en droit de trouver ça cher. Fin 2023, Sam Altman a affolé la presse économique en estimant que la course à l'AGI nécessitera 5 à 7 *trillions* de dollars d'investissements – soit 7 % du PIB mondial. Aujourd'hui, l'IA générative est une solution qui recherche désespérément un problème, une lubie qui coûte beaucoup d'argent, ne sert pratiquement à rien et réclame toujours plus de matière première.

L'intelligence artificielle est un écocide

Qu'elle trouve ou non des usages, l'IA est un gouffre énergétique. Chaque image générée par un logiciel comme Midjourney ou DALL-E consomme autant d'électricité qu'une charge complète de téléphone^[38]. Les cartes graphiques consomment quatre fois plus d'électricité qu'un serveur classique, et chaque système en utilise des dizaines de milliers à la fois. Amazon, Microsoft et Google construisent désormais ce que l'industrie appelle des *hyperscalers*, des entrepôts réfrigérés de la taille de plusieurs terrains de foot, qui hébergent au moins 5000 serveurs et peuvent consommer autant d'électricité que 50 000 foyers états-uniens. Microsoft ouvre un nouveau centre de données tous les trois jours. La demande mondiale explose, et la consommation d'énergie est à l'avenant. Selon l'Agence internationale de l'énergie^[39], celle-ci devrait doubler d'ici 2026. À l'échelle mondiale, le continent de la donnée consommera alors autant d'électricité que le Japon et ses 125 millions d'habitants.

Malgré les promesses de croissance verte de Big Tech, l'électricité décarbonée ne suffira pas. La folie de l'intelligence artificielle est donc un miracle pour les producteurs de combustibles fossiles qui cherchaient désespérément de nouvelles sources de revenus dans une économie qui s'est engagée à réduire sa dépendance envers ces sources d'énergie. Voilà que le conglomérat de la tech leur propose des milliards de dollars d'investissements et fait retarder la fermeture de centrales à charbon^[40]. Alléluia! Le lobby du gaz naturel s'émerveille du «marché émergent» de l'IA^[41], Amazon achète un centre de données alimenté à l'énergie nucléaire, Microsoft embauche des spécialistes de l'atome pour préparer les siens et Sam Altman investit dans des *start-up* de réacteurs portables, comme Oklo, spécialement conçus pour alimenter des centres de données. L'avenir de l'IA est atomique.

Et ce n'est que le début. Pour construire GPT-5 et ses successeurs, Microsoft et OpenAI seraient en train de concevoir un complexe appelé Stargate^[42], à la fois centre de données et supercalculateur, dont la construction prendra six ans et coûtera à elle seule dix milliards de dollars. Cette abomination informatique et écologique, dont le coût matériel total se situerait autour des 100 milliards de dollars, consommerait jusqu'à 5 gigawatts d'électricité – à titre de comparaison, 1 gigawatt permet d'alimenter 750 000 foyers aux États-Unis pendant un an. Au risque de vexer Microsoft et OpenAI, on peut raisonnablement prédire qu'investir 100 milliards de dollars dans une machine à *bullshit* géante ne la rendra ni intelligente ni consciente. Sam Altman, jamais satisfait du présent, semble persuadé du contraire. Il affirme maintenant que la voie vers l'AGI passe par la fusion nucléaire, une avancée technologique qui promet une énergie propre et illimitée. D'ailleurs, Altman préside aussi Helion, une *start-up* de réacteurs à fusion nucléaire sous contrat avec nul autre que Microsoft, et s'en vante en annonçant une «intelligence et une énergie illimitées^[43]», sans trop parler des conflits d'intérêts illimités.

Contrairement aux métaphores neutralisantes que nous servent les géants de la Silicon Valley, l'intelligence artificielle ne vit pas dans les nuages du *cloud*. Son empreinte matérielle ne se limite pas à l'électricité. Elle dépend d'une infrastructure toxique planétaire qui brûle des

carburants, extrait des métaux rares, consomme de l'eau douce et rejette du dioxyde de carbone dans l'atmosphère. Une interaction de 20 à 50 questions avec ChatGPT consomme 500 millilitres d'eau *potable*. Sachant que ChatGPT a accumulé 100 millions d'utilisateurs quotidiens, on a vite un ordre d'idées du gaspillage. Selon la même étude^[44], la consommation mondiale d'eau de l'industrie de l'IA pourrait dépasser celle du Danemark, en 2027.

Demandez aux *chatbots* Claude, Gemini ou ChatGPT où ils vivent, et ils insisteront sur leur immatérialité. En réalité, leur empreinte matérielle est considérable. L'IA n'est pas née de la côte d'Elon ou de Sam, elle naît tous les jours dans les mines de cobalt en République démocratique du Congo, dans les champs de lithium à ciel ouvert du salar d'Uyuni, dans les usines de traitement de terres rares en Mongolie intérieure^[45]. Dans la «vue multidimensionnelle^[46]» proposée par le chercheur en médias numériques Eryk Salvaggio, elle se révèle comme un assemblage planétaire sur trois niveaux – composants électroniques, réseaux d'information et de relations sociales – qui extrait à la fois de la donnée informatique, des matières premières physiques et du labour humain.

Il est essentiel de démanteler le récit dématérialisant de l'IA. Non, des entités informatiques conscientes ne flottent pas dans l'éther de la donnée et du *cloud*. Dans nos imaginaires, l'IA doit faire surgir des images de hangars fumants, de cheminées d'évaporation de centrales nucléaires, de lacs pollués et de terres ravagées. Critiquer correctement la technique, c'est d'abord mettre à jour les représentations que nous nous en faisons. Il nous faut un point de vue surplombant, des diagrammes, des cartes et des maquettes^[47]. Ouvrons le *Contre-atlas de l'intelligence artificielle*^[48] et observons le plan anatomique^[49] conçu par la chercheuse australienne Kate Crawford. Penchons-nous sur la chaîne de production mondialisée du *compute* dessinée par l'AI Now Institute^[50], qui montre comment les composants s'assemblent et s'échangent aux quatre coins de la planète. Jetons pour de bon les androïdes nacrés et les nuages cotonneux de données dans la décharge de nos imaginaires: l'intelligence artificielle est une industrie lourde, qui menace la pérennité du vivant de façon extrêmement palpable.

L'intelligence artificielle est un pillage

Centre d'une industrie fondamentalement extractive, qui transforme l'activité humaine en information monétisable, la Silicon Valley a plus que jamais soif de données personnelles et biométriques, et bafoue de plus en plus les lois nationales et internationales qui protègent les contenus numériques. Après vingt-cinq ans de laisser-faire gouvernemental aux États-Unis, les plateformes de la Big Tech sont devenues de gigantesques aspirateurs à textes, sons, photos et vidéos, habituées à opérer au mépris des lois. Si Google, Microsoft et Amazon dépensent joyeusement des milliards de dollars pour l'expansion des modèles actuels, c'est aussi parce qu'ils espèrent pouvoir les utiliser pour capter et analyser plus de données que jamais auparavant.

Ce qui se passe depuis le début de l'année 2023 n'a pas de précédent en matière de volume et d'agressivité: c'est un véritable carnage, une razzia numérique indiscriminée, avec les chalutiers de la Silicon Valley qui profanent et ravagent l'océan de données mondiales. À tel point que l'impensable est arrivé: les modèles ont digéré l'intégralité du contenu du web, et sont désormais à court de données publiques. Alors, pour entraîner son *chatbot* Gemini, Google va jusqu'à piquer les données personnelles des utilisateurs de Gmail. Meta, OpenAI et Google entraînent leurs logiciels sur la base de données Books3, composée de 183 000 livres numériques piratés et sous copyright^[51]. En Californie, un recours collectif en justice accuse Google et OpenAI de piller les données de «millions» d'individus^[52].

Puisque les gouvernements laissent faire, la société civile commence à contre-attaquer. Une lettre ouverte du syndicat de l'édition états-unien, signée par 8500 auteurs et autrices, réclame que les entreprises d'IA paient pour l'exploitation de leur travail «sans consentement, crédit ou compensation^[53]». Reddit, StackOverflow et Twitter/X demandent aux entreprises d'IA de les payer pour récolter leurs données^[54]. Un collectif d'artistes a mis au point et distribué un outil, Nightshade, qui «empoisonne» les logiciels d'IA qui tenteraient d'accaparer leur travail^[55]. Des collectifs de codeurs, d'artistes visuels, d'auteurs et

autrices de *fan fiction* et de youtubeurs se découvrent un esprit luddite, et se révoltent contre l'opération de pillage menée par les entreprises d'IA. La presse commence à parler d'une «révolte de la donnée^[56]». Pour empêcher la Big Tech de s'approprier leurs œuvres, les scénaristes et acteurs d'Hollywood ont organisé la plus longue grève de l'histoire de l'industrie^[57]. Et ils ont gagné!

Une guérilla juridique est engagée. Scarlett Johansson, la banque d'images Getty, le *New York Times*, le média d'investigation *The Intercept*, les auteurs John Grisham et George R.R. Martin, la guilde des auteurs et une coalition de majors du disque: à ce jour, l'industrie de l'IA est engagée dans une vingtaine de procès pour vol, non-respect du droit d'auteur et utilisation illégale de contenu. Nous vivons peut-être un moment décisif, similaire à la bataille entre les ayants droit de l'industrie culturelle et les premiers sites de partage de fichiers comme Napster, dont l'issue a profondément changé l'économie numérique^[58].

Le soulèvement des damnés de la donnée, dont le grondement monte jusqu'aux fenêtres des législateurs, inquiète la Silicon Valley. En octobre 2023, un porte-parole du fonds d'investissement Andreessen Horowitz déclarait que facturer «le coût réel ou potentiel du droit d'auteur aux créateurs des modèles d'IA aura [aurait] pour conséquence de les tuer ou d'entraver leur développement^[59]». Boo hoo! Cette déclaration est ahurissante: le plus gros investisseur de l'IA californienne reconnaît que l'activité qu'il subventionne doit *nécessairement* être illégale pour être rentable. Encore plus formidable: pour OpenAI, le respect du droit d'auteur empêcherait ses logiciels d'exister^[60]. Pour le gang des cleptomane de la Silicon Valley, le respect des droits d'auteur est bel et bien une menace existentielle.

L'intelligence artificielle est une bulle financière

La question n'est pas de savoir si elle éclatera, mais quand, et avec quelle violence. Le marché de l'IA n'a aucun sens. Malgré l'absence de perspectives de profit à court terme et des coûts de fonctionnement astronomiques, malgré une empreinte écologique terrifiante, les fabricants de logiciels d'IA ont reçu 25 milliards de dollars de capital-

risque au premier trimestre de 2024 pour développer leur activité, après une année 2023 bouclée à 42 milliards de dollars^[61]. Il y avait alors 1 812 *start-up* pour se partager le butin, soit 40 % de plus qu'en 2022^[62]. Celles-ci ne produisent rien d'autre que du vent, mais leur valeur en Bourse est en moyenne 83 fois supérieure à leur revenu annuel estimé^[63].

La Big Tech aussi se porte très bien, merci pour elle. La capitalisation boursière des «Sept Magnifiques» de la Silicon Valley équivaut désormais aux bourses du Royaume-Uni, du Canada et du Japon réunies^[64]. Microsoft, actuellement l'entreprise la plus cotée au monde, vaut autant que la Bourse de Toronto. OpenAI a atteint les 2 milliards de dollars de revenus en 2023, sans générer le moindre profit. Le grand gagnant reste le fabricant de cartes graphiques Nvidia, leader mondial de puissance de calcul, qui a vu ses profits augmenter de 580 % et sa capitalisation boursière augmenter de 1000 milliards de dollars d'une année à l'autre.

