

Défis communs en matière de droits de l'homme soulevés par les différentes applications des neurotechnologies dans le domaine biomédical

Rapport commandé par le Comité de bioéthique (DH-BIO) du Conseil de l'Europe

Auteur : Marcello Ienca^{1,2}

¹ Département des Sciences et Technologies de la Santé, ETH Zurich, Zurich, Suisse.

² Collège des Humanités, EPFL, Lausanne, Suisse.

Table des matières

<i>Résumé exécutif</i>	2
1. Introduction : La révolution neurotechnologique	5
1.1. Histoire et taxonomie de la neurotechnologie	5
1.1.1. Imagerie cérébrale.....	5
1.1.2. L'électroencéphalographie intracrânienne	8
1.1.3. Interfaces cerveau-machine pour aider le contrôle moteur et la communication.	9
1.1.4. Neurostimulation.....	12
1.1.5. Neurotechnologie omniprésente	13
1.2. Quelles informations la neurotechnologie peut-elle lire ?	16
1.3. Les informations que la neurotechnologie ne peut pas lire	20
2. Défis éthiques	23
2.1. Neuroéthique et neurodroit	23
2.2. Vie privée	24
2.3. Autonomie, agentivité et responsabilité	28
2.4. Justice	29
3. Les défis des droits de l'homme	31
3.1. Lacunes dans les cadres internationaux des droits de l'homme	32
3.2. Introduction de la notion de « neurodroits »	37
3.3. Définition des « neurodroits »	39
3.4. Ancrer les « neurodroits » dans l'histoire des idées	39
3.4.1. Liberté de pensée et de conscience	39
3.4.2. Vie privée	42
3.4.3. Intégrité mentale	44
3.4.4. Identité personnelle.....	47

3.5. Une taxonomie conceptuelle des « neurodroits »	48
3.5.1. Famille thématique : la liberté de pensée	49
3.5.2. Famille thématique : la vie privée	51
3.5.3. Famille thématique : l'intégrité mentale	53
3.5.4. Famille thématique : l'identité personnelle et la continuité psychologique.....	55
3.5.5. Principes éthiques annexes	56
4. Les « neurodroits dans la politique internationale »	58
5. Questions ouvertes et avenir des « neurodroits »	60
6. Conclusions et options politiques	64
Annexe 1 : Glossaire	68
Remerciements	69
Références	69

Résumé exécutif

Les neurotechnologies sont des technologies émergentes qui établissent une connexion avec le cerveau humain, et grâce auxquelles l'activité neuronale humaine peut être enregistrée et/ou modifiée. Ces technologies ouvrent de nouvelles possibilités pour explorer et influencer le cerveau humain, ou pour communiquer avec lui. Les neurotechnologies médicales offrent la possibilité d'aider les personnes souffrant de troubles neurologiques ou psychiatriques tels que la maladie de Parkinson, la démence, les accidents vasculaires cérébraux et les troubles dépressifs majeurs. Les systèmes neurotechnologiques non médicaux fournissent de nouveaux outils et de nouvelles méthodes pour surveiller et moduler l'activité cérébrale chez les sujets sains, ainsi que pour leur permettre d'interagir avec les dispositifs numériques. Pouvoir intervenir efficacement et en toute sécurité sur le cerveau humain grâce aux neurotechnologies est une frontière scientifique qui doit être franchie pour le bien de l'humanité. Cependant, cela soulève dans le même temps des défis éthiques et juridiques majeurs. La neuroéthique et le neurodroit sont les deux principaux domaines de recherche qui traitent respectivement des questions éthiques et juridiques soulevées par notre capacité toujours plus grande à agir sur le cerveau par le biais des neurotechnologies.

Au cours de la dernière décennie, les études philosophico-juridiques dans les domaines de la neuroéthique et du neurodroit ont accordé une importance croissante à l'analyse normative des défis éthico-juridiques que posent les sciences de l'esprit et du cerveau en matière de droits, de libertés et d'obligations. Cette entreprise d'analyse des implications éthiques et juridiques des neurosciences porte le nom de « neurodroits ». Les « neurodroits » peuvent être définis comme les principes éthiques, juridiques, sociaux ou naturels de liberté ou de droit dans ce qui touche au domaine cérébral et mental d'une personne. Il s'agit donc des règles normatives fondamentales régissant la protection et la préservation du cerveau et de l'esprit humains. Dans leur version la plus répandue, les « neurodroits » sont caractérisés comme une catégorie émergente de droits de l'homme destinés à protéger l'espace cérébral et mental de la personne.

La réflexion sur les « neurodroits » a fait l'objet d'une large couverture dans les médias grand public, et est devenue une thématique courante du discours public consacré à la

neuroéthique. En outre, elle est en train de devenir un outil réglementaire en politique internationale. Pourtant, plusieurs défis méta-éthiques, éthico-normatifs, juridico-philosophiques et pratiques doivent encore être résolus pour que les « neurodroits » puissent être utilisés comme des instruments efficaces dans la gouvernance mondiale des neurotechnologies, et être intégrés de manière adéquate dans le droit international des droits de l'homme. Dans le but de relever ces défis, ce rapport tente de fournir une analyse éthico-normative, historique et conceptuelle complète sur les « neurodroits ». En particulier, ce rapport poursuit un quintuple objectif : (i) fournir une vue d'ensemble des neurotechnologies biomédicales actuelles et futures ; (ii) reconstituer l'histoire des « neurodroits » et les situer dans l'histoire plus large des idées ; (iii) résumer les initiatives politiques en cours concernant les « neurodroits » dans le paysage politique international actuel ; (iv) aborder de manière proactive certains défis éthico-juridiques non résolus ; et (v) identifier les domaines prioritaires pour orienter la poursuite de la réflexion universitaire et du travail politique dans ce domaine.

Les conclusions de ce rapport suggèrent que les « neurodroits » reflètent des intérêts humains fondamentaux qui sont profondément ancrés dans l'histoire des idées. Ces droits introduisent des spécificités normatives quant à la protection du domaine cérébral et mental de la personne. Ces spécificités ne s'inscrivent pas dans une répétition des cadres de droits de l'homme déjà existants, mais ajoutent un nouveau niveau fondamental de protection normative. Cela corrobore l'idée que les êtres humains jouissent d'un ensemble général de droits les protégeant contre certains types d'interférences visant leur cerveau et leur esprit, y compris contre des interférences liées à l'utilisation abusive des neurotechnologies. Dans une large mesure, les conclusions du présent rapport corroborent également la thèse, plus forte sur le plan normatif, selon laquelle les droits et libertés fondamentaux relatifs au cerveau et à l'esprit humains doivent être considérés comme le substrat fondamental de tous les autres droits et libertés.

Cette vue d'ensemble montre qu'il n'existe pas encore de consensus total au sujet des limites normatives et conceptuelles des « neurodroits », pas plus que sur la terminologie à employer. Des divergences existent quant à la manière dont ces droits sont interprétés, nommés et conceptuellement articulés. Néanmoins, un certain degré de convergence se dessine autour de la désignation de trois grandes familles de « neurodroits ». Tout d'abord, la nécessité de dispositions spécifiques pour protéger les informations privées contenues dans le cerveau semble faire l'objet d'un haut degré d'acceptation et de reconnaissance. Le droit à la vie privée mentale semble être le candidat le plus solide conceptuellement pour assumer ce rôle. Deuxièmement, le droit à l'intégrité mentale semble avoir le plus haut degré d'ancrage juridique. Bien qu'il y ait quelques variations dans l'interprétation de ce droit, il existe un consensus théorique total sur la nécessité de protéger la personne contre les dommages psychologiques et les interférences mentales. Troisièmement, il y a eu différentes propositions de « neurodroits » visant à préserver et promouvoir la liberté de l'esprit humain, et empêcher ainsi toute manipulation extérieure. Il s'agit notamment d'interprétations qui cherchent à faire évoluer le droit à la liberté de pensée, le droit à la liberté cognitive et le droit à l'identité personnelle.

À côté de cela, des droits positifs tels que la promotion de la justice et de l'égalité – garantir, par exemple, un accès égalitaire aux neurotechnologies biomédicales et promouvoir

le bien-être des patients sur la base du principe éthique de bienfaisance – ont jusqu'à présent occupé un rôle secondaire dans le débat sur les « neurodroits ».

L'introduction des « neurodroits » dans le cadre des droits de l'homme peut nécessiter d'ajouter de nouveaux protocoles aux instruments existants, ou même de créer de nouveaux instruments multilatéraux spécifiquement dédiés à la neuroéthique et au neurodroit. Dans les deux cas, certaines questions éthiques, méta-éthiques et juridiques fondamentales doivent être résolues afin de surmonter des problèmes tels que l'inflation des droits, et de fournir aux « neurodroits » une justification normative adéquate. Il s'agit par exemple de présenter des tests et examens justifiant l'introduction des « neurodroits », de clarifier la relation entre « neurodroits » moraux et « neurodroits » juridiques, et d'harmoniser les « neurodroits » avec les instruments normatifs existants.

La Convention du Conseil de l'Europe pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine (Convention d'Oviedo) offre une plate-forme et une base normative idéaux pour protéger et promouvoir les « neurodroits ». Étant donné qu'elle met l'accent sur l'interdiction de l'utilisation abusive des innovations dans le domaine de la biomédecine, sur la protection de la dignité et de l'identité de tous les êtres humains et sur la garantie du respect de leur intégrité et de leurs libertés fondamentales, la Convention est parfaitement adaptée pour consacrer des « neurodroits » tels que le droit à l'intégrité mentale, le droit à l'identité personnelle et le droit à la liberté cognitive.

Comprendre, traiter et améliorer le cerveau et l'esprit humains constitue l'un des grands défis scientifiques de notre époque. Atteindre ces objectifs tout en préservant la justice, les droits fondamentaux et la dignité humaine, telle est la tâche parallèle de l'éthique et du droit. Les « neurodroits » seront probablement un outil utile pour accomplir cette tâche.

1. Introduction : La révolution neurotechnologique

La *neurotechnologie* est le terme générique couramment employé pour décrire un ensemble large et hétérogène de méthodes, de systèmes et d'instruments qui établissent une connexion avec le cerveau humain, et grâce auxquelles on peut enregistrer et/ou modifier l'activité neuronale. En d'autres termes, nous pouvons appeler neurotechnologie toute technologie permettant d'explorer, d'influencer ou d'intercommuniquer avec le cerveau humain.

Au cours des trois dernières décennies, l'innovation technologique et les découvertes scientifiques dans les domaines des neurosciences et de l'ingénierie biomédicale, alliées aux progrès concomitants de la modélisation informatique et des logiciels d'apprentissage automatique utilisés pour l'analyse des données, ont conduit à accumuler les progrès en matière de neurotechnologie. Ces progrès ont permis l'avènement d'un large éventail d'applications cliniques touchant à la prévention, au diagnostic, au traitement et aux soins des personnes atteintes de troubles neurologiques et psychiatriques. En raison du rythme soutenu de ces progrès, des auteurs ont baptisé ce processus du nom de « révolution neurotechnologique » (Scott, 2013). En outre, la réduction des coûts du matériel, ainsi que l'augmentation de la taille du marché des neurotechnologies, ont récemment permis la diffusion d'applications neurotechnologiques dans différents domaines extra-cliniques de l'activité humaine, comme la justice, la défense, la communication et l'industrie de consommation. Cette diffusion en dehors du cadre purement clinique s'est accompagnée d'une réflexion accrue, d'un débat, et, plus récemment, d'une délibération sur les implications éthiques, juridiques et sociales de la neurotechnologie. En raison des progrès technologiques et de la diffusion sociale des applications neurotechnologiques, l'intérêt du public pour la neurotechnologie s'est en effet accru ces dernières années.

Le présent rapport poursuit un quintuple objectif. Premièrement, il fait le point sur les systèmes et méthodes neurotechnologiques actuels et futurs. Deuxièmement, il donne un aperçu critique des implications éthiques et juridiques actuelles et à venir des neurotechnologies, en mettant l'accent sur leurs implications en matière de droits de l'homme, notamment sur l'émergence de la catégorie des « neurodroits ». Troisièmement, il résume les initiatives politiques en cours sur les « neurodroits » dans le paysage politique international. Quatrièmement, il aborde de manière proactive certains défis éthiques et juridiques encore non résolus, qui sont liés à l'introduction des « neurodroits » dans le cadre des droits de l'homme. Enfin, il identifie des domaines prioritaires pour orienter la poursuite de la réflexion académique et du travail politique vers ces derniers.

La première section de ce rapport présente un résumé de l'histoire de la neurotechnologie, une vue d'ensemble des systèmes et des méthodes neurotechnologiques actuels, et une évaluation critique des capacités et des limites de ceux-ci.

1.1.Histoire et taxonomie de la neurotechnologie

1.1.1. Imagerie cérébrale

L'histoire de la neurotechnologie débute à la fin du XIXe siècle. Dans la seconde moitié du XIXe siècle, différentes techniques ont été mises au point pour détecter les composantes de l'activité cérébrale. En 1873, le physiologiste italien Camillo Golgi découvre une méthode d'imprégnation chromatographique qui lui permet de visualiser les neurones au microscope.

Deux ans plus tard, le physiologiste anglais Richard Canton utilise un instrument électromécanique appelé galvanomètre (du nom de l'anatomiste italien Luigi Galvani) pour observer les impulsions électriques provenant des surfaces du cerveau d'un lapin et d'un singe, tous deux vivants. Quarante-neuf ans plus tard, en 1924, le psychiatre allemand Hans Berger enregistre le premier électroencéphalogramme (EEG) humain. Grâce à un instrument de sa propre invention, Berger a pu observer et mesurer l'activité électrique du cerveau humain et découvrir les ondes alpha, c'est-à-dire les oscillations neuronales situées dans la gamme de fréquences de 8 à 12 Hz, qui se sont plus tard avérées très importantes dans les processus de *neurofeedback*. Depuis, l'EEG, en tant que méthode de surveillance électrophysiologique, est abondamment utilisé en clinique comme en recherche pour enregistrer l'activité électrique du cerveau. Cette activité électrique y est enregistrée de manière non invasive, c'est-à-dire depuis l'extérieur du crâne : des électrodes sont placées sur le cuir chevelu et mesurent les fluctuations de tension issues des flux d'ions qui parcourent l'intérieur des neurones, et qui, ce faisant, génèrent un courant électrique.

Parallèlement à l'électricité, le magnétisme a aussi contribué à dévoiler l'activité cérébrale aux scientifiques. En 1971, le chimiste américain Paul Lauterbur s'est servi de gradients de champ magnétique et d'une technique de rétroprojection pour créer des images d'objets physiques. Les premières images obtenues à l'aide de cette technique ont été publiées en 1973 dans la revue *Nature*. C'était la première fois que des champs magnétiques et des ondes radio étaient utilisés pour recréer des images de corps physiques (Lauterbur, 1973). Lauterbur a donné à sa méthode le nom de « zeugmatographie », terme qui a ensuite été remplacé par l'expression plus courante « **imagerie par résonance magnétique** » (**IRM**). L'IRM s'est rapidement révélée très utile pour visualiser les structures internes du corps, par essence difficiles à observer. Il était déjà clair à cette époque que le cerveau humain était un sujet de prédilection. La toute première IRM du cerveau humain fut réalisée en 1982 par le scientifique américain John Schenck à l'aide d'un aimant de 1,5 tesla suffisamment puissant pour voir à travers le corps humain et en obtenir des images claires et haute résolution.

