

**Colloque du Samedi 13 avril 2019 :
La génération Y et le défi des technosciences**

Troisième partie : technosciences et technologie

Tout l'être humain est-il algorithmable ?

GROUPE 4 STUDENT EX MACHINA 2018-2019

Dylan LEDANOIS – TBS

Nolwenn ARMOGATHE – IEP Sociologie

Rachel DELPIERRE – UT1 Droit du Numérique

Jean-Baptiste DAVID – UT2 Philosophie

Raphaël MATHON – ISAE Supaero

Benjamin ANIZON – UT3 Biosanté

Eleazar MBOCK – UT3 Biosanté

TUTEUR : Michel Rouffet

Présentation par le tuteur

Notre groupe comprenait 7 étudiants de disciplines très variées : philosophie, biologie, ingénieur, juriste, commerce, sociologie. Et il a fallu d'abord nous mettre d'accord sur les termes « algorithmable » et « être humain »

Puis nous avons sélectionné collectivement 6 thématiques déclinant le sujet, et 2 exemples concrets l'illustrant et avons discuté du contenu de ces 8 parties : pas toujours simple de comprendre des développements parfois spécialisés et avec des langages spécifiques. Mais c'était bien là tout l'intérêt de l'exercice

Enfin, les étudiants se sont positionnés pour rédiger chaque partie sans oublier l'introduction et la conclusion... et aussi la façon de présenter ce travail au colloque. Je dois avouer que j'ai été vraiment bluffé par la motivation des étudiants (5 réunions de travail, aucun absent !) leur capacité d'écoute et d'analyse et la qualité de l'exposé au colloque, tout cela en 4 mois

J'espère très vivement que leur travail sera reconnu dans leurs cursus pédagogique, à un moment où l'on sent bien que, pour avancer, le monde a besoin qu'on croise les disciplines... et que les humains réapprennent à s'écouter et à se comprendre

Table des matières

THEME 1 : Où en est la connaissance sur les fonctionnements cognitifs du cerveau, permettant leur mise en algorithme ? (R. Mathon et B. Anizon)

D'une organisation régionalisée et spécialisée à une vision connexionniste

L'imagerie à l'aide de l'approche systémique.

Les réseaux de neurones et leurs fonctionnements, une influence pour l'intelligence artificielle

Les deux premières couches d'un réseau de neurones convolutif

De l'intelligence artificielle vers un ordinateur neuronal

THEME 2 : Dans quelle mesure sait-on créer ou modifier des cellules vivantes de l'humain ? (B. Anizon, E. Mbock)

Exemple des bases azotées / implémenter de l'épigénétique sur les algorithmes

Historique des techniques pour modifier des cellules : jusqu'où peut-on modifier les cellules humaines et avec quelle précision ?

THEME 3 : Un robot est-il capable d'adopter des comportements émotionnels ? (D. Ledanois)

THEME 4 : Un algorithme peut-il rendre compte du langage humain dans toutes ses expressions ? (J-B. David)

THEME 5 : Les algorithmes peuvent-ils traduire correctement le fonctionnement de différentes sociétés et dans quels buts ? (N. Armogathe)

EXEMPLE 1 : La justice peut-elle être rendue par des robots ? (R. Delpierre)

I. La subjectivité des algorithmes auto-apprenant

II. Deux approches différentes en droit pénal

III. L'exemple de l'appréciation de l'originalité en droit de la propriété littéraire et artistique.

EXEMPLE 2 : Quelle fiabilité pour la conduite automatique des avions? (R. Mathon)

CONCLUSION: Émergence et mécanisme de la conscience biologique et algorithmique (B. Anizon)

BIBLIOGRAPHIE

Introduction

Si l'on considère l'éventualité de programmer un esprit, on se rend compte très vite de certaines difficultés. En effet, même si l'on arrive à faire agir un programme ou un robot de la même façon que nous, il restera un doute : le robot pense-t-il ? Cette question est largement traitée dans la science-fiction, mais tombe souvent en aporie. Dans ces œuvres, l'algorithme est présenté comme un être humain, il lui ressemble en tout point, parfois on ne sait même pas qu'il s'agit d'une machine, on ne l'apprend qu'au dénouement. Pourtant, on va naturellement faire une distinction entre la machine et l'homme, cela se remarque par le fait que nous doutons de l'existence de leur psyché alors même que nous ne doutons pas de celles des êtres humains.

En fait, cela s'explique par le fait que nous percevons l'homme de deux manières. D'une part, nous sommes des êtres conscients, nous avons chacun une connaissance immédiate de notre existence par la pensée et en même temps nous savons que nous pensons. Il y a une sorte de boucle mystérieuse de laquelle nous ne pouvons sortir et qui semble définir notre individualité, notre existence en tant qu'humain. D'autre part, nous voyons des êtres qui nous ressemblent physiquement, avec qui on parle et avec qui nous formons une sorte de communauté dont nous dépendons tous. Ces êtres sont les autres êtres humains, ceux qu'on appelle *alter-ego* (l'autre moi). Partant, nous prêtons à l'*alter-ego* cette subjectivité qui semble définir l'homme, alors même que nous n'avons jamais accès à la subjectivité d'autrui. « Je » vois des êtres qui me ressemblent en tous points, or « je » pense, donc ces êtres doivent penser eux aussi. Nous pouvons distinguer alors deux conceptions de l'être humain. Il y a, premièrement, l'homme en tant que sujet, qui pense et qui n'est représenté par « mon » expérience. Il y a, deuxièmement, l'homme en tant qu'objet, c'est-à-dire l'autre, que nous pouvons percevoir et donc étudier par les sciences.

Par ailleurs, si l'on comprend « algorithmable » comme le fait d'écrire ce qu'est l'homme dans un langage informatique, il peut être plus facile de retourner la question. On se demandera s'il est possible de créer un algorithme qui, par son fonctionnement, peut être considéré comme un être humain. Autrement dit, le programme est-il humanisable ?

Au-delà du simple fantasme de l'être humain artificiel, la question que pose l'algorithme ici est notre capacité à prédire les actions et les comportements humains. En effet, si l'on définit l'algorithme comme une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre un problème ou d'obtenir un résultat, et que l'on cherche à le comparer avec l'être humain, alors il est nécessaire d'avoir un modèle de comparaison. Or cette comparaison n'est

possible que si le modèle est à l'image de l'algorithme, c'est-à-dire sans ambiguïté. S'en suit alors deux conséquences. Premièrement, ce modèle doit être décrit dans un langage formel, qui permettrait une éventuelle traduction informatique. Deuxièmement, les actions et comportements de ce modèle humain doivent être connus et au moins partiellement mesurables, dans le but d'évaluer les résultats de l'algorithme. Autrement dit, ce qu'interroge l'éventualité d'un programme humanisable c'est la capacité des sciences modernes à produire un savoir sur l'homme. Il ne s'agit pas d'affirmer ou d'infirmer la vérité de ces sciences mais bien d'explorer les pistes d'une modélisation globale de l'être humain. C'est donc l'homme en tant qu'objet qui est mis à l'épreuve en premier lieu.

Dans *les Mots et les Choses*, Michel Foucault fait une analyse historique des sciences humaines dans laquelle il souligne un point crucial pour nous. Il explique justement que ce qui a permis l'émergence des sciences humaines vient de cette considération de l'homme en tant qu'objet extérieur et empirique. Il s'agit alors de l'homme en tant qu'il travaille, vit et parle. Ces trois domaines d'actions sont à prendre au sens large, ce qui est important c'est qu'il s'agisse d'éléments extérieurs à l'expérience du sujet. On voit l'homme travailler, vivre et parler et c'est cela que l'on va mesurer et étudier.

Durant la période moderne, trois nouvelles épistémès, c'est-à-dire trois nouveaux cadres et méthodes scientifiques sont nés. Il s'agit alors de l'économie politique, de la biologie et de la philologie, chacune étudiant respectivement le travail, la vie et la parole. C'est à partir de ces trois nouveaux champs de la connaissance qu'émergent à leur tour les sciences humaines pendant le XIX^e siècle. Apparaissent alors respectivement la sociologie, la psychologie et l'étude des mythes et littératures. Chacune de ces sciences, humaines ou non ont construit leurs théories à partir desquelles on peut envisager la modélisation d'un programme humain.

Nous pouvons imaginer que la chose se ferait en deux temps. Premièrement, la biologie, l'économie et la philologie pourraient nous donner des fondations pour construire l'algorithme. Deuxièmement, la psychologie, la sociologie et l'étude littéraires pourraient nous permettre de tester l'algorithme. Enfin, soulignons que si ces trois facettes du comportement humain sont étudiées séparément, il est probable qu'elles soient imbriquées les unes dans les autres, ce qui signifie que nous ne pouvons pas constituer cet être informatique de façon partielle. L'être humain travaille parce qu'il vit et parle parce qu'il travaille en groupe, le programme doit faire de même. Il doit pouvoir vivre et mourir, travailler pour survivre et communiquer avec ses semblables.

Pour approfondir notre réflexion, nous nous appuyerons sur les questions suivantes :

1. Où en est la connaissance sur les fonctionnements cognitifs du cerveau permettant leur mise en algorithme ?
2. Dans quelle mesure sait-on créer ou modifier des cellules vivantes de l'humain ?
3. Un robot est-il capable d'adopter des comportements émotionnels ?
4. Un algorithme peut-il rendre compte du langage humain dans toutes ses expressions ?
5. Des algorithmes peuvent-ils traduire correctement le fonctionnement de différentes sociétés et dans quels buts ?

THEME 1 : Où en est la connaissance sur les fonctionnements cognitifs du cerveau, permettant leur mise en algorithme ? (R. Mathon et B. Anizon)

D'une organisation régionalisée et spécialisée à une vision connexionniste

L'hypothèse de fonctions cognitives en fonction de la localisation cérébrale n'a pas attendu l'imagerie médicale pour être émise. En effet au début du XVIII^e siècle la phrénologie s'attarde à relier la forme du crâne à des traits de caractères tels que la combativité, l'amour conjugal ou encore la vénération pour Dieu. En s'essayant de mesurer la voûte crânienne l'inventeur de cette pseudo-science, Franz Joseph Gall, ne put valider scientifiquement son hypothèse, souffrant de biais expérimentaux trop nombreux. Il fallut un demi-siècle pour que cette théorie disparaisse et soit reprise avec précaution par le Dr. Broca à propos des localisations fonctionnelles. Brillant médecin et anatomiste, ce dernier identifie grâce à des patients aphasiques une zone touchée située dans le cortex frontal inférieur gauche.

L'aphasie motrice étant un trouble de production du langage, cette zone lésée est selon lui indispensable à la parole, devenant ainsi une des premières régions associées à une fonction précise. De cette association fonctionnelle est née la cartographie des régions cérébrales par le Dr. Brodmann en 1909 qui catégorise le cortex en fonction de l'architecture neuronale. Cette carte cérébrale permet de définir les différentes aires fonctionnelles et s'est vu affinée au cours du siècle. La théorie principale de la localisation des fonctions cérébrales est passée ainsi d'une vision floue à un modèle scientifique résistant à la revue par les pairs.