Les entreprises qui bénéficient le plus de cette manne ne sont pas celles qui vendent ou intègrent les systèmes d'IA à leur *business*, mais celles qui fournissent le matériel. Si l'IA est une ruée vers l'or, ses vainqueurs sont les fabricants de pioches. Mais Nvidia est peut-être aussi le canari au fond de la mine, le premier signe du coup de grisou économique. En avril 2024, l'action de l'entreprise a brutalement chuté de 20 %, et les Sept Magnifiques ont perdu 1000 milliards de dollars de capitalisation boursière dans son sillage. Pour certains, c'est une simple entaille dans la courbe verticale du marché; pour d'autres, c'est la trajectoire parabolique normale d'une industrie qui n'avait finalement rien d'autre à vendre que des promesses^[65].

L'heure est aux pronostics. La bulle de l'IA explosera-t-elle comme la bulle *dot-com*, durant laquelle une poignée d'entreprises du web naissant a fait quintupler la valeur du Nasdaq entre 1995 et 2000? Le fabricant de puces Nvidia est-il, comme le pense le *Financial Times*^[66], le nouveau Cisco, qui vendait lui aussi des pelles et des pioches aux orpailleurs numériques de l'an 2000? L'IA est-elle comme le Bitcoin, décrit ainsi par la journaliste Nastasia Hadjaji^[67]: un cheval de Troie technique pour une politique d'extrême droite? Finira-t-elle comme le métavers, ce

marché imaginaire en réalité virtuelle que l'oracle McKinsey évaluait à 5000 milliards de dollars potentiels d'ici 2030^[68] et dont plus personne ne parle aujourd'hui? S'effondrera-t-elle comme les NFT, ces actifs spéculatifs virtuels totalement absurdes, ou comme le web3, cet empilement de *buzzwords* qui promettait de changer le monde sans jamais préciser comment?

Arrêtons pour de bon de donner du crédit aux entrepreneurs de la tech. Depuis le début de la décennie 2020, le technocapitalisme ne fonctionne plus que par vagues d'hallucinations successives, suivies de (très) brèves périodes de lucidité. La Silicon Valley semble bloquée dans un *trip* d'acide qui ne redescend pas, et dont l'IA n'est que la plus récente hallucination. En face, les investisseurs de Wall Street ont eux aussi perdu la raison et se comportent «comme au casino^[69]» (dixit Warren Buffet), accros au jeu et aux politiques d'emprunt à taux zéro de la banque centrale états-unienne. Il y aurait de quoi rire, sauf que chaque fois qu'une bulle explose en Californie, c'est toute la société civile qui paie l'addition. Pour reprendre le constat de Cory Doctorow, les bulles financières de la tech existent en deux variétés^[70]: celles qui laissent des ruines et celles qui ne laissent qu'un trou béant.

Le problème structurel de l'IA, c'est son coût. Contrairement aux sites d'*e-commerce* et aux premiers réseaux sociaux, que presque n'importe qui pouvait créer depuis chez lui, sur un serveur personnel (du moins dans un premier temps) et via une connexion ADSL, les grands modèles de langage comme GPT-4 coûtent extraordinairement cher à construire, entraîner et entretenir. Leur existence dépend d'himalayas de financement et d'infrastructures matérielles gigantesques. Si le robinet à milliards se tarit, le centre de données s'arrête, et adieu les machines à générer du vide. Et que restera-t-il, une fois la poussière du *crash* retombée? Pas grand-chose, sinon un paquet de prophéties périmées, des milliers de cartes graphiques H100 obsolètes et des multinationales californiennes toujours plus profondément fourrées dans nos États.

L'intelligence artificielle est un autoritarisme

Dans l'ombre des robots conversationnels hallucinés et de l'AGI divine, une autre «intelligence artificielle», celle des algorithmes d'analyses de *big data*, a rapidement trouvé des usages dans le contrôle politique des plus vulnérables au service des plus puissants. En France, aux Pays-Bas, en Australie^[71], les systèmes automatisés appliqués aux aides sociales ont accru la surveillance et multiplié les pénalités imposées aux plus pauvres, exacerbant l'injustice et l'inhumanité administrative, et détruisant des vies. La même logique est à l'œuvre dans toutes les sphères publiques – santé, éducation, police, justice – où des algorithmes opaques et biaisés sont déployés pour modéliser, évaluer et assigner des individus. Pas une semaine ne passe sans qu'un gouvernement néolibéral déploie un robot conversationnel dans une administration publique.

En France, qu'Emmanuel Macron rêvait jadis de transformer en «*start-up nation*», l'IA «générative» est entrée en force dans les administrations publiques en 2024 «pour rendre l'action publique plus simple et plus efficace^[72]». Comme incarnation de la simplicité, l'administration a bidouillé Albert, un robot conversationnel «souverain» qui répondra aux demandes des usagers et pourra à terme automatiser la retranscription d'audiences judiciaires, le dépôt de plaintes ou les comptes rendus médicaux. Comme un symbole, en décembre 2023, une enquête du *Monde*^[73] révélait que l'algorithme (anonyme) de la Caisse nationale des allocations familiales, chargé d'attribuer à chacun de ses 13 millions d'allocataires un «score de risque» de fraude (qui est en réalité un score de suspicion), ciblait et pénalisait les ménages aux plus faibles revenus. Il ne s'agissait pas d'un bogue, c'est le fonctionnement normal du système. Ne nous y trompons pas: le *chatbot* Albert, malgré son prénom sympa, est lui aussi conçu pour faire la chasse aux pauvres, simplement, efficacement et sans jour de congé. Malheur aux faux positifs, l'indifférence administrative les attend.

On l'a vu: l'IA est essentiellement une technique de classement. Elle ne propose que des solutions bêtement techniques à des problèmes sociaux. Faisant mine d'oublier que classer, c'est dominer, comme le dit la sociologue et féministe Christine Delphy^[74], elle crée des états d'exception, recalcule sans cesse le droit de chacun à exister^[75].

Lorsqu'elle tombe entre les mains du gouvernement, elle impose ce que Marion Fourcade et Kieran Healy appellent une «société ordinale^[76]», bâtie sur des classements et des notes, une ségrégation et une «méritocratie» par le score. Elle transforme notre monde social en un modèle réduit statistique qui pénalise systématiquement les mal-notés. Ce sont celles et ceux que Marion Fourcade appelle le «lumpenscoretariat» du capitalisme numérique, les marginaux et les déclassés, qui n'entrent pas dans les cases du formulaire et se retrouvent sans ressources face à une discrimination individuelle sur mesure. Ce qu'introduit l'IA, c'est une nouvelle ligne de front de la lutte des classes.

Pour citer Dan McQuillan, l'IA est aussi un «thatchérisme informatique^[77]». Son déploiement accélère la précarisation du travail, la privatisation des services publics, l'état de surveillance et la dégradation des relations sociales. Pendant que les prophètes pointent du doigt la menace impossible d'une superintelligence cosmique, ici et maintenant, dans le privé et le public, des logiciels d'analyse de données dégradent les conditions d'existence d'individus et de groupes sociaux entiers.

La majorité de la population n'utilise pas d'intelligence artificielle, mais elle y est déjà assujettie. Tous les jours, des systèmes automatiques interviennent sur notre réalité et déterminent notre accès à certaines ressources, de manière à la fois opaque et invisible. Certains sont relativement inoffensifs – les algorithmes de recommandations de YouTube ou Spotify, qui agencent le contenu de nos flux d'actualité –, d'autres déterminent des conditions d'existence – l'algorithme qui détermine notre éligibilité à un prêt bancaire, une assurance, une aide sociale. Aucun d'entre eux n'est fiable, transparent ou interprétable. Nous vivons tous et toutes à la merci de l'erreur de calcul sans recours.

Dans les mains des chasseurs de coûts des entrepôts ou des *start-up*, le logiciel devient un contremaître qui surveille les travailleurs et maximise l'extraction de leur labeur. Interface muette et soi-disant neutre entre ceux qui décident et ceux qui subissent, elle déshumanise un peu plus les rapports de classe. Après avoir fait ses preuves chez le précurseur Amazon, exemple le plus frappant de cette gestion panoptique^[78] où les embauches, les cadences, les objectifs journaliers et les licenciements sont fixés par des machines, le *management* algorithmique s'étend^[79]

désormais de la grande distribution aux tours de La Défense, en passant par les coursiers d'Uber et Deliveroo. En attendant la société de l'oisiveté dans des Disneyland orbitaux, des millions de microtravailleurs, qui forment le prolétariat de l'IA cartographié par le chercheur Antonio Casilli^[80], travaillent sous la scrutation inflexible des algorithmes. Partout où il est utilisé, le logiciel est le faux-nez du patron et de l'actionnaire, dont la seule ambition est d'augmenter les marges de profit aux dépens des travailleurs.

Enfin, l'industrie de l'IA produit son propre type de prolétariat, une internationale de travailleurs^[81] qui collectent, organisent et annotent des ensembles de données. Cette tâche répétitive représente 80 % du temps requis pour le développement d'un modèle comme GPT-4^[82]. Jeff Bezos, qui a nommé sa plateforme de microtravail Amazon Mechanical Turk en référence au célèbre canular du Turc mécanique du XVIII^e siècle, sait très bien qu'il vend de l'«intelligence artificielle artificielle». L'industrie de la «fauxtomatisme» représentait un marché de deux milliards de dollars en 2023. Sa capitale est Nairobi, mais ce continent invisible, qui se déplace au gré des courants de la mondialisation, s'étend aux Philippines, au Mexique, au Venezuela, en Inde et jusque dans les camps de réfugiés au Liban. En janvier 2023, *Time* révélait qu'une armée de travailleurs kenyans, payés entre 1,30 et 2 dollars de l'heure, modérait les réponses fournies par ChatGPT^[83], au détriment de leur santé mentale et émotionnelle^[84]. Quand les travailleurs pauvres se font rares, les exploiters deviennent créatifs: en Finlande, ce sont les détenus qui sont enrôlés pour faire le sale boulot de l'IA^[85]. Ce n'est pas un bogue, c'est une architecture: c'est ce que la Silicon Valley appelle l'apprentissage par renforcement basé sur l'intervention humaine (RLHF) et qu'elle considère comme la grande avancée technique du moment.

Le gigantesque outil de navigation de l'information que le chercheur Matteo Pasquinelli appelle «nooscope» repose sur l'exploitation continue d'un prolétariat invisible^[86]. Comme tout système intrinsèquement capitaliste, il est «accro à l'expansion géographique» et à la division internationale du travail. Aux États-Unis les *data scientists* et les *hacker houses* communales, les PDG-messies milliardaires et les montagnes de

capital-risque; dans les pays du Sud, les cols bleus qui font réellement fonctionner la machine, pour un à deux dollars de l'heure. Cette distribution produit des effets inattendus et révélateurs: ChatGPT et les autres robots conversationnels affectionnent particulièrement le verbe *to delve* (fouiller), rarement utilisé aux États-Unis, mais très présent dans l'anglais nigérian^[87]. Pire encore, dans un vertige digne de Philip K. Dick, les êtres humains payés à la tâche pour surveiller des logiciels qui simulent l'intelligence humaine... automatisent leur travail grâce à ces logiciels^[88]. Au terminus du simulacre, les machines parlent seules aux machines.

L'intelligence artificielle est une discrimination

À sa tête, l'industrie de l'IA occidentale emploie majoritairement des hommes blancs de moins de 40 ans^[89]. Ce qui ne va pas sans poser de «problèmes^[90]». Cette population homogène, en position dominante dans la hiérarchie sociale, construit des systèmes informatiques structurellement biaisés à l'encontre des femmes et des populations non blanches. Depuis le début des années 2010, les scandales autour de l'apprentissage machine, particulièrement dans le champ de la vision par ordinateur, se sont multipliés. En 2015, l'algorithme de reconnaissance faciale de Google identifie les personnes noires comme des «gorilles^[91]». Depuis, rien n'a changé: «Google a un passif de biais vis-à-vis des femmes noires», explique Safiya Noble, sociologue et autrice d'*Algorithms of Oppression* (Les algorithmes de l'oppression). La seule fois où l'entreprise a tenté d'intervenir sur le code pour plus de diversité, son logiciel Gemini s'est mis à créer des soldats nazis noirs, latinos et asiatiques^[92].