Toutefois, disposer de cartes anatomiques du cerveau ne suffit pas pour observer les *mécanismes* cérébraux, c'est-à-dire les processus et les fonctions du cerveau. Pour réaliser de telles observations, il a fallu développer de nouvelles neurotechnologies. Grâce aux recherches pionnières du physiologiste Angelo Mosso, on avait déjà émis l'hypothèse, à la fin du XIXe siècle, que le flux sanguin du cerveau était intimement relié à la fonction cérébrale. Mosso avait observé que, lorsque des patients effectuaient des tâches cognitives comme des calculs mathématiques, les pulsations cérébrales augmentaient localement. Vers la fin des années

La **fonction cérébrale** désigne le fonctionnement des circuits neuronaux dans le cerveau. Il s'agit des manifestations neuronales d'états mentaux tels que les pensées, les expériences et les actions : compte tenu des nombreuses répétitions d'une pensée, d'une expérience ou d'une action, il est possible d'utiliser des méthodes statistiques pour déterminer de manière fiable quelles zones du cerveau ont subi un changement de signal magnétique, c'est-à-dire quelles zones du cerveau

1980, grâce à l'IRM haute résolution, des chercheurs du Massachusetts General Hospital ont confirmé et documenté cette augmentation du débit sanguin cérébral dans les zones soumises à une plus grande activité neuronale. La méthode qu'ils ont utilisée exploite les images du cortex visuel primaire : ils ont calculé les volumes sanguins cérébraux obtenus

Figure 1- Fonctionnement du cerveau

dans le cadre d'une stimulation, puis les ont comparés aux volumes obtenus dans le cadre d'un état non stimulé. Ils ont ainsi pu réaliser des cartes fonctionnelles du cerveau, qui montrent une activité accrue dans les zones soumises à une stimulation. Presque à la même époque, Seiji Ogawa, de l'Université de Tokyo au Japon, a montré qu'il était possible de détecter les différentes propriétés magnétiques de l'hémoglobine, dans sa forme oxygénée comme désoxygénée, en utilisant l'IRM. En exploitant les propriétés de cette technique, on peut en effet observer les variations du signal magnétique au sein du cerveau. Grâce à cette nouvelle méthode, connue aujourd'hui sous le nom d'imagerie dépendant du niveau d'oxygène sanguin ou « *blood-oxygen-level dependent* » (BOLD) en anglais, une nouvelle technique appelée **imagerie par résonance magnétique fonctionnelle (IRMf)** a été dérivée de l'IRM. L'IRMf permet de mesurer l'activité cérébrale de manière indirecte, c'est-à-dire en partant des marqueurs indirects que sont les réponses hémodynamiques (i.e. le débit sanguin cérébral). Les techniques actuelles d'IRMf permettent de localiser l'activité cérébrale, de visualiser graphiquement les schémas d'activation cérébrale et de déterminer leur intensité par le biais d'un code couleur. Ces techniques sont désormais utilisées à des fins variées, par exemple pour évaluer les risques préopératoires, pour détecter des anomalies fonctionnelles en cartographiant les zones du cerveau (détecter, par exemple, une asymétrie hémisphérique gauche-droite dans les régions du langage et de la mémoire), ou encore pour observer la récupération post-AVC ou postopératoire des patients. L'IRMf a également été utilisée pour surveiller les effets des thérapies médicamenteuses et comportementales sur les patients, ou même pour aider au diagnostic d'affections neurologiques comme la dépression ou la maladie d'Alzheimer (de Vos *et al.*, 2018 ; Li, Xu, & Lu, 2018).

Depuis la naissance de l'IRMf, une révolution neurotechnologique a eu lieu. Au cours des trente dernières années, le nombre, la variété et le degré de précision des outils technologiques capables de visualiser, d'enregistrer ou même de modifier l'activité cérébrale n'ont cessé d'augmenter. Au cours des années 1990, souvent qualifiées de « décennie du cerveau », l'utilisation des techniques d'imagerie cérébrale a connu un essor spectaculaire. Outre l'EEG, l'IRM et l'IRMf, un certain nombre d'autres techniques se sont rajoutées au fil du temps, notamment la tomographie par émission de positons (TEP), la magnétoencéphalographie (MEG) et la spectroscopie proche infrarouge fonctionnelle (fNIRS). La TEP, qui est aujourd'hui l'une des principales techniques de la médecine nucléaire, est utilisée pour observer les processus métaboliques dans l'organisme, y compris dans le cerveau. Elle détecte les paires de rayons gamma indirectement émis par un radioligand (une substance biochimique radioactive) capable d'émettre des positons (antiparticules d'électrons), et qui est introduit dans l'organisme par le biais d'une molécule biologiquement active appelée « traceur radioactif ». L'application de cette technique à l'étude du cerveau repose sur l'hypothèse que les zones de radioactivité accrue sont associées à une activité cérébrale accrue. Comme dans le cas de l'IRMf, la TEP mesure de manière indirecte le flux sanguin des différentes parties du cerveau, flux qui est généralement considéré comme lié à l'activité neuronale. La TEP est aujourd'hui utilisée à différentes fins de diagnostic, notamment pour rechercher des tumeurs cérébrales, pour diagnostiquer la maladie d'Alzheimer, ou encore pour la préparation pré-chirurgicale de l'épilepsie.

La MEG, en revanche, permet de cartographier l'activité cérébrale en enregistrant, à l'aide de magnétomètres très sensibles, les champs magnétiques produits par les courants

électriques naturellement présents dans le cerveau. Mesurée pour la première fois par le physicien David Cohen en 1968, le processus de MEG était à l'origine réalisé dans une pièce blindée magnétiquement afin de réduire le bruit de fond magnétique. Les capteurs MEG d'aujourd'hui (environ trois cents) sont placés à l'intérieur d'un casque, qui couvre la majeure partie de la tête de l'individu. De cette manière, on peut recueillir les mesures MEG plus rapidement et plus efficacement. Plus récemment, des prototypes de casques MEG portatifs ont été développés : il s'agit de systèmes MEG qui peuvent être portés comme un casque et qui permettent un mouvement libre et naturel lors du balayage, le patient n'étant alors plus obligé de rester immobile dans un scanner (Boto et al., 2018).

La fNIRS a été découverte par Frans Jobsis en 1977. Il avait observé que les tissus biologiques étaient relativement transparents à la lumière dans la région du proche infrarouge, c'est-à-dire entre 700 et 1300 nanomètres, de sorte que la transmission d'une quantité suffisante de photons à travers les organes du corps permettait de les surveiller *in situ* (Jobsis, 1977). Dans cette région du proche infrarouge, l'hémoglobine – dans ses deux principales variantes, l'oxyhémoglobine (O₂Hb) et la désoxyhémoglobine (HHb) – présente un mode d'absorption dépendant de l'oxygène. Elle joue donc le rôle de chromophore (substance responsable de l'absorption des rayonnements électromagnétiques dans le champ visible) au sein des tissus biologiques qui absorbent la lumière dans cette région du proche infrarouge. Comme elle ne nécessite pas de matériel particulièrement encombrant ni lourd, la fNIRS est désormais utilisée dans une grande variété d'appareils portatifs et d'applications, et permet, par exemple, de suivre l'activité cérébrale générée par le comportement d'un individu au cours de sa vie quotidienne.

Les neurotechnologies telles que l'EEG, la fNIRS (et peut-être, dans un avenir proche, la MEG et l'IRMf portatives) transforment la technologie de l'imagerie cérébrale en la rendant utilisable à grande échelle. Cela se traduit par le développement de capteurs miniaturisés, fonctionnant sur batterie et sans fil, qui peuvent mesurer l'activité cérébrale en accompagnant l'individu dans ses déplacements (Ayaz, Izzetoglu, Izzetoglu, & Onaral, 2019).

Enfin, au cours des dix dernières années, **l'optogénétique** a été de plus en plus utilisée en neurosciences « pour étudier la manière dont des types de cellules spécifiques contribuent aux fonctions cérébrales et aux états de troubles cérébraux » (Boyden, 2015). Cette technique consiste à utiliser la lumière pour contrôler des neurones génétiquement modifiés en amont afin que les canaux ioniques qu'ils contiennent deviennent particulièrement sensibles à la lumière. Bien que les outils d'optogénétique ne soient pas encore tout à fait au point et n'aient actuellement qu'une application limitée dans la pratique clinique, on peut de plus en plus s'en servir comme outils de connectivité neuronale et de phénotypage cellulaire, aussi bien que comme outils d'enregistrement et d'imagerie cérébrale. Il a ainsi été avancé que « les résultats émergeant de l'utilisation de l'optogénétique dans les neurosciences fondamentales, et de la neurotechnologie dans son ensemble, fourniront dans les années à venir des informations variées sur de nouvelles molécules à cibler pour développer des médicaments, de nouveaux circuits visant à stimuler électriquement le cerveau, de nouveaux protocoles de médecine régénérative, et d'autres stratégies pour aider à réparer le cerveau » (Boyden, 2015).

1.1.2. L'électroencéphalographie intracrânienne

Contrairement à toutes les techniques précédentes, l'électroencéphalographie intracrânienne ne mesure pas l'activité cérébrale depuis l'extérieur, mais depuis l'intérieur du

crâne. La forme la plus courante d'électroencéphalographie intracrânienne est l'électrocorticographie (ECoG). Bien que cette technique utilise des électrodes pour enregistrer l'activité électrique (elle est en cela similaire à l'électroencéphalographie classique), ces électrodes ne sont pas placées sur le cuir chevelu, mais directement sur la surface exposée du cerveau, c'est-à-dire sur le cortex lui-même. L'ECoG, qui nécessite une intervention neurochirurgicale, n'est pas faite pour étudier le cerveau de sujets sains – car le risque de dommages collatéraux serait trop élevé comparé aux bénéfices potentiels –, mais elle est largement utilisée sur les modèles animaux et sur les humains dans un cadre neurochirurgical. Dans une procédure d'exploration pré-chirurgicale, elle sert par exemple à localiser et à déterminer les limites des foyers pathologiques, notamment dans les zones les plus sujettes aux phénomènes épileptiques. En termes de qualité de mesure, le fait de pouvoir placer des électrodes directement sur le cortex rend l'ECoG beaucoup plus précise, informative et détaillée que l'EEG, et ouvre une fenêtre de tir beaucoup plus large pour accéder aux processus cognitifs humains.

En 2017, une entreprise privée du nom de Neuralink a commencé à travailler sur un système d'électroencéphalographie intracrânienne capable de lire les informations du cerveau. Le système se fonde sur des sondes, composées principalement de polyimide et disposant d'un conducteur mince en or ou en platine. Les sondes doivent être insérées dans le cerveau à l'issue d'un processus automatisé réalisé par un robot chirurgical. Chaque sonde se compose d'une zone de fils contenant des électrodes capables de localiser les signaux électriques dans le cerveau, ainsi que d'une zone sensorielle où le fil interagit avec un système électronique pour amplifier le signal cérébral avant qu'il ne soit capté. Chaque système comprend jusqu'à 3072 électrodes par formation. Cette technologie a jusqu'à présent été testée sur des modèles animaux (rongeurs, suidés et singes), mais l'entreprise a prévu de commencer prochainement des expériences sur l'homme.

1.1.3. Interfaces cerveau-machine pour aider le contrôle moteur et la communication

À l'origine, les techniques d'imagerie cérébrale ont été développées, et sont d'ailleurs encore largement mises en œuvre, dans le domaine du diagnostic. Le but est de détecter des anomalies dans le cerveau des patients, comme des lésions, des tumeurs ou des indicateurs d'épisodes épileptiques. Cependant, depuis la fin des années 1960, une branche de la neurotechnologie s'est développée dans le but non seulement de « lire » l'activité cérébrale, mais aussi de l'exploiter pour créer un canal de communication direct entre l'utilisateur et des ordinateurs numériques. Ce domaine de recherche est connu sous le nom d'interface cerveau-machine (ICM), et doit en partie sa naissance aux investissements massifs du ministère américain de la Défense, ainsi qu'à des recherches approfondies menées sur des modèles animaux. En 1969, un neurone du cerveau d'un singe a été connecté avec succès à un dispositif extérieur placé devant le visage de ce singe. Le dispositif était capable de se déplacer en fonction de l'activité du neurone : un mouvement se déclenchait lorsque le singe pensait activement à quelque chose. Après plusieurs séances d'entraînement, l'animal avait appris à activer intentionnellement un de ses neurones afin de déclencher l'activité du dispositif extérieur. Cependant, on date généralement l'invention officielle des interfaces cerveau-machine de 1973, soit quatre ans plus tard, lorsque Jacques Vidal publia un article sur la communication cerveau-machine intitulé « Toward Direct Brain-Computer Communication »

(Vidal & JJ, 1973). Dans cet article, le terme anglais « *brain-computer interface* », ou BCI, apparaît pour la première fois. L'article présente également un prototype de canal direct de communication entre le cerveau et un ordinateur. Les premières démonstrations des BCI de Vidal s'intéressaient au contrôle des mouvements dans une simulation de dédale. Quinze ans plus tard, des chercheurs de l'université de Skopje ont réussi pour la première fois à utiliser une BCI pour contrôler un robot grâce à l'activité cérébrale.

Dans les années 1970 et 1980, différentes études ont réussi à établir des canaux de communication de plus en plus fiables entre l'activité cérébrale d'animaux non humains et les systèmes informatiques. Vers la fin des années 1990, les premières interfaces invasives pour les humains souffrant de handicaps moteurs ont été développées. Des chercheurs de l'université Emory d'Atlanta, dirigés par Philip Kennedy et Roy Bakay, ont été les premiers à installer chez un être humain un implant cérébral capable de produire des signaux de qualité suffisante pour simuler les mouvements souhaités par le patient. Le patient, un Américain de 50 ans, souffrait de ce que l'on appelle le syndrome d'enfermement. L'implant a été installé dans le cerveau du patient en 1998, et lui a permis d'apprendre à contrôler un curseur d'ordinateur grâce à sa propre activité cérébrale.

Syndrome d'enfermement : État neurologique dans lequel le patient est alerte et conscient, mais incapable de bouger ou de communiquer verbalement en raison d'une paralysie complète de la quasi-totalité des muscles volontaires du corps, à l'exception des mouvements verticaux des yeux et de leur clignement. On a généralement confirmation que le patient est conscient grâce à sa capacité à comprendre le langage verbal et à répondre à ses interlocuteurs par des mouvements oculaires, qui sont en fait sa seule fenêtre sur le monde. De plus, l'EEG

Figure 2- Le syndrome de l'enfermement

Au cours des deux décennies suivantes, plusieurs groupes de recherche ont accompli de grands progrès dans la transformation des données générées par les électrodes du cerveau en actions effectuées par un ordinateur. Les BCI d'aujourd'hui permettent aux patients tétraplégiques d'utiliser des tablettes, d'écrire des courriels, de se servir d'un chat, de diffuser de la musique ou de regarder des vidéos sur YouTube – et ce uniquement grâce à leur activité cérébrale. Récemment, un patient tétraplégique a pu jouer l'Hymne à la joie de Beethoven en commandant directement par la pensée une application de simulation de piano (Nuyujukian et al., 2018). Outre les curseurs et les tablettes, on trouve également, parmi les appareils numériques pouvant désormais être directement contrôlés par le cerveau, les bras robotisés (par exemple, pour les patients qui ont été amputés d'un membre), des fauteuils roulants électroniques, voire des robots humanoïdes.

Les BCI modernes peuvent être divisées en deux types : **invasives** et **non invasives**. Les BCI invasives enregistrent les signaux du cerveau grâce à une implantation chirurgicale de réseaux d'électrodes dans le cerveau, ou en étant directement connectées au système nerveux central. Les BCI non invasives créent une interface entre les signaux cérébraux et des technologies d'imagerie cérébrale telles que l'électroencéphalographie (EEG), qui enregistre l'activité cérébrale grâce à des électrodes placées à l'extérieur du crâne. Les BCI invasives et non invasives ont néanmoins pour point commun qu'elles établissent une interaction directe entre le cerveau de l'utilisateur et un dispositif informatique. On présente généralement cette interaction sous la forme d'un cycle en quatre étapes.

La première phase est celle de **l'entrée** de données, c'est-à-dire la génération, par un individu, d'une activité cérébrale spécifique en réponse à un stimulus. Cette activité cérébrale est générée lorsque l'individu est placé dans un certain état cognitif ou effectue une certaine tâche mentale. Par exemple, lorsqu'un utilisateur de BCI contrôle un fauteuil roulant électronique, une matrice contenant tous les choix d'itinéraires possibles est présentée sur l'interface que l'utilisateur regarde. Un modèle d'activation cérébrale fréquemment utilisé dans les BCI sont les potentiels évoqués (PE), c'est-à-dire les réponses cérébrales qui sont le résultat direct d'un événement sensoriel, cognitif ou moteur spécifique. Parmi ces PE, un intérêt croissant s'est formé autour de l'onde P300, généralement déclenchée lors de la prise de décision (Fazel-Rezai et *al.*, 2012). Dans notre exemple, lorsque l'itinéraire souhaité arrive sur l'interface, les signaux cérébraux de l'utilisateur contiendront un signal P300 qui pourra être détecté par la BCI.

La deuxième phase est celle de la **mesure et de l'enregistrement de l'activité cérébrale**. Au cours de cette phase, les schémas d'activité cérébrale de l'utilisateur générés lors d'un processus cognitif ou lors de l'exécution d'une tâche mentale sont détectés et mesurés par l'interface. Par exemple, lorsqu'un utilisateur de BCI choisit une certaine option d'itinéraire (un point d'arrivée précis ou une instruction de tourner à gauche), la BCI peut détecter en temps réel l'onde P300 générée par le cerveau de cet utilisateur. La mesure enregistrée peut être mise en œuvre de différentes manières selon le type de BCI utilisé. Le type le plus courant de BCI non invasive s'appuie sur l'électroencéphalographie (EEG) ; il existe également d'autres systèmes de mesure, comme la spectroscopie fonctionnelle dans le proche infrarouge (fNIRS) ou des solutions invasives comme l'électrocorticographie (ECoG).