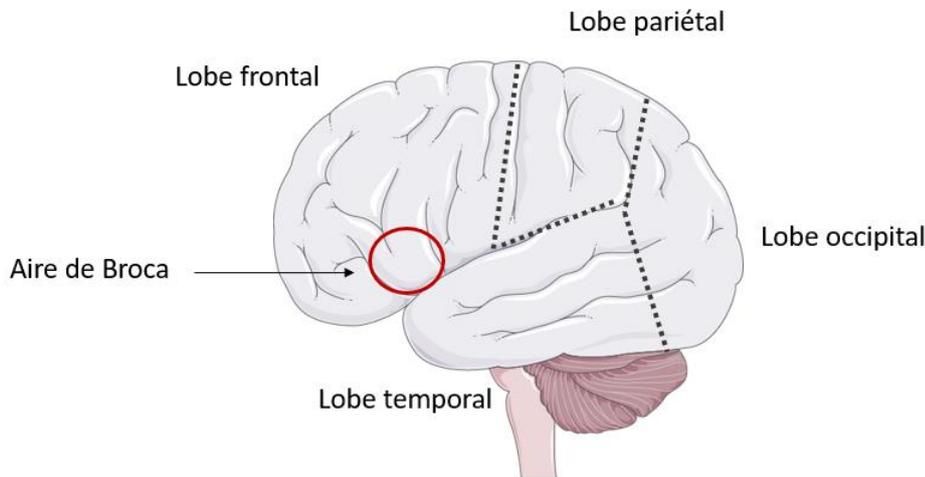


Schéma d'un cerveau (Banque d'image Servier)

La principale méthode d'investigation d'une fonction cérébrale au XIX^e et au début du XX^e siècle est l'étude de lésion cérébrale. Le cas de Phineas Gage en 1848 devient un cas d'école en neurologie. Il travaille dans la construction de voies ferrées, une erreur d'inattention lors de manipulations d'explosifs déclenche un accident. Une barre de fer pénètre sa joue gauche et ressort par le sommet de son crâne. Il survécut et ne souffrit pas de handicaps neurologiques. Néanmoins son comportement changea du tout au tout, il n'était plus lui-même : grossier, d'humeur changeante et capricieux, il ne put garder son travail. Cette découverte permit de comprendre que des processus cognitifs complexes tels que les émotions dépendent d'un substrat neurobiologique (neurones). Les études anatomiques de Ramón y Cajal sur la structure du système nerveux en 1906 viennent compléter les connaissances sur les processus cognitifs.

L'imagerie à l'aide de l'approche systémique.

De manière parallèle à ces études structurales, l'utilisation de rayonnements ionisants (rayons X) en radiologie à la fin du 19^{ème} siècle permet d'obtenir des images du corps humain afin d'établir un diagnostic. Néanmoins il fallut attendre la moitié du 20^{ème} siècle pour avoir une utilisation efficace des rayonnements non ionisants utilisés en IRM. Cette méthode d'imagerie s'est ensuite complexifiée pour étudier de manière précise les régions cérébrales activées lors de la réalisation de tâches précises. L'IRM fonctionnelle utilise le signal BOLD (*blood oxygen level dependant*) pour localiser les zones cérébrales consommant plus d'oxygène. L'utilisation de l'imagerie médicale permet avec l'aide des manipulations animales sur différents modèles (chat, souris, singe) de préciser

des fonctions aux différentes aires cérébrales. A la fin du 20ème siècle des outils permettant une analyse multi-échelle (Etude de l'expression génique, transgénèse, étude des propriétés d'activation des neurones et de leurs connexions) précisent le modèle cérébral utilisé par les neuroscientifiques. Ce modèle passe ainsi d'un confinement de régions spécialisées pour une fonction (cortex visuel, hippocampe pour la mémoire...) à un modèle plastique en constante dynamique. Cette capacité plastique du réseau cérébral est inhérente aux neurones et aux connexions qu'elles peuvent avoir en entrées et sorties. Ainsi la configuration du réseau à l'échelle locale aura un effet à l'échelle globale.

Les réseaux de neurones et leurs fonctionnements, une influence pour l'intelligence artificielle

Le cerveau est composé d'environ 100 milliards de neurones organisés au cours du développement, suivant l'expression interne de ses gènes ainsi que l'environnement pour migrer et se différencier en neurones spécialisés. Les types de contacts formés par un neurone A dépendent donc de l'environnement, des neurones à côté du neurone A qui pourraient établir des contacts. Nous pouvons voir ci-dessous un type d'interaction dont le "feedback" consistant en une activation du neurone gris qui va activer le neurone noir qui l'inhibe en retour, c'est une boucle de rétroaction négative.

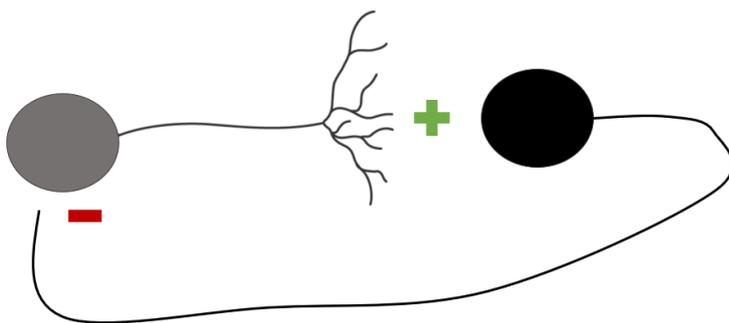


Schéma de boucle de rétroaction négative (feedback)

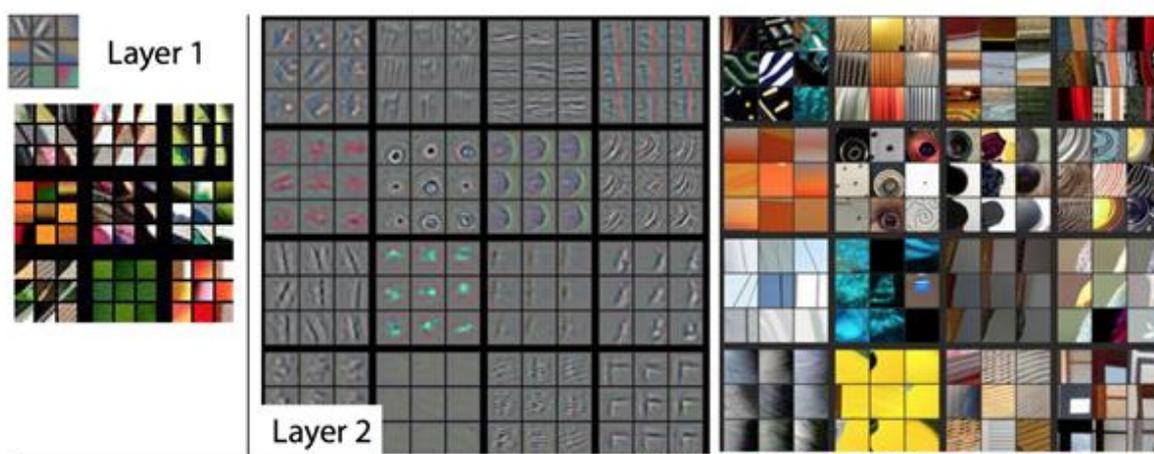
Les réseaux inhibiteurs comme l'action du neurone noir sur le neurone gris vont permettre de générer des effets non linéaires, c'est à dire que l'on ne peut que difficilement prédire le résultat à la fin sauf en connaissant les détails précis des connexions. Cette propriété des connexions neuronales a été reprise pour fonder et améliorer des algorithmes en réseaux de neurones artificiels. Deux autres propriétés ont servi à l'amélioration de ces algorithmes, des couches de neurones avec des projections convergentes et des poids modifiables attribués à chaque neurone.

C'est en 1943 que deux chercheurs en neurologie, McCulloch et Pitts, proposèrent pour la première fois un modèle mathématique d'un neurone artificiel. Ce premier modèle (de type *feed-forward*), extrêmement puissant par sa simplicité, est en fait constitué d'une somme pondérée de stimuli externes, à

laquelle est affectée une fonction seuil. Ce neurone formel, constitue alors la brique élémentaire de réseaux de neurones interconnectés.

Un algorithme de réseau de neurones est susceptible d'apprendre automatiquement un certain nombre d'exemples, puis de généraliser (d'une façon empirique) cet apprentissage à d'autres échantillons, sans qu'il faille à aucun moment décrire une quelconque représentation symbolique de la connaissance, et sans que le système ne se préoccupe jamais de cet aspect du problème. Ces algorithmes sont, pour cette raison, particulièrement efficaces pour l'identification d'anomalies (détection de fraudes), pour la reconnaissance de motifs (maintenance prédictive), ou la classification (images, musique). Cette approche présente également l'avantage d'avoir une capacité à la généralisation. Lorsque le réseau a appris certains motifs, il est capable de reconnaître ces mêmes motifs légèrement déformés, sans qu'il soit nécessaire de refaire un apprentissage.

Deux leviers peuvent être employés afin de développer un réseau de neurones. Le premier est le choix des données d'entrée. Un réseau de données est en mesure de manipuler toute information stockée en format numérique. Néanmoins, la donnée brute ne permet pas toujours d'extraire suffisamment de détails pertinents pour reproduire un comportement humain. Dans la vision par ordinateur, une des stratégies est de diviser l'image en plusieurs découpages, sur lesquels sont appliqués des filtres issus de la science du traitement du signal (convolution), avant de connecter ensemble les sorties. Le réseau pourra alors donner la probabilité que l'image appartienne à une certaine catégorie. Le second levier est ensuite la dimension du réseau.



Les deux premières couches d'un réseau de neurones convolutif

Une stratégie consiste à élargir le réseau, en subdivisant l'image initiale en un plus grand nombre d'échantillons à filtrer. Une autre réside dans l'ajout d'un plus grand nombre de couches successives de traitement, augmentant ainsi la profondeur du réseau (Deep Learning). Le temps de calcul nécessaire au

traitement est alors un facteur limitant, particulièrement dans des applications embarquées où le fonctionnement doit-être proche du temps-réel (véhicules autonomes, caméras de surveillance).

De l'intelligence artificielle vers un ordinateur neuronal

L'algorithme n'est pas le seul à être impacté par les avancées de la connaissance des processus cognitifs. La recherche se tourne également sur le développement de nouveaux moyens de calculs, et nouvelles architectures de processeurs, plus adapté à la structure des réseaux de neurones. En effet, au-delà du développement de nouvelles méthodes de résolution de problèmes, IBM a proposé en 2014 un processeur dont l'assemblage de ses transistors a été adapté pour constituer un réseau neuronal d'unités de traitement, pouvant être utilisé pour la reconnaissance d'image, de voix et de formes. La puce TrueNorth compte 1 million de neurones et 256 millions de synapses, reliés par un circuit électrique. La mémorisation, la communication et le calcul logique s'effectue sur chacun de ses 4096 cœurs, contrairement à un processeur conventionnel dont chacune de ces fonctions est séparée. Cette adaptation du support de l'algorithme au modèle du réseau neuronal permet déjà d'afficher des réductions très intéressantes de consommation d'énergie.

La nécessité de repenser la machine est effectivement partie d'un constat simple. Si un ordinateur classique devait faire tourner une simulation en temps réel du fonctionnement du cerveau et de ses 100 trillions de synapses, cela nécessiterait une puissance de 12GW, soit l'équivalent de 5 centrales nucléaires. Le cerveau humain n'en consomme que 20W. Les initiatives innovantes sur l'architecture du processeur et les technologies de semiconducteurs permettront peut-être de réduire l'impact énergétique des datacenters.