Il n'y a pas que Google. Aujourd'hui encore, la reconnaissance faciale automatisée ne sait pas différencier deux personnes noires – pour la machine, elles sont toutes pareilles^[93]! Selon une étude du MIT publiée en 2018, le taux d'erreur d'identification tournait autour de 1 % pour les hommes blancs; pour les femmes noires, il était de 34 %^[94]. Les systèmes se trompent 10 à 100 fois plus souvent avec les individus noirs

et asiatiques qu’avec les individus blancs^[95]. Dans les algorithmes utilisés par la police et par la justice, le seul fait d’être noir est un facteur de risque criminel^[96]. Aux États-Unis, le déploiement de l’intelligence artificielle a installé un véritable continuum de l’injustice raciale, désormais largement automatisé.

Année après année, les fabricants d’IA ont répété que c’était la faute des données mal labellisées ou pas assez diverses, mais certainement pas de leurs beaux algorithmes neutres et apolitiques. Grâce aux travaux de Ruha Benjamin, de Joy Buolamwini et de l’Algorithmic Justice League, mais aussi à ceux de Meredith Whittaker, Sarah Myers West et Kate Crawford pour l’AI Now Institute^[97], on sait que c’est faux. Cette discrimination raciale n’est ni un hasard ni un bogue: les systèmes d’IA font le travail pour lequel ils sont conçus – un travail d’identification où plane le spectre de la phrénologie, une pseudoscience raciste du XIX^e siècle qui affirmait que l’intelligence se lisait dans la forme du crâne^[98]. Dans une société structurellement raciste, ils encodent et automatisent les asymétries raciales. Et ça ne se résoudra pas avec une base de données «inclusive».

En 2024, rien n’a changé. Tous les générateurs d’images actuels amplifient les stéréotypes de genre^[99] et de race. Tous les générateurs d’images sont orientalistes^[100]. De la même manière, dans une société structurellement patriarcale, où les femmes représentent 22 % des professionnels de l’IA (et ne captent que 2 % du capital-risque), les machines à *bullshit* créées, entraînées et testées par les hommes produisent des récits misogynes à la chaîne^[101]. Sur LinkedIn et Instagram, les systèmes de modération automatique vendus par Microsoft et Google censurent et objectifient les photos de femmes, automatiquement labellisés comme de la pornographie^[102].

Ces systèmes sont tellement doués pour reproduire des délires xénophobes, misogynes et toxiques que l’extrême droite française, qui nage déjà dans un monde imaginaire, s’en sert pour inonder les réseaux sociaux d’images artificielles racistes^[103]. Comme l’idéologie d’extrême droite, l’IA échoue à représenter le monde. Elle ne fonctionne que par archétypes et biais, par catégorisations *a priori*. Lorsqu’elle confond les visages noirs entre eux et qu’elle fait de l’ethnicité un handicap dans le

système judiciaire, lorsqu'elle sexualise par défaut les corps féminins, lorsqu'elle représente uniquement des Mexicains en sombrero et des Japonaises en tenue de geisha, elle produit une différence – elle rappelle à ses sujets la distance qui les sépare de la «norme» masculine et blanche, éternellement désirable. Elle est en cela, pour reprendre Yarden Katz^[104], une «prothèse pour le maintien de l'ordre social racial et l'avancée des projets capitalistes et impérialistes».

Déployée dans l'espace public avec la promesse terrifiante d'identifier des attitudes jugées «suspectes», comme c'est désormais le cas en France sous prétexte de la sécurisation de la ville de Paris pour les Jeux olympiques, elle devient un auxiliaire du biopouvoir – un outil qui dicte aux corps subalternes, perpétuellement suspects, où aller, comment se déplacer, quelles zones et quelles attitudes proscrire. Des systèmes aussi violents, aussi excluants et aussi puissants devraient être interdits, comme le suggèrent Amnesty International^[105], la Quadrature du Net^[106] et une galaxie d'ONG de défense des droits fondamentaux; ils sont au contraire en train de se normaliser.



La maison coloniale de Monticello, en Virginie, ne ressemble à aucune autre. Elle a été dessinée par Thomas Jefferson, qui y a vécu une partie de sa vie. Le père fondateur des États-Unis y exploite 200 esclaves, tout en militant publiquement pour l'abolition de l'esclavage. La maison est entièrement conçue autour de cette contradiction, que Jefferson résout grâce à l'interface technique. Lorsqu'il reçoit des intellectuels libéraux pour débattre de la philosophie des Lumières, ses invités voient des pans de mur bouger tout seuls pour faire apparaître les plats, comme par magie. En réalité, un système de monte-plats permet de cacher les esclaves en coulisses. La maison est un trompe-l'œil, comme la façade progressiste de l'esclavagiste Jefferson. Elle est une scène sur laquelle des élites bourgeoises s'extasient devant le *deus ex machina*, pendant que des esclaves invisibles actionnent des poulies à la force des bras. Deux siècles plus tard, dans la salle à manger de Monticello, nous

débattons à nouveau d'une technique qui n'existe pas, bernés par une interface qui masque la brutalité du réel.

Dès que l'on s'extrait du récit religieux de l'intelligence artificielle, une seule chose apparaît: la violence algorithmique. La violence omniprésente, systémique, banale, structurelle. La violence comme grammaire et comme modèle d'affaires. Le mythe de l'intelligence artificielle dissimule une entreprise globale de déshumanisation de l'individu, d'atomisation du collectif et de stérilisation de la vie politique par leurs réductions à des modèles statistiques. Résistons à la tentation de décrire au conditionnel les éventuels bénéfiques de la super-IA: admettons que ce «futur» est un mirage publicitaire, alors que la violence de ces systèmes est non seulement réelle, mais empiriquement démontrée. Ces logiciels, produits par des patrons réactionnaires et déployés sur un système en bout de course de plus en plus tenté par le fascisme, ne vont pas améliorer nos conditions d'existence. Soyons comme les tisserands anglais du XIX^e siècle, destructeurs de machines Jacquard. Ne refusons pas la machine; refusons la précarisation, la déqualification, la paupérisation, la violence éternelle de la caste patronale et de l'administration néolibérale qu'elle tente de camoufler.

Changeons de perspective. Au lieu des prophéties invraisemblables de «parrains de l'IA» et des capital-risqueurs en roue libre, écoutons les témoignages des corps à la marge, de celles et ceux qui subissent déjà largement la violence algorithmique. Adoptons l'approche critique radicale, féministe et décoloniale portée par les technocritiques contemporaines. Celles qui éprouvent la violence algorithmique sont les mieux placées pour en percevoir les rouages, et produire les contre-récits de l'IA indispensables à son désamorçage. Elles savent, et nous savons désormais, que l'IA n'est ni intelligente, ni artificielle, ni futuriste, ni bénéfique à l'humanité. Si la Silicon Valley essaie de nous vendre l'apocalypse, c'est parce que son projet technique, économique et politique en est une – la domination ou l'autodestruction, également totales. L'IA n'est pas la fin du monde, mais les capitalistes autoritaires qui la développent veulent sans aucun doute la fin du monde social. Elle n'annonce pas l'avènement du posthumain, mais instaure déjà, par petites touches, le régime de l'inhumain statistique. Face à cette négation

du réel, notre réponse ne peut être qu'une affirmation technocritique.
Une intelligence organique, irréductible et insaisissable.

Remerciements

Fabriquer un livre pour la première fois est une montagne qui ne se gravit qu'à plusieurs, encordé aux critiques et aux bienveillances.

Ma première pensée va donc à Mark Fortier de Lux Éditeur, qui s'est lancé avec moi dans cette drôle d'ascension, un après-midi de juin, et a passé six mois à me partager sa passion pour l'étrange et merveilleux métier d'auteur. Elle va ensuite à Alexandre Sánchez, dont le regard impitoyable et judicieux a ciselé mon texte indécis. Merci pour le temps et l'attention que vous m'avez consacrés, et d'avoir fait de ce livre ce qu'il est, collectivement.

Merci ensuite, pêle-mêle, à Pauline pour ses relectures implacables, sa chasse aux sigles et ses approbations qui structurent, surtout dans la tempête et les doutes; à Thibaud et Camille pour avoir ouvert une île à un auteur monomaniac au fond de la mine; merci à toutes les universitaires, chercheuses, technocritiques, militantes, combattantes et luddites, géantes de la pensée révolutionnaire, féministe et décoloniale de la technique, sur les épaules desquelles repose cet essai. Merci à Myriam, ma *partner in crime*, dont l'amour, le cœur et l'intellect m'envoient défier les gravités du monde. Merci à Paul, mon frère, qui ne cesse de me montrer comment devenir. Je vous aime. À l'heure de remettre ce livre se jouent des drames qui m'incitent à ajouter ce cri au milieu de mes remerciements: «*Free Palestine.*» Cela va sans dire, cet ouvrage est garanti 100 % organique: son auteur luddite n'a utilisé aucun outil d'automatisation, pas même un correcteur automatique.

Notes et références

État des lieux

- [1] «Pause Giant AI Experiments», Future of Life Institute, 22 mars 2023.
- [2] Pilli.Lalitha Dura Lavanya, «Brief History of Deep Learning», *Medium*, 17 juillet 2023.
- [3] John Markoff, «Scientists See Promise in Deep-Learning Programs», *The New York Times*, 23 novembre 2012.
- [4] Sebastien Bubeck *et al.*, «Sparks of Artificial General Intelligence: Early Experiments with GPT-4», *arXiv*, 22 mars 2023.
- [5] Nitasha Tiku, «The Google Engineer Who Thinks the Company's AI Has Come to Life», *The Washington Post*, 11 juin 2022.
- [6] Tad Friend, «Sam Altman's Manifest Destiny», *The New Yorker*, 3 octobre 2016.
- [7] Ibid.
- [8] Sam Altman, «Machine Intelligence, Part 1», blogue personnel, 25 février 2015.
- [9] Steven Levy, «What OpenAI Really Wants», *Wired*, 5 septembre 2023.
- [10] Sam Altman, «Successful People», blogue personnel, 8 mars 2013.
- [11] Cade Metz, «The ChatGPT King Isn't Worried, but He Knows You Might Be», *The New York Times*, 31 mars 2023.
- [12] Matt Reynolds, «The AI Doomsday Bible Is a Book About the Atomic Bomb», *Wired*, 8 août 2023.
- [13] Billy Perrigo, «DeepMind's CEO Helped Take AI Mainstream. Now He's Urging Caution», *Time*, 12 janvier 2023.
- [14] «The Economic Potential of Generative AI: The Next Productivity Frontier», McKinsey Digital, 14 juin 2023.
- [15] Steven Levy, «Microsoft's Satya Nadella Is Betting Everything on AI», *Wired*, 13 juin 2023.
- [16] Cecilia Kang, «How Sam Altman Stormed Washington to Set the A.I. Agenda», *The New York Times*, 7 juin 2023.
- [17] «Oversight of A.I.: Rules for Artificial Intelligence», US Senate Committee, 16 mai 2023.
- [18] Rishi Iyengar, «OpenAI's CEO Goes on a Diplomatic Charm Offensive», *Foreign Policy*, 20 juin 2023.
- [19] Sam Altman, Greg Brockman et Ilya Sutskever, «Governance of Superintelligence», OpenAI, 22 mai 2023.