Afin que la BCI puisse les utiliser et générer à partir d'elles les sorties attendues (c'est-à-dire les résultats souhaités par l'utilisateur), les données neuronales brutes mesurées au cours de la deuxième phase doivent d'abord être décodées, puis classées. Ce processus de **décodage et de classification** a généralement lieu lors de la troisième phase du cycle de la BCI. Au cours de cette phase, les données sont traitées pour « nettoyer » les signaux cérébraux, c'est-à-dire pour augmenter le rapport signal/bruit (la mesure de la force du signal souhaité par rapport au bruit de fond) et pour filtrer les aspects les plus pertinents de chaque signal. Ce traitement est nécessaire, car il permet de distinguer les caractéristiques pertinentes du signal des caractéristiques non pertinentes, notamment du bruit de fond dû à l'activité cérébrale sous-jacente et qui n'est pas dirigée vers l'exécution de cette tâche mentale spécifique. Dans ce processus de décodage et de classification, l'intelligence artificielle joue un rôle fondamental. Les algorithmes d'apprentissage machine sont en effet entraînés à décoder correctement les signaux neuronaux enregistrés et amplifiés, ainsi qu'à classer correctement les données qui en résultent.

Une fois décodés, les signaux peuvent être traduits en **sortie**. La sortie est l'exécution, grâce au contrôle permis par la BCI, de l'action initialement prévue, souhaitée ou jugée bénéfique par l'utilisateur (dans notre exemple, tourner à gauche en fauteuil roulant). Les dispositifs pouvant être contrôlés de la sorte sont par exemple des dispositifs motorisés (des fauteuils roulants électroniques et des membres robotisés), des dispositifs de détection, ou différentes applications logicielles et matérielles (y compris des applications pour smartphone). À la fin de chaque cycle, l'utilisateur constate la sortie d'information issue du cycle qui vient

de s'achever (par exemple, percevoir que le fauteuil roulant tourne à gauche), et le cycle suivant peut alors commencer.

1.1.4. Neurostimulation

La neurotechnologie permet non seulement de lire l'activité cérébrale, mais aussi de *réécrire* les signaux neuronaux. En effet, non seulement le cerveau émet des signaux électriques qui peuvent être enregistrés directement ou indirectement par des machines, mais il est également capable de recevoir des signaux électriques provenant de l'intérieur ou de l'extérieur de lui-même. Ce type de neurotechnologie capable de moduler l'activité cérébrale est généralement appelé « **neurostimulation** » ou « neuromodulation » selon son mode de fonctionnement, car il stimule électromagnétiquement certaines régions du cerveau. La neurostimulation procède en appliquant des électrodes sur le cerveau, sur la moelle épinière ou sur un nerf périphérique. Elle peut fonctionner de deux manières différentes : soit en stimulant activement les nerfs pour produire une réponse biologique naturelle, soit en appliquant directement sur la zone cérébrale d'action des doses infimes d'agents pharmaceutiques ciblés.

La neurostimulation et la neuromodulation peuvent être invasives ou non invasives. Parmi les méthodes invasives, la plus courante – et la plus prometteuse – est la **stimulation cérébrale profonde** (SCP). La SCP est la principale forme de traitement chirurgical utilisée pour traiter les symptômes de troubles neurologiques affectant la motricité, comme la maladie de Parkinson et la dystonie. Il s'agit donc d'une pratique invasive qui nécessite l'implantation d'électrodes à l'intérieur du cerveau. Ces électrodes sont reliées à des fils qui en constituent l'extension et qui sont acheminés sous la peau, derrière l'oreille du patient et le long de son cou. Ces extensions sont à leur tour reliées à un boîtier générateur d'impulsions (un dispositif similaire à un stimulateur cardiaque), généralement placé sous la peau de la poitrine ou au-dessus de l'estomac. Une télécommande permet au patient d'allumer et d'éteindre le générateur d'impulsions. Lorsque l'appareil est allumé, les électrodes stimulent à haute fréquence la zone cérébrale concernée. Cette stimulation modifie certains des signaux électriques du cerveau à l'origine des symptômes de la maladie de Parkinson, notamment des tremblements. Au cours des deux dernières décennies, la recherche en neurotechnologie a montré que la SCP pouvait réduire les symptômes et améliorer la qualité de vie des personnes atteintes de la maladie de Parkinson à un stade avancé, et dont les symptômes ne peuvent plus être atténués par des médicaments. Bien que la SCP soit généralement implantée lors de stades avancés de la maladie de Parkinson, des études récentes ont montré que l'utilisation précoce de la stimulation cérébrale profonde pouvait également avoir des effets bénéfiques. Dans certains cas, la SCP est capable d'éliminer le tremblement et permet au patient de retrouver les fonctions motrices nécessaires pour réaliser ses activités quotidiennes comme se laver, manger, boire, lire, etc. Les systèmes SCP de dernière génération s'appuient sur des ordinateurs intelligents interconnectés avec le cerveau, qui ajustent automatiquement le niveau de stimulation en fonction des symptômes de chaque personne. C'est ce qu'on appelle la **SCP en boucle fermée**. Avec la SCP traditionnelle en boucle ouverte, un neurologue doit ajuster manuellement les paramètres de stimulation tous les 3 à 12 mois après l'implantation de la SCP. En revanche, dans les systèmes de SCP en boucle fermée de dernière génération, la programmation des paramètres de stimulation est effectuée automatiquement par l'appareil, qui sait quel biomarqueur est mesuré.

Comme le courant électrique peut traverser le crâne, la neurostimulation et la neuromodulation peuvent également être mises en œuvre de manière non invasive. Cette propriété est largement exploitée par un large éventail de technologies que l'on désigne sous le nom de stimulation électrique transcrânienne. C'est le cas de la stimulation transcrânienne à courant continu (tDCS), et de la stimulation transcrânienne à courant alternatif (tACS). À l'origine, la tDCS et la tACS ont toutes deux été développées pour traiter des patients souffrant de lésions cérébrales ou de troubles psychiatriques comme la dépression. En raison de leur nature non invasive, ces deux technologies ont été testées en parallèle pour savoir si elles amélioreraient certaines fonctions cognitives chez des individus en bonne santé. Au cours des dernières années, certaines études ont mis en évidence que la stimulation électrique transcrânienne générait bel et bien des améliorations cognitives, modérées, mais néanmoins significatives. L'une des plus importantes de ces études a été menée par les neuroscientifiques Robert Reinhart et John Nguyen de l'université de Boston, qui ont utilisé une combinaison d'EEG et de la tACS pour surveiller, puis stimuler de manière non invasive le cerveau de deux groupes de participants. Le premier groupe était composé de personnes âgées de 20 à 29 ans. Le second groupe comprenait des personnes beaucoup plus âgées, de 60 à 76 ans. En comparant grâce à l'EEG les performances de ces deux groupes sur des tâches de travail et des tâches faisant spécifiquement appel à la mémoire, les chercheurs ont constaté que les individus plus âgés étaient plus lents et moins précis pour se souvenir des objets qu'ils avaient vus, ou pour identifier les différences cachées entre deux images presque identiques. Le groupe plus âgé présentait également une synchronisation moindre de l'activité cérébrale qui a lieu entre deux parties du cerveau fortement impliquées dans les processus de mémorisation, à savoir le cortex frontal et le cortex temporal. Après que leur cortex frontal et le cortex temporal ont été stimulés pendant vingt-cinq minutes à l'aide de la tACS, les sujets âgés ont vu s'améliorer non seulement la synchronisation de leur activité cérébrale, mais aussi leurs performances aux tests de mémoire, au point que les résultats qu'ils obtenaient étaient comparables à ceux du groupe des vingtenaires. Cet effet a persisté plus de cinquante minutes après la stimulation, avant que l'activité cérébrale et les performances du groupe plus âgé ne reviennent à des niveaux normaux. En utilisant ce type de stimulation, les chercheurs ont non seulement démontré qu'ils pouvaient reconnecter ou resynchroniser les circuits de mémoire, mais aussi qu'ils pouvaient faire temporairement revenir certains aspects de l'horloge biologique de la mémoire humaine au niveau qu'ils avaient quarante ans avant (Reinhart & Nguyen, 2019). Des résultats de ce type démontrent la possibilité, au moins en principe, d'utiliser les neurotechnologies non seulement pour décoder l'activité cérébrale et contrôler des objets par la pensée, mais également pour améliorer les capacités cognitives humaines au-delà de leur seuil de normalité biologique (seuil qui est à comprendre statistiquement parlant) : un phénomène généralement désigné sous l'expression d'« **amélioration cognitive** » (Bostrom & Sandberg, 2008).

1.1.5. Neurotechnologie omniprésente

Les neurotechnologies ont été originellement développées, et sont encore principalement mises en œuvre, dans le contexte de la médecine clinique et de la recherche en neurosciences. Toutefois, ces dernières années, un certain nombre d'applications neurotechnologiques ont fait leur entrée sur le marché, et sont désormais intégrées dans des appareils grand public destinés à des utilisateurs en bonne santé qui les utilisent à différentes

fins non cliniques. On regroupe généralement l'ensemble de ces applications non invasives, évolutives et possiblement omniprésentes de la neurotechnologie sous le terme générique de « **neurotechnologie omniprésente** » ou « *pervasive neurotechnology* » en anglais (Ienca, Haselager et Emanuel, 2018). Cette notion est empruntée à celle, plus largement utilisée, d'« informatique ubiquitaire ». Aujourd'hui, on trouve parmi les applications de la neurotechnologie omniprésente : les BCI pour contrôler des appareils ou le *neuromonitoring* en temps réel, les systèmes de conduite de véhicules s'appuyant sur des neurosenseurs, les outils d'entraînement cognitif, la stimulation électrique et magnétique du cerveau, les objets connectés portatifs utilisés pour le bien-être mental, et les systèmes de réalité virtuelle.

La plupart de ces applications utilisent des enregistrements EEG ou d'autres techniques non invasives pour surveiller l'activité électrique du cerveau à différentes fins, comme le *neuromonitoring* (évaluation en temps réel du fonctionnement cérébral), l'entraînement neurocognitif (utilisation de certaines bandes de fréquences pour améliorer le fonctionnement neurocognitif) et le contrôle à distance des objets. Ainsi, les BCI reposant sur l'EEG sont de plus en plus utilisées comme accessoires portatifs pour toute une série d'activités quotidiennes, notamment les jeux, le divertissement et le contrôle à distance des smartphones. Par exemple, les sociétés Emotiv, Neurosky et Muse proposent une large gamme de casques sans fil, utilisables au quotidien, à connecter aux smartphones et ordinateurs personnels compatibles. Le contrôle cérébral peut donc être utilisé pour contrôler à distance différents types d'appareils et entreprendre différentes activités, notamment les jeux ou d'autres formes de divertissement, le marketing, l'autocontrôle et la communication. Ainsi, grâce à la neurotechnologie omniprésente, utiliser des interfaces cerveau-machine pour contrôler des appareils ne se limite plus au domaine clinique. Récemment, par exemple, le constructeur automobile Nissan a présenté une sorte d'interface cerveau-voiture : cette interface détecte les ondes cérébrales du conducteur afin de surveiller sa capacité d'attention, et donc d'anticiper et de prévenir les éventuelles baisses d'attention dues à la fatigue ou au sommeil. L'objectif du système est de permettre au véhicule de réagir quelques fractions de seconde plus vite que ce que le temps de réaction naturel du conducteur aurait permis. Selon le constructeur automobile, cette technologie peut améliorer non seulement la sécurité, mais aussi la qualité de la conduite, et ce même chez les conducteurs expérimentés. Un autre dispositif appelé UDrone, fabriqué par la société chinoise EEGSmart, permet de contrôler un drone grâce à l'activité cérébrale. L'UDrone est un petit appareil léger, il est équipé d'une caméra de huit mégapixels et peut rester en l'air pendant sept minutes environ. L'activité cérébrale de l'utilisateur est lue par EEG et traduite en un signal sans fil grâce auquel il est possible de contrôler l'appareil.

La possibilité d'un contrôle cérébral non invasif a également attiré l'attention du secteur de la communication mobile. Plusieurs grandes entreprises, dont Apple et Samsung, intègrent désormais des gadgets neuronaux dans les accessoires qui accompagnent leurs principaux produits. Par exemple, des accessoires pour iPhone tels que le casque XWave permettent déjà aux utilisateurs de se connecter directement aux iPhones compatibles grâce à une lecture de leurs propres ondes cérébrales. Parallèlement, des prototypes de la prochaine génération de Samsung Galaxy Tabs et d'autres appareils mobiles ou portatifs ont été testés : leur contrôle serait assuré par l'activité cérébrale via une BCI reposant sur l'EEG (Powell, Munetomo, Schlueter et Mizukoshi, 2013). À la lumière de ces tendances, certains experts ont prédit que les dispositifs neuronaux seront progressivement préférés par les humains, quand ils voudront

interagir avec les ordinateurs, au clavier, à l'écran tactile, à la souris et aux dispositifs de commande vocale.

Les dispositifs d'imagerie cérébrale et les BCI ne sont pas les seuls à entrer dans la catégorie des neurotechnologies omniprésentes : différents stimulateurs cérébraux électriques appartiennent également à cette catégorie. Contrairement aux dispositifs d'imagerie cérébrale, les stimulateurs neuronaux ne sont pas utilisés dans le but principal d'enregistrer ou de décoder l'activité cérébrale, mais plutôt pour stimuler ou moduler électriquement l'activité cérébrale. Les appareils de stimulation transcrânienne à courant continu (tDCS) portatifs et faciles d'utilisation constituent la forme la plus répandue de stimulateur neuronal grand public. Ils sont utilisés dans le cadre de différentes formes de consommation, et visent à optimiser les performances cérébrales dans des tâches cognitives variées, en fonction de la région cérébrale stimulée. Récemment, la stimulation magnétique transcrânienne (TMS) – une technique magnétique qui stimule pendant un bref laps de temps de petites régions du cerveau à des fins diagnostiques et thérapeutiques – a également évolué vers des dispositifs portatifs, qui se sont avérés efficaces dans le traitement de la migraine (Lefaucheur et al., 2014).

Depuis 2018, le géant technologique Facebook consacre une équipe de 60 ingénieurs à la construction d'une interface cerveau-machine qui permettra aux utilisateurs de rédiger des messages sur le réseau social en utilisant uniquement leur esprit, s'abstrayant ainsi des commandes clavier et vocales. L'équipe prévoit de plus d'utiliser des techniques d'imagerie optique pour scanner le cerveau à une fréquence de cent fois par seconde afin de détecter le langage silencieux dans la tête d'un individu, et de le transposer ensuite en un texte écrit. Regina Dugan, responsable de la division Building 8 au sein du secteur Recherche & Développement de Facebook, a expliqué que le but était de permettre aux utilisateurs de taper cent mots par minute, soit cinq fois plus vite que manuellement sur un portable. Cette technologie, baptisée « interface cérébrale directe » par Facebook, combinera des algorithmes d'apprentissage machine chargés de détecter le langage à des systèmes d'imagerie cérébrale optique à résolution spatiale avancée, ainsi qu'à des dispositifs cérébraux de nouvelle génération. Parmi les autres entreprises actives dans le secteur de la neurotechnologie omniprésente, on peut également citer Neuralink et Kernel. Kernel a mis au point de nombreux logiciels et dispositifs de suivi de l'activité cérébrale, comme le logiciel « Sound ID », un programme qui, à partir des seules données cérébrales de l'utilisateur, peut indiquer quelles paroles ou quelle chanson il est en train d'écouter, ou bien les dispositifs « Flux » et « Flow ».

En résumé, si les neurotechnologies ont permis, au cours des dernières décennies, de déchiffrer certains mystères du cerveau humain, de mieux le comprendre d'un point de vue scientifique et de s'appuyer sur lui à des fins médicales, la décennie qui vient de s'ouvrir verra probablement les neurotechnologies devenir omniprésentes, participer à des dimensions variées de notre vie et devenir de plus en plus efficaces dans la modulation neuronale de notre psychologie et de notre comportement. S'il convient de se réjouir des progrès continus réalisés dans le développement des neurotechnologies, les implications éthiques et juridiques de la révolution neurotechnologique doivent être prises en compte de manière précoce et proactive. À mesure que les neurotechnologies progressent, il est ainsi crucial d'évaluer notre cadre actuel des droits de l'homme afin de savoir s'il est bien équipé, sur le plan conceptuel comme sur le plan normatif, pour relever les nouveaux défis posés par l'intersection du cerveau, de la

machine et de la société ; pour fournir des orientations aux chercheurs et aux acteurs de l'innovation ; et pour assurer la protection des individus et des différents groupes.

1.2. Quelles informations la neurotechnologie peut-elle lire ?

Bien qu'étant très différentes en termes de fonctionnalité, d'applicabilité et de mode d'utilisation, les différentes neurotechnologies décrites ci-dessus ont du moins une caractéristique commune : elles peuvent enregistrer des données quantitatives concernant la structure, l'activité et le fonctionnement du cerveau humain. Il s'agit notamment, d'une part de mesures directes de la structure, de l'activité et/ou de la fonction du cerveau (par exemple, le courant neuronal ou la somme des signaux bioélectriques de l'EEG), et d'autre part d'indicateurs fonctionnels indirects (par exemple, le flux sanguin dans l'IRMf et la fNIRS). On peut appeler « **données du cerveau humain** » ces données quantitatives sur la structure, l'activité et la fonction du cerveau humain. Les données du cerveau humain peuvent révéler des informations sur l'état de santé d'une personne (par exemple, sa santé neurologique ou psychiatrique) et il est possible, dans une certaine mesure, d'en tirer des déductions sur les processus mentaux.