THEME 2 : Dans quelle mesure sait-on créer ou modifier des cellules vivantes de l'humain ? (B. Anizon, E. Mbock)

Exemple des bases azotées / implémenter de l'épigénétique sur les algorithmes

Si toutefois la création d'un organisme synthétique complet à l'échelle d'un corps humain n'est pas d'actualité, les études sur le fonctionnement du vivant sont fournies en exemples de manipulation de cellules vivantes allant jusqu'à

l'ajout de nouvelles fonctions. Une cellule eucaryote peut se définir comme un espace limité par une membrane plasmique contenant du cytosol un liquide dans lequel baigne des organites dont le noyau fait partie, c'est l'entité élémentaire de toute forme de vie elle peut grandir se reproduire, mourir et possède des qualités qui lui sont propres, à l'image des cellules sanguines ou encore les cellules nerveuses.

Le noyau de la cellule renferme l'ADN support du génome, il correspond à l'ensemble du matériel génétique d'un organisme transmis par ses parents, il se scinde en deux grandes parties chez l'Homme, en génome nucléaire et mitochondriale, et en méta génome (flore intestinale...). Ce matériel génétique se compose de gènes- un gène constitue une unité fonctionnelle par la présence d'une séquence promotrice, précédant une séquence transcrite, qui donnera naissance à un produit qui est l'ARN, c'est ce qu'on appelle l'expression d'un gène (la transcription). Le génome humain se compose de près de 21000 gènes, dont l'expression est programmée, rythmée au cours du développement embryonnaire (gène Hox) mais aussi tout au long de la vie.

La transcription se fait par l'activation dudit gène, autrement dit un gène peut être activé ou désactivé. Ce mécanisme dépend de la structure de la chromatine (ADN avec protéines) et son organisation en 3D dans le noyau, et fait intervenir des régulations que l'on nomme épigénétiques. L'ADN (chromatine) peut donc être fermé ou ouvert, on parle euchromatine (ouvert) et hétérochromatine (fermé), Le produit de cette transcription est un ARN, il en existe de différentes sortes, et c'est par ce produit que les gènes peuvent interagir entre eux. De façon non exhaustive dans le bestiaire des ARN on y retrouve une classification simple avec les ARNs codants (3% - 5%) responsables de ce que l'on voit (phénotype visible de la synthèse protéique) et les non codants responsables de la régulation (<2%) et de la traduction (95%).

La manipulation du vivant est aujourd'hui au cœur des laboratoires de recherche sur le vivant, de la biochimie aux neurosciences, les exemples, de protéines « chimères » conférant des propriétés nouvelles aux matériaux jusqu'à l'optogénétique qui permet l'activation par la lumière, de groupes de neurones ciblés dans le cerveau, ne manquent pas. Ces prouesses sont rendues possibles grâce aux avancées dans le domaine de l'édition du génome et de la transgénèse. Mais si la modification des gènes du vivant passe par l'ajout, la délétion ou la réécriture du génome, le code génétique lui reste inchangé à savoir qu'il se compose de quatre lettres, quatre bases azotées ATCG ayant une complémentarité simple A avec T et C avec G, ce système est présent dans tout le vivant et constitue le code de base du vivant. Cette dernière affirmation tend à être démentie du fait de la création par l'homme de nouvelles bases azotées, de nouvelles lettres du code génétique. Dans une étude de 2014, deux de ces nouvelles lettres « d5SICS » et son complémentaire « dNaM », ont été rajouté

en plus des bases ATCG dans un plasmide qui a été répliqué dans une bactérie (E. Coli) modifié par transgénèse, pour permettre l'entrée des deux nouvelles bases. En 2015 et 2017 des études similaires ont été menées mais cette fois-ci en plus de l'ajout de bases azotées supplémentaires, ont été incorporés des ARNt (ARN de transfert), spécifique acide aminé synthétique servant à la traduction et donc à la synthèse protéique. Ces études ont démontré la faisabilité de l'ajout de lettres supplémentaire dans le code génétique, sans pour autant altérer le fonctionnement basal de l'organisme hôte mais en plus en y ajoutant des fonctionnalités supplémentaires.

Ces nouveaux outils que sont les bases non naturelles ouvrent des perspectives disruptives dans le domaine de la xénobiologie, dans le but de créer des nouveaux médicaments ou même de nouveaux matériaux, voire de nouveaux organismes s'affranchissant des lois qui régissent le vivant tel que nous le connaissons. Mais cela nous pose aussi la question du pourquoi du comment au cours de l'évolution seule les bases A T C G ont été retenues ?

Une autre grande possibilité offerte par ces outils est l'opportunité de construire un modèle de la cellule vivante de A à Z par ordinateur en implémentant des fonctions biologiques au niveau subcellulaire l'une après l'autre et bâtir ainsi une cellule vivante virtuelle pouvant prédire le devenir d'une cellule vivante réel. Il existe déjà des programmes informatiques modélisant avec une assez grande fidélité des fonctions biologiques

Une étude de 2015 présente un nouveau modèle pour le développement de créatures artificielles à partir d'une seule cellule. Le modèle vise à fournir une abstraction plus biologiquement plausible de la morphogenèse et du processus de spécialisation, que l'organogenèse suit. Il repose sur trois éléments principaux : un système de physique cellulaire simulant la division et la dynamique d'adhésion intercellulaire, un cycle cellulaire simplifié offrant aux cellules la possibilité de sélectionner des actions telles que la division, la quiescence, la différenciation ou l'apoptose et, enfin, un mécanisme de spécialisation cellulaire quantifiant la capacité à remplir différentes fonctions. Un réseau de régulation génique artificiel évolué est utilisé comme contrôleur cellulaire. Comme preuve de concept, deux expériences sont présentées dans lesquelles la morphologie d'un organisme multicellulaire est guidée par les faiblesses des cellules et leur efficacité à remplir différentes fonctions sous l'effet de contraintes environnementales.

A une échelle des gènes, des algorithmes aussi existent et présentent une bonne robustesse, en 2013 une étude a implémenté la notion épigénétique dans les réseaux artificiels de régulation génique, qui sont des modèles informatiques qui s'inspirent des réseaux biologiques de régulation génique. Depuis leur création, ils ont été utilisés pour déduire des connaissances sur la régulation des gènes et

en tant que méthodes de calcul. Il a été démontré que ces modèles informatiques possèdent des propriétés généralement trouvées dans le monde biologique, telles que la robustesse et l'auto-organisation. Récemment, cette étude décrit un nouveau modèle, le réseau de réglementation de l'épigénétique artificielle (AERN), qui s'appuie sur les modèles existants en ajoutant une couche de contrôle épigénétique. Les résultats démontrent que les AERN sont plus aptes à contrôler plusieurs trajectoires opposées lorsqu'ils sont appliqués à une tâche de contrôle du chaos dans un système dynamique conservateur, ce qui suggère que les AERN constituent un domaine de recherche intéressant, pour modéliser la régulation génique.

Historique des techniques pour modifier des cellules : jusqu'où peut-on modifier les cellules humaines et avec quelle précision ?

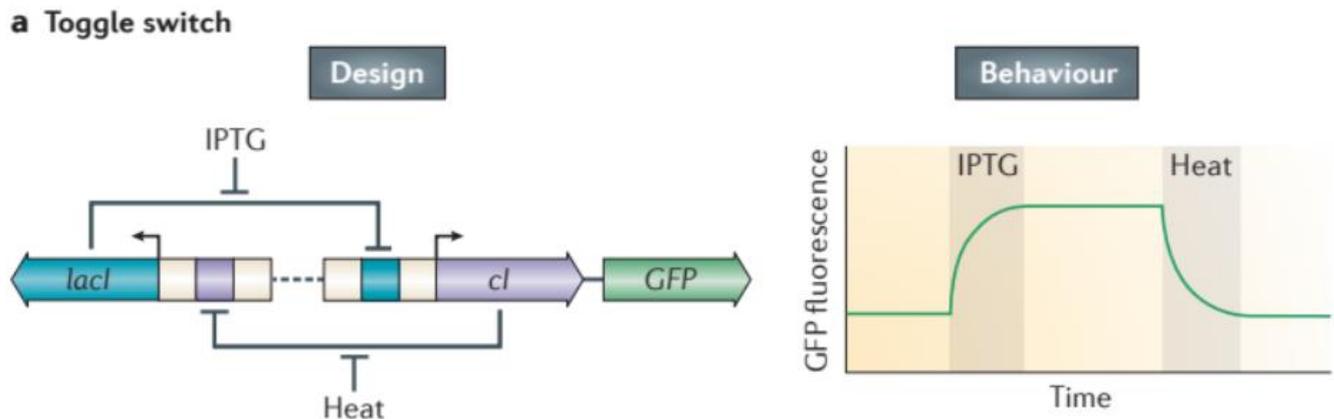
Le fait de manipuler des cellules vivantes remonte à des millénaires dès lors que l'on a commencé à faire une sélection végétale en agriculture il y a 10 000 ans. Néanmoins lorsqu'on parle de modifications géniques il n'est pas question de sélection naturelle ni d'évolution soumise à des facteurs de stress mais plutôt des techniques modernes d'ingénieries géniques et cellulaires. Le début de cette aire commence en 1961 lorsque des systèmes régulateurs de gènes répondant à l'environnement sont découverts (Monod, J Jacob. *Cold Spring Harb. Symp. Quant. Biol.*, 1961). Ces systèmes sont intéressants à identifier car ils constituent la base de l'expression génique, le squelette de tous les processus cellulaires.

Dans les années 80 une forte amélioration des techniques de séquençage du génome permet de générer plusieurs bases de données spécialisées et générales de l'ensemble du génome humain et de milliers d'autres espèces animales et végétales. Cette aire de l'« omics » correspond à l'identification de l'ADN (génomique), l'ARN (transcriptomique) et les protéines (protéomique). La transgénèse (insertion d'ADN étranger dans un organisme) et les techniques de clonage moléculaire vont rendre possible la manipulation des gènes à une tout autre échelle. Cette démocratisation de la manipulation génique et l'accumulation de nombreuses données vont développer un type précis d'ingénierie, le reverse-engineering.

Ce dernier se définit par l'examen des composants d'un système où l'on peut comprendre ses fonctions intégrées. En clair on induit des perturbations d'un réseau cellulaire qui vont permettre de décrire les relations entre les différentes particules de ce réseau et à son échelle globale.

Quel est le lien avec la création et la modification de cellules vivantes ? La compréhension des mécanismes réglant précisément le comportement biologique d'une cellule (*E. coli* par exemple) ouvre la voie à l'élaboration du

gène design. Les cellules incorporent un gène construit par ingénierie génétique influençant les circuits cellulaires de telle manière à produire une fonction. Par exemple une construction génique de base est l'interrupteur, exprimant un gène dans une conformation lorsqu'on injecte une molécule et un autre lorsqu'on augmente la température (graphe *behaviour* : avec l'IPTG le gène s'exprime et à Heat le gène est inhibé). Le design du gène est juste un schéma de l'organisation du gène et de la séquence.



Un exemple plus complexe est le *quorum sensing*, un mécanisme bactérien détectant et réagissant à une densité de population en synchronisant l'expression de gènes. Cette technique peut implémenter un rythme circadien pour l'expression d'un gène.

L'objectif de la création de cellules vivantes est de contrôler les différentes phases de croissance, division, différenciation, spécialisation ou encore de mort cellulaire.