- [20] António Guterres, «Secretary-General's Opening Remarks at Press Briefing on Policy Brief on Information Integrity on Digital Platforms», site de l'Organisation des Nations Unies, 12 juin 2023.
- [21] Tweet de la Commission européenne, Twitter/X, 14 septembre 2023, 8 h 19.
- [22] David Kushner, «How A.I. Could Reincarnate Your Dead Grandparents – Or Wipe out Your Kids», *Rolling Stone*, 4 septembre 2023.
- [23] Marc Andreessen, «Why AI Will Save the World», Andreessen Horowitz, 6 juin 2023.
- [24] Cade Metz *et al.*, «Ego, Fear and Money: How the A.I. Fuse Was Lit», *The New York Times*, 3 décembre 2023.
- [25] «Intelligence artificielle: 67 % des Français la perçoivent comme une menace, selon un sondage», *Sud Ouest*, 13 juin 2023.
- [26] Hélène Jouan, «À Montréal, l'un des pères de l'intelligence artificielle alerte sur une menace existentielle pour l'homme», *Le Monde*, 2 juin 2023.
- [27] «2022 Expert Survey on Progress in AI», *AI Impacts*, 4 août 2022.
- [28] Günther Anders, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, Paris, Encyclopédie des nuisances / Ivrea, 2002 [1956], p. 37.
- [29] Clémence Mary, «Après l'éco-anxiété, l'«IA-anxiété»?», *Libération*, 19 juin 2023.
- [30] Timothy Morton, *Hyperobjects: Philosophy and Ecology after the End of the World*, Minneapolis, University of Minnesota Press, coll. «Posthumanities», 2013.
- [31] Melanie Mitchell, «Why AI Is Harder than We Think», *arXiv*, 26 avril 2021.
- [32] Morgane Tual, «Stephen Hawking et Elon Musk réclament l'interdiction des “robots tueurs”», *Le Monde*, 27 juillet 2015.
- [33] Emily M. Bender *et al.*, «On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?», ACM Digital Library, 1er mars 2021.
- [34] Noam Chomsky, «The False Promise of ChatGPT», *The New York Times*, 8 mars 2023.
- [35] Matteo Wong, «AI Doomerism Is a Decoy», *The Atlantic*, 2 juin 2023.
- [36] Lee Vinsel, «You're Doing It Wrong: Notes on Criticism and Technology Hype», *Medium*, 1er février 2021.
- [37] Tweet de Suresh Venkatasubramanian, Twitter/X, 1er avril 2023, 12 h 50.
- [38] Tweet de Suresh Venkatasubramanian, Twitter/X, 1er avril 2023, 12 h 51.
- [39] Sam Altman, «The Merge», blogue personnel, 7 décembre 2017.
- [40] David F. Noble, *The Religion of Technology: The Divinity of Man and the Spirit of Invention*, New York, Alfred A. Knopf, 1997, p. 9.
- [41] Nicolas Papernot, «SaTML 2023 – Timnit Gebru – Eugenics and the Promise of Utopia through AGI», YouTube, 16 février 2023.
- [42] Émile P. Torres utilise les pronoms non binaires *they/them*.

Chapitre 1

- [1] Julian S. Huxley, *Religion Without Revelation*, New York, Harper & Brothers, 1927.
- [2] Julian S. Huxley, *New Bottles for New Wine*, Londres, Chatto and Windus, 1957, p. 17.
- [3] Max More, «The Philosophy of Transhumanism», dans Max More et Natasha Vita-More (dir.), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, Chichester, Wiley-Blackwell, 2013, p. 3-17.
- [4] Natasha Vita-More, «The Transhumanist Manifesto», Humanity+, s. d. [1983].
- [5] Anna Wiener, «Only Human», *The New Republic*, 16 février 2017.
- [6] Ed Regis, «Meet The Extropians», *Wired*, 1er octobre 1994.
- [7] Jim McClellan, «The Tomorrow People», *UK Observer*, 26 mars 1995.
- [8] Ibid.
- [9] Steve Mollman, «Billionaire Peter Thiel Says He's Signed up to Be Cryogenically Preserved When He Dies So He Can Be Revived in the Future – But He's "Not Convinced It Works"», *Fortune*, 5 mai 2023.
- [10] Eliezer Yudkowsky, «You Only Live Twice», *LessWrong*, 12 décembre 2008.
- [11] Frank Bi, «The Eternal Promise», *The Verge*, s. d. [6 novembre 2015].
- [12] Nick Bostrom, *The Transhumanist FAQ: A General Introduction*, Wilton (CT), World Transhumanist Association, 2003.
- [13] Benoît Georges, «Nick Bostrom: "Le but ultime de l'intelligence artificielle doit être la disparition du travail"», *Les Échos*, 26 juin 2017.
- [14] Nick Bostrom, «Existential Risk Prevention as Global Priority», *Global Policy*, vol. 4, no 1, février 2013.
- [15] Bostrom, *The Transhumanist FAQ*, op. cit.
- [16] World Economic Forum, «Will the Future be Human? – Yuval Noah Hariri», YouTube, 25 janvier 2018.
- [17] Steve Wilstein, «Laid-Back Summitry: "Hot Tub Diplomacy" Takes the Chill off East-West Relations», *Los Angeles Times*, 15 février 1987.
- [18] Carolyn Che, *Work Pray Code: When Work Becomes Religion in Silicon Valley*, Princeton, Princeton University Press, 2022.
- [19] SXSW, «Elon Musk on Human Life Beyond Earth», YouTube, 25 février 2019.
- [20] Robert Hart, «Forget Musk's Martian Ambition – Jeff Bezos Thinks Humans Should Live In Giant Cylindrical Space Stations», *Forbes*, 15 décembre 2023.
- [21] Betsy Isaacson, «Silicon Valley Is Trying to Make Humans Immortal – and Finding Some Success», *Newsweek*, 3 mai 2015.
- [22] Mike Brown, «Mark Zuckerberg: Humans Living Past 100 "Will Be Pretty Normal"», *Inverse*, 11 novembre 2016.
- [23] Anthony Cuthbertson, «Billionaire Trump Supporter Peter Thiel Denies Being a Vampire», *The Independent*, 2 novembre 2018.
- [24] Mick Brown, «Peter Thiel: The Billionaire Tech Entrepreneur on a Mission to Cheat Death», *The Telegraph*, 19 septembre 2014.

- [25] Ariana Eunjung Cha, «Tech Titans' Latest Project: Defy Death», *The Washington Post*, 4 avril 2015.
- [26] Antonio Regalado, «A Contrarian in Biotech», *MIT Technology Review*, 16 mars 2015.
- [27] Eleanor Garth, «Global Longevity Investment Hit \$5.2 Billion in 2022», *Longevity*, 31 janvier 2023.
- [28] Christina Farr, «Silicon Valley Techies Are Turning to a Cheap Diabetes Drug to Help Them Live Longer», CNBC, 23 mars 2019.
- [29] Ellen Huet, «Silicon Valley's Quest to Live Forever Has Many Warring Factions», Bloomberg, 19 décembre 2023.
- [30] Compte Twitter/X de Balaji.
- [31] Sam Altman, «Helion», blogue personnel, 5 novembre 2021.
- [32] Tad Friend, «Sam Altman's Manifest Destiny», *The New Yorker*, 3 octobre 2016.
- [33] Sam Altman, «The Merge», blogue personnel, 7 décembre 2017.
- [34] Tweet de Sam Altman, Twitter/X, 4 février 2022, 12 h 13.
- [35] Josh Marcus, «Elon Musk Supports Eliminating Voting Rights for People without Children», *The Independent*, 4 juillet 2023.
- [36] Miles Klee, «Elon Musk All but Endorses the Great Replacement Conspiracy Theory», *Rolling Stone*, 5 janvier 2024.
- [37] Taylor Telford, «Elon Musk Wants to Get inside Your Head. Literally», *The Washington Post*, 17 juillet 2019.
- [38] Joshua Benton, «Elon Musk's Anti-Semitic, Apartheid-Loving Grandfather», *The Atlantic*, 20 septembre 2023.
- [39] Ira Basen, «In Science We Trust», CBC, 28 juin 2021.
- [40] Catherine Jewell, «L'éthique, la technologie et l'avenir de l'humanité», *OMPI Magazine*, août 2018.
- [41] Peter Singer, «Famine, Affluence, and Morality», *Philosophy & Public Affairs*, vol. 1, no 3, printemps 1972.
- [42] Nat Hentoff, «A Professor Who Argues for Infanticide», *The Washington Post*, 11 septembre 1999.
- [43] Will Crouch, «Doctor, NGO Worker, or Something Else Entirely? Which Careers Do the Most Good?», présentation à Oxford, 4 avril 2019.
- [44] Linsey McGoey, «Elite Universities Gave Us Effective Altruism, the Dumbest Idea of the Century», *Jacobin*, 19 janvier 2023.
- [45] Tweet du *New Yorker*, Twitter/X, 9 août 2022, 12 h 44.
- [46] «What to Read to understand Effective Altruism», *The Economist*, 24 août 2022.
- [47] William MacAskill, «Effective Altruism and the Current Funding Situation», 80,000 Hours, 16 mai 2022.
- [48] Gideo Lewis-Kraus, «The Reluctant Prophet of Effective Altruism», *The New Yorker*, 8 août 2022.

- [49] Linda Kinstler, «The Good Delusion: Has Effective Altruism Broken Bad?», *The Economist*, 15 novembre 2022.
- [50] Tweet d'Émile P. Torres, Twitter/X, 1er mai 2024, 23 h 08.
- [51] Nicholas Beckstead, «On the Overwhelming Importance of Shaping the Far Future», dissertation, State University of New Jersey, mai 2013.
- [52] Karthik Tadepalli, «Valuing Lives Instrumentally Leads to Uncomfortable Conclusions», Effective Altruism Forum, 4 septembre 2022.
- [53] William MacAskill, «Longtermism», Effective Altruism Forum, 25 juillet 2019.
- [54] Nick Bostrom, «Existential Risks: Analyzing Human extinction Scenarios and Related Hazards», *Journal of Evolution and Technology*, vol. 9, mars 2002.
- [55] McGoey, «Elite Universities Gave Us Effective Altruism, the Dumbest Idea of the Century», *loc. cit.*
- [56] Christine Emba, «Why “Longtermism” Isn’t Ethically Sound», *The Washington Post*, 5 septembre 2022.
- [57] Tweet d'Elon Musk, Twitter/X, 2 août 2022, 1 h 15.
- [58] Fiona Harvey, «Boris Johnson’s Climate Speech Annotated: What He Said and What He Meant», *The Guardian*, 23 septembre 2021.

Chapitre 2

- [1] Peter Thiel, «The Education of a Libertarian», *Cato Unbound*, 13 avril 2009.
- [2] Tad Friend, «Tomorrow’s Advance Man», *The New Yorker*, 11 mai 2015.
- [3] Max Chafkin, *The Contrarian: Peter Thiel and Silicon Valley’s Pursuit of Power*, New York, Penguin, 2021.
- [4] Ben Horowitz, «Politics and The Future», Andreessen Horowitz, 14 décembre 2023.
- [5] Margaret O’Mara, «The Church of Techno-Optimism», *The New York Times*, 28 septembre 2019.
- [6] Asma Mhalla, «Les géants de la technologie sont aussi des acteurs politiques, idéologiques, militaires», *L’invité de 7 h 50 du week-end*, France Inter, 17 février 2024; *Technopolitique. Comment la technologie fait de nous des soldats*, Paris, Seuil, coll. «Documents», 2024.
- [7] Douglas Rushkoff, «“We Will Coup Whoever We Want!”: The Unbearable Hubris of Musk and the Billionaire Tech Bros», *The Guardian*, 25 novembre 2023.
- [8] Katja Grace *et al.*, «Thousands of AI Authors on the Future of AI», prépublication, janvier 2024.
- [9] «Sizing the Price: What’s the Real Value of AI for Your Business and How Can You Capitalise?», PriceWaterhouseCooper, 2017.
- [10] Ajeya Cotra, «Draft Report on AI Timelines», *AI Alignment Forum*, 18 septembre 2020.
- [11] Eric Lamarre *et al.*, «A Generative AI Reset: Rewiring to Turn Potential into Value in 2024», McKinsey, 4 mars 2024.