Informations sur le cerveau et la santé mentale

Comme nous l'avons vu, les techniques d'imagerie cérébrale peuvent être utilisées pour obtenir des images de la morphologie (par exemple, avec l'IRM), de la fonction (par exemple, avec l'IRMf), du métabolisme (par exemple, avec la TEP) ou du contenu moléculaire (par exemple, avec la spectroscopie RM) du cerveau humain. Les données générées par ces techniques d'imagerie cérébrale peuvent être utilisées

Un biomarqueur est « une caractéristique objectivement mesurée et interprétée, selon le cas, comme un indicateur de processus biologiques normaux, de processus pathogènes ou bien de réponses pharmacologiques à une intervention thérapeutique » (Biomarkers

Figure 3- Définition du biomarqueur

pour identifier des biomarqueurs de maladies cérébrales, que l'on appelle **biomarqueurs d'imagerie cérébrale**. Par exemple, si l'on s'intéresse aux biomarqueurs d'imagerie cérébrale de la maladie d'Alzheimer (MA), on peut tenter de mesurer le dépôt de bêta-amyloïde avec la TEP de l'amyloïde, ou le niveau d'atrophie du cerveau et de l'hippocampe avec l'IRM (Bateman et al., 2012). Les biomarqueurs d'imagerie cérébrale peuvent également être utilisés pour identifier les signes précoces du déclin cognitif (Gordon et al., 2018). Par exemple, les biomarqueurs d'imagerie cérébrale de la sclérose en plaques (SEP) comprennent le comptage et le volume des lésions développées, des lésions en développement et des trous noirs. Quant aux biomarqueurs d'imagerie cérébrale de l'accident vasculaire cérébral ischémique, il peut s'agir, lors de la phase aiguë, du volume de la pénombre ischémique, que les techniques de résonance magnétique peuvent estimer. Plus récemment, on a utilisé les données d'imagerie cérébrale pour identifier les signes de troubles mentaux comme la dépression et la schizophrénie (Castanheira, Silva, Cheniaux, & Telles-Correia, 2019). Les biomarqueurs de troubles neurologiques, psychiatriques et mentaux sont cependant des données sensibles, car elles peuvent toucher au cœur même d'un être humain, à savoir à sa santé et sa vie mentales.

Informations sur les états mentaux : Au niveau neurobiologique, les données cérébrales sont les manifestations les plus directes des états mentaux. Les neurotechnologies

actuelles, notamment les techniques non invasives, ne sont pas encore capables de *décoder les pensées*. Cela signifie que ces technologies sont incapables de fournir un aperçu complet, détaillé et en temps réel des schémas neuronaux sous-tendant les processus cognitifs. Cependant, elles permettent déjà, grâce à une méthode connue sous le nom d'*inférence inverse* (Poldrack, 2011), de déduire le lancement de certains processus mentaux (par exemple, des processus perceptifs et cognitifs) de schémas spécifiques d'activation du cerveau. On utilise pour cela des méthodes invasives comme des méthodes non invasives d'enregistrement (et de manipulation) des circuits neuronaux. Chez les animaux de laboratoire, des études ont montré qu'il était possible de décoder la perception visuelle et de la manipuler avec une grande précision (Carrillo-Reid, Han, Yang, Akrouh, & Yuste, 2019 ; Marshel et al., 2019). Dans les études menées sur des sujets humains, les chercheurs ont utilisé des scans IRMf et des signaux d'électrocorticographie à haute densité pour décoder avec précision l'imagerie mentale et le langage silencieux (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013 ; Kay, Naselaris, Prenger, & Gallant, 2008). Des travaux récents sur les enregistrements EEG intracrâniens de l'activité cérébrale liée à la parole ont par ailleurs atteint une précision remarquable dans l'identification des schémas d'activité cérébrale liés au langage intérieur (Moses, Leonard, Makin, & Chang, 2019). Les techniques dites ML ont quant à elles contribué à améliorer l'analyse des processus cognitifs, également enregistrés à partir de mesures EEG (Hubbard, Kikumoto, & Mayr, 2019 ; Omurtag, Aghajani, & Keles, 2017). De plus, l'IRMf est capable de montrer des zones soumises à une activité accrue liée aux processus de mémoire. Une technique d'analyse relativement récente basée sur des algorithmes intelligents, appelée analyse de motifs multi-voxels (MVPA), a montré quant à elle qu'elle pouvait détecter des motifs spécifiques d'activité cérébrale apparaissant lorsque les individus sont en train de se souvenir d'expériences particulières, et donc décoder des informations liées à la mémoire. La MVPA est une méthode statistique qui traite les données d'IRMf par le biais d'un algorithme informatique programmé pour apprendre automatiquement les schémas neuronaux associés à des pensées ou à des expériences spécifiques. L'une des premières applications de la MVPA à la mémoire humaine a été relaté par une publication parue dans la revue *Science* en 2005 (Polyn, Natu, Cohen et Norman, 2005). Dans cette étude, des chercheurs de l'Université de Pennsylvanie et de Princeton ont demandé aux sujets de leur expérience d'imaginer une série de personnes célèbres, de lieux et d'objets courants ; une IRMf était conduite en même temps. À partir des données recueillies, les chercheurs ont entraîné un algorithme intelligent à identifier les schémas d'activité associés à chacune des trois catégories (personnes célèbres, lieux ou objets). Par la suite, on a invité les sujets à se souvenir des images qu'ils avaient précédemment visualisées. L'algorithme s'est alors montré capable de reconstruire les traces neuronales spécifiques à chaque catégorie d'images, et a prédit ces dernières de manière précise quelques secondes avant chaque réponse. Par exemple, avant que le sujet ne nomme une personne célèbre, l'algorithme identifiait le schéma d'activité « personne célèbre » dans le flux de données IRMf, un schéma qui correspondait à l'activation d'une zone du cortex utilisée dans le traitement des visages. De même, avant que le sujet ne désigne une chaise, l'algorithme a pu détecter le schéma d'activité « objet » dans les données neuronales, et ainsi de suite.

Par ailleurs, au cours des quinze dernières années, les techniques algorithmiques visant à décoder les souvenirs se sont affinées, et sont désormais capables, dans une certaine mesure, de deviner les souvenirs d'une personne à partir de l'analyse des données de son cerveau. Dans

une étude, des chercheurs ont pu identifier, reconstruire et différencier les souvenirs visuels de personnes qui avaient regardé un épisode précis d'une série télévisée. Sur la base des seules données neuronales, les chercheurs ont ainsi pu révéler des détails mnémoniques très précis, comme le fait de savoir si la scène a été filmée en extérieur ou en intérieur, ou celui de savoir si le protagoniste était présent ou non. Bien que les sujets se soient souvenus de la même scène en la décrivant avec des mots différents, les traces neuronales dans leur cerveau, décodées par l'algorithme, correspondaient à des schémas d'activation très similaires (Chen et *al.*, 2017).

Dans le décodage de la mémoire, la mémoire visuelle joue assurément un rôle privilégié. En 2017, des chercheurs américains ont réussi, à partir de la seule activité cérébrale, à décoder le contenu visuel d'images perçues par les sujets de leur expérience. En utilisant des algorithmes d'apprentissage profond, une approche de l'IA qui tente de calquer le fonctionnement du cerveau, les chercheurs ont construit un modèle de la façon dont le cerveau humain encode les informations. Ils ont invité les sujets à regarder des centaines de courtes vidéos pendant que, dans le même temps, une machine IRMf mesurait l'activité neuronale de leur cortex visuel et d'une autre zone de leur cerveau. En parallèle, un réseau de neurones artificiel chargé de traiter des images a été entraîné à associer les images vidéo à l'activité cérébrale. Lorsqu'on a demandé aux sujets de regarder de nouvelles vidéos, l'algorithme a pu prédire avec précision les zones du cerveau qui seraient activées, et même visualiser quelles caractéristiques des informations visuelles étaient traitées par chaque zone du cortex. Sur le même principe que l'expérience précédente, un autre réseau de neurones était quant à lui capable de deviner avec un taux de précision de 50 % quelle catégorie d'images le sujet était en train de regarder parmi les 15 catégories possibles. Les chercheurs ont aussi montré que, grâce à l'IA, il était non seulement possible de déchiffrer les souvenirs, mais aussi de dévoiler les images mentales du cerveau d'un individu. Pour ce faire, ils ont entraîné un réseau de neurones à reconstruire partiellement, en convertissant les signaux cérébraux en pixels, le contenu visuel de ce qu'un participant de l'expérience voyait.

Outre les phénomènes de mémoire, on peut également utiliser les techniques décrites ci-dessus pour déduire d'autres types d'informations mentales. Un exemple notable est celui des **intentions cachées**. Grâce à la neurotechnologie et à l'IA, différents groupes de scientifiques ont réussi, ces dernières années, à identifier les processus conscients, mais aussi ce que l'on a appelé les intentions cachées d'une personne. Ainsi, on serait capable de prédire ses actions futures. Des études de ce type ont montré à plusieurs reprises que l'activité neuronale inconsciente est première chronologiquement, et peut influencer les décisions libres des êtres humains. Par exemple, des neuroscientifiques dirigés par le psychologue britannique John-Dylan Hynes ont affirmé être capables de décoder, à partir de l'activité cérébrale de certains participants à leur étude, des informations exactes sur les actions que ces individus avaient l'intention d'accomplir. La tâche en question consistait à choisir d'additionner ou de soustraire deux nombres, et de garder cette intention secrète pendant quelques secondes. Pendant ce court intervalle, les scientifiques ont pu déterminer avec une précision de 70 % laquelle des deux actions (addition ou soustraction) les sujets avaient l'intention d'effectuer en secret (Bles & Haynes, 2008). Des résultats similaires ont été obtenus avec succès quand des choix moteurs et des choix de raisonnement étaient en jeu : ils ont bel et bien pu être prédits à partir de signaux cérébraux. Par exemple, des chercheurs du Bernstein Center for Computational Neuroscience de Berlin ont montré que le résultat d'une décision libre

d'additionner ou de soustraire des nombres pouvait être décodé en analysant l'activité neuronale des cortex préfrontal et pariétal médians quatre secondes avant que le participant ne fasse son choix en son for intérieur. Ces signaux qui précèdent le choix conscient suggèrent que la préparation inconsciente des choix ne se limite pas à la préparation motrice. Au contraire, des décisions impliquant des échelles d'abstraction très différentes semblent pouvoir être anticipées à partir d'une activité cérébrale antérieure (Soon, He, Bode, & Haynes, 2013). Des études de ce type, toutes inspirées de recherches menées au début des années 1980 par le physiologiste Benjamin Libet, ouvrent des scénarios très inquiétants concernant le concept de **libre arbitre** et tous les autres concepts éthico-juridiques qui en découlent.

Grâce aux techniques de ML, les scanners cérébraux peuvent aujourd'hui être utilisés pour observer ou prédire les intentions et les souvenirs qui sont en jeu quand on sollicite un choix binaire (addition/soustraction) dans le cadre de protocoles expérimentaux bien définis. Mais ils peuvent également être utilisés pour décoder des **préférences** plus générales. Une étude américaine a montré que les scans IRMf permettent de déduire avec succès les opinions politiques des utilisateurs en identifiant les différences fonctionnelles dans le cerveau des électeurs démocrates et celui des électeurs républicains (Schreiber et al., 2013). De même, les préférences en matière de mode de vie ont été associées à des différences fonctionnelles présentes dans le cerveau des hommes et des femmes. D'autres études entraînent même des algorithmes d'apprentissage machine afin de prédire, à partir de données neuronales, des préférences très privées telles que l'orientation sexuelle (Safron et al., 2018), le désir de consommer des substances illicites comme la cocaïne, ou encore le plaisir du jeu (Kober et al., 2016).

La possibilité d'identifier de manière non invasive de telles particularités mentales à partir de différences cérébrales fonctionnelles a suscité un intérêt particulier en dehors du domaine biomédical, notamment à des fins de marketing. Déjà au début des années 2000, McClure et ses collègues avaient utilisé l'IRMf pour mettre en évidence des différences fonctionnelles (activation accrue dans le cortex préfrontal dorsolatéral et l'hippocampe) présentes dans le cerveau des personnes buvant consciemment du Coca Cola par rapport aux mêmes personnes buvant la même boisson mais sans étiquette, donc sans savoir qu'il s'agissait de Coca Cola. Leurs résultats ont montré que les stratégies de marketing (par exemple, la présence de la marque Coca Cola) peuvent entraîner des réponses différentes dans le cerveau des consommateurs (McClure et al., 2004). Cela signifie que lorsque nous buvons un Coca Cola dans une bouteille où l'étiquette de la marque est visible, nous aimons davantage cette boisson que lorsque nous la buvons dans une bouteille sans étiquette, même si la boisson est exactement la même. Ces résultats ont conduit à la création d'une branche des neurosciences à l'intersection avec le marketing, appelée **neuromarketing**. Ce domaine de recherche s'est rapidement développé au cours de la dernière décennie. Aujourd'hui, plusieurs multinationales comme Google, Disney, CBS et MacDonald's emploient des services de recherche en neuromarketing pour mesurer les préférences et les impressions des consommateurs vis-à-vis de leurs publicités ou de leurs produits. En outre, un certain nombre de sociétés spécialisées dans le neuromarketing, dont EmSense, Neurosence, MindLab International et Nielsen, utilisent régulièrement des techniques d'imagerie cérébrale, principalement l'IRMf et l'EEG, mais aussi la « *steady state topography* » (SST) ou même des mesures physiologiques (par

exemple, la réponse galvanique de la peau), pour étudier, analyser et prédire le comportement des consommateurs.

Même les **rêves** se sont avérés être, dans une certaine mesure, décodables algorithmiquement grâce à la neurotechnologie. En 2013, des chercheurs du Computational Neuroscience Laboratory de Kyoto ont montré qu'ils pouvaient décoder les rêves en se basant uniquement sur l'activité cérébrale d'individus endormis. En développant une approche de décodage neuronal s'appuyant sur des modèles d'apprentissage machine, les chercheurs ont pu détecter, classer et prédire le contenu des images visuelles rêvées par les sujets de l'expérience, dont l'activité cérébrale était enregistrée par une IRMf (Horikawa *et al.*, 2013).

Les souvenirs, les images mentales, les intentions et les rêves sont autant de processus essentiels de la vie cognitive humaine dont le fonctionnement et le contenu sont de plus en plus accessibles à l'étude empirique. En combinant les neurotechnologies et l'intelligence artificielle, il est désormais possible de décoder différentes composantes du contenu informationnel extrêmement riche du cerveau. En raison de ce potentiel informationnel, ces technologies ont souvent été envisagées comme une forme de « lecture du cerveau » (Haynes, 2011), en s'inspirant ainsi de l'analogie entre le déchiffrement des informations et des états mentaux à partir de données neuronales, et l'interprétation d'un texte écrit par la lecture. Certaines de ces études, comme nous l'avons vu, ont même réussi à atteindre un degré suffisant de solidité épistémique pour pouvoir construire des modèles prédictifs.

En résumé, ces dernières années, la richesse quantitative et qualitative des enregistrements neuronaux s'est progressivement et rapidement améliorée. Ce processus d'amélioration a été rendu possible par l'amélioration matérielle des machines, mais aussi, et surtout, par l'amélioration des techniques d'analyse – ce pour quoi l'intelligence artificielle a joué, et continue de jouer, un rôle clef.

1.3. Les informations que la neurotechnologie ne peut pas lire

Les neurotechnologies actuelles, notamment les techniques non invasives, ne sont pas encore capables de *décoder les pensées*. Cela signifie que les neurotechnologies, à leur niveau actuel de développement technologique, ne sont pas capables de fournir un compte rendu complet, détaillé, intelligible et en temps réel des schémas neuronaux sous-tendant des processus mentaux bien spécifiques comme les souvenirs ou les émotions. Pour décoder les pensées, il faudrait donc être capable de révéler le contenu des états mentaux, que l'on appelle le **contenu mental**. Le contenu mental désigne la façon dont un certain état mental en vient à être tel qu'il est. Par exemple, le contenu mental de la mémoire d'une personne est ce dont cette personne se souvient réellement. Le contenu mental de la perception d'une personne est ce qui est réellement perçu par cette personne, etc. En philosophie de l'esprit, on avance ainsi qu'il existe deux principaux types de contenus mentaux : les contenus conceptuels, et les contenus non conceptuels.

Le contenu conceptuel est le **contenu sémantique** d'un état mental. Il est considéré comme analogue au contenu que nous pouvons trouver dans les mots, les expressions et les phrases du langage verbal. Les états mentaux à contenu conceptuel sont également appelés « attitudes propositionnelles ». Le contenu conceptuel d'un état mental comme un souvenir est une proposition particulière qui peut en principe être exprimée par une phrase. Par exemple, si une personne se souvient à un moment précis d'un événement passé, comme « la nuit de son

18^e anniversaire, sa mère portait un chemisier blanc », il faudrait, pour décoder le contenu sémantique de cet état mental, élaborer une neurotechnologie capable d'identifier dans les signaux cérébraux de cette personne en pleine réminiscence les corrélats neuronaux d'une proposition affirmant que la nuit de son 18^e anniversaire, sa mère portait un chemisier blanc. Dans l'idéal, cette neurotechnologie devrait même être capable de transposer telle quelle la proposition « le soir de mon 18^e anniversaire, ma mère portait un chemisier blanc » à partir des seules données cérébrales.