Depuis les années 2000 d'énormes progrès ont eu lieu tant en création de nouveaux circuits que dans la découverte de processus biologiques. L'implémentation de portes logiques telles qu'AND, NOR et OR pour l'expression génique permet de complexifier les constructions de réseaux afin d'obtenir différents patterns de cellules. Simplement c'est l'expression d'un gène A uniquement avec B ou bien l'expression de A sans B. Ces notions étant à la base de l'élaboration d'algorithmes nous pouvons bien imaginer que ces découvertes nous rapprochent de travaux liant les cellules et des composants informatiques (Tamsir A et al. Nature, 2011). Récemment l'élaboration de techniques de manipulation génétique tel que CRISPR Cas (venant d'un système de défense bactérien) réduit les coûts et établit une approche holistique des réseaux. Ce développement est à l'origine de la compréhension des mouvements et des formes cellulaires. Néanmoins la variabilité inhérente à la manipulation biologique ralentit le développement de la biologie synthétique. La création d'un organisme nouveau totalement synthétique n'est pas encore à la portée de cette science (Cameron et al. Nature Reviews, 2014).

THEME 3 : Un robot est-il capable d'adopter des comportements émotionnels ? (D. Ledanois)

Si la science-fiction nous inonde de robots perfectionnés au point de posséder une conscience, des émotions, et par lequel on pourrait les confondre à des êtres humains à la manière de séries tel que la série *Westworld* ou le jeu *Detroit Become Human*, est-ce que nous nous en sommes proches ou ceci n'est qu'un rêve illusoire ?

À l'heure actuelle, même si la philosophie a défini ce qu'était la conscience, la science n'a pas réussi à la comprendre, comme le déclare le chercheur au laboratoire d'informatique de Paris VI : « On ne sait pas qu'est la conscience, on n'en connaît pas les fondements. On n'est donc pas capable de créer une machine consciente ».

Concernant les émotions, l'intelligence artificielle arrive à reconnaître les différentes émotions tels que la surprise, la joie ou la tristesse. En effet, dans l'entreprise Angus.ai, dont les co-fondateurs de cette start-up Gwennaël Gâté et Aurélien Moreau sont d'anciens directeurs logiciels chez Software Robotics où ils avaient développé les « capacités perceptives » des robots Nao et Pepper c'est-à-dire la faculté de pouvoir reconnaître les émotions via une série de capteurs et logiciels. Chez Angus.ai, ils ont par ailleurs amélioré ce système en le rendant possible d'accès sans robot. On fait ingérer à l'intelligence artificielle des milliers et des milliers d'expressions afin qu'il puisse les reconnaître.

Certaines raisons se dégagent pour prouver qu'une machine est meilleure qu'un être humain pour les émotions : La première est que les humains n'ont que trop peu investi dans l'intelligence émotionnelle alors que c'est dans ce domaine-là qu'on a tout investi pour des robots comme Nao et Pepper. En effet, nous naissons avec trois cerveaux : le cerveau reptilien (gère les fonctions vitales de l'organisme : fréquence cardiaque, respiration, température corporelle etc.), le cerveau limbique (concerne les émotions) et le néocortex (le langage, la pensée abstraite, l'imagination et même la conscience). Dès le plus jeune âge, l'éducation privilégie de muscler le néocortex, ainsi en grandissant il est plus entraîné que le cerveau limbique.

La deuxième est la puissance de la machine qui peut repérer jusqu'au plus petit paramètre afin d'établir un diagnostic par rapport à l'état émotionnelle de la personne en face de lui, alors que chez l'humain, une grande part de ce que nous voyons est géré par notre inconscient et donc nous n'en avons pas conscience. En effet, dans son livre *Cerveau augmenté, Homme diminué*, le chercheur en épistémologie franco-argentin Miguel Benasayag, estime que la quantité globale d'apprentissage indirect pouvait représenter jusqu'à 70% de l'information mémorisée, ne laissant qu'une petite partie à notre conscience.

La troisième concerne directement les besoins des humains et des robots. En faisant une comparaison des deux, on se rend directement compte que les humains n'ont pas que les besoins des autres à satisfaire, ils ont aussi les leurs, contrairement aux robots dont la fonction première est directement la satisfaction d'autrui.

Mais ce que la machine reconnaît n'est-il pas que le sommet de l'iceberg ? Les émotions, c'est bien plus qu'un simple rire ou une surprise, la machine ne peut pas reconnaître ce qui est au fond du cœur humain : le doute, l'amour ou la haine. L'algorithme de la machine ne détecte pas forcément une expression mensongère, en s'entraînant tel un acteur de théâtre ou de cinéma, on peut donner l'illusion d'un franc sourire alors que ce n'est pas du tout la réalité. Si la machine ne peut pas reconnaître et comprendre tous les types d'émotions d'un être humain, comment peut-on lui faire ressentir des émotions à autrui ?

Ainsi, à défaut de pouvoir réellement ressentir les émotions, elle peut néanmoins les simuler afin de transmettre des émotions aux êtres humains. Et sur ce domaine-là, Catherine Pelachaud, directrice du CNRS et à Télécom Paris-Tech, pense que la machine en est capable. En effet, toujours selon elle : « le ressenti est du domaine de l'homme, et ça doit le rester ! Une machine est là pour pallier des besoins. Pour cela, la simulation peut suffire. »

De plus, il ne faut pas oublier tout le côté du comportement non-verbal qui apporte énormément à la communication, cela permet de rendre plus convaincant nos propos, et peut nous faire oublier qu'on a en face de nous une machine. Mais si on inscrit dans l'algorithme des comportements non verbaux comme bouger les mains quand on parle, ou se mordre la lèvre quand on est en colère, qu'est qui différenciera une machine d'une autre ? Chaque homme a ses propres comportements non-verbaux, les machines pourront-elles en être ainsi ou seront-elles limitées dans leurs choix ? Les comportements non verbaux changent aussi de culture en culture, devront-on faire des robots spécialement pour le Japon ou la France ?

Enfin, il faut aussi tenir compte du réalisme de l'humanoïde. Si le visage du robot est un bloc de métal comme dans le film d'animation *Wall-e*, il aura du mal à exprimer une émotion mais s'il est très réaliste comme dans la série *Humans* il pourra réussir à tromper l'observateur humain. Mais entre ces deux extrêmes, on tombe dans ce qu'on appelle « la vallée de l'étrange ». C'est un concept, créé dans les années 1970 par le roboticien japonais Masahiro Mori, exposant que les visages ressemblant le plus possible à celui d'un humain soient ceux qui angoissent le plus les vrais humains. Ainsi, des études ont démontré que des prothèses mécaniques comme les lames des coureurs handisports sont mieux acceptées des patients que des prothèses ressemblant précisément au membre manquant.

THEME 4 : Un algorithme peut-il rendre compte du langage humain dans toutes ses expressions ? (J-B. David)

Lorsqu'on parle de langage en informatique, on fait souvent référence à l'écriture du programme, à la syntaxe du code. Or, ici le terme langage renvoie à quelque chose de bien plus large que le code ou la syntaxe. On parlera alors de langage dès qu'un système de signes est utilisé dans le cadre d'une communication. Ainsi, chaque signe ou ensemble de signes (signifiant) renvoient à un sens (signifié) dans le cadre d'une sémantique.

D'autre part, l'aspect communicationnel du langage renvoie nécessairement à une intercompréhension, dans le sens où il y a langage seulement s'il y a plusieurs personnes, animaux ou objets qui communiquent. Dans ce cas, le langage humain désigne spécifiquement ces systèmes lorsqu'ils sont développés, utilisés et compris par et pour des êtres humains. Par exemple, une langue comme le français ou l'anglais, la signalisation routière par panneaux ou les émojis sont des langages humains. Ouvrir la définition du langage, que l'on a tendance à réduire à l'utilisation écrite ou orale d'une langue, est ici nécessaire pour se poser des questions plus larges. Le programme peut-il communiquer avec l'être humain ? Si oui, cette communication peut-elle se faire à travers un langage humain ? Si non, pourquoi ?

Si l'on considère la communication comme un échange d'informations entre deux entités (humaines ou non), alors on peut relever plusieurs formes de communications entre humains et algorithmes. Premièrement, il y a la commande du programme via une interface. Par exemple, pour la plupart des algorithmes, il existe une console qui est une interface utilisateur simple qui permet une communication avec le programme. L'utilisateur entre une commande dans un langage formel et, en fonction des commandes, le programme affiche le(s) résultat(s) correspondant(s). Pourtant, cette forme de communication ne va que dans un seul sens : l'utilisateur ordonne et le programme exécute. On peut alors étendre ce schéma à presque toutes les communications entre humains et programmes, où ce dernier n'est jamais plus qu'un simple outil. Plus récemment, de nouveaux algorithmes ont mis en lumière une autre forme de communication.

Il y a donc, deuxièmement, des programmes capables de s'adapter au comportement de leur utilisateur et d'agir en fonction de lui. La grande majorité de ces programmes se sont développés grâce à internet, en raison des données fournies par le grand nombre d'utilisateurs et de leur utilité. En effet, il est bien plus utile d'avoir une expérience personnalisée quand le contenu proposé est vaste et varié, comme sur YouTube, où environ 400h de vidéos sont mises en

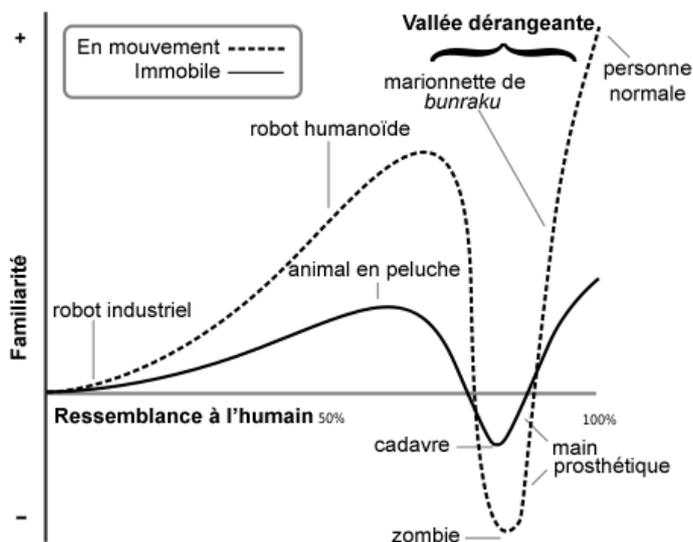
lignes chaque minute. Il faut donc un programme qui oriente nos choix, et ces derniers fonctionnent grâce au « *machine learning* ». En fonction de ce que vous regardez, l'algorithme va construire un profil qui vous correspond et, en comparant ce profil à d'autres, va vous proposer certaines vidéos ou au contraire en écarter d'autres. Autrement dit, le programme fait une suggestion sans même que l'utilisateur ne l'ait demandé. La communication se fait donc dans deux sens : l'utilisateur envoie des informations en cliquant sur les boutons, en faisant certaines recherches et le programme répond par des suggestions de contenu. Le programme n'est plus juste un outil mais un conseiller qui participe activement dans la communication. On pourrait dire, que ce deuxième type de programme n'est que l'extension du premier puisqu'on est toujours dans le cas où un utilisateur entre des commandes via l'interface et reçoit des résultats. Mais ce serait oublier deux spécificités de ces algorithmes. Premièrement, le programme utilise d'autres informations que celles données par l'utilisateur. Autrement dit, il rajoute de la nouveauté dans la communication et n'est donc pas totalement passif. Deuxièmement, l'expérience utilisateur est différente en fonction de qui utilise le programme. Par exemple, si l'utilisateur cherche « ENT » dans Google, les résultats proposés seront différents en fonction de sa localisation et de ses recherches antérieures.