- [12] George Steer *et al.*, «AI Boom Drives Global Stock Markets to Best First Quarter in 5 Years», *Financial Times*, 28 mars 2024.
- [13] Achille Mbembe, *Brutalisme*, Paris, La Découverte, 2020, p. 23.
- [14] Dan McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*, Bristol, Bristol University Press, 2022, p. 73.
- [15] Marc Andreessen, «The Techno-Optimist Manifesto», Andreessen Horowitz, 16 octobre 2023.
- [16] «World Child Hunger Facts», *World Hunger News*, s. d. [14 juillet 2018].
- [17] Chris Katje, «AI Taking Jobs Could Benefit Economy, Says Marc Andreessen: “It’s A Process Of Change And People Get Freaked Out”», *Benzinga*, 11 juillet 2023.
- [18] Connor Axiotes, Nikhil Woodruff et Scott Santens, «Basic Income and AI-induced Unemployment», Adam Smith Institute, 3 mai 2023.
- [19] Paige McGlaflin et Joseph Abrams, «Elon Musk Says AI Will Remove Need for Jobs and Create “Universal High Income.” But Workers Don’t Want to Wait for Robots to Get Financial Relief», *Fortune*, 6 novembre 2023.
- [20] Martin-Brehm Christensen *et al.*, *La loi du plus riche. Pourquoi et comment taxer les plus riches pour lutter contre les inégalités*, Oxford, Oxfam International, janvier 2023.
- [21] Alexandre Lacroix, «Alain Damasio-Antoinette Rouvroy. Passer entre les mailles», *Philosophie magazine*, 25 septembre 2019.
- [22] Cédric Durand, *Technoféodalisme. Critique de l’économie numérique*, Paris, La Découverte, coll. «Zones», 2020.
- [23] David Moscrop, «Are We Transitioning From Capitalism to Silicon Serfdom? An Interview with Yanis Varoufakis», *Jacobin*, 18 février 2024.
- [24] Hubert Guillaud, «Réguler le “capitalisme de commandement”», blogue d’Hubert Guillaud, 16 novembre 2023.
- [25] Piotr Smolar, «Elon Musk, imprévisible acteur de la guerre en Ukraine», *Le Monde*, 9 septembre 2023.
- [26] Cat Zakrzewski, «Democrats Probe Musk’s SpaceX, Examining Russia’s Alleged Starlink Use», *The Washington Post*, 7 mars 2024.
- [27] Tom Simonite, «3 Years After the Project Maven Uproar, Google Cozies to the Pentagon», *Wired*, 18 novembre 2021.
- [28] Sarah Emerson, «OpenAI’s Lobbyists Are Targeting the Pentagon and Other Defense Agencies», *Forbes*, 7 février 2024.
- [29] Kara Swisher, «How Trump Pushed Silicon Valley off the Rails», *The Washington Post*, 15 février 2024.
- [30] Evan Osnos, «Can Mark Zuckerberg Fix Facebook Before It Breaks Democracy?», *The New Yorker*, 10 septembre 2018.
- [31] Kate Losse, «I Was Zuckerberg’s Speechwriter. “Companies Over Countries” Was His Early Motto», *Vox*, 16 avril 2018.
- [32] Alan Jacobs, «The Ambiguous Utopia of Iain M. Banks», *The New Atlantis*, été 2009.

- [33] Alexandre Piquard, «Jeff Bezos rêve d'envoyer l'humanité dans l'espace», *Le Monde*, 11 juin 2021.
- [34] Barton Gellman, «Peter Thiel Is Taking a Break From Democracy», *The Atlantic*, 9 novembre 2023.
- [35] Ariana Eunjung Cha, «Tech Titans' Latest Project: Defy Death», *The Washington Post*, 4 avril 2015.
- [36] Thom Geier, «Peter Thiel Thinks “Star Trek” Is “Communist” While “Star Wars” Is “Capitalist”», *The Wrap*, 11 janvier 2017.
- [37] George Packer, «No Death, No Taxes», *The New Yorker*, 20 novembre 2011.
- [38] Maureen Dowd, «Peter Thiel, Trump's Tech Pal, Explains Himself», *The New York Times*, 11 janvier 2017.
- [39] Kara Swisher, *Burn Book: A Tech Love Story*, New York, Simon and Schuster, 2024, p. 54.
- [40] Mark O'Connell, «“A Managerial Mephistopheles”: Inside the Mind of Jeff Bezos», *The Guardian*, 3 février 2021.
- [41] James Pogue, «Inside the New Right, Where Peter Thiel Is Placing His Biggest Bets», *Vanity Fair*, 20 avril 2022.
- [42] Mencius Moldbug, «The Case Against Democracy: Ten Red Pills», *Unqualified Reservations*, 24 avril 2017.
- [43] Tweet d'Elon Musk, Twitter/X, 17 mai 2020, 1 h 44.
- [44] Zack Beauchamp, «Accelerationism: The Obscure Idea Inspiring White Supremacist Killers Around the World», *Vox*, 18 novembre 2019.
- [45] Andreessen, «The Techno-Optimist Manifesto», *loc. cit.*
- [46] Hallam Stevens, «A New Silicon Valley Manifesto Reveals the Bleak, Dangerous Philosophy Driving the Tech Industry», *The Conversation*, 5 novembre 2023.
- [47] Tad Friend, «Sam Altman's Manifest Destiny», *The New Yorker*, 3 octobre 2016.
- [48] Nick Bostrom, «Apology for an Old Email», site web personnel, 9 janvier 2023.
- [49] Jamelle Bouie, «Why an Unremarkable Racist Enjoyed the Backing of Billionaires», *The New York Times*, 12 août 2023.
- [50] Quinn Slobodian, «The Rise of the New Tech Right», *The New Statesman*, 13 septembre 2023.
- [51] Yarden Katz, «Intelligence Under Racial Capitalism», *Monthly Review*, vol. 74, no 4, septembre 2022, p. 32-52.
- [52] Matt Reynolds, «Elon Musk Is Totally Wrong about Population Collapse», *Wired*, 6 octobre 2022.
- [53] «The Fertility Sector Is Booming», *The Economist*, 17 juillet 2023.
- [54] Sophie Alexander, Dana Hull et Graham Starr, «Musk Is Funding Fertility, Population Research in Texas Project», *Bloomberg*, 14 août 2023.
- [55] Ishaan Tharoor, «Elon Musk Raises the Specter of “White Genocide”», *The Washington Post*, 1er août 2023.

[56] Greg Sargent, «Elon Musk Pushes a Vile, Toxic Hate Video – and Exposes His Own Scam», *The New Republic*, 21 mars 2024.

[57] Mark O’Connell, «“A Managerial Mephistopheles”: Inside the Mind of Jeff Bezos», *The Guardian*, 3 février 2021.

[58] Marc Andreessen, «The Techno-Optimist Manifesto», Andreessen Horowitz, 16 octobre 2023.

[59] David F. Noble, *The Religion of Technology: The Divinity of Man and the Spirit of Invention*, New York, Alfred A. Knopf, 1997, p. 2.

[60] Melia Robinson, «Inside the Most Expensive Town in America, Where Tech Moguls Live In Multimillion-Dollar Mansions and the Average Household Income Is over \$450,000», *Business Insider*, 14 février 2019.

[61] Conor Dougherty et Erin Griffith, «The Silicon Valley Elite Who Want to Build a City From Scratch», *The New York Times*, 25 août 2023.

[62] Melia Robinson, «A Silicon Valley Billionaire’s Dream Of a Floating Libertarian Utopia May Have Finally Been Killed», *Business Insider*, 8 mars 2018.

[63] Vitalik Buterin, «My Techno-Optimism», site web personnel, 27 novembre 2023.

[64] James Dale Davidson et lord William Rees-Mogg, *The Sovereign Individual: The Coming Economic Revolution, How to Survive and Prosper in It*, Londres, MacMillan, 1997.

[65] Gil Duran, «The Tech Baron Seeking to Purge San Francisco of “Blues”», *The New Republic*, 26 avril 2024.

[66] Tim Urban, «How (and Why) SpaceX Will Colonize Mars», *Wait But Why*, 16 août 2015.

[67] Paris Marx, «Jeff Bezos and His Billionaire Space Fantasy», *Jacobin*, 14 décembre 2019.

[68] Jordan S. Carroll, «To Understand Elon Musk, You Have to Understand This ’60s Sci-Fi Novel», *Jacobin*, 29 mai 2022.

[69] Marx, «Jeff Bezos and His Billionaire Space Fantasy», *loc. cit.*

[70] Jonathan Taplin, «How Musk, Thiel, Zuckerberg, and Andreessen – Four Billionaire Techno-Oligarchs – Are Creating an Alternate, Autocratic Reality», *Vanity Fair*, 22 août 2022.

Chapitre 3

[1] Katja Grace, «AI is not an Arms Race», *Time*, 31 mai 2023.

[2] «Senator Schumer Opening Statement», Washington, 8e AI Insight Forum, 6 décembre 2023.

[3] Kevin Roose, «Silicon Valley Confronts a Grim New A.I. Metric», *The New York Times*, 6 décembre 2023.

[4] Tyler Austin Harper, «The 100-Year Extinction Panic Is Back, Right on Schedule», *The New York Times*, 26 janvier 2024.

[5] Jeffrey Herf, *Le modernisme réactionnaire. Haine de la raison et culte de la technologie aux sources du nazisme*, Paris, L’échappée, coll. «Versus», 2018.

- [6] Sam Altman, «Machine Intelligence, Part 1», blogue personnel, 25 février 2015.
- [7] Dan Milmo, «Elon Musk Launches AI Startup and Warns of a “Terminator Future”», *The Guardian*, 13 juillet 2023.
- [8] Kevin Roose, «Inside the White-Hot Center of A.I Doomerism», *The New York Times*, 11 juillet 2023.
- [9] Will Douglas Heaven, «How Existential Risk Became the Biggest Meme in AI», *MIT Technology Review*, 19 juin 2023.
- [10] Émile P. Torres, «Human Extinction: A Brief Guided Tour of the Book», *Medium*, 15 juillet 2023.
- [11] Bill Joy, «Why The Future Doesn’t Need Us», *Wired*, 1er avril 2000.
- [12] Nick Bostrom, «The Future of Humanity», *Geopolitics, History, and International Relations*, vol. 1, no 2, 2009, p. 41-78.
- [13] Nick Bostrom, «Existential Risks: Analyzing Human extinction Scenarios and Related Hazards», *Journal of Evolution and Technology*, vol. 9, mars 2002.
- [14] Nick Bostrom, «Existential Risk Prevention as Global Priority», *Global Policy*, vol. 4, no 1, février 2013.
- [15] «What We Do», Berkeley Existential Risk Initiative, s. d.
- [16] Maxime Stauffer *et al.*, *Existential Risk and Rapid Technological Change: Advancing Risk-informed Development*, Genève, United Nations Office for Disaster Risk Reduction, 2023.
- [17] Tweet de la Commission européenne, Twitter/X, 14 septembre 2023, 8 h 15.
- [18] Sirkka Heinonen, «All We Need Is Futures Resilience – The Limits to Growth Revisited», Club de Rome, 6 septembre 2023.
- [19] Matt McFarland, «Elon Musk: “With Artificial Intelligence, We Are Summoning the Demon”», *The Washington Post*, 24 octobre 2014.
- [20] Nick Bostrom et Eliezer Yudkowsky, «The Ethics of Artificial Intelligence», dans Keith Frankish et William M. Ramsey (dir.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, 2014, p. 316-334.
- [21] Lauren Jackson, «What if A.I. Sentience Is a Question of Degree?», *The New York Times*, 12 avril 2023.
- [22] Rodrigo Ochigame, «The Invention of Ethical AI», *The Intercept*, 20 décembre 2019.
- [23] Michelle Donelan, «Ministerial Foreword», dans *Introducing the AI Institute*, Londres, Controller of His Majesty’s Stationery Office, 2023, p. 5.
- [24] Eliezer Yudkowsky, «Pausing AI Developments Isn’t Enough. We Need to Shut it All Down», *Time*, 29 mars 2023.
- [25] Nick Bostrom, «The Future of Human Evolution», dans Charles Tandy (dir.), *Death and Anti-Death*, t. 2, *Two Hundred Years After Kant, Fifty Years After Turing*, Ann Arbor, Ria University Press, coll. «Death and Anti-Death», 2004.
- [26] Nick Bostrom, «What is a Singleton?», *Linguistic and Philosophical Investigations*, vol. 5, no 2, 2006, p. 48-54.