Le **contenu non conceptuel**, en revanche, est un contenu mental qui n'est pas exprimé sous la forme d'une proposition, mais qui relève de l'expérience vécue, de la qualité d'une chose, de la phénoménologie. Selon une longue tradition de pensée philosophique et scientifique, le contenu non conceptuel est le contenu généralement constitutif des sentiments, des expériences et des sensations. Alors que le contenu conceptuel peut être sémantiquement évalué de manière assez directe, cela n'est pas possible pour le contenu non conceptuel. Les états mentaux au contenu non conceptuel sont communément désignés sous le nom d'états « qualitatifs », « expérientiels » ou « phénoménaux ». Le contenu de ces états mentaux est de nature qualitative, expérientielle ou phénoménale, et n'est donc jamais identifiable à une simple proposition. Par exemple, si une personne vit une certaine expérience qualitative quand elle sent le parfum d'une glycine, il faudra reconnaître les corrélats neuronaux de cette expérience qualitative unique (et non d'autres expériences) dans les données neuronales de la personne si l'on veut décoder ce contenu non conceptuel. Dans l'idéal, cette expérience qualitative telle que la neurotechnologie l'a identifiée devrait aussi, au moins en principe, pouvoir être reproduite dans le cerveau d'une autre personne. Une fois que les corrélats neuronaux de cette expérience précise ont été identifiés, recréer les mêmes corrélats dans le cerveau d'un autre individu devrait, en principe, faire vivre la même expérience qualitative à ce dernier.

Au stade du développement de la neurotechnologie en 2021, le décodage du contenu mental, qu'il soit conceptuel ou non, n'est pas possible. Cependant, certaines études récentes montrent que des éléments de contenu conceptuel peuvent être décodés à partir de l'activité cérébrale humaine. Jack Gallant et son équipe de l'université de Berkeley ont analysé la manière dont les cerveaux de sept individus placés sous IRMf réagissaient au visionnage de 129 minutes d'extraits de bandes-annonces de films et d'autres sources (Huth et *al.*, 2016). Ils avaient en amont créé un algorithme de décodage qui s'est montré capable, à partir des signaux d'activité cérébrale mesurés par l'IRMf, de décoder des informations détaillées au sujet des catégories d'objets et des actions apparaissant dans les films. Bien que cet algorithme ne soit pas suffisamment puissant pour décoder le contenu sémantique complet des informations mentales, il est suffisamment puissant pour décoder avec précision la présence ou l'absence de catégories sémantiques générales (par exemple, animal *vs* structure), de catégories spécifiques (par exemple, un chien ou un mur) et d'actions (par exemple, parler *vs* courir). En d'autres termes, l'algorithme n'est pas capable de retranscrire une représentation mentale telle que « c'est Fuffy, mon chien adoré », mais il peut décoder le contenu mental « c'est un chien » ou même « ce chien court ».

Il faut considérer de façon réaliste les limites actuelles du décodage mental par la neurotechnologie, afin d'éviter que le grand public ait des attentes irréalistes, et guider une gouvernance qui puisse s'appuyer sur des faits prouvés. Il convient toutefois de souligner que les limites actuelles de la neurotechnologie en matière de décodage mental sont contingentes,

c'est-à-dire qu'elles ne sont pas une fatalité. En effet, étant donné que les états mentaux résultent de l'activité neuronale et sont élaborés par elle, il n'y a aucune raison logique de penser que le contenu mental soit destiné à rester à jamais indécodable. Au fur et à mesure que les dispositifs et les méthodes neurotechnologiques s'amélioreront, la capacité des systèmes neurotechnologiques à décoder le contenu mental s'améliorera en conséquence.

Outre les limites actuelles qui touchent le décodage du contenu mental, il convient également de souligner que de nombreuses neurotechnologies actuellement disponibles au grand public ont une précision limitée. Il a été observé que de nombreux fabricants de produits neurotechnologiques grand public faisaient de la publicité en vantant des effets (par exemple, l'amélioration du bien-être mental) soit non fondés, soit vaguement corroborés par des preuves scientifiques (Wexler & Reiner, 2019). Encore une fois, cependant, cette limite à la précision des produits de consommation actuels est une limite contingente, sans doute causée par la nouveauté de ce secteur industriel, ainsi que par la vérification assez sporadique de ces systèmes. À mesure que les entreprises de neurotechnologie grand public améliorent leurs dispositifs, et que davantage de données cérébrales deviennent disponibles pour mener des analyses et entraîner les algorithmes, il est donc plausible de prédire que la précision et la solidité scientifique des produits neurotechnologiques grand public augmenteront en conséquence.

2. Défis éthiques

2.1. Neuroéthique et neurodroit

Depuis les années 1990, le développement continu des neurotechnologies et leur application croissante dans le secteur biomédical ont suscité de nombreuses réflexions sur les implications éthiques et juridiques liées à l'exploration et/ou à la modification du cerveau humain. Ces réflexions ont donné naissance à deux nouveaux champs d'investigation normative situés à l'intersection des neurosciences, de la bioéthique, de la médecine et du droit : la neuroéthique et le neurodroit.

Le mot « neuroéthique » a été inventé par William Safire en 2002. Il l'a originellement défini comme « l'examen de ce qui est bien et mal, bon et mauvais dans le traitement, le perfectionnement ou l'invasion importune et la manipulation inquiétante du cerveau humain » (Safire, 2002). La même année, Adina Roskies a proposé de diviser le domaine en deux branches intimement liées : l'éthique des neurosciences et les neurosciences de l'éthique. Elle définit la première comme un cadre moral visant à réguler, ordonner et guider le comportement dans la recherche neuroscientifique. Quant à la seconde, elle la définit au contraire comme l'étude empirique de la manière dont la moralité en tant que telle trouve son origine dans le cerveau humain et est suivie ou non par celui-ci (Roskies, 2002). Dans ce rapport, nous nous concentrerons quasi exclusivement sur l'éthique des neurosciences, bien que certaines réflexions puissent tout aussi bien valoir pour la neuroscience de l'éthique.

L'apparition du terme « neurodroit » a précédé d'une dizaine d'années celle du terme « neuroéthique », puisqu'il a été inventé par J. Sherrod Taylor et *al.* en 1991 afin de désigner la collaboration croissante entre neuropsychologues et avocats au sein du système de justice pénale (Taylor, Harp, & Elliott, 1991). Cependant, il s'agissait d'une définition très restreinte, qui pouvait difficilement englober le large spectre des études neurolégales modernes. Plus tard, sa signification a été élargie pour désigner l'ensemble de l'espace de réflexion situé à l'intersection des neurosciences et du droit (Shen, 2016).

Une étape historique pour la neuroéthique et le neurodroit a été franchie, en mai 2006, lorsque la Société internationale de neuroéthique (INS) a été fondée, dans le prolongement d'une réunion qui s'était tenue la même année à Asilomar en Californie. L'INS a joué un rôle central dans l'institutionnalisation de la neuroéthique et du neurodroit comme disciplines universitaires à part entière.

Tout au long des années 1990 et au début des années 2000, quatre grandes thématiques ont dominé l'opinion publique et la réflexion académique consacrée à la neuroéthique et au neurodroit : l'admissibilité éthique de l'amélioration des fonctions cognitives (en particulier par le biais des nootropes) (Farah et *al.*, 2004 ; Turner & Sahakian, 2006) ; les implications philosophico-juridiques de la neuroscience qui étudie le libre arbitre, avec un accent particulier mis sur les notions de responsabilité morale et de culpabilité légale (Fins, 2004 ; Moreno, 2003 ; Pereboom & Caruso, 2002) ; l'éthique de l'imagerie cérébrale et du déchiffrement du contenu de l'esprit (Farah, 2002) (Illes, Kirschen, & Gabrieli, 2003 ; Illes et *al.*, 2004) ; et la validité, ainsi que l'admissibilité, des preuves neuroscientifiques devant les tribunaux.

De plus, depuis le début des années 2000, un cinquième domaine complémentaire d'investigation neuroéthique et neurojuridique est apparu. Il entreprend d'examiner les défis éthico-juridiques des neurosciences et des neurotechnologies au prisme des catégories

normatives que sont les droits et les devoirs (voir 3.2). Cette manière d'analyser les implications éthiques et juridiques des neurosciences est aujourd'hui connue sous le nom de « neurodroits ».

Les implications éthiques des neurosciences et des neurotechnologies peuvent dès lors être regroupées en trois grandes thématiques : vie privée ; autonomie, agentivité et responsabilité ; justice. Dans ce qui suit, nous allons fournir une analyse concise de chacune de ces trois catégories.

2.2. Vie privée

La **protection de la vie privée** est une préoccupation éthique majeure qui s'impose à nous face à la collecte, à la diffusion et au traitement des données cérébrales. Si le traitement de n'importe quelle donnée humaine pose des problèmes de respect de la vie privée, on estime que le traitement des données cérébrales pose de nouveaux défis à la notion de respect de la vie privée, et ce pour quatre raisons principales : le contrôle conscient limité de l'individu sur ses propres enregistrements cérébraux, la protection du for intérieur, la richesse informationnelle, et le risque de neurodiscrimination.

Tout d'abord, il faut rappeler que la vie privée est à la fois un droit et une capacité. En tant que telle, elle repose sur la capacité consciente d'un individu à filtrer le flux de données et à en isoler intentionnellement les informations qu'il considère comme privées. Les données cérébrales, en revanche, échappent le plus souvent à ce contrôle conscient et ne peuvent donc pas toujours être intentionnellement isolées. Parmi les types d'informations auxquels les neurotechnologies peuvent accéder, on trouve non seulement les processus cérébraux conscients, mais aussi les processus subconscients (par exemple, la cognition inconsciente ou les états affectifs subconscients), sur lesquels un individu a, par définition, un contrôle conscient limité, voire nul. Par exemple, lorsqu'un participant à une expérience d'imagerie cérébrale donne son accord pour que son activité cérébrale soit enregistrée dans le cadre de l'objectif de l'étude indiqué dans le formulaire de consentement éclairé, il est toujours possible, en principe, de collecter des informations subconscientes redondantes sans que ce participant en soit conscient, ou bien sans son autorisation. Contre cette objection, on pourrait faire valoir que lorsqu'un individu consent à autoriser l'acquisition de ses données cérébrales, il accepte aussi de renoncer à protéger ses informations mentales, du moins dans une certaine mesure.

Toutefois, dans les scénarios où la collecte de données cérébrales serait soit obligatoire (par exemple, dans le secteur militaire ou dans le cadre du travail), soit très avantageuse (voir, par exemple, le projet de Facebook de rendre la dactylographie cérébrale plus rapide que l'écran tactile), le risque de partager ses données sous la pression d'une contrainte explicite ou implicite est bien concret. En outre, il est possible d'enregistrer des données redondantes, c'est-à-dire des données liées à une fonction et/ou une structure cérébrale, et de les utiliser pour une finalité autre que celle à laquelle l'individu a explicitement consenti. Cela augmente donc le risque de voir violée l'exigence de *limitation de la finalité*, à savoir le principe selon lequel des données collectées pour une finalité déterminée ne doivent pas être utilisées pour une autre finalité incompatible. Cette exigence est protégée par le Règlement général sur la protection des données (RGPD) de l'UE, ¹qui stipule que les données à caractère personnel doivent être

¹ Les réflexions sur la vie privée et la protection des données contenues dans ce rapport sont principalement inspirées du Règlement général sur la protection des données (RGPD) de l'UE,

collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités (article 5, paragraphe 1, point b) du RGPD).

Si le problème du contrôle conscient limité est commun à d'autres types de données (par exemple, les données génétiques), il acquiert une plus grande complexité éthique dans le domaine neuronal. En effet, les données cérébrales ne peuvent pas être séparées entre données traitées d'une part, et système générant ces données et prenant les décisions pour les traiter (le cerveau humain) d'autre part. Ienca et Andorno ont appelé cela le « problème de l'inception » (Ienca & Andorno, 2017b). En outre, contrairement à d'autres mesures physiologiques, on peut considérer que les données cérébrales disposent d'un contenu sémantique : leur nature est donc propositionnelle. Comme nous l'avons vu précédemment, la possibilité d'un décodage sémantique modéré (le décodage du contenu conceptuel) a déjà été démontrée par des expériences effectuées grâce à l'imagerie cérébrale (Huth et al., 2016). Par conséquent, l'acquisition non autorisée des données cérébrales d'une personne, par exemple au titre de preuve devant un tribunal, ou après avoir été exigées par les forces de l'ordre lors d'un interrogatoire, peut violer le droit de garder le silence, c'est-à-dire le principe juridique qui garantit à tout individu le droit de refuser de répondre aux questions des agents de l'autorité publique ou des fonctionnaires d'un tribunal. Dans les enquêtes criminelles, le fait que l'activité cérébrale génère des contenus sémantiques et des attitudes propositionnelles soulève par exemple la question de savoir si les résultats de cette activité doivent être considérés comme une preuve physique ou comme un témoignage. Le premier type de preuve désigne uniquement des éléments matériels, comme des cheveux, des échantillons de sang, des empreintes digitales ou d'autres matériaux biologiques. Le second est constitué des déclarations ou des paroles prononcées par l'accusé, par une victime ou par des témoins. La reconnaissance officielle de la capacité des données cérébrales à servir, de manière fiable, de témoignage devant les tribunaux, est actuellement très rare.

Deuxièmement, les informations cérébrales sont liées à ce que le philosophe du XVII^e siècle René Descartes appelait le « *locus internus* », c'est-à-dire le for intérieur d'une personne (également appelé « *forum internum* »). Dans ce for intérieur, on trouve les informations non exprimées par la personne, ses préférences préconscientes, ses souvenirs, ses intentions cachées et ses croyances. Ces informations sont dites intérieures parce qu'elles existent uniquement dans l'esprit de cette personne précise, éventuellement sous une forme propositionnelle – même si cette personne ne les extériorise pas dans le monde extérieur par la parole, l'écriture ou par tout autre comportement. À ce titre, il a été avancé que les informations cérébrales constituaient le dernier ressort de l'individu en matière de confidentialité des informations qui le concernent, : elles comprennent en effet les comportements non exécutés, ainsi que tout discours intérieur ou toute autre action non externalisée (Ienca & Andorno, 2017b). En principe, la vie privée de l'esprit pouvait donc être préservée même si le comportement individuel était constamment surveillé (suivi des activités, technologie numérique personnelle, auto-quantification ou simple observation). En revanche, la collecte et le traitement de données cérébrales permettent une certaine latitude d'accès aux informations mentales, et ce même en

car il s'agit de l'une des lois sur la protection des données les plus complètes au monde, et car elle a également un impact extraterritorial sur d'autres systèmes juridiques.

l'absence de comportement observable. Cela peut donc remettre en question l'intimité de l'esprit ou ce que l'on appelle la **vie privée mentale**.

Troisièmement, la **richesse informationnelle** des enregistrements cérébraux. Ces derniers peuvent en effet contenir sous forme codée des informations hautement privées sur les individus dont ils proviennent, comme certaines caractéristiques prédictives sur leur santé, leurs attitudes et leurs états mentaux. Comme nous l'avons vu précédemment, de nombreuses études menées depuis le début du troisième millénaire ont démontré qu'il était possible, à partir des données cérébrales d'une personne (enregistrées via EEG, IRMf ou autres), de décoder des contenus mentaux comme des intentions cachées (Haynes et *al.*, 2007), des informations dissimulées (Bles & Haynes, 2008), des images naturelles (Kay et *al.*, 2008), des expériences visuelles (Nishimoto et *al.*, 2011) et les éléments inconscients présidant à la prise de décisions libres (Bode et *al.*, 2011). Certaines études ont réussi à atteindre un degré suffisant de solidité épistémique pour construire des modèles prédictifs. Par exemple, deux études célèbres ont utilisé l'IRMf pour prédire le flux de conscience (Haynes & Rees, 2005) et les choix d'une personne sur sa préparation motrice, mais aussi, et de manière surprenante, sur ses intentions abstraites. Ces études ont soulevé un cri d'alarme légitime au sein de la communauté scientifique, car la conscience, l'intentionnalité et le libre choix sont des composantes essentielles de la faculté que l'on appelle habituellement le « libre arbitre » dans la tradition théologique et philosophique. Bien que le débat reste ouvert quant à la possibilité de décoder, non seulement les corrélats neuronaux des informations mentales, mais aussi leur contenu réel, il est indéniable que, ces dernières années, la richesse quantitative et qualitative des enregistrements neuronaux s'est progressivement et rapidement améliorée. Ce processus d'amélioration a été permis par l'amélioration des machines, mais aussi et surtout par l'amélioration des techniques d'analyse, dans laquelle l'intelligence artificielle a joué, et continue de jouer, un rôle clef. Dans un avenir proche, décoder des informations mentales privées devrait devenir de plus en plus envisageable grâce aux progrès concomitants de la technologie des capteurs, de la résolution spatiale des enregistrements et des techniques d'apprentissage machine visant à reconnaître des formes et à extraire des caractéristiques.