Depuis qu'Alan Turing a formalisé le test du « jeu de l'imitation » en 1950, nombreux sont les informaticiens qui ont essayé de construire des programmes capables de parler dans une langue humaine. Ainsi sont nés les agents conversationnels (ou *chatbots* en anglais), qui là encore se distinguent les uns des autres par leurs fonctionnements. Il y a, premièrement, ceux qui utilisent des schémas de conversations sous formes d'arbres. Ces algorithmes nous intéressent peu car ils sont de fait très limités, même si certains sont capables de repérer des mots clefs dans les questions posées pour proposer la réponse la plus cohérente possible. Il y a, deuxièmement, ceux qui utilisent le « *machine learning* » et apprennent petit à petit des conversations qu'ils ont avec les divers utilisateurs. Ils sont alors censés s'améliorer pour, à terme, parler comme un véritable être humain. On pourrait donc penser que ce sont ces algorithmes-là qui auraient une chance de passer le test de Turing. Pourtant, malgré les nombreuses tentatives, aucun d'entre eux n'y est arrivé. Pour comprendre ce qui ne fonctionne pas avec ces programmes, il est nécessaire d'en essayer plusieurs soi-même.

En fait, même si certaines parties des conversations sont cohérentes, il arrive souvent un moment où ce que dit le robot ne correspond pas. Il peut y avoir plusieurs raisons à cela : le programme peut avoir mal compris la formulation de la phrase ou le double sens d'un mot par exemple. Or, dès qu'il y a des erreurs, l'illusion disparaît. On est alors complètement conscient qu'il s'agit d'un robot et on est incapable de lui prêter une once d'humanité. En outre, plus le robot est

performant par ailleurs, plus l'erreur crée un malaise chez l'utilisateur. On se retrouve ainsi dans la même situation que celle décrite par l'hypothèse de la « Vallée dérangeante » proposée par Masahiro Mori en 1970. Selon les observations du roboticien, plus un robot imite les traits humains, plus ses imperfections nous paraissent monstrueuses. Ainsi, on sera plus familier avec un robot aux traits simples comme les « Cozmo » de chez Anki, qu'avec un robot qui essaie de ressembler à un être humain comme « Sophia » de chez Hanson Robotics. Si l'on applique cette hypothèse au langage, alors on comprend pourquoi un agent conversationnel qui cherche à reproduire le langage humain peut créer le malaise chez l'utilisateur. C'est d'ailleurs à partir de ce constat que les entreprises qui utilisent les dialogueurs, pour orienter leurs clients, utilisent des programmes plus simples, qui proposent moins de fonctionnalités mais qui fonctionnent à coup sûr.

Mais alors, pourquoi les algorithmes sont incapables de communiquer dans une langue, de la même façon que le ferait un être humain ? Pour le comprendre, nous allons aborder une expérience de pensée proposée par John Searle que l'on appelle « la chambre chinoise ». Admettons que l'on vous demande de répondre par écrit à des messages dans une langue que vous ne connaissez pas. Pour ce faire, on vous donne un manuel dans lequel on vous dit quoi répondre en fonction des messages reçus. Le test se passe très bien et l'interrogateur est persuadé que vous parlez sa langue, pourtant ce n'est pas le cas. Vous ne comprenez pas un seul des mots que vous avez écrits.



Ainsi, on comprend pourquoi le test de Turing est fondamentalement inutile. Ce n'est pas parce que la machine vous fait croire qu'elle parle votre langue que c'est le cas. Savoir quoi répondre à une phrase ne veut pas dire que l'on comprend la phrase ou la réponse. La question de savoir si cela fait sens ou non pour l'algorithme reste un mystère, même si l'imitation est convaincante. En fait

ce que souligne Searle ici c'est que le langage humain n'a pas de sens pour un algorithme. Et c'est tout à fait normal, car une langue se forme à partir de besoins purement humains. Par exemple, les verbes « manger », « respirer » ou « dormir » ont un sens pour nous parce qu'on a besoin de manger, respirer et dormir pour vivre et que, donc, ces mots sont reliés à des expériences vécues. Or, la plupart de ce que le langage humain exprime ne peut avoir de sens pour un programme qui se contenterait de répondre à des messages textuels puisqu'il ne peut pas relier les mots à autre chose. Il faut que l'algorithme ou le robot ait besoin de communiquer avec les êtres humains et que ses besoins soient proches des nôtres pour espérer qu'il apprenne un langage humain. Or, dans le cadre des agents conversationnels, ce n'est jamais le cas : le programme est limité à sa simple fonction conversationnelle et c'est pour ça qu'il échoue.

Pour conclure, un algorithme est capable de communiquer avec un être humain et de s'adapter dans le cadre de cette communication, mais cette dernière ne peut pas se faire dans un langage humain. En effet, nos langues ne font pas sens pour l'algorithme qui se limiterait à l'analyse de textes. En revanche, nous pourrions proposer l'hypothèse suivante : les signes font sens pour le programme lorsqu'ils sont reliés à sa fonction première. Autrement dit, un clic a plus de sens pour l'algorithme de YouTube qu'une phrase simple pour Mitsuku. Attention tout de même aux conclusions hâtives, nous ne disons pas que l'algorithme de YouTube pense. En revanche, si les algorithmes sont dotés d'une forme, même primitive, de psyché, alors ce qui fait sens pour eux doit être relié à leur fonction, à leurs agissements. Donc, pour qu'un algorithme ou un robot maîtrisent un jour un langage humain, il est nécessaire qu'il ait des besoins équivalents aux nôtres et qu'ils agissent de façon comparable à nos actions.

THEME 5 : Les algorithmes peuvent-ils traduire correctement le fonctionnement de différentes sociétés et dans quels buts ? (N. Armogathe)

La possibilité d'algorithmiser intégralement l'humain est loin d'être effective. L'écart reste conséquent entre le techniquement possible, le fantasmé et le fantasmable. Pour autant, grâce aux progrès scientifiques, l'algorithmisation de l'homme est pensable. Il est de fait impératif de discuter des applications scientifiques et technologiques possibles. Nous allons étudier dans cette partie les modalités selon lesquelles les algorithmes peuvent traduire adéquatement le fonctionnement des sociétés. Quels sont les effets des dispositifs produits par la techno-médecine sur les corporéités individuelles et sociétales ?

En sociologie, les algorithmes sont définis comme des dispositifs ayant « *la capacité de capturer, d'orienter, de déterminer, d'intercepter, de modeler, de contrôler et d'assurer les gestes, les conduites, les opinions et les discours des*

êtres vivants » (Agamben, 2006). Ce sont des objets de délégation des responsabilités individuelles visant à optimiser les comportements sociaux. Les dispositifs algorithmiques renvoient à l'action intelligente (efficace, efficiente et économique), de laquelle dépend la santé de l'économie marchande. Le processus de quantification est lui-même encouragé par le principe d'évaluation, qui structure les rapports sociaux et influence les politiques publiques et privées. En vertu de l'idéologie néolibérale, les individus sont premièrement des agents économiques qu'il est nécessaire d'orienter. Les algorithmes se doivent de corriger ce qui fait problème social. Ainsi, les voitures intelligentes ont été créées pour éviter les accidents. Cependant, se posent de nouveaux problèmes liés aux questions de responsabilités juridiques.

Historiquement, l'idée de l'homme-machine provient de la théorisation organisationnelle et scientifique taylorienne. Produits par et pour l'humain, les algorithmes ont radicalement transformé les fonctions et le fonctionnement de nos sociétés modernes. Devenus « capital humain » dans le raisonnement capitaliste et managérial, les humains sont considérés en termes de gestion de stock et de flux. Les mécanismes cérébraux sont perçus comme malléables et traitables. Le développement des sciences cognitives est allé dans ce sens, avec pour postulat que la pensée est un flux d'information traitable, qui permet l'exécution à la chaîne de mouvements. Désormais les investissements en recherche se tournent vers le *deep learning* pour reproduire un système organisé imitant le fonctionnement du cerveau. Les individus étant déjà pensés en terme mécanique, la création de leurs semblables artificiels ne serait qu'une suite logique. En parallèle, la quantification de soi (benchmarking personne) est encouragée et devient la mesure d'évaluation de la valeur d'un individu. Au nom de la performance, chacun a désormais une chance (qui fait foi d'injonction sociale) pour rationaliser son temps, son corps, sa vie. Les individus, désireux d'être maîtres de leur optimisation, incorporent les algorithmes et délèguent à l'outil la mesure de leurs actions par rapport à leurs intentions.

Quelle est alors l'intelligence algorithmique ? Le concept d'intelligence est éminemment politique. Il renvoie à la détermination de ce que les individus sont capables de faire et recevoir. Sociologiquement, il est incontestable, que l'intelligence a toujours servi des enjeux de domination sur l'homme (sur les femmes ou les êtres considérés de « races » inférieures) et sur la nature (plantes et animaux). Dans son artificialité, l'intelligence renvoie à la prétention de l'homme à dominer la nature par sa création dans la similitude. Cependant, au niveau de la médecine, les algorithmes sont porteurs de grandes promesses et les évolutions techniques liées au séquençage d'ADN et du GWAS laissent à penser à une démocratisation de la « techno-médecine ». L'avènement des algorithmes dans le champ du numérique et de la science - notamment statistique, a conduit à une révolution épistémologique (Cardon, 2015). Premièrement, l'humain est

réceptacle et transmetteur de données aux sciences cognitives. Deuxièmement, l'ingénieur travaille avec pour postulat que l'évolution biologique humaine est arrivée à son terme. Il ambitionne ainsi de réparer et d'assurer l'évolution de l'espèce via l'usage de nouveaux dispositifs technologiques (les NBIC - nanotechnologies, biotechnologies, informatique et sciences cognitives). Il est attendu que les algorithmes réalisent des synthèses de données procurées par les Big Data pour la compréhension et la prédiction de pathologies sociales et médicales. Il ne fait nul doute que la construction de l'algorithme est sociale. Celle-ci est intrinsèquement normative, donc participe à la création d'un cadre contraignant. Enfin, le système algorithmique répond à une dualité organisationnelle entre d'un côté la possible émancipation de la condition humaine et de l'autre celui de la surveillance.

Étudiant ce phénomène, l'anthropologue Pascal Picq voit dans les technosciences la présence de l'hyper-capitalisme désireux de terrasser la mort. D'autant que les maladies, et la mort elle-même sont considérées dans nos sociétés occidentales comme des anomalies. Il a établi une corrélation entre la recherche dans l'artefact de l'immortalité et la conquête de l'espace (possible surpopulation). Enfin, il met en exergue un paradoxe dans la résolution des problèmes par l'intelligence artificielle : plus un problème est récent plus il est aisé à résoudre. Ce qui rappelle évidemment le caractère construit et situé des algorithmes et d'un homme-artificiel qui ne serait que le reflet des pratiques et représentations de l'homme-moderne. Son caractère universel poserait de nouveaux problèmes de domination. Les usages sociaux de tels dispositifs donnent à réfléchir quant aux modalités d'évaluation éthique des algorithmes en termes de transparence, sens, sécurité et confiance. Comment vérifier qu'ils soient utilisés éthiquement ? Comment ne pas imposer un universalisme technique et prendre en compte la pluralité des variables comportementales et des schèmes de pensée ? Comment éviter une monopolisation des moyens de production des algorithmes et donc de l'algorithmisation de l'homme ? Dans cette perspective, le label ADEL (*Algorithm Data Ethics Label*) s'attache à étudier les valeurs intrinsèques (les données), valeurs de gestion (cycle de vie des données) et valeurs d'usage (qualité des finalités et résultats) pour définir le degré d'éthique d'un algorithme. Sociologiquement seule une pluralité de systèmes d'évaluation permettrait de mettre en évidence le caractère construit de la morale et du champ des possibilités offert aux individus, comme en ce qui concerne les enjeux de modifications de génomes.