- [27] Nick Bostrom, «The Vulnerable World Hypothesis», *Global Policy*, vol. 10, no 4, novembre 2019, p. 455-476.
- [28] Ibid.
- [29] Ibid.
- [30] Tweet de Sam Altman, Twitter/X, 3 février 2023, 16 h 28.
- [31] Sam Altman, Greg Brockman et Ilya Sutskever, «Governance of Superintelligence», OpenAI, 22 mai 2023.
- [32] Andrew Marantz, «Among the A.I. Domsayers», *The New Yorker*, 11 mars 2024.
- [33] Colin Allen, Iva Smit et Wendell Wallach, «Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches», *Ethics and Information Technology*, vol. 7, no 3, septembre 2005, p. 149-155.
- [34] Robert M. Geraci, «Spiritual Robots: Religion and Our Scientific View of the Natural World», *Theology and Science*, vol. 4, no 3, 2006, p. 229-246.
- [35] Markus Anderljung *et al.*, «Frontier AI Regulation: Managing Emerging Risks to Public Safety», OpenAI, 6 juillet 2023.
- [36] Will Knight, «A Letter Prompted Talk of AI Doomsday. Many Who Signed Weren't Actually AI Doomers», *Wired*, 17 août 2023.
- [37] Billy Perrigo, «Meta's AI Chief Yann LeCun on AGI, Open-Source, and AI Risk», *Time*, 13 février 2024.
- [38] Melanie Mitchell, «Debates on the Nature of Artificial General Intelligence», *Science*, 21 mars 2024.
- [39] Shazeda Ahmed *et al.*, «Building the Epistemic Community of AI Safety», Social Science Research Network, 22 novembre 2023.
- [40] remember, Andrea_Miotti, «Full Transcript: Elizer Yudkowsky on the Bankless Podcast», *LessWrong*, 23 février 2023.

Chapitre 4

- [1] Émile P. Torres, «Against Longtermism», *Aeon*, 19 octobre 2021.
- [2] Ibid.
- [3] «Comment fonctionne l'assurance-chômage, dotée de 45 milliards de budget annuel?», France Info, 9 novembre 2023.
- [4] Benjamin Todd, «Is Effective Altruism Growing? An Update on the Stock of Funding vs People», 80,000 Hours, 28 juillet 2021.
- [5] «The "Effective Altruism" Movement Is Louder than It Is Large», *The Economist*, 10 janvier 2024.
- [6] Parmy Olson, «There's Too Much Money Going to AI Doomers», *The Washington Post*, 16 août 2023.
- [7] Sarah Emerson, «OpenAI's Lobbyists Are Targeting the Pentagon and Other Defense Agencies», *Forbes*, 7 février 2024.
- [8] Karen Hao et Charlie Warzel, «Inside the Chaos at OpenAI», *The Atlantic*, 19 novembre 2023.

- [9] Nick Robins-Early, «Oxford Shuts Down Institute Run by Elon Musk-Backed Philosopher», *The Guardian*, 20 avril 2024.
- [10] «Computation Used to Train Notable Artificial Intelligence Systems», *Our World in Data*, 2024.
- [11] Jack Clark et Raymond Perrault (dir.), *Artificial Intelligence Index Report 2024*, Stanford, Institute for Human-Centered AI, Université Stanford, 2024, p. 64.
- [12] Alex Heath, «Inside the Hunt for AI Chips», *The Verge*, 11 août 2023.
- [13] «The Big Tech Company Leading in AI Acquisitions», CB Insights, 27 septembre 2023.
- [14] «Generative AI Raises Competition Concerns», Federal Trade Commission, 29 juin 2023.
- [15] Markus Anderljung *et al.*, «Frontier AI Regulation: Managing Emerging Risks to Public Safety», OpenAI, 6 juillet 2023.
- [16] Jyoti Mann, «Silicon Valley Is Going Wild over Satya Nadella’s “Baller” Comments Putting OpenAI in Its Place», *Business Insider*, 22 mars 2024.
- [17] Madhumita Murgia, «AI Academics under Pressure to Do Commercial Research», *Financial Times*, 13 mars 2019.
- [18] Clark et Perrault (dir.), *Artificial Intelligence Index Report 2024*, *op. cit.* p. 30.
- [19] Joseph Menn et Naomi Nix, «Big Tech Funds the Very People Who Are Supposed to Hold It Accountable», *The Washington Post*, 7 décembre 2023.
- [20] Christopher Newfield, «How to Make “AI” Intelligent; or, The Question of Epistemic Equality», *Critical AI*, vol. 1, nos 1-2, 1er octobre 2023.
- [21] Dave Paresh et Jeffrey Dastin, «Google Told Its Scientists to “Strike a Positive Tone” in AI Research», Reuters, 23 décembre 2020.
- [22] Kevin Roose, «A Conversation With Bing’s Chatbot Left Me Deeply Unsettled», *The New York Times*, 17 février 2023.
- [23] Hans Cools, Baldwin Van Gorp et Michael Opgenhaffen, «Where Exactly between Utopia and Dystopia? A Framing Analysis of AI and Automation in US Newspapers», *Journalism*, vol. 25, no 1, p. 3-21, 2024; Ching-Hua Chan, Wanhsiu Sunny Tsai et Su Yeon Cho, «Framing Artificial Intelligence in American Newspapers», Conférence AAAI/ ACM, janvier 2019.
- [24] Émile P. Torres, «Selling “Longtermism”: How PR and Marketing Drive a Controversial New Movement», *Salon*, 10 septembre 2022.
- [25] Julia Wise, Sky Mayhew et Aaron Gertler, «Advice for Responding to Journalists», document de travail, mars 2020.
- [26] Nirit Weiss-Blatt, «The AI Panic Campaign – part 1», *AI Panic*, 15 octobre 2023.
- [27] Guillaume Dandurand, Fenwick McKelvey et Jonathan Roberge, «L’IA profite d’une couverture partielle des médias», *The Conversation*, 8 juin 2023.
- [28] Guillaume Dandurand *et al.*, *Entraîner l’actualité. La couverture canadienne du cycle d’engouement pour l’IA (2012-2021)*, Québec, Institut national de la recherche scientifique, 2023.

- [29] Pauline Gourlet *et al.*, «Approcher l'IA par l'analyse des récits médiatiques», dans *Prises, mé.prises, em.prises, dé.prises, re.prises... Les formes de participation au développement de l'IA en France*, 2023.
- [30] Romain Campenon, «Intelligence artificielle: les salariés inquiets et méfiants», LearnThings, 2 février 2024.
- [31] Ibid.
- [32] Ibid.
- [33] Ibid.
- [34] Harshawn Ratanpal, «Federal Lobbying on Artificial Intelligence Grows as Legislative Efforts Stall», *OpenSecrets News*, 4 janvier 2024.
- [35] Mared Gwyn Jones, «L'industrie technologique dépense plus de 100 millions d'euros par an en lobbying numérique auprès de l'UE», Euronews, 12 septembre 2019.
- [36] Adam Satariano et Matina Stevis-Gridneff, «Big Tech Turns Its Lobbyists Loose on Europe, Alarming Regulators», *The New York Times*, 14 décembre 2020.
- [37] Brendan Bordelon, «How a Billionaire-Backed Network of AI Advisers Took over Washington», *Politico*, 13 octobre 2023.
- [38] Brendan Bordelon, «Key Congress Staffers in AI Debate Are Funded by Tech Giants like Google and Microsoft», *Politico*, 3 décembre 2023.
- [39] Paul Sawers, «OpenAI, Google, Microsoft and Anthropic Form Body to Oversee Safe "Frontier AI" Development», *TechCrunch*, 26 juillet 2023.
- [40] Ibid.
- [41] Devin Coldewey, «Top AI Companies Visit the White House to Make "Voluntary" Safety Commitments», *TechCrunch*, 21 juillet 2023.
- [42] Anderljung *et al.*, «Frontier AI Regulation», *loc. cit.*
- [43] Cat Zakrzewski, «This Agency Is Tasked with Keeping AI Safe. Its Offices Are Crumbling», *The Washington Post*, 6 mars 2024.
- [44] Ashley Belanger, «Feds Appoint "AI Doomer" to Run AI Safety at Us Institute», *Ars Technica*, 17 avril 2024.
- [45] Beatrice Nolan, «Ex-OpenAI Researcher Says There's a 50 % Chance AI Development Could End in "Doom"», *Business Insider*, 3 mai 2023.
- [46] Laurie Clarke, «How Silicon Valley Doomers Are Shaping Rishi Sunak's AI Plans», *Politico*, 14 septembre 2023.
- [47] Ian Hogarth, «We Must Slow Down the Race to God-like AI», *Financial Times*, 13 avril 2023.
- [48] Agence France-Presse, «Elon Musk Issues Warning about "Humanoid Robots" in AI "Age of Abundance"», France 24, 2 novembre 2023.
- [49] Mark Scott *et al.*, «Inside the Shadowy Global Battle to Tame the World's Most Dangerous Technology», *Politico*, 26 mars 2024.
- [50] «The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023», site web du gouvernement du Royaume-Uni, 1er novembre

2023.

[51] «Byte by Byte: How Big Tech Undermined the AI Act», Corporate Europe Observatory, 17 novembre 2023.

[52] Billy Perrigo, «Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation», *Time*, 20 juin 2023.

[53] «Byte by Byte», *loc. cit.*

[54] *Ibid.*

[55] Rishi Bommasani *et al.*, «Do Foundation Model Providers Comply with the Draft EU AI Act?», Center for Research on Foundation Models, 15 juin 2023.

[56] Peggy Corlin, «La double casquette de Cédric O, lobbyiste de Mistral et conseiller du gouvernement pour l'IA générative», *La Lettre*, 16 novembre 2023.

[57] Alexandre Piquard, «Arthur Mensch, de Mistral AI: “En intelligence artificielle, nous pouvons rivaliser avec Google ou OpenAI”», *Le Monde*, 26 février 2024.

[58] Tweet de Luca Bertuzzi, Twitter/X, 26 février 2024, 14 h 25.

[59] «Trojan Horses: How European Startups Teamed up with Big Tech to Gut the AI Act», Corporate Europe Observatory, 11 mars 2024.

[60] Alex Hern, «OpenAI Leaders Call for Regulation to Prevent AI Destroying Humanity», *The Guardian*, 24 mai 2023.

[61] Mustafa Suleyman et Eric Schmidt, «Mustafa Suleyman and Eric Schmidt: We Need an AI Equivalent of the IPCC», *Financial Times*, 19 octobre 2023.

[62] Jason Hausenloy, Andrea Miotti et Claire Dennis, «Multinational AGI Consortium (MAGIC): A Proposal for International Coordination on AI», *Conjecture*, 13 octobre 2023.

[63] Steven Overly, «It's Time to Regulate AI like Cars and Drugs, Top Microsoft Exec Says», *Politico*, 13 septembre 2023.

[64] Rachel Griffin, «EU Platform Regulation in the Age of Neo-Illiberalism», Social Science Research Network, 29 mars 2024.

Chapitre 5

[1] «Electronic “Brain” Teaches Itself», *The New York Times*, 13 juillet 1958.

[2] Melanie Mitchell, «Why AI Is Harder than We Think», *arXiv*, 28 avril 2021.

[3] Cory Doctorow, «The Real AI Fight», *Medium*, 27 novembre 2023.

[4] Sebastien Bubeck *et al.*, «Sparks of Artificial General Intelligence: Early Experiments with GPT-4», *arXiv*, 13 avril 2023.

[5] Will Douglas Heaven, «AI Is Wrestling with a Replication Crisis», *MIT Technology Review*, 12 novembre 2020.