En raison de leur richesse informationnelle, les données cérébrales ont également une dimension **biométrique**. Les signaux cérébraux permettent en effet de distinguer ou de retracer l'identité d'un individu, et sont liés à celui-ci. Certains enregistrements cérébraux (par exemple, les signaux EEG) peuvent être utilisés comme identifiant biométrique unique, similaire en cela aux empreintes digitales ou à l'ADN. En 2007, des informaticiens britanniques ont développé un cadre biométrique basé sur l'EEG pour vérifier automatiquement l'identité (Palaniappan & Mandic, 2007). Depuis lors, de nombreux systèmes biométriques non intrusifs, basés sur l'EEG, ont été développés pour reconnaître, authentifier et identifier des personnes. Cependant, contrairement à d'autres informations identifiantes, les ondes cérébrales peuvent en théorie être enregistrées à des fins biométriques à l'insu de la personne (par exemple, comme activité secondaire dans une étude de neuromarketing), donc en l'absence d'une réelle capacité de la personne à consentir à la collecte et à l'utilisation de ces informations. À mesure que le marché des dispositifs portatifs basés sur l'EEG se développe, et en l'absence de toute capacité réelle d'obtenir un consentement éclairé quant au traitement des données qu'ils génèrent, de nouvelles réponses pour protéger les données cérébrales doivent être établies.

Enfin, le traitement des données cérébrales, notamment des biomarqueurs d'imagerie cérébrale, engendre un risque de « **neurodiscrimination** », c'est-à-dire d'une discrimination fondée sur les signatures neurales d'une personne (qui pourraient indiquer, par exemple, une prédisposition à la démence) ou bien sur sa santé mentale, ses traits de personnalité, ses performances cognitives, ses intentions et ses états émotionnels. La neurodiscrimination est, à bien des égards, semblable à la discrimination génétique, qui intervient lorsqu'une personne est traitée différemment par son employeur ou sa compagnie d'assurance au motif qu'elle est porteuse d'une mutation génétique causant ou augmentant le risque d'une maladie héréditaire. Tout comme le risque de discrimination génétique est une préoccupation de plus en plus pressante en raison de la disponibilité croissante des tests génétiques, dont des tests génétiques acheteables directement par le consommateur (Chapman, Mehta, Parent, & Caplan, 2019), la neurodiscrimination deviendra probablement elle aussi une préoccupation éthique de plus en plus pressante en raison de la disponibilité croissante de la neurotechnologie omniprésente, dont font partie des dispositifs et appareils neuronaux directement acheteables par le consommateur. La discrimination génétique est explicitement interdite par l'article 11 (chapitre IV) de la Convention d'Oviedo, qui stipule que « toute forme de discrimination à l'égard d'une personne en raison de son patrimoine génétique est interdite ». Il pourrait peut-être devenir nécessaire d'inclure une disposition similaire pour interdire un type de discrimination qui se fonderait sur les caractéristiques neurobiologiques d'une personne. Une telle disposition peut contribuer à renforcer la protection de la dignité de tous les êtres humains, ainsi qu'à promouvoir leurs droits et leurs libertés sans discrimination (article 1). Si notre société attache de l'importance à l'accès aux soins de santé pour les personnes en bonne santé comme pour les malades, pour les personnes neurodominantes comme pour les personnes neurodiverses, nous devrions soutenir largement des interdictions strictes de toute neurodiscrimination dans le cadre de l'assurance maladie, y compris de l'assurance maladie de l'employeur, car elles peuvent porter atteinte aux principes éthiques garantissant des soins de santé universels, ainsi que l'équité des systèmes de soins de santé.

Les pratiques de gestion des données neurotechnologiques posent également des problèmes de confidentialité. Les individus concernés peuvent en effet perdre le contrôle de leurs données cérébrales de plusieurs façons :

- (i) en consentant à la collecte de leurs données sans être conscient et sans être suffisamment informé ;
- (ii) en donnant un consentement éclairé au traitement de leurs données pour une certaine finalité, mais en restant inconscient des utilisations ultérieures de leurs données à des fins différentes (y compris leur récupération par des tiers) ;
- (iii) en étant contraint à la collecte de leurs données cérébrales (par exemple, une obligation venant de l'employeur, ou au cours d'un interrogatoire de police) ;
- (iv) à cause d'un accès non autorisé à ces données à des tiers ;
- (v) à la suite d'un vol de données.

Ienca et Haselager (2016) ont passé en revue les différents défis de sécurité posés par les BCI et identifié plusieurs vulnérabilités qui pourraient être exploitées par des acteurs malveillants pour obtenir illégalement l'accès aux informations cérébrales en piratant ces BCI.

2.3. Autonomie, agentivité et responsabilité

L'utilisation croissante de l'apprentissage machine et, plus généralement, de l'intelligence artificielle pour optimiser le fonctionnement des BCI, a également des conséquences sur les notions éthiques d'autonomie, d'agentivité et de responsabilité. Par exemple, Haselager (2013) a fait remarquer que, lorsque le contrôle d'une BCI dépend en partie de composants algorithmiques intelligents, il peut devenir difficile de discerner si la sortie comportementale en résultant a réellement été effectuée par l'utilisateur. Cette difficulté introduit un principe d'indétermination au sein du processus cognitif chapeautant le cycle de vie d'une action de sa conception (ou intention) jusqu'à son exécution. Par conséquent, ceci crée également une incertitude sur l'attribution de la responsabilité à l'auteur de l'action.

Ce principe d'indétermination pourrait remettre en cause la notion de responsabilité individuelle, remise en cause dont les répercussions sur le droit pénal et le droit des assurances sont évidentes. De plus, il pourrait générer un sentiment d'aliénation chez l'utilisateur, notamment chez un individu vulnérable comme c'est le cas d'un patient neurologique. Imaginons par exemple qu'un patient souffrant de tétraplégie utilise une BCI fortement enrichie en composants intelligents visant à extraire, décoder et classifier des informations : comment sera-t-il possible de déterminer quelles composantes des actions du patient sont attribuables à la volonté du patient, et lesquelles sont attribuables à l'IA ? Cette question est particulièrement ardue, comme mentionné ci-dessus, dans des circonstances où l'attribution de la responsabilité a une signification juridique. Plus généralement, Thompson a observé que l'utilisation des BCI est problématique pour le droit pénal. La raison en est que le droit pénal exige qu'une personne ne puisse être jugée pénalement responsable que si elle a satisfait à l'exigence suivante fondatrice de l'acte de culpabilité : « l'agent a accompli un comportement (convenablement spécifié) ». En revanche, les agents qui agissent sur le monde à l'aide d'interfaces cerveau-machine n'accomplissent manifestement aucun comportement, de sorte que s'ils commettent des crimes à l'aide de BCI, statuer pour savoir si leur action est assimilable à un acte de culpabilité n'est pas évident (Thompson, 2019).

En outre, il est possible que la place centrale de ces composants intelligents dans le fonctionnement des BCI puisse affecter l'expérience subjective, et donc l'identité personnelle. Cette hypothèse a récemment obtenu une première confirmation empirique dans une étude qualitative consacrée à l'expérience personnelle des patients utilisant des BCI (Gilbert, Cook, O'Brien, & Illes, 2019 ; Gilbert, Goddard, Viaña, Carter, & Horne, 2017). Il convient toutefois de noter que si l'IA peut brouiller les aspects subjectifs de l'identité personnelle, une BCI optimisée par l'IA, prise dans son ensemble, peut considérablement améliorer la capacité de l'utilisateur à agir dans un environnement donné, en particulier lorsqu'elle est utilisée pour le contrôle moteur par un patient souffrant d'une déficience motrice grave. Il est donc difficile de déterminer si, dans l'absolu, les BCI intelligentes peuvent accroître ou non l'autonomie de l'utilisateur. Au contraire, il est nécessaire d'évaluer au cas par cas, et de déterminer dans quelles circonstances, dans quels intervalles de temps et dans quels domaines mentaux ou physiques on détecte un changement (positif ou négatif) dans l'autonomie de l'utilisateur. Pour réaliser ces évaluations, il est important d'acquérir non seulement des informations quantitatives et objectives (par exemple, des mesures mathématiques ou des observations comportementales), mais aussi des informations qualitatives et subjectives. Dans cette dernière

catégorie, on trouve en particulier les auto-évaluations de l'utilisateur, considérées comme une fenêtre d'accès à l'expérience à la première personne de l'utilisateur (Ferretti & Ienca, 2018).

Comme nous l'avons vu, avec l'augmentation des utilisations non cliniques des BCI, un autre défi éthique qui se profile sera celui de l'amélioration cognitive ou de toute autre amélioration cérébrale, c'est-à-dire toute augmentation fonctionnelle du système nerveux. Alors que les applications cliniques des BCI visent à restaurer les fonctions motrices ou cognitives chez les personnes souffrant de déficiences physiques ou cognitives, comme les survivants d'un accident vasculaire cérébral, les applications de BCI visant à l'amélioration cérébrale pourraient, dans un avenir proche, conduire à des performances cognitives ou physiques supérieures à ce que l'on observe habituellement chez les personnes en bonne santé. Il sera alors urgent de discuter des types d'amélioration à autoriser et dans quelles circonstances les autoriser. Aujourd'hui déjà, il existe un vaste écosystème d'entreprises privées qui commercialisent des BCI non invasives auprès d'un nombre sans cesse croissant d'utilisateurs en bonne santé, et cela à des fins comme l'auto-quantification, l'entraînement cognitif, le *neurogaming* (des jeux vidéo contrôlés par le cerveau et utilisés à des fins récréatives ou compétitives) et la polysomnographie. Certaines entreprises, dont Emotiv qui est basée à San Francisco en Californie, affirment publiquement (bien que sans preuve scientifique solide) être en mesure de « contribuer à améliorer le bien-être et la productivité » des utilisateurs en bonne santé cognitive. Par ailleurs, les BCI dédiées au contrôle moteur permettent déjà, non seulement d'amplifier les capacités existantes, mais aussi d'acquérir des facultés autrement absentes chez l'être humain : en particulier, le contrôle télépathique de dispositifs robotiques comme les drones ou d'autres véhicules semi-autonomes. Ce mode d'interaction homme-machine est largement répandu dans l'industrie des transports et le secteur militaire. Bientôt, la diffusion de ces applications et la validation scientifique de leurs mécanismes de fonctionnement rendront incontournable un questionnement éthique sur l'augmentation humaine.

Enfin, l'utilisation malveillante de la neurotechnologie par des tiers, notamment par le biais d'interventions extérieures détournant le contrôle des systèmes neurotechnologiques utilisés par un individu, ont des conséquences sur l'autonomie, l'agentivité et la responsabilité. Il a été expérimentalement démontré que ces neurotechnologies pouvaient être piratées par des acteurs malveillants qui veulent s'en octroyer le contrôle. Ce type d'attaque pourrait avoir des conséquences délétères pour la victime, notamment en conduisant à l'extraction non consentie de ses informations mentales, à l'expropriation du contrôle conscient de ses membres robotisés, ou encore à des blessures physiques et psychologiques graves causées par l'augmentation intentionnelle de l'intensité de la neurostimulation (Chaudhary & Agrawal, 2018 ; Ienca & Haselager, 2016 ; Pugh, Pycroft, Sandberg, Aziz, & Savulescu, 2018 ; Pycroft et al., 2016).

2.4. Justice

Les conséquences de la neurotechnologie en matière de justice touchent à la détermination des conditions d'un accès juste et équitable aux avantages qu'elle peut fournir. Ienca (2018) a désigné sous le nom de « démocratisation » de la neurotechnologie toute approche de gouvernance conçue pour universaliser et distribuer uniformément les avantages potentiels induits par la neurotechnologie (Ienca, 2019a). Distribuer également et équitablement les avantages potentiels induits par la neurotechnologie est de première importance s'agissant des neurotechnologies médicales comme les implants DBS utilisés pour

traiter la maladie de Parkinson, les technologies d'assistance intelligentes utilisées pour traiter la démence, les interfaces neurales compensant la perte de la fonction motrice, et les technologies de réhabilitation cérébrale. Ces dernières années, cependant, des auteurs ont commencé à réfléchir aux conséquences que peuvent avoir sur la justice et l'équité des neurotechnologies conçues pour améliorer les fonctions cognitives (Ienca, Shaw, & Elger, 2019).

Par ailleurs, le problème des disparités mondiales en matière d'accès à la neurotechnologie à des fins médicales est une autre question éthique essentielle pour la justice mondiale, car de nombreux États du Sud à faible revenu ne disposent pas d'infrastructures de diagnostic et de soins suffisantes (Palk *et al.*, 2020).

3. Les défis des droits de l'homme

Alors que la neurotechnologie progresse et ouvre de nouvelles possibilités pour surveiller et contrôler les fonctions cognitives, affectives, conatives et physiques, l'incertitude règne quant à savoir si, et comment, le droit doit faire face à ces avancées. En particulier, on peut se demander si les questions éthiques posées par les tendances émergentes en matière de neurotechnologie peuvent être entièrement traitées par les acteurs même de la neurotechnologie. Ces dernières années, plusieurs experts ont affirmé que la complexité des défis éthiques soulevés par la neurotechnologie (et l'IA) ne pouvait être régulée que par le biais d'orientations éthiques professionnelles, de bonnes pratiques et de l'autorégulation. Au contraire, il a été avancé ailleurs qu'il faudrait, pour instituer une bonne gouvernance sur ces questions, une révision, voire une réforme radicale de concepts juridiques présents notamment dans le droit civil, le droit commercial, le droit pénal et la philosophie du droit. Alors que la littérature savante a porté une attention significative aux applications émergentes de la neurotechnologie dans le contexte du droit pénal, ou bien à l'utilisation croissante des preuves neuroscientifiques dans les tribunaux, une faible attention a été accordée aux conséquences du développement de la neuroscience et de la neurotechnologie sur le droit des droits de l'homme. Pourtant, cette composante du discours tenu sur la neuroéthique et le neurodroit est particulièrement pertinente, car la nature universelle du cadre des droits de l'homme pourrait fournir une base solide pour fonder ce que Boire appelle une « jurisprudence de l'esprit » (Boire, 2001).

Étant donné que la fonction cérébrale et les facultés mentales recourent plusieurs domaines de l'activité humaine, il est peu probable qu'une approche unique de la gouvernance des neurotechnologies puisse être efficace. Par conséquent, tout cadre d'une gouvernance mondiale des neurotechnologies devrait englober plusieurs niveaux : neuroéthique, droit souple, innovation responsable, et réglementation contraignante. Plus important encore, étant donné la place centrale du cerveau dans la vie humaine, les défis normatifs de la neurotechnologie et de l'IA devraient être fondés sur les cadres des droits de l'homme.

Il n'entre pas dans l'objectif de ce rapport de discuter des différentes théories des fondements des droits de l'homme, ni de prendre position à leur sujet. Aux fins du présent rapport, on adoptera donc une conception pratique et large des droits de l'homme, comme celle proposée par Beitz (2011, p. 109). Selon Beitz, les droits de l'homme sont « des exigences dont l'objet est de protéger des intérêts individuels urgents contre des dangers prévisibles (« menaces communes ») auxquels les individus sont exposés au cours des circonstances typiques de leur vie dans un ordre mondial moderne composé d'États » (Beitz, 2011). De manière générale, on peut dire que l'objectif des droits de l'homme est d'assurer les conditions négatives et positives nécessaires pour qu'un individu puisse mener une vie minimalement bonne (Fagan, 2015).

On a pu voir que les défis éthiques posés par les BCI et autres neurotechnologies nous incitaient à aborder une question éthique et socio-juridique fondamentale : *déterminer si, ou dans quelles conditions, il est légitime d'accéder à l'activité neuronale d'une personne ou d'interférer avec elle* (Ienca, 2017). Cette question doit être posée à différents niveaux de juridiction, notamment au niveau des droits de l'homme (Ienca & Andorno, 2017). La raison pour cela provient d'un triple fait : d'abord, comme nous l'avons expliqué, l'activité neuronale se définit comme le substrat fondamental de l'identité personnelle et, par conséquent, de la

responsabilité morale et juridique. Par conséquent, le déchiffrement et la manipulation de l'activité neuronale au moyen de techniques neurotechnologiques pourraient, en toute logique, avoir des répercussions sans précédent sur l'identité personnelle des utilisateurs, et introduire un élément d'obscurcissement, voire d'indétermination, dans l'attribution de la responsabilité morale et juridique. Deuxièmement, l'activité cérébrale est détectable chez tout être humain, indépendamment de son genre, de son sexe, de sa nationalité, de son ethnie ou de son appartenance politique ou religieuse. Enfin, comme indiqué plus haut, les données cérébrales comportent non seulement des informations électrophysiologiques, mais aussi des informations mentales. Selon une tradition qui va d'Homère à John Milton et jusqu'à Virginia Woolf, le domaine de l'esprit et les informations qui lui sont associées doivent être protégés en tant que domaine privé *par excellence*, et être considéré comme le dernier territoire inaccessible à l'érosion de la confidentialité de l'information et à l'intrusion insidieuse de la société des données.