Une réflexion sur la science et la création algorithmique de l'humain ne peut mettre de côté le problème de l'eugénisme. Loin d'avoir disparus ou d'être particulier à ceux se revendiquant d'idéologie raciste, il existe des formes d'eugénisme dans les discours scientifiques ou de sens commun : conscients ou non, radicaux ou euphémisés. L'eugénisme se construit autour de ce qui est

considéré socialement comme anormal donc déviant biologiquement. Il renvoie à l'idée que l'espèce humaine puisse être améliorée par le fait de décourager les individus porteurs de caractères héréditaires indésirables à se reproduire. À titre d'exemple, le progrès scientifique a rendu possible le diagnostic génomique complet d'un fœtus par une simple prise de sang. Sous certaines conditions il est possible de détecter si un enfant est porteur de trisomie. Il en résulte que la plupart des grossesses concernant un enfant trisomique sont interrompues. Ce phénomène social, par sa pratique, est accepté et institutionnalisé. Cela donne à s'interroger sur ce qui sera socialement écarté dans l'algorithmisation de l'homme et sur ce qui fera office d'injonction social. Dans ce sens, le dispositif CRISPR propose de modifier les génomes et de les corriger, de traiter et prévenir les défaillances engendrant des maladies. Comment penser alors ce qui est marginal biologiquement ? Quel degré d'homogénéité ? Les formes de sélections naturelles et sociales ne sont jamais absentes des sociétés, mais se transforment et s'adaptent selon l'éthique en vigueur. Tout groupement humain tend à un universel fantasmé et irréel pour homogénéiser ce qui le constitue.

Les biais humains lors des processus de programmation, constituent des dangers qui faussent les fonctionnements intellectuels des machines. Qu'en serait-il pour un homme entièrement algorithmé ? Quelle « marge sociale d'erreur » ? Les individus qui construisent les algorithmes transfèrent leurs représentations. Et il n'est aucunement possible de nier le caractère sexiste et raciste de nos sociétés. C'est ainsi que Benjamin Bayart, président de la FDN déclare que « Le problème ce n'est pas l'ordinateur, c'est l'intelligence humaine. Et ce n'est pas un problème de chercheur en informatique, c'est un problème de sociologie ».

Enfin, Cardon nous donne à voir que « vue depuis les algorithmes, la société ne repose plus sur de grands systèmes de déterminations, mais elle est une sorte de microphysique des comportements et des interactions que des capteurs placés à bas niveau savent décoder ». Les algorithmes se focalisent sur ce que les individus font, plutôt que sur ce qu'ils aspirent à faire. Pris dans les flux du contrôle machinique, l'individu de la société moderne est défini par son passé qui lui-même détermine sa reproduction future immuable : « le comportementalisme algorithmique, c'est ce qui reste de l'habitus lorsqu'on a fait disparaître les structures sociales ». Comme l'a formulé Deleuze, l'individu des algorithmes n'est alors plus qu'un *dividu* amputé de ce qui fait de lui un être humain.

Déjà, dans les dispositifs dont nous bénéficions, les individus doivent prouver à l'algorithme qu'ils désirent changer. La machine, comme la société, tend à entretenir ce qu'elle veut être immuable : les comportements et idées. Comment penser algorithmiquement le progressisme ?

Les risques de *l'algocratie* sont nombreux, entre la mise à l'écart de la singularité (les signaux faibles), l'hyper-conformisante prédiction, et l'eugénisme sous-tendu. Les algorithmes sont la parfaite traduction des représentations sur ce qu'est et/ou ce que devrait être la société concernée. Le problème étant que ces représentations sont celles des individus qui *détiennent les moyens de production* (qui produisent). Aux vues des éléments aujourd'hui, il est certain que les algorithmes s'inscrivent dans l'idéologie néo-libérale et transfèrent massivement des idées politiquement situées, en termes de classe, genre, race. Visant à *machiniser* l'homme, il est naturel de s'interroger sur l'impact sociétal que cela pourrait avoir sur la *valeur* des humains dont la vie est organisée autour de l'idée de performance. Enfin, en vertu des lois économiques, ce qui serait un *marché de l'homme algorithmé* tendrait à un monopole, alors qui serait gagnant ?

In fine, il faut assurer d'urgence des mobilisations éthiques, philosophiques et politiques pour que les citoyens aient les moyens de décider, car la mutation introduite par la techno-médecine est une mutation irréversible.

EXEMPLE 1 : La justice peut-elle être rendue par des robots ? (R. DELPIERRE)

Afin de concevoir si tout l'être humain est algorithmable, il est intéressant de se demander la question de la possibilité de coder son pouvoir de décision. Les algorithmes auto-apprenants pourraient être une bonne solution pour remplacer le juge, on prendrait alors un ensemble de décision de justice ou de cas et on appliquerait les solutions qui se dégagent de l'étude profonde de tous ces cas. Cependant il a été mis en évidence que ces algorithmes étaient souvent biaisés du fait de la mauvaise qualité des données qui leur sont donnés (I). C'est aussi cette subjectivité qui pose des contraintes dans l'idée de vouloir un jour remplacer le juge pénal par un robot même si cela est parfois possible (II) ou encore l'impossibilité de donner au robot le pouvoir souverain d'apprécier subjectivement l'originalité d'une œuvre de l'esprit (III).

I. La subjectivité des algorithmes auto-apprenant

Il est venu à l'idée en pratique de mettre en place des algorithmes auto-apprenants sur la base des décisions de justice. L'idée qui a émergé à travers le globe, a été mise en lumière par l'exemple de l'algorithme COMPAS, l'idée était de réunir dans une base de données une multitude de décisions de justice pour calculer différents éléments comme la récidive. Cet

algorithme crée par la société Northpointe aux États-Unis s'est avéré biaisé. En effet les individus noirs étaient deux fois plus notés comme possibles récidivistes que les individus blancs. Il s'est avéré que la qualité des données est extrêmement importante dans la mise en place de ces algorithmes. Dès lors que les données sont biaisées les résultats le sont aussi. On se retrouve donc avec des résultats racistes qui reflètent le racisme des données et qui ne reflètent finalement pas la réalité.

On va donc avoir un problème de subjectivité dans la mise en place des résultats, les résultats sont soumis à la subjectivité de chaque décision de justice. Il y a donc une variable sociétale forte dans la collection des décisions. Il est difficile de mettre en place des résultats valables sur la base de données qui datées.

Pour avoir des résultats fiables, il faudrait donc avoir des données dites parfaites, qui ne sont en aucun cas biaisées. Il faudrait donc être sûrs que les données ne sont marquées d'aucune forme de discrimination et cela s'avère compliqué à vérifier, puisque c'est en réalité lorsque l'algorithme apprend seul des décisions et qu'il trouve des solutions biaisées que l'on se rend compte du côté biaisé des données.

II. Deux approches différentes en droit pénal

L'objectivité du juge / radar automatique

Dans le cas des contraventions, la décision peut être purement objective, comme dans le cas des radars, dans ce cas, le seul critère pris en compte est la vitesse. En effet, le radar saisit la différence de vitesse entre la vitesse autorisée et la vitesse réelle. Ce traitement se fait entièrement par le biais d'un algorithme qui pose à chaque situation sa peine en fonction de l'excès de vitesse constaté. Aucun critère humain n'est alors pris en compte dans la décision mais un recours devant le juge en contestation de la contravention peut être possible. Dans le cas des contraventions, ce sont les décisions les moins graves, en effet aucune peine de prison n'est encourue au titre de la contravention. Si on prend cependant le cas des excès de vitesse aggravés, la peine devient délictuelle et le traitement automatisé n'est alors plus privilégié. Dès lors que l'élément à prendre en compte est purement objectif et que la peine n'est pas grave, le recours à l'algorithme semble tout à fait possible et ne semble poser aucun problème.

La subjectivité de la culpabilité et de la peine en droit pénal

Cependant, dans des cas plus graves, c'est à dire les délits et les crimes, la question d'un juge immatériel est plus complexe. En droit pénal il existe deux parties à la décision du juge, d'abord la décision de culpabilité,

coupable ou non coupable, et ensuite la peine. En matière de culpabilité il faut imaginer remplacer un juge ou un jury par un algorithme. Il faudrait réussir à lister tous les critères subjectifs qui sont pris en compte dans la prise de décision du juge. Les décisions sont une accumulation de choix subjectifs liés à différents faits, témoignages ou preuves et c'est donc la prise en compte de tous ces éléments qui permet d'emporter l'intime conviction du juge ou des membres d'un jury. Il est difficile de concevoir qu'un algorithme puisse prendre le même genre de décision relative à la culpabilité.

Ensuite il y a la question des peines, il pourrait y avoir deux possibilités, soit qu'il existe pour chaque infraction une peine, mais en matière de droit pénal, une tentative avait été faite de donner à chaque infraction sa peine sans possibilité de personnalisation mais il s'était alors avéré que les jurys en matière criminelle considéraient que les peines étaient trop sévères par rapport aux infractions réellement commises et préféraient alors déclarer les individus non coupables. Le principe de personnalisation des peines ou encore de modulation de la peine permet ainsi d'adapter la peine à la situation mais aussi à l'individu. On va alors prendre en compte beaucoup de critères. Certains sont des critères purement objectifs comme l'âge et d'autres sont des critères plus subjectifs comme la dangerosité de l'individu. Le juge prend en compte l'ensemble de ces éléments de manière globale pour trouver une peine adaptée, on ne sait cependant pas à quel degré il prend en compte les différents critères, il doit bien entendu pouvoir justifier des raisons principales qui ont motivé sa peine mais il n'a pas à donner précisément à quel degré il prend en compte chaque élément, c'est dès lors sur cette base là qu'il est difficile de donner à un algorithme le même rôle puisqu'il faudrait donner à l'algorithme un ordre d'importance des éléments alors qu'on ne connaît jamais réellement cette importance tant les décisions des juges sont partiellement subjectives.

III. L'exemple de l'appréciation de l'originalité en droit de la propriété littéraire et artistique.

En matière de propriété littéraire et artistique, un des éléments essentiels à la protection d'une œuvre de l'esprit est le critère de son originalité. Il n'y a pas vraiment de définition d'originalité de l'œuvre, dans certains cas, on va présumer l'originalité de certaines œuvres classiques comme la littérature, la musique et les beaux-arts et dans ces cas-là l'algorithme pourrait la présumer mais pour les autres œuvres, le critère de l'originalité revient à l'unique appréciation du juge du fond, c'est à dire à son jugement personnel et subjectif, et est même exclu de celle de la Cour

de Cassation. Cette originalité serait basée sur trois définitions pour le Professeur Laure Marino « l'empreinte de la personnalité de l'auteur, la marque de l'apport intellectuel de l'auteur et l'expression des choix libres et créatifs de l'auteur ». Or, l'appréciation souveraine du juge fait que l'on ne connaît absolument pas les critères sur lesquels celui-ci se base pour rendre sa décision. En effet il définit de manière purement subjective si l'œuvre peut refléter la personnalité de son auteur, or il est impossible de donner des détails à l'algorithme pour qu'il puisse prendre la même décision. On ne pourrait pas donner à un algorithme des éléments purement objectifs pour qu'il puisse prendre sa décision. Quels éléments pourrait-on alors donner à un algorithme ou un robot pour qu'il donne un avis humain et personnel sur l'originalité d'une œuvre de l'esprit ? Il existe beaucoup de questions, comme celle de l'appréciation de l'originalité d'une œuvre, pour lesquelles la réponse numérique n'est pas à ce jour et ne sera peut-être jamais adaptée.