[6] Hilary Putnam, «Brains in a Vat», dans *Reason, Truth and History*, Cambridge, Cambridge University Press, 1981, p. 1-21.

[7] Rowan Jacobsen, «Brains Are Not Required When It Comes to Thinking and Solving Problems – Simple Cells Can Do It», *Scientific American*, 1er février 2024.

[8] Alexis T. Baria et Keith Cross, «The Brain Is a Computer Is a Brain: Neuroscience's Internal Debate and the Social Significance of the Computational

Metaphor», *arXiv*, 18 juillet 2021.

[9] Will Douglas Heaven, «AI Hype Is Built on High Test Scores. Those Tests Are Flawed», *MIT Technology Review*, 30 août 2023.

[10] Maggie Harrison Dupré, «ChatGPT Is Just an Automated Mansplaining Machine», *Futurism*, 2 août 2023.

[11] Vectara, «Hallucinations Leaderboard», Github, 14 mai 2024.

[12] Emily M. Bender et Alex Hanna, «“AI” Hurts Consumers and Workers – and Isn’t Intelligent», *Tech Policy Press*, 4 août 2023.

[13] Baldur Bjarnason, «The LImentalist Effect: How Chat-Based Large Language Models Replicate the Mechanisms of a Psychic’s Con», *Software Crisis*, 4 juillet 2023.

[14] Bob Knorpp, «Predicted 25 % Drop In Search Volume Remains Unclear», *Datos*, 23 avril 2024.

[15] Ted Chiang, «ChatGPT Is a Blurry JPEG of the Web», *The New Yorker*, 9 février 2023.

[16] Geoffrey A. Fowler, «I Tried the New Google. Its Answers Are Worse», *The Washington Post*, 1er avril 2024.

[17] Will Oremus, «AI Chatbot Got Election Info Wrong 30 Percent of Time, European Study Finds», *The Washington Post*, 15 décembre 2023.

[18] Arvind Narayanan et Sayash Kapoor, «Will AI Transform Law?», *AI Snake Oil*, 24 janvier 2024.

[19] Lizzie O’Leary, «How IBM’s Watson Went From the Future of Health Care to Sold Off for Parts», *Slate*, 31 janvier 2022.

[20] *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*, Genève, Organisation mondiale de la santé, 2024, p. 28.

[21] Sylvain Trinel, «Le chatbot insulte un client: le transporteur DPD désactive son IA en urgence», BFM TV, 22 janvier 2024.

[22] Kyle Melnick, «Air Canada Chatbot Promised a Discount. Now the Airline Has to Pay It», *The Washington Post*, 18 février 2024.

[23] Joy Buchanan, Stephen Hill et Olga Shapoval, «ChatGPT Hallucinates Non-existent Citations: Evidence from Economics», *The American Economist*, vol. 69, no 1, novembre 2024, p. 80-87.

[24] Cory Doctorow, «Humans Are Not Perfectly Vigilant», *Pluralistic*, 1er avril 2024.

[25] Shane Snider, «Data Leaders Say “AI Paralysis” Stifling Adoption: Study», *InformationWeek*, 15 décembre 2023.

[26] Colleen McClain, «Americans’ Use of ChatGPT Is Ticking Up, but Few Trust Its Election Information», Pew Research Center, 26 mars 2024.

[27] François Cadelon *et al.*, «How People Can Create – and Destroy – Value with Generative AI», Boston Consulting Group, 21 septembre 2023.

[28] «Organizations Bullish on AI Adoption Despite Losing an Average of \$406M Each Year Due to Underperforming AI Models», *Business Wire*, 20 mars 2024.

- [29] Gary Marcus, «Evidence That Lims Are Reaching a Point of Diminishing Returns – And What That Might Mean», *Marcus on AI*, 13 avril 2024.
- [30] Ashley Stewart, «OpenAI Admits GPT-4 May Be Worse on Some Tasks After Users Complained the Model Is “Lazier” and “Dumber”», *Business Insider*, 30 juillet 2023.
- [31] Carl Franzen, «The AI Feedback Loop: Researchers Warn of “Model Collapse” as AI Trains on AI-Generated Content», *VentureBeat*, 12 juin 2023.
- [32] Ilya Shumailov *et al.*, «The Curse of Recursion: Training on Generated Data Makes Models Forget», *arXiv*, 27 mai 2023.
- [33] Kevin Roose, «A.I. Has a Measurement Problem», *The New York Times*, 15 avril 2024.
- [34] Anissa Gardizy et Aaron Holmes, «Amazon, Google Quietly Tamp Down Generative AI Expectations», *The Information*, 12 mars 2024.
- [35] MacKenzie Sigalos et Ryan Browne, «Openai’s Sam Altman Says Human-Level AI Is Coming but Will Change World Much Less Than We Think», CNBC, 16 janvier 2024.
- [36] Katyanna Quach, «Now That You’ve All Tried It... ChatGPT Web Traffic Falls 10 %», *The Register*, 7 juillet 2023.
- [37] Will Oremus, «AI Chatbots Lose Money Every Time You Use Them. That Is a Problem», *The Washington Post*, 5 juin 2023.
- [38] Melissa Heikkilä, «Making an Image with Generative AI Uses as Much Energy as Charging Your Phone», *Technology Review*, 1er décembre 2023.
- [39] Eren Çam *et al.*, *Electricity 2024: Analysis and Forecast to 2026*, Paris, International Energy Agency, 2024, p. 8.
- [40] Saijel Kishan et Josh Saul, «AI Needs So Much Power That Old Coal Plants Are Sticking Around», *Bloomberg*, 25 janvier 2024.
- [41] Myles McCormick, Jamie Smyth et Amanda Chu, «AI Revolution Will Be Boon for Natural Gas, Say Fossil Fuel Bosses», *Financial Times*, 1er avril 2024.
- [42] Anissa Gardizy et Amir Efrati, «Microsoft and OpenAI Plot \$100 Billion Stargate AI Supercomputer», *The Information*, 29 mars 2024.
- [43] Tweet de Sam Altman, Twitter/X, 1er mai 2022, 12 h 10.
- [44] Pengfei Li *et al.*, «Making AI Less “Thirsty”: Uncovering and Addressing the Secret Water Footprint of AI Models», *arXiv*, 6 avril 2023.
- [45] Cécile Bontron, «Rare-Earth Mining in China Comes at a Heavy Cost for Local Villages», *The Guardian*, 7 août 2012.
- [46] Eryk Salvaggio, «AI as Multidimensional Infrastructure», Siegel Family Endowment, 17 novembre 2022.
- [47] Taller Estampa, Cartography of Generative AI, affiche, s. l., 2024.
- [48] Kate Crawford, *Contre-atlas de l’intelligence artificielle*, Paris, Zulma, 2022 [2021].
- [49] Kate Crawford et Vladan Joler, *Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data and Planetary Resources*, New York / Novi Sad, AI Now Institute / Share Lab, 7 septembre 2018.

- [50] Jai Vipra et Sarah Myers West, «Computational Power and AI», AI Now Institute, 27 septembre 2023.
- [51] Alex Reisner, «These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech», *The Atlantic*, 25 septembre 2023.
- [52] Gerrit De Vynck, «ChatGPT Maker OpenAI Faces a Lawsuit Over How It Used People’s Data», *The Washington Post*, 28 juin 2023.
- [53] Authors Guild, «Open Letter to Generative AI Leaders», lettre ouverte, 18 juillet 2023.
- [54] Mike Isaac, «Reddit Wants to Get Paid for Helping to Teach Big A.I. Systems», *The New York Times*, 18 avril 2023.
- [55] Melissa Heikkilä, «This New Data Poisoning Tool Lets Artists Fight back against Generative AI», *MIT Technology Review*, 23 octobre 2023.
- [56] Sheera Frenkel et Stuart A. Thompson, «“Not for Machines to Harvest”: Data Revolts Break Out Against A.I.», *The New York Times*, 15 juillet 2023.
- [57] Dani Anguiano et Lois Beckett, «How Hollywood Writers Triumphed over AI – And Why It Matters», *The Guardian*, 1er octobre 2023.
- [58] Dave Karpf, «An AI Future Is Much Shakier than You Think», *Foreign Policy*, 31 décembre 2023.
- [59] Frank Landymore, «AI Investors Horrified at the Suggestion That They Pay Artists for Using Their Copyrighted Work», *Futurism*, 10 novembre 2023.
- [60] Benj Edwards, «OpenAI Says It’s “Impossible” to Create Useful AI Models without Copyrighted Material», *Ars Technica*, 9 janvier 2024.
- [61] Yuvraj Malik, «Global Venture Capital Investment near 5-Year Low in Q1, Pitchbook Data Shows», Reuters, 3 avril 2024.
- [62] Kyle Wiggers, «Investors Are Growing Increasingly Weary of AI», *TechCrunch*, 15 avril 2024.
- [63] Stephanie Palazzolo, «AI Valuations May Be Coming Down to Earth; A Glimpse of OpenAI Search Engine», *The Information*, 2 avril 2024.
- [64] Torsten Sløk, «The Magnitude of the AI Bubble», Apollo Academy, 9 janvier 2024.
- [65] John Rapley, «Nvidia’s Falling Share Price Shows the AI Reckoning Has Begun», *UnHerd*, 22 avril 2024.
- [66] June Yoon, «AI Hype Has Echoes of the Telecoms Boom and Bust», *Financial Times*, 14 février 2024.
- [67] Olivier Clairouin, «“No crypto”, de Nastasia Hadjadji: Le mensonge du bitcoin», *Le Monde*, 26 mai 2023.
- [68] Homayoun Hatami *et al.*, «A Ceo’s Guide to the Metaverse», McKinsey, 24 janvier 2023.
- [69] Will Daniel, «Warren Buffett Says the Stock Market Is Increasingly “Casino-Like” – And Young Investors Need to Remember This “One Fact of Financial Life” to Avoid the Mess», *Fortune*, 26 février 2024.
- [70] Cory Doctorow, «Cory Doctorow: What Kind of Bubble is AI?», *Locus Mag*, 18 décembre 2023.

- [71] Morgan Meaker, «The Fraud-Detection Business Has a Dirty Secret», *Wired*, 7 mars 2023.
- [72] «Intelligence artificielle: l'État s'engage pour rendre l'action publique plus simple, plus efficace au bénéfice des Français», communiqué de presse no 1452, Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique, 12 décembre 2023.
- [73] Gabriel Geiger *et al.*, «Profilage et discriminations: enquête sur les dérives de l'algorithme des caisses d'allocations familiales», *Le Monde*, 4 décembre 2023.
- [74] Christine Delphy, *Classer, dominer. Qui sont les «autres»?», Paris, La Fabrique, 2008.*
- [75] Hubert Guillaud, «Le risque fasciste de l'IA», blogue personnel, 18 janvier 2024.
- [76] Marion Fourcade et Kieran Healy, *The Ordinal Society*, Cambridge (MA), Harvard University Press, 2024.
- [77] Dan McQuillan, «AI as Algorithmic Thatcherism», blogue personnel, 21 décembre 2008.
- [78] Jarrow Insights, *La vie dans le panoptique d'Amazon: une étude internationale auprès des travailleurs d'Amazon*, Nyon, UNI Global Union, 2023.
- [79] Catherine Quignon, «Dans les entreprises, l'essor du management par les algorithmes», *Le Monde*, 20 février 2022.
- [80] Antonio A. Casilli, *En attendant les robots. Enquête sur le travail du clic*, Paris, Seuil, coll. «La couleur des idées», 2019.
- [81] Andrew Deck, «The Workers at the Frontlines of the AI Revolution», *Rest of World*, 11 juillet 2023.
- [82] James Muldoon *et al.*, «A Typology of Artificial Intelligence Data Work», *Big Data & Society*, vol. 11, no 1, janvier-mars 2024.
- [83] Billy Perrigo, «Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic», *Time*, 18 janvier 2023.
- [84] Niamh Rowe, «“It's Destroyed Me Completely”: Kenyan Moderators Decry Toll of Training of AI Models», *The Guardian*, 2 août 2023.
- [85] Morgan Meaker, «These Prisoners Are Training AI», *Wired*, 11 septembre 2023.
- [86] Vladan Joder et Matteo Pasquinelli, «The Nooscope Manifested: AI as an Instrument of Knowledge Extractivism», *AI & Society*, vol. 36, no 4, 2021, p. 1263-1280.
- [87] Alex Hern, «TechScape: How Cheap, Outsourced Labour in Africa Is Shaping AI English», *The Guardian*, 16 avril 2024.
- [88] Devin Coldewey, «Mechanical Turk Workers Are Using AI to Automate Being Human», *TechCrunch*, 14 juin 2023.
- [89] Hern, «TechScape», *loc. cit.*
- [90] Kate Crawford, «Artificial Intelligence's White Guy Problem», *The New York Times*, 25 juillet 2016.