Alors que la neurotechnologie est susceptible d'affecter des droits de l'homme comme la vie privée, la liberté de pensée, le droit à l'intégrité mentale, la non-discrimination, le droit à un procès équitable ou le principe de non-incrimination, la plupart des dispositions du droit international des droits de l'homme ne font aucune référence explicite aux neurosciences et/ou à la neurotechnologie. Contrairement à d'autres développements biomédicaux, qui ont déjà fait l'objet d'efforts normatifs nationaux et internationaux, la neurotechnologie reste encore largement une *terra incognita* pour le droit des droits de l'homme. Cependant, le constat des conséquences potentielles des neurosciences et des neurotechnologies sur les propriétés inhérentes aux êtres humains réclame une réponse rapide et applicable de la part du droit des droits de l'homme.

3.1. Lacunes dans les cadres internationaux des droits de l'homme

Les instruments relatifs aux droits de l'homme, tels que la Déclaration universelle des droits de l'homme des Nations unies (DUDH, 1948) et le Pacte international relatif aux droits civils et politiques (PIDCP, 1966), ont été rédigés avant que le cerveau humain ne devienne l'objet de mesures (à l'exception près de quelques techniques comme l'EEG), avant qu'il ne soit soumis à l'analyse informatique et avant qu'il ne soit directement modifiable par l'intermédiaire du *neurofeedback* ou de la neurostimulation. Ces instruments ne définissaient donc pas explicitement les conditions d'accès au cerveau humain et aux informations qui peuvent en être extraites, ni les conditions que doit respecter une intervention neurotechnologique pour préserver la dignité humaine et les droits de l'homme (Ienca et *al.*, 2021).

Ceci étant dit, la DUDH énonce malgré tout une liste de « droits fondamentaux et de libertés fondamentales », et affirme leur caractère universel en tant que droits inhérents, inaliénables et applicables à tous les êtres humains. À y regarder de plus près, certains des droits et libertés contenus dans la DUDH font implicitement référence à la protection du cerveau et de l'espace mental (par exemple, l'article 18 sur la « liberté de pensée »). Toutefois, ils n'ont pas pu énoncer explicitement des scénarios mettant en jeu certains risques spécifiques rendus possibles par les neurotechnologies, ni fournir un cadre correctement articulé pour protéger le cerveau et l'esprit humains.

Dans les décennies qui suivirent l'adoption de la DUDH, de nouvelles déclarations ont été rédigées afin de protéger les droits de l'homme face aux effets potentiellement néfastes des avancées technologiques. Par exemple, les progrès de la génétique, notamment le développement des technologies de séquençage et d'édition du génome, ont été abordés par l'UNESCO dans sa Déclaration internationale sur les données génétiques humaines (2003). Cette déclaration reconnaît aux données génétiques humaines une « spécificité » qui tient à leur nature sensible, car « elles peuvent indiquer des prédispositions génétiques concernant des individus », et aussi car « ce pouvoir prédictif peut être plus étendu que ne l'indiquent les évaluations faites au moment de l'obtention des données » (Préambule). À partir de ce statut particulier, la Déclaration définit ensuite les conditions régissant une utilisation légitime des données génétiques humaines, en mettant l'accent sur le consentement libre et éclairé, la prévention de la discrimination et de la stigmatisation, la protection de la vie privée et de la confidentialité, et le partage équitable des bénéfices induits avec l'ensemble de la société.

Si on les compare aux données génétiques, les données cérébrales manquent de garanties explicites, et les instruments des droits de l'homme ne les protègent pas autant que d'autres données. Autre exemple, la Convention d'Oviedo du Conseil de l'Europe (1997) contient des dispositions explicites sur les technologies et pratiques biomédicales tels que les tests génétiques prédictifs, le génie génétique, la procréation médicalement assistée, la recherche sur les embryons *in vitro*, ou le prélèvement d'organes et de tissus sur des donneurs vivants à des fins de transplantation. Cependant, elle ne contient aucune disposition relative à la neurotechnologie, et ne fait aucune référence à la protection du cerveau et de l'esprit humains. De plus, dans le paysage du droit international humanitaire, il n'existe aucun traité international spécifique traitant des éventuelles applications militaires de la neurotechnologie. Et ce, malgré le fait que les traités de désarmement, comme la Convention sur l'interdiction des armes biologiques (CIAB) de 1972, abordent les applications au domaine militaire d'autres biotechnologies, comme les armes biologiques et les armes à toxines, et interdisent leur développement, leur production, leur acquisition, leur transfert, leur stockage et leur utilisation.

Par conséquent, aborder la question des droits de l'homme en relation avec les neurotechnologies – et, plus largement, en relation avec le cerveau et l'esprit humains – apparaît comme un défi de plus en plus fondamental. En 2017, Ienca et Andorno ont réalisé une évaluation éthico-juridique des droits de l'homme à l'ère des neurosciences et des neurotechnologies (Ienca & Andorno, 2017b). Ces auteurs ont mené une analyse parallèle et comparative des tendances émergentes en matière de neurotechnologie, ainsi que des dispositions de droits de l'homme liées à la protection du cerveau et de l'esprit humains que contiennent les instruments existants en matière de droits de l'homme, tels que la DUDH des Nations unies (1948), la Charte des droits fondamentaux de l'Union européenne (2000) et la Déclaration universelle sur la bioéthique et les droits de l'homme de l'UNESCO (2005). Leur analyse a permis de conclure que les droits de l'homme existants sont nécessaires, mais qu'ils ne sont peut-être pas suffisants sur le plan normatif – ou, du moins, pas assez souples pour répondre aux problèmes nouveaux soulevés par la neurotechnologie. Pour cette raison, les auteurs ont également fait valoir que « les possibilités ouvertes par les développements neurotechnologiques et leur application à divers aspects de la vie humaine obligeront à reconceptualiser certains droits de l'homme, voire à créer de nouveaux droits pour protéger les personnes contre des dommages potentiels » (Ienca et Andorno, 2017b).

Des lacunes normatives ont donc été décelées dans le droit supranational et international. Ienca et *al.* (2021) ont ainsi donné un aperçu des lacunes actuelles du cadre juridique en vigueur. Ils ont observé qu'à l'évidence, il n'existait actuellement, dans le droit supranational et international, aucun cadre contraignant de gouvernance spécifiquement dédié au cerveau humain et aux informations qui en sont issues. Or, toutes les données générées par les systèmes neurotechnologiques devraient être considérées comme des données à caractère personnel, telles que définies par des instruments comme le RGPD de l'Union européenne, juridiquement contraignant, les Lignes directrices de l'OCDE de 2013 sur la protection de la vie privée, non contraignantes, et la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. Ces instruments définissent les données à caractère personnel comme toute information sur une personne physique identifiée ou identifiable (art. 4 du RGPD ; art. 1 des Lignes directrices de l'OCDE sur la protection de la vie privée ; art. 2a de la Convention du Conseil de l'Europe).

Cependant, il convient de prendre garde à plusieurs points spécifiques si on doit considérer une telle définition à l'aune du progrès neurotechnologique. Premièrement, les instruments de protection des données comme le RGPD ne sont pas toujours applicables si les données obtenues par les systèmes neurotechnologiques sont anonymisées (Ienca et *al.*, 2021) – et ceci malgré le fait que la difficulté technique qu'implique l'anonymisation des données cérébrales laisse la voie ouverte à une possible désanonymisation. La possibilité que des données cérébrales anonymisées redeviennent identifiables est réelle au vu du type de technologies traitant ces données, ainsi que de la précieuse richesse informationnelle de celles-ci. Des chercheurs ont ainsi démontré qu'il était expérimentalement possible de ré-identifier des sujets en s'appuyant sur des mesures électrophysiologiques ou des données fournies par l'imagerie cérébrale. Il est même possible de deviner des états émotionnels présents ou de prédire des comportements futurs à partir de ces données cérébrales, ainsi que de décoder des informations sensibles à partir de l'activité neuronale des sujets ou de leurs phénotypes numériques (Omurtag et *al.*, 2017 ; Schwarz et *al.*, 2019).

L'article 4, n° 15 du RGPD définit également un type spécifique de données personnelles appelées données de santé, qui sont définies comme « des données relatives à la santé physique ou mentale d'une personne physique ». À première vue, les informations dérivées du cerveau et de l'esprit d'une personne constituent des données de santé, dans la mesure où elles peuvent être utilisées pour émettre des conclusions sur l'état de santé physique et mentale d'une personne physique. Cette interprétation est soutenue par l'article 29, qui en profite pour préciser que les données de santé en tant que catégorie comprennent également les informations relatives à la capacité intellectuelle ou émotionnelle d'une personne (Kohnstamm, 2011). Cela impliquerait que les données cérébrales devraient bénéficier de la plus haute protection (cf. article 9, RGPD). Cependant, la question de savoir si toutes les données cérébrales peuvent être considérées comme des données de santé, et donc traitées comme sensibles comme l'exige le RGPD, n'est pas tranchée. Ainsi, il a pu être avancé que les données cérébrales générées par les neurotechnologies grand public n'étaient pas des « données de santé », et disposaient donc potentiellement d'une moindre protection par rapport aux données prélevées dans le cadre clinique, ces dispositifs grand public n'étant pas concernés par les régimes de réglementation des dispositifs médicaux (Rainey et *al.*, 2020). Cela signifie que les données cérébrales collectées par le biais de neurotechnologies non cliniques, comme les BCI

grand public, sont sous-protégées, et que de possibles violations de la vie privée mentale, ainsi que des risques de neurodiscrimination, peuvent survenir.

Deuxièmement, il est exceptionnellement difficile de sauvegarder le droit à l'oubli d'une personne lorsque l'on a affaire à des utilisateurs de neurotechnologies. Le droit à l'oubli désigne le droit qu'a une personne de demander à un responsable de gestion des données de supprimer ses données personnelles des répertoires de données concernés. Comme nous l'avons vu précédemment, les données cérébrales échappent souvent au contrôle conscient et peuvent facilement être désanonymisées. Par conséquent, les personnes concernées peuvent ne pas être conscientes des données qui sont collectées à leur sujet, ni à quelles fins elles sont collectées. En outre, même si une personne est initialement capable de consentir au traitement des données et qu'elle peut les faire supprimer ensuite, le responsable de la gestion de ces données ou d'autres personnes peuvent toujours user de ces données supprimées pour établir un lien entre celles-ci et la personne dont elles sont issues. Dans le cas de données cérébrales contenant des informations inconscientes, le responsable de leur gestion pourrait même conserver des données dont la personne n'a pas conscience (Ienca et *al.*, 2021). La nature même des données cérébrales pourrait ainsi compromettre la capacité des personnes concernées à exercer sur elles leur droit d'accès, de modification et de suppression. Par exemple, Greenberg a fait remarquer que toute personne ne possédait pas nécessairement un ordinateur suffisamment puissant pour traiter les données d'une BCI (Greenberg, 2019). De même, la suppression des données cérébrales pourrait diminuer considérablement la précision des modèles ML générés avec ces données. La conséquence en est que le droit à l'oubli aurait du mal à s'appliquer aux BCI à cause de l'impact qu'une telle suppression de données aurait sur la précision des modèles prédictifs pilotés par l'IA (*ibid.*).

Troisièmement, bien que les lois internationales sur la protection des données, à l'instar du RGPD, soient essentielles pour protéger les données personnelles, le niveau de protection qu'elles offrent aux données cérébrales semble insuffisant : ces lois autorisent en effet certaines exemptions aux droits des personnes lorsque les données concernées sont traitées à des fins de recherche ou de statistique (telles que définies à l'article 5 (1) (b) RGPD). C'est également le cas lorsque la recherche est menée par une entité privée, comme une entreprise de neurotechnologie grand public. Étant donné que les données cérébrales sont en grande partie traitées par des entités qui se réclament de fins scientifiques et statistiques, cela signifie qu'une partie du traitement des données cérébrales par des acteurs publics comme par des acteurs privés (par exemple, des agences gouvernementales ou des entreprises de neurotechnologie grand public) inclut des exemptions aux principales règles de protection des données. En effet, il est difficile de déterminer avec certitude à quelles conditions le privilège d'utiliser les données cérébrales, dont jouit la recherche scientifique, s'appliquerait également aux données cérébrales collectées en dehors du domaine biomédical. C'est pourquoi Ienca et *al.* (2021) ont appelé à mettre en place une transparence plus grande sur les différentes finalités de la recherche, et à permettre aux individus d'intervenir dans la gestion ultérieure de leurs propres données au cas où elles soient utilisées à des fins indésirables (Ienca et *al.*, 2021).

Quatrièmement, comme nous l'avons vu précédemment, les données cérébrales peuvent porter atteinte à un autre principe normatif fortement associé à la confidentialité des informations et à l'autonomie personnelle, à savoir la *limitation de la finalité*. Les catégories sensibles de données à caractère personnel (y compris les données de santé) ne peuvent être

collectées qu'à des fins spécifiques qui doivent être précisées au moment où la personne concernée donne son consentement. Par exemple, selon le RGPD, l'impératif de limitation de la finalité exige que les données à caractère personnel soient collectées pour des finalités spécifiques, explicites et légitimes, et ne soient pas utilisées par la suite d'une manière qui serait incompatible avec ces finalités (article 5, paragraphe 1, point b), du RGPD). Cependant, la mise en œuvre de cette exigence de limitation de la finalité dans le domaine des données cérébrales est particulièrement difficile : en effet, les neurotechnologies actuelles ne peuvent pas isoler en amont les données nécessaires à la finalité de la myriade de signaux cérébraux enregistrés par le dispositif (problème de rapport signal/bruit), qui comprend également des processus subconscients. Cela signifie qu'une grande quantité d'informations redondantes peut être collectée – dont la personne concernée n'a, dans une large mesure, pas conscience. Des mesures de sécurisation des données, comme la confidentialité différentielle, ont été proposées pour garder un certain équilibre entre le consentement d'une part, et l'intrication de finalités précises au sein d'un traitement global d'autre part. Toutefois, il est difficile d'adapter ces mesures à la limite de ré-identifiabilité prévue par la loi (Ienca et *al.*, 2021). En outre, les outils techniques de filtrage sélectif – comme des anonymiseurs attachés à l'interface cerveau-machine (Chizeck & Bonaci, 2014) et les systèmes se fondant sur l'identification par radiofréquence (RFID) pour identifier l'activités cérébrale en temps réel et de manière sécurisée (Ajrawi et *al.*, 2021), en sont à leurs premiers stades de développement. Compte tenu de ces facteurs, il peut être plus difficile, pour les individus, d'exercer un contrôle sur leurs données cérébrales que cela ne l'est pour d'autres catégories de données.

La prévention de la neurodiscrimination est un autre domaine où la couverture juridique est déficiente. Bien que « toute forme de discrimination fondée sur le patrimoine génétique d'une personne » soit interdite par la Convention du Conseil de l'Europe pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine (Convention d'Oviedo), aucun mécanisme de protection n'interdit explicitement la neurodiscrimination. Ceci est problématique, car le potentiel qu'ont les données cérébrales de révéler des caractéristiques sensibles est très élevé (Rainey, Bublitz, Maslen, & Thornton, 2019). Il deviendrait donc possible de discriminer des individus sur la base de leurs caractéristiques neurologiques et/ou psychologiques.

Enfin, Ienca et *al.* (2021) ont fait valoir que les garanties fournies par les réglementations sur la protection des données ne sont pas forcément bien applicables aux données prélevées au niveau d'un groupe. Cette incapacité à faire évoluer les mécanismes de protection de manière adéquate fait peser un double risque sur la vie privée des groupes. Tout d'abord (A), des tiers peuvent émettre des déductions sur les caractéristiques d'un groupe de personnes en se fondant sur une ou plusieurs spécificités inhérentes à leurs données cérébrales et partagées par tous les individus du groupe. Il peut s'agir, par exemple, d'un temps de réaction plus lent que la moyenne à des tests cognitifs, ou bien d'une augmentation de l'activité cérébrale en fonction de l'adoption de certains comportements en ligne. De plus (B), à partir de leurs données cérébrales, même anonymisées, les individus peuvent être catégorisés à leur insu comme faisant partie d'un groupe jusque-là insoupçonné (Ienca et *al.*, 2021). Cela peut, là encore, entraîner une neurodiscrimination à leur encontre.

3.2. Introduction de la notion de « neurodroits ».

Comme nous l'avons vu précédemment, la recherche en neuroéthique s'est principalement concentrée sur des sujets tels que l'admissibilité éthique de l'amélioration cognitive grâce aux nootropes, les implications philosophico-juridiques de la neuroscience du libre arbitre, l'éthique de l'imagerie cérébrale (notamment concernant le déchiffrement de l'esprit), ainsi que la validité et l'admissibilité des preuves neuroscientifiques dans les tribunaux.

Depuis le début des années 2000, un nouveau domaine d'investigation neuroéthique et neurojuridique est apparu. Il a entrepris d'examiner les défis éthico-juridiques des neurosciences et des neurotechnologies au prisme des catégories normatives que sont les droits et les devoirs. Le travail de Boire et Sententia sur la notion de « liberté cognitive » (Boire, 2001 ; Sententia, 2004) est une étape pionnière dans cette direction. La liberté cognitive y est définie comme « le droit et la liberté de contrôler sa propre conscience et son mécanisme de pensée électrochimique » (Sententia, 2004, p. 227). Il convient de noter que ce champ d'investigation neuroéthique a émergé dans le prolongement des débats dominants susmentionnés en neuroéthique et en neurodroit. Boire, par exemple, a développé ses réflexions sur la liberté cognitive dans le contexte des débats qui ont cours sur l'éthique de l'imagerie cérébrale et de la lecture des pensées (Boire, 2001). De même, Sententia a développé sa définition et son analyse normative de la liberté cognitive en faisant le point sur le débat consacré, au sein de la réflexion neuroéthique, à l'amélioration cognitive (Sententia, 2004).

Le point de départ de leurs analyses, si on le compare aux travaux antérieurs menés sur la neuroéthique, est cependant différent : il est de nature théorico-normative. Les deux auteurs affirment ainsi que le concept de liberté cognitive doit être interprété, non pas comme une simple description neurophilosophique ou comme un *desideratum moral*, mais comme un « droit fondamental » (Sententia, 2004, p. 223). Ce point de vue est bien illustré par l'argument de Sententia selon lequel les progrès neurotechnologiques nécessitent une analyse de haut niveau qui parte des « droits individuels inscrits dans notre constitution démocratique », ainsi que par son affirmation selon laquelle la liberté cognitive « est le substrat nécessaire pour quasiment toutes les autres libertés » (Sententia, 2004, 227). Dans les années 2010, cette conception d'une liberté cognitive fondée sur les droits a été approfondie par l'analyse de Farahany portant sur le quatrième amendement de la Constitution des États-Unis et sur la clause d'auto-incrimination du cinquième amendement (Farahany, 2012). En outre, elle a été vigoureusement réaffirmée par la thèse de Bublitz selon laquelle le recours aux « interventions sur l'esprit en dehors des contextes thérapeutiques » doit pousser le droit à reconnaître la liberté cognitive (que lui appelle « autodétermination mentale ») comme un « droit humain fondamental » qui « garantit la souveraineté d'un individu sur son esprit » (J.-C. Bublitz, 2013). Bien qu'aucun des auteurs susmentionnés n'ait fait usage de ce terme, ce corpus d'études a jeté les bases du domaine de réflexion né à l'intersection de la neuroéthique et du neurodroit, et qui est aujourd'hui de plus en plus connu dans le grand public sous le nom de « neurodroits ». Ce domaine de réflexion a introduit un nouvel angle pour examiner les défis éthico-juridiques qui se posent dans les sciences de l'esprit et du cerveau, à savoir l'approche en termes de droits – droits juridiques comme droits au sens philosophique ou droits moraux –, de libertés et d'obligations.

Le terme « neurodroit » a été introduit pour la première fois par Ienca et Andorno en avril 2017 dans un article (Ienca & Andorno, 2017a) complétant leur analyse éthico-juridique

des droits de l'homme à l'ère des neurosciences et des neurotechnologies (Ienca & Andorno, 2017b). Cette analyse est parvenue à la conclusion que les droits de l'homme déjà existants, bien qu'ils soient nécessaires, pouvaient être normativement insuffisants pour répondre aux questions soulevées par la neurotechnologie. Pour cette raison, les auteurs ont fait valoir que « les possibilités ouvertes par les développements neurotechnologiques et leur application à divers aspects de la vie humaine nécessiteront une reconceptualisation de certains droits de l'homme, voire la création de nouveaux droits pour protéger les personnes contre des dommages potentiels » (Ienca & Andorno, 2017b). En d'autres termes, ils ont présenté les « neurodroits » comme une catégorie émergente de droits fondamentaux portant sur la protection du cerveau et de l'esprit d'une personne.

Ils ont notamment identifié quatre « neurodroits » qui, selon eux, pourraient revêtir une grande importance dans les décennies à venir : le droit à la liberté cognitive (qu'ils interprètent comme *Sententia et Bublitz*), le droit à la vie privée mentale, le droit à l'intégrité mentale et le droit à la continuité psychologique. Cet article a suscité un vif débat dans les médias et dans la communauté universitaire : de nombreux auteurs ont exprimé leur soutien à cette proposition des « neurodroits », mais il y eut aussi certaines voix dissidentes. Entre autres, Cascio a approuvé la proposition, mais s'est demandé en parallèle si les « neurodroits » devaient être considérés comme des droits légaux de l'esprit ou bien comme des droits légaux de la personne (Cascio, 2017). En outre, il s'est interrogé de manière critique sur les limites des « neurodroits », en particulier dans le cas des mineurs. À peu près au même moment, Pizzetti a soutenu, dans une lettre adressée à l'UNESCO, que les quatre droits neuronaux identifiés par Ienca et Andorno pourraient constituer les éléments constitutifs d'une « Déclaration universelle sur les neurosciences et les droits de l'homme » (Pizzetti, 2017). En revanche, Nawrot a critiqué la proposition et a émis des doutes sur la capacité des « neurodroits » à « réconcilier l'intrusion technologique dans notre château intérieur » (à comprendre au sens figuré comme faisant référence au cerveau et à l'esprit humains) avec le concept de « liberté de pensée » et le « fondement d'un État démocratique régi par le droit » (Nawrot, 2019). Sommaggio & Mazzocca ont également étudié plus avant la relation entre droits de l'homme et liberté cognitive (Sommaggio & Mazzocca, 2020). Ils en ont conclu que la notion de liberté cognitive fournissait le terrain conceptuel nécessaire à la construction d'une « déclaration des neurodroits de l'homme ».

Environ six mois plus tard, le débat sur les « neurodroits » a été relancé et amplifié par un article publié dans *Nature* par une équipe de vingt-cinq chercheurs dirigée par Rafael Yuste et Sara Goering (Yuste et al., 2017). Les auteurs ont identifié quatre domaines de préoccupation liés aux neurotechnologies et à l'IA : la vie privée et le consentement, l'agentivité et l'identité, l'augmentation, et enfin les biais et les préjugés. Pour chacun de ces quatre domaines de préoccupation, ils ont fait valoir que des « clauses protégeant ces droits (appelés neurodroits) » devraient être ajoutées aux traités internationaux (ivi). Cet article a eu une influence considérable sur l'opinion publique. En faisant passer le discours sur les « neurodroits » de l'analyse éthico-juridique à l'élaboration de politiques, cette proposition a eu un grand impact sur des réformes législatives menées au niveau national, notamment au Chili. Bien que la sémantique, la justification théorique et la caractérisation normative de ces droits n'aient pas été entièrement abordées dans le bref article original, cette proposition a été présentée plus en détail quelques années plus tard par le même groupe (Goering et al., 2021), ainsi que par Yuste,

Genser et Hermann (Yuste, Genser, & Herrmann, 2021). En outre, le travail de Yuste a abouti à la création de la *Neurorights Initiative* à l'Université de Columbia – le premier groupe de réflexion institutionnel sur les « neurodroits » –, puis, en collaboration avec des partenaires européens et nord-américains, à celle du *Neurorights Network*, premier réseau international de chercheurs travaillant sur les « neurodroits », et dont les membres se comptent actuellement sur quatre continents.

3.3. Définition des « neurodroits ».

Dans ce rapport, les « **neurodroits** » sont définis comme les principes éthiques, juridiques, sociaux ou naturels de liberté ou de droit qui touchent au domaine cérébral et mental d'une personne. Il s'agit donc des règles normatives fondamentales visant à protéger et préserver le cerveau et l'esprit humains. Par conséquent, l'étude des « neurodroits » constitue un sous-domaine de la recherche neuroéthique et neurojuridique. Comme l'indique cette définition, les « neurodroits » sont des droits complexes et protéiformes, qui sont généralement interprétés comme des droits moraux (c'est-à-dire des droits au sens philosophique) et comme des droits juridiques. L'étude des « neurodroits » est une voie d'investigation privilégiée pour savoir si le cadre existant des droits de l'homme est suffisant pour bien prendre en compte les questions posées par les neurotechnologies, ou si, au contraire, l'on doit envisager de créer de nouveaux droits de l'homme dédiés au champ neuro-cognitif afin de réguler les neurotechnologies.

3.4. Ancrer les « neurodroits » dans l'histoire des idées

Les « neurodroits » ne viennent pas de nulle part. Dans l'histoire de la philosophie et de la pensée juridique et politique, il est possible de voir dans plusieurs édifices conceptuels des antécédents historiques, et des fondements conceptuels des « neurodroits ». On peut identifier en particulier trois grandes familles conceptuelles : la liberté de pensée et de conscience, le droit à la vie privée, et le droit à l'intégrité mentale. Examinons-les en détail.

3.4.1. Liberté de pensée et de conscience

La thèse selon laquelle l'esprit humain et ses processus cognitifs sont libres est omniprésente dans l'histoire des idées. Bien que la liberté de pensée n'ait pas été particulièrement encouragée, et ait même été légalement persécutée pendant la majeure partie de l'histoire de l'humanité, une tradition culturelle visant à protéger cette liberté fondamentale traverse toute l'histoire de la pensée.

L'une des toutes premières traces de l'idée de liberté de pensée remonte à la dynastie Maurya, qui régnait sur une grande partie du sous-continent indien au III^e siècle avant notre ère. En particulier, au cours de la seconde moitié du siècle, l'empereur indien Ashoka le Grand émit des édits promouvant le respect de la « liberté de conscience » (Luzzatti, 2006). En Europe, pendant la démocratie athénienne, les tensions politiques continues entre démocratie et oligarchie ont conduit à l'émergence du terme de *parrhēsia*. Bien que la traduction littérale la plus proche en français moderne soit la « franchise », la *parrhēsia* a souvent été invoquée par les partisans de la démocratie comme synonyme de « liberté d'expression » pendant les

périodes où le risque d'oligarchie menaçait la démocratie, par exemple pendant la guerre du Péloponnèse (431-404 avant notre ère).

Au Ier siècle avant notre ère, Paul de Tarse a quant à lui réfléchi aux limites de la liberté de pensée. Dans sa première lettre aux Corinthiens, il se demande jusqu'à quel point la liberté [*eleutheria*] de quelqu'un devrait être jugée par la conscience [*suneideseos*] d'un autre (10:29, Collins & Harrington, 1999). Dans la philosophie chrétienne, la notion de liberté de pensée est souvent étroitement liée à la notion de *liberum arbitrium*, que l'on a l'habitude de traduire en français par « *libre arbitre* ». Cependant, alors que la liberté de pensée et de conscience constituait, dans cette conception, un principe normatif (souvent lié à un engagement politique en faveur de la tolérance religieuse), le concept de libre arbitre était à l'origine compris comme le constat descriptif qu'aucune nécessité ne pesait sur la volonté humaine. Cette définition descriptive du libre arbitre est enracinée dans la philosophie grecque de l'Antiquité tardive, en particulier chez les stoïciens. Par exemple, le philosophe stoïcien Épictète (50-vers 135 de notre ère), considérait comme un « fait que rien ne nous empêche de faire ou de choisir quelque chose » (Long, 2002).

Tout au long de la Renaissance, plusieurs concepts de nature similaire sont apparus. Par exemple, au XVIIe siècle, le pasteur et théologien puritain Roger Williams a inventé la notion de « liberté d'âme », à savoir l'idée que Dieu a doté les êtres humains du droit inné de choisir leur foi (Gaustad, 2001). Cette notion est ensuite devenue la « liberté de religion » ou « liberté de culte », actuellement protégée par l'article 18 de la DUDH. La Déclaration des droits des États-Unis contient également une garantie, dans le premier amendement, énonçant qu'il est interdit d'édicter des lois interférant avec la religion « ou qui en interdisent le libre exercice ».

Le poète John Milton, contemporain de Williams, a quant à lui utilisé l'expression « liberté de l'esprit » pour désigner le droit et la capacité des gens à protéger leur esprit des interférences extérieures (Milton, 1791). Milton a été parmi les premiers penseurs à introduire l'idée que l'esprit humain était le dernier refuge de la liberté personnelle et de l'autodétermination. Au XIXe siècle, cette idée a été reprise et développée par John Stuart Mill, qui a affirmé que « [s]ur lui-même, sur son propre corps et sur son esprit, l'individu est souverain » (Mill, 1859, p. 12). Comme l'a observé Martah Farah, cette souveraineté de l'individu sur lui-même se fonde sur « une intuition de la liberté individuelle [...] qui n'était même pas refusée aux prisonniers : la liberté de penser ses propres pensées et d'avoir sa propre personnalité » (Farah, 2002, p. 23). Enfin, au XXe siècle, la romancière Virginia Woolf a réaffirmé cette idée en écrivant : « Il n'y a pas de porte, ni de serrure, ni de verrou que vous puissiez placer sur la liberté de mon esprit » (Woolf, 1929). Cette conception de l'esprit comme lieu ultime de la liberté individuelle a eu une grande influence sur le débat concernant les « neurodroits ». Ainsi, Sententia s'est fait l'écho de cette tradition en affirmant que « le droit et la liberté de contrôler sa propre conscience et ses processus de pensée électrochimiques constitue le substrat nécessaire pour quasiment toutes les autres libertés » (Sententia, 2004). De même, Levy a déclaré que « si nous avons le droit à une sphère de liberté dans laquelle nous avons le droit de faire ce que nous voulons, notre esprit doit être inclus dans cette sphère » (Levy, 2007 ; 38, p. 179).

Comme on l'a indiqué précédemment, cette liberté de pensée au sens normatif est protégée par la Déclaration universelle des droits de l'homme (DUDH), juridiquement contraignante pour les États membres du Pacte international relatif aux droits civils et

politiques (PIDCP). Le droit à la liberté de pensée est explicité à l'article 18, qui stipule ce qui suit :

Toute personne a droit à la liberté de pensée, de conscience et de religion ; ce droit implique la liberté de changer de religion ou de conviction, ainsi que la liberté de manifester sa religion ou sa conviction, seule ou en commun, tant en public qu'en privé, par l'enseignement, les pratiques, le culte et l'accomplissement des rites (article 18).

La liberté de pensée est également protégée par l'article 9 de la Convention européenne des droits de l'homme (CEDH, 1950). Cependant, la CEDH donne une définition relativement restreinte de la liberté de pensée, en l'associant étroitement à la liberté de conscience et de religion.

En revanche, le Comité des droits de l'homme des Nations unies (CDH) a souligné que la portée du droit à la liberté de pensée est « vaste et profonde ; elle englobe la liberté de pensée sur tous les sujets » (Comité des droits de l'homme des Nations unies, 1993). En d'autres termes, bien que la DUDH établisse un lien entre liberté de pensée et liberté de religion, elle ne réduit pas la première à la seconde, car la liberté de pensée doit être considérée comme ayant une portée plus étendue et plus profonde. En outre, le CDH a précisé qu'il convient de bien distinguer la « liberté de pensée, de conscience, de religion ou de conviction » de la « liberté de manifester sa religion ou sa conviction » (Comité, 1993). En outre, il précise que la DUDH « n'autorise aucune limitation, quelle qu'elle soit, à la liberté de pensée et de conscience ou à la liberté d'avoir ou d'adopter une religion ou une conviction de son choix. Ces libertés sont protégées de manière inconditionnelle » (ivi).

Dans le débat sur les « neurodroits », Ienca et Andorno ont souligné à nouveau cette distinction entre liberté de pensée et liberté de manifester une pensée ou une croyance. Ils ont affirmé que la liberté cognitive protégeait la sphère de la pensée avant même que la pensée ne s'extériorise ou ne se manifeste par la parole, l'écriture ou le comportement. À ce titre, selon eux, la liberté cognitive est chronologiquement antérieure à toute autre liberté (Ienca & Andorno, 2017b) et complète la liberté d'expression, la liberté de la presse et la liberté de réunion.

Aux États-Unis, la protection de la liberté de pensée est fréquemment associée au premier amendement (Richards, 2015). Bien qu'il ne mentionne pas explicitement la liberté de pensée, les tribunaux américains ont explicitement fait référence à un « droit à la liberté de pensée du premier amendement » (*Doe vs City of Lafayette, Indiana*, 2003). La Cour suprême des États-Unis a quant à elle déclaré qu'« au cœur du premier amendement se trouve l'idée qu'un individu doit être libre de croire comme il l'entend » (*Abood vs Detroit Board of Education*, 1977).

La liberté de pensée a souvent été vue comme le précurseur et l'ancêtre d'autres libertés telles que la liberté de religion et