EXEMPLE 2 : Quelle fiabilité pour la conduite automatique des avions ? (R. MATHON)

La mobilité fait partie des grands axes de développement d'algorithmes intelligents. Les voitures et transports en commun autonomes ne font plus partie du monde de la science-fiction. Quelle prospective pour ces moyens de locomotion, contrôlés par des algorithmes ? Peut-on algorithmiser le pilote de ligne ?

Afin de décrire le degré d'autonomie des véhicules, la *National Highway Traffic Safety Administration*, agence fédérale américaine chargée de la sécurité routière, a établi une classification comportant 5 niveaux. Le niveau 0, le plus bas, correspond à un véhicule sur lequel le pilote réalise toutes les tâches de la conduite. Le niveau 5 correspond à une conduite totalement autonome, sans condition, avec option pour le conducteur de reprendre le contrôle. Aujourd'hui, la majorité des véhicules sont dits de niveau 1 et disposent de fonctions automatisées (ABS et ESP pour augmenter la sécurité des véhicules, régulateur de vitesse pour le confort du pilote). Les véhicules récents, disposant de fonctions automatisées combinées (régulateur de vitesse adaptatif, *park assist*, centrage sur la voie), sont désormais de niveau 2.

Mais l'utilisation d'algorithmes intelligents dans nos voitures ne

s'arrête pas là. Google a débuté le développement de son système de pilotage automatique dès le mois d'octobre 2010. Aujourd'hui, l'algorithme a accumulé une expérience de 16 millions de kilomètres de conduite sur les routes américaines, ainsi que plus de 10 milliards de kilomètres sur simulateur. Depuis fin 2018, les habitants de la ville de Phoenix dans l'Arizona peuvent déjà profiter d'un service de taxis autonomes. Accepter la circulation de véhicules autonomes sur nos routes ne pourra cependant se faire qu'avec une réduction nettement visible des accidents, tout en améliorant la mobilité et la qualité de vie des utilisateurs.

Si les algorithmes intelligents se multiplient sur les voitures, il en est aussi de même sur les avions. Le développement de systèmes de navigation dans les cockpits a débuté dans les années 1960. Le Concorde a été le premier avion de ligne à disposer de commandes de vol électrique. Pour la première fois, un circuit électronique assurait la commande des actionneurs de l'avion, en réponse aux ordres du pilote et des informations fournies par les différents capteurs. Aujourd'hui, les systèmes de commande des avions permettent d'étendre le domaine de vol au-delà de leur capacité intrinsèque, et d'améliorer les qualités de pilotage. Les architectes systèmes déterminent alors le compromis entre stabilité naturelle de l'avion, et les avantages tirés de ces lois de commandes ; la réduction de la consommation étant très souvent privilégiée.

C'est dans ce cadre qu'a été développé le système anti-décrochage MCAS (*Maneuvering Characteristics Augmentation System*) des Boeing 737 Max. Ce logiciel, suspecté d'être impliqué dans le crash de deux avions (Vol 610 Lion Air, vol 302 Ethiopian Airlines), a été conçu pour compenser l'effet d'une modification du design des Boeing 737. Les nouveaux moteurs, consommant moins, mais étant plus volumineux, ont été déplacés, créant un moment cabreur devant être corrigé à l'aide du MCAS. Une maîtrise insuffisante de cette loi de commande par les pilotes, associée à une défaillance capteur, ont amené l'équipage à perdre le contrôle de leur avion.

En parallèle, les constructeurs aéronautiques continuent de développer les systèmes d'aide à la navigation qui équiperont le cockpit du futur. Il est attendu que le nombre de pilotes dans l'avion, fixé à 2 pour les court-courriers, et pouvant atteindre 4 pour les plus long-courrier, soit progressivement réduit. Avec un ensemble de systèmes assistant le pilote et lui conférant plus d'attention au guidage et à la navigation de l'appareil,

l'avion du futur n'aura besoin que d'un seul pilote à bord. Déjà aujourd'hui, les avions suivent un plan de vol inscrit dans le système de navigation, et peuvent réaliser des atterrissages sans intervention des pilotes. Néanmoins, l'incident du MCAS de Boeing souligne l'un des défis majeurs des systèmes de transport autonomes : la difficile prédiction du comportement d'un système en interaction avec l'humain.

La gestion du trafic aérien est également impactée par les algorithmes d'analyse et d'aide à la prise de décision. L'espace aérien de demain est en cours de modernisation, par l'utilisation de nouveaux concepts opérationnels, tels qu'une augmentation du partage d'informations entre tous les acteurs (opérateurs aériens, aéroports, prestataires de service de navigation aérienne), une intégration plus poussée entre les systèmes sol et embarqués, et un emploi généralisé de systèmes de navigation par satellite. Ces nouvelles briques technologiques permettront d'optimiser les trajectoires empruntées par les avions et les séquences de roulage en aéroport, réduisant la consommation de kérosène, l'émission de CO2 et les nuisances sonores.

L'EASA, Agence européenne de la sécurité aérienne, entame ainsi la phase de déploiement de son nouveau programme de gestion du trafic aérien, SESAR (*Single European Sky Air Traffic Management Research*). L'architecture de ce système autorisera le déploiement par la suite de nouvelles briques technologiques, de traitement et d'analyse de ces grands volumes de données produites, afin d'apporter plus de sûreté et d'efficacité par l'usage d'algorithmes intelligents.

Un remplacement total des pilotes et des contrôleurs aériens par une intelligence artificielle est-elle pour autant envisageable ? Certains algorithmes apprenants, tels que les réseaux de neurones, restent encore bien souvent des boîtes noires, dont le raisonnement ne peut pas être reconstruit. Même si la recherche s'oriente sur la programmation d'intelligences artificielles explicables, permettant d'accorder notre confiance en un système autonome, l'usage de l'IA dans des chaînes de décisions critiques telle que la conduite de l'avion ne pourra pas passer l'étape de certification par les agences internationales de sécurité aérienne avant le dépassement d'un certain nombre de barrières technologiques.

CONCLUSION : Émergence et mécanisme de la conscience biologique et algorithmique (B. Anizon)

La conscience est un terme utilisé dans certains domaines scientifiques et littéraires tels que la biologie et la philosophie. En médecine la définition de conscience est utilisée dans le contexte de vigilance et d'éveil. Une personne consciente est réveillée et peut agir en fonction de son environnement. Néanmoins cette définition ne saisit pas les processus complexes de conscience ayant pourtant lieu chez des patients atteints du syndrome d'enfermement (*locked-in syndrom*). Ce dernier empêche toute communication avec l'extérieur due à une paralysie quasi-totale du corps ; il peut néanmoins communiquer à l'aide du clignement de ses paupières. Ce syndrome abordé dans l'ouvrage autobiographique *Le scaphandre et le Papillon* de Jean Dominique Bauby relate sa vie et comment il appréhende son monde intérieur, ses souvenirs de sa vie passée et expériences durant son enfermement. Ce cas a sensibilisé la communauté scientifique au sujet des personnes que l'on pensait être en état végétatif mais qui en réalité ne peuvent pas communiquer. La conscience abordée ici sera donc la capacité à représenter et manipuler mentalement des informations (une sorte de « théâtre mental »), ainsi que la capacité à pouvoir réfléchir sur nos propres pensées, une « métacognition ». (Dehaene et al. *Science*, 2017 ; S. Dehaene. *Le code de la conscience*, 2014)

Récemment des recherches en neurosciences ont permis de ramener ce sujet complexe d'actualité grâce à des techniques d'imageries et des études de réseaux neuronaux complexes. Identifier les corrélats neuronaux de la conscience est un nouvel objectif de la recherche de pointe en neurosciences et plusieurs articles publiés ont pu d'ores et déjà la faire avancer (Stanislas Dehaene à Neurospin). En étudiant la conscience, S. Dehaene a d'abord étudié ce qui n'est pas un processus conscient, les processus inconscients. En effet la majorité des informations perceptives sont d'abord traitées inconsciemment par le cerveau, et ce n'est qu'au bout d'un passage à travers différents « centres de traitements » que l'information peut être manipulée consciemment. La vision est un très bon exemple pour comprendre cette chaîne : la lumière réfléchiée sur un arbre est perçue par les photorécepteurs situés sur la rétine, allant ensuite par le nerf optique vers le cortex visuel primaire. Chaque information visuelle que l'on perçoit consciemment est une construction de notre cerveau créée à partir de la chaîne de traitement de l'information façonnée au cours de l'évolution.

Cette évolution de la vie est d'ailleurs un bon moyen pour s'imaginer comment la conscience s'est développée. A-t-elle surgi du développement du cortex pré-frontal uniquement chez les primates, ou s'est-elle complexifiée au cours du temps, en lien avec d'autres systèmes perceptifs et processus cognitifs comme la

vision et le langage ? Une théorie récente montre que la fonction originelle de la conscience est le mouvement d'un organisme vers de la nourriture. Être conscient de son environnement et de ses états internes (avoir faim) serait la première étape vers l'élaboration d'une conscience. Il faut aussi une étape de perception à cette conscience primaire, comme la vision, qui va permettre de se mouvoir sans dépendre uniquement d'un seul sens. Lorsqu'un prédateur est en vue il est possible de s'enfuir rapidement. La représentation mentale de ce prédateur et de l'environnement permettra de réagir plus rapidement et de garder en mémoire les informations indispensables à sa survie. Un dernier aspect à cette conscience est la perception du temps dans cette représentation mentale. La gratification différée est un bon moyen de savoir si les animaux ont une conscience du temps. Un exemple est la capacité de poulets adultes à résister à manger un repas placé devant eux s'ils s'attendent à recevoir ensuite un plus grand repas. Cela montre que des animaux que l'on ne pense capables d'avoir une conscience complexe adoptent des comportements complexes que même des humains adultes ont du mal à effectuer. Enfin, le langage a permis de complexifier ces représentations de l'environnement et de les communiquer, de transmettre le savoir et le faire perpétuer. Il a permis une rapidité et flexibilité dans la manière de transmettre les informations manipulées dans notre théâtre mental.

En arrivant à cette conscience complexe que nous possédons aujourd'hui, l'hypothèse de pouvoir la modéliser semble impossible vu ce qu'il nous reste à découvrir de notre cerveau. Les algorithmes de type réseaux de neurones, spécialisés dans la reconnaissance de visages ne simulent que des mécanismes inconscients. Même lorsque les algorithmes peuvent être qualifiés de « créatifs » comme l'algorithme spécialisé du jeu de Go, ce ne sont que des mécanismes « simples » de communications entre neurones. En effet les propriétés des neurones biologiques ont été à l'origine de l'élaboration et sont toujours un modèle d'amélioration des algorithmes de type réseaux de neurones. Comme abordée précédemment, la génération d'effets non linéaires par l'établissement de réseaux inhibiteurs, des couches aux projections convergentes et des poids propres à chaque connexion synaptique ont permis l'optimisation des algorithmes. Néanmoins il n'est pas possible de raisonner qu'à l'échelle d'une connexion neuronale lorsqu'on aborde les corrélats neuronaux de la conscience.