- [91] Alistair Barr, «Google Mistakenly Tags Black People as “Gorillas,” Showing Limits of Algorithms», *The Wall Street Journal*, 1er juillet 2015.
- [92] Adi Robertson, «Google Apologizes for “Missing the Mark” after Gemini Generated Racially Diverse Nazis», *The Verge*, 21 février 2024.
- [93] Thaddeus L. Johnson et Natasha N. Johnson, «Police Facial Recognition Technology Can’t Tell Black People Apart», *Scientific American*, 18 mai 2023.
- [94] Steve Lohr, «Facial Recognition Is Accurate, if You’re a White Guy», *The New York Times*, 9 février 2018.
- [95] Natasha Singer et Cade Metz, «Many Facial-Recognition Systems Are Biased, Says U.S. Study», *The New York Times*, 19 décembre 2019.
- [96] Julia Angwin *et al.*, «Machine Bias», *ProPublica*, 23 mai 2016.
- [97] Sarah Myers West, «Discriminating Systems: Gender, Race, and Power in AI – Report», AI Now Institute, 1er avril 2019.
- [98] Catherine Stinson, «The Dark Past of Algorithms That Associate Appearance and Criminality», *American Scientist*, vol. 109, no 1, janvier-février 2021.
- [99] Leonardo Nicoletti et Dina Bass, «Humans Are Biased. Generative AI Is Even Worse», *Bloomberg*, 9 juin 2023.
- [100] Victoria Turk, «How AI Reduces the World to Stereotypes», *Rest of World*, 10 octobre 2023.
- [101] Daniel van Niekerk *et al.*, *Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models*, Paris/ Ljubljana, UNESCO/IRCAI, 2024.
- [102] Gianluca Mauro et Hilke Schellman, «“There Is No Standard”: Investigation Finds AI Algorithms Objectify Women’s Bodies», *The Guardian*, 8 février 2023.
- [103] Marie Tomaszewski, «Avec l’IA, l’extrême-droite donne forme à ses fantasmes», *Streetpress*, 5 avril 2024.
- [104] Yarden Katz, *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*, New York, Columbia University Press, 2020.
- [105] Giulia Torino et Mathew Mahmoudi, «Ban The Scan: Inside the NYPD’s Surveillance Machine», Amnesty International, 2022.
- [106] «Nos arguments pour faire interdire la vidéosurveillance automatisée», La Quadrature du Net, 23 janvier 2023.

Déjà parus dans la collection «Futur proche»

Atossa Araxia Abrahamian

Citoyennetés à vendre. Enquête sur le marché international des passeports

Marco Armiero

Poubelloccène. Chroniques de l'ère des déchets

Normand Baillargeon et Jean-Marc Piotte (dir.)

Au bout de l'impasse, à gauche. Récits de vie militante et perspectives d'avenir
Franco «Bifo» Berardi Tueries. *Forcenés et suicidaires à l'ère du capitalisme absolu*

Gaétan Breton

La dette. Règlement de comptes

Gaétan Breton

Faire payer les pauvres.
Éléments pour une fiscalité progressiste

Gaétan Breton

Tout doit disparaître. Partenariats publicprivé et liquidation des services publics

Jean Bricmont

L'impérialisme humanitaire.
Droit humanitaire, droit d'ingérence, droit du plus fort?

Pierre-Luc Brisson

L'âge des démagogues. Entretiens avec Chris Hedges

Noam Chomsky

Comprendre le pouvoir

Noam Chomsky

Futurs proches. Liberté, indépendance et impérialisme au XXI^e siècle

Noam Chomsky

La lutte ou la chute!

Pourquoi il faut se révolter contre les maîtres de l'espèce humaine

Noam Chomsky

L'optimisme contre le désespoir. Entretiens avec C.J. Polychroniou

Noam Chomsky

Le précipice.

Entretiens avec C.J. Polychroniou

Noam Chomsky

Qui mène le monde?

Noam Chomsky et Vijay Prashad

Le retrait. La fragilité de la puissance des États-Unis: Irak, Libye, Afghanistan

Noam Chomsky et Marv Waterstone

Les conséquences du capitalisme. Du mécontentement à la résistance

Gabriella Coleman

Anonymous. Hacker, activiste, faussaire, mouchard, lanceur d'alerte

Pierre Dardot, Haug Guéguen, Christian Laval et Pierre Sauvêtre

Le choix de la guerre civile.

Une autre histoire du néolibéralisme

Mitchell Dean et Daniel Zamora

Le dernier homme et la fin de la révolution. Foucault après Mai 68

Francis Dupuis-Déri (dir.)

Québec en mouvements. Idées

et pratiques militantes contemporaines

Chris Hedges

L'empire de l'illusion. La mort de la culture et le triomphe du spectacle

Chris Hedges

Les fascistes américains. La droite chrétienne à l'assaut des États-Unis

Chris Hedges

La mort de l'élite progressiste

Edward S. Herman et David Peterson

Génocide et propagande. L'instrumentalisation politique des massacres

Institut de recherche et d'informations socioéconomiques (IRIS)

Dépossession: une histoire économique du Québec contemporain.

Tome 1: les ressources

Institut de recherche et d'informations socioéconomiques (IRIS)

Dépossession: une histoire économique du Québec contemporain.

Tome 2: les institutions publiques

Am Johal et Matt Hern

Réchauffement planétaire et douceur de vivre. Road trip en territoire pétrolifère

Razmig Keucheyan

Hémisphère gauche. Une cartographie des nouvelles pensées critiques

Naomi Klein

Le choc des utopies. Porto Rico contre les capitalistes du désastre

Naomi Klein

Dire non ne suffit plus.

Contre la stratégie du choc de Trump

Naomi Klein

La maison brûle. Plaidoyer pour un New Deal vert

Naomi Klein

No logo. La tyrannie des marques

Naomi Klein

Tout peut changer. Capitalisme et changement climatique

Naomi Klein et Rebecca Steffof

Comment tout peut changer.

Outils à l'usage de la jeunesse mobilisée pour la justice climatique et sociale

Andrea Langlois et Frédéric Dubois (dir.)

Médias autonomes. Nourrir la résistance et la dissidence

Julien Lefort-Favreau

Le luxe de l'indépendance. Réflexions sur le monde du livre

Linda McQuaig

Les milliardaires. Comment les ultra-riches nuisent à l'économie

Luc Rabouin

Démocratiser la ville. Le budget participatif: de Porto Alegre à Montréal

Sherene H. Razack

La chasse aux Musulmans. Évincer les Musulmans de l'espace politique

Jeremy Scahill

Le nouvel art de la guerre. Dirty Wars

Jeremy Scahill et l'équipe de *The Intercept*

La machine à tuer. La guerre des drones

Samir Shaheen Hussain

Plus aucun enfant autochtone arraché. Pour en finir avec le colonialisme médical canadien

Tom Slee

Ce qui est à toi est à moi. Contre Airbnb, Uber et autres avatars de l'«économie du partage»

Nick Srnicek

Capitalisme de plateforme. L'hégémonie de l'économie numérique

Olúfémí O. Táíwò

L'élite cannibale. Comment les puissants se sont approprié les luttes identitaires (et tout le reste)

Astra Taylor

Démocratie.com. Pouvoir, culture et résistance à l'ère des géants de la Silicon Valley

Laura Tripaldi

Gender Tech. Ce que la technologie fait au corps des femmes

Harsha Walia

Frontières et domination.

Migrations, capitalisme et nationalisme

Lesley W. Wood

Mater la meute. La militarisation de la gestion policière des manifestations

La mise en page est de Jolin MASSON

La révision du texte est de Isabelle PAUZÉ

Le epub a été créé par claudebergeron.com

Lux Éditeur
C.P. 83578 BP Garnier
Montréal, Québec
H2J 4E9
www.luxediteur.com

Dans la Silicon Valley, terre traditionnellement fertile pour les spiritualités excentriques, un nouveau récit se répand. Les *start-up* de l'intelligence artificielle vendent désormais aux investisseurs l'imminence de la fin des temps. L'IA deviendrait si intelligente qu'elle en serait divine, capable de nous sauver comme de nous anéantir. Sous son influence, l'industrie de la tech tout entière bascule dans un discours aux accents religieux. On ne vend plus le progrès, mais la métamorphose. On ne vend plus le futur, mais la fin de l'histoire.

Certains les appellent les doomers – les catastrophistes – ou encore les techno-oligarques. Pour les besoins de ce livre, nous les appellerons les «prophètes». Cette congrégation se compose majoritairement d'hommes blancs, issus des élites universitaire et entrepreneuriale. Ils viennent de Stanford, Oxford, Harvard ou du MIT. Ils sont ingénieurs, chercheurs, philosophes, patrons, investisseurs et milliardaires, et pas une semaine ne passe sans que ces prophètes apparaissent dans la presse pour prononcer un oracle. Dans un inquiétant amalgame d'autoritarisme et d'ultracapitalisme, ils veulent nous faire croire en leur toute-puissance pour mieux imposer leur pouvoir. Ce faisant, ils représentent une réelle menace pour la société civile et nos libertés.

Journaliste indépendant spécialiste des nouvelles technologies, Thibault Prévost travaille depuis dix ans pour divers médias français, notamment *Arrêt sur images*.

Table des Matières

Couverture	1
Page de titre	2
Crédits	3
Table des matières	4
État des lieux	7
L'ère des perroquets automatiques	9
Tous vers l'intelligence artificielle générale	12
Sam Altman, rock star du lobbying	16
La bataille des prophéties	19
Un air de déjà-vu	22
Keep Calm and Reclaim	23
1. La Terre sainte du transhumanisme	29
Ultralibéralisme et transhumanisme californien	30
Nick Bostrom, le prophète déguisé en prof	32
Dépasser les frontières, l'obsession de la Silicon Valley	35
L'altruisme efficace: une machine à justifier le (techno)capitalisme	41
Vers le long-termisme, où l'humain n'a plus d'importance	46
2. Futuritarismes	50
Vers un capitalisme de commandement	52
Retour vers le futur	58
Lumières noires sur la Silicon Valley	62
Vers l'infini et l'au-delà	68
3. L'apocalypse selon l'IA	73
De la catastrophe naturelle à l'apocalypse artificielle	76
Le totalitarisme contre l'extinction	81
À la recherche des Tables de l'IA	87
Superalignement et super-bullshit	90
4. Du bon usage des prophéties	93
Lutte fratricide dans la Silicon Valley	94

Mainmise du capital sur la technoscience	98
La fabrique de la sidération	101
L'AI governance, l'intoxication des institutions	105
Réguler le monopole, c'est réguler l'outil	112
5. Systèmes d'exploitation	116
L'intelligence artificielle est un culte	117
L'intelligence artificielle ne fonctionne (toujours) pas	119
L'intelligence artificielle est un écocide	123
L'intelligence artificielle est un pillage	126
L'intelligence artificielle est une bulle financière	127
L'intelligence artificielle est un autoritarisme	129
L'intelligence artificielle est une discrimination	133
Remerciements	138
Notes et références	139
Déjà parus dans la collection «Futur proche»	157
Quatrième de couverture	163