Organiser le cerveau humain informatiquement en différents sous-systèmes spécialisés (langage, vision, moteur, mémoire) peut être efficace. Il ressort un problème de cette architecture. Comme il le souligne dans son article, le chercheur S. Dehaene montre que l'architecture organisée de ces sous-systèmes va induire un problème dans le choix d'une action (Dehaene et al. *Science*, 2017). S'il doit choisir entre différentes possibilités il va converger vers une seule décision en prenant non seulement son environnement en compte mais aussi le

temps (gratification différée). Ainsi la conscience algorithmique se devra d'avoir une architecture complexe pouvant se représenter mentalement les informations provenant de son environnement et de son état interne de manière globale. Ces informations se devront d'être accessibles à différents systèmes collaborant entre eux. La conscience ne serait pas contenue dans un système mais émergerait de l'interaction des sous-systèmes.

La dernière étape de la simulation de la conscience est la méta cognition, la cognition de notre cognition. Différents processus chez l'homme sont étudiés comme représentant cette méta-cognition, la confiance de nos décisions, la détection de l'erreur ou encore la capacité d'attribuer un doute ou une confiance à nos souvenirs. Cette dernière nous permet d'optimiser nos stratégies pour mémoriser un cours par exemple. Il faudrait que l'algorithme ait conscience de ses propres décisions et de savoir avec quelle probabilité il peut attribuer une confiance à ses choix. Un algorithme appelé Pathnet utilise un algorithme génétique pour apprendre quel chemin dans son réseau neuronal est le mieux approprié pour une tâche donnée. Cette architecture flexible optimise les résultats en fonction de son environnement et pourrait être la première étape pour un algorithme conscient.

Nous sommes conscients d'avoir laissé de côté l'aspect philosophique de la conscience. Différentes théories et doctrines existent mais les deux principales sont illustrées par Chalmers et Dennett. Chalmers évoque le fait que la conscience est un élément fondamental de la nature, universelle de la même façon que l'espace ou la masse. Chaque atome de l'univers possède un degré de conscience. Une cellule comme un neurone a ainsi un niveau de conscience propre, qui se complexifie dans un cerveau humain (David Chalmers, Comment expliquer la conscience ?). L'autre doctrine abordée par Dennett est l'illusionnisme et déclare que toutes les impressions subjectives de conscience (qualia) que nous ressentons n'existent pas.

Nous avons étudié plusieurs aspects des possibilités et limites de la mise en algorithme de l'être humain. De sa définition philosophique, biologique et informatique, la création d'un tel algorithme requiert de nouveaux moyens de structure informatique inspirés de la neurobiologie. L'invention de ces structures se pourrait être mêlée à des organismes synthétiques, reproduisant des mécanismes logiques en utilisant des molécules biochimiques. Une fois achevée, l'algorithme aura-t-il une émotion ou ne pourra-t-il que les reconnaître sur nos visages et nos réactions physiologiques ? Comprendra-t-il le sens de ces émotions ou sera-t-il bloqué, obligé de répondre à la fonction mise dans son code ? Les limites des comportements du robot se retrouveront ainsi dans ses fondements même, son code et son langage. La question de la difficulté du sens de la langue pour le robot est très intéressante. Il faudrait étudier l'émergence d'un langage, comment se sont développées les bases de nos langues afin de

comprendre comment résoudre le problème de « la chambre chinoise ». Le robot pensera-t-il dans notre langue ou dans la sienne, qui lui donne du sens et qui se rapproche de sa perception en tant que machine ?

Enfin même les algorithmes actuels provoquent des problèmes éthiques, juridiques et sociétaux auxquelles nous sommes confrontés. Le cas de la justice et de la subjectivité inhérente au code et des données d'entrées est une application sensible en droit pénal. Comment créer un juge informatique qui prendrait en compte les éléments objectifs (âge, faits) autant que certains subjectifs (comportement de l'individu). Comment reproduire un jugement en âme et conscience ? L'algorithme ne pourra effectuer cette tâche que lorsqu'il aura développé une sensibilité à ces éléments subjectifs. A une plus grande échelle l'ensemble des décisions de justice de ces algorithmes seront-ils tous les mêmes ou bien tous auront une subjectivité différente développée grâce à leurs expériences ? Introduire un apprentissage dépendant de l'expérience peut éviter de tomber dans l'algocratie. On réduit les biais dus aux transferts de représentations de la personne construisant l'algorithme.

Il est nécessaire de prendre en compte la sociologie dans l'élaboration des algorithmes. Elle permettra de guider le développement d'une intelligence artificielle possédant le moins de « défauts » humains identifiés pour l'instant et d'entraîner le moins d'impacts négatifs sur la population (eugénisme, racisme, hyper-conformisme). Les algorithmes se doivent ainsi d'être étudiés par des initiatives comme l'ADEL (*Algorithm Data Ethics Label*) et d'autres organismes agissant en amont et en aval de la création de l'algorithme. Dernièrement des accidents avec des avions de Boeing révèlent des erreurs dans le code d'un algorithme de stabilisation s'activant lorsque l'avion se redresse du au poids des moteurs (Nom de l'algorithme : MCAS). Certains pilotes ne connaissaient même pas l'existence de cet algorithme, ce qui laisse deviner le peu de précaution avec lequel nous pensons notre rapport à l'algorithme. C'est un outil puissant, qui nous permet déjà d'avancer grandement dans tous les domaines. Néanmoins comme tout outil, son utilisation se doit d'être faite avec un principe de précaution. Il nous faut être tous sensibilisés à ses forces, ses limites et savoir comment l'améliorer autour d'un cadre multidisciplinaire (éthique, juridique, philosophique, scientifique, politique ...).

BIBLIOGRAPHIE :

- A Semi-Synthetic Organism that Stores and Retrieves Increased Genetic Information* Yorke Zhang¹, Jerod L. Ptacin², Emil C. Fischer¹, Hans R. Aerni², Carolina E. Caffaro², Kristine San Jose², Aaron W. Feldman¹, Court R. Turner², and Floyd E. Romesberg¹ ¹Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA ²Synthorx, Inc., La Jolla, California 92037, USA, *Nature*. 2017
- Recoded organisms engineered to depend on synthetic amino acids* Alexis J. Rovner^{1,2}, Adrian D. Haimovich^{1,2,*}, Spencer R. Katz^{1,2,*}, Zhe Li^{1,2}, Michael W. Grome^{1,2}, Brandon M. Gassaway^{2,3}, Miriam Amiram^{1,2}, Jaymin R. Patel^{1,2}, Ryan R. Gallagher^{1,2}, Jesse Rinehart^{2,3}, and Farren J. Isaacs^{1,2} ¹Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA ²Systems Biology Institute, Yale University, West Haven, CT 06516, USA ³Department of Cellular and Molecular Physiology, Yale University, New Haven, CT 06520, USA, *Nature*. 2015
- The Expanded Genetic Alphabet* Dr. Denis A. Malyshev and Prof. Floyd E. Romesberg*, ^a a Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037 USA *Chem Int Ed Engl*. 2015
- Genetic alphabet expansion biotechnology by creating unnatural base pairs* Kyung Hyun Lee, Kiyofumi Hamashima, Michiko Kimoto and Ichiro Hirao *current opinion in biotechnology*, 2018
- A semi-synthetic organism with an expanded genetic alphabet*, Denis A. Malyshev¹, Kirandeep Dhami¹, Thomas Lavergne¹, Tingjian Chen¹, Nan Dai², Jeremy M. Foster², Ivan R. Correia Jr ² & Floyd E. Romesberg¹ *Nature*, 2014
<http://2018.igem.org/Team:Toulouse-INSA-UPS/Description>
- Self-organization of Symbiotic Multicellular Structures* Jean Disset, Sylvain Cussat-Blanc, Yves Duthen *HAL*, 2014
- The incorporation of epigenetics in artificial gene regulatory networks* Author links open overlay panel Alexander P. Turner ad Michael A. Lonesad Luis A. Fuente ad Susan Stepney bd Leo S. D. Caves cd Andy M. Tyrrell ad *Elsevier*, 2013
https://www.lemonde.fr/pixels/article/2015/10/12/intelligence-artificielle-une-machine-peut-elle-ressentir-de-l-emotion_4787837_4408996.html
https://services.la-croix.com/webdocs/pages/longform_robots/index.html
- Masahiro Mori, « *La vallée de l'étrange* », *Gradhiva. Revue d'anthropologie et d'histoire des arts*, n° 15, 16 mai 2012, p. 26–33 (DOI [10.4000/gradhiva.2311](https://doi.org/10.4000/gradhiva.2311)). (traduction fr. Isabel Yaya)
- exemples : « Spring », « Ask Mona » et « Weabot »
- J. R. Searle, “Minds, Brains and programs”, *The Behavioral and Brain Sciences*, vol. 3, Cambridge University Press, 1980, tr. fr. “Esprits, cerveaux et programmes”, in D. Hofstadter, D. Dennett, *Vues de l'Esprit*, Paris, Interéditions, 1987, pp. 354-373
- Cardon Dominique, *À quoi rêvent les algorithmes : nos vies à l'heure des big data*, 2015.

Déchaux, Jean-Hugues. « *La procréation à l'ère de la révolution génomique* », *Esprit*, vol. octobre, no. 10, 2017, pp. 113-129

France Culture, *Papiers*, n°26, vol. octobre-décembre 2018, Les révolutions de l'intelligence.

« *Les avancées biotechnologiques conduisent-elles toujours à un progrès pour les patients ? Un regard éthique.* », Clefs CEA, n°67, 13 décembre 2018, Contexte - Technologie - Domaines d'application – Perspective.

Matières à Penser, Les grands espoirs en médecine : Modifier un gène et guérir, René Frydman, diffusé le 14 février 2019, France Culture: <https://www.franceculture.fr/emissions/matieres-a-penser/les-grands-espoirs-en-medecine-45-modifier-un-gene-et-guerir>

La Méthode Scientifique, Génomique : comment déchiffrer le vivant ?, animé par Nicolas Martin, diffusé le 20 novembre 2017, France Culture: <https://www.franceculture.fr/emissions/la-methode-scientifique/la-methode-scientifique-lundi-20-novembre-2017>

L'origine de la conscience - Comment les choses inconscientes sont devenues conscientes : Kurzgesagt : in a nutshell

<https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>

<https://www.boeing.com/commercial/737max/737-max-software-updates.page>

<https://www.ecologique-solidaire.gouv.fr/sesar-volet-technologique-du-ciel-unique-europeen>

<https://www.intel.fr/content/www/fr/fr/it-managers/explainable-ai.html>

Bibliographie secondaire :

Foucault, Michel, *Les mots et les choses*, Gallimard, Paris, 1966

Turing, Alan. « Computing Machinery and Intelligence ». *Mind*, 1950. Oxford University Press, vol. 59, n° 236, octobre 1950. (Traduction : Alan Ross Anderson (dir.) (trad. Patrice Blanchard), *Pensée et machine*, Champ Vallon, 1983)

<https://www.suricats-consulting.com/humain-robot-intelligence-artificielle-emotion-robotique/>

<https://www.lesechos.fr/intelligence-artificielle/veille-technologique/0301990677687-robots-et-emotions-lequation-impossible-2201016.php>

<https://www.lesechos.fr/idees-debats/cercle/cercle-177129-quand-robot-rime-avec-emotion-lia-au-service-de-la-relation-client-2138150.php>

https://www.huffingtonpost.fr/laurent-alexandre/sequencage-adn_b_3266107.html

<http://lifecodexx.com/en/expectant-mothers/prenatest/>

<https://iatranshumanisme.com/transhumanisme/intelligence-artificielle-dimensions-socio-economiques-politiques-et-ethiques/>

https://www.lemonde.fr/campus/article/2018/04/25/les-neurosciences-vont-elles-faire-de-l-homme-une-machine_5290332_4401467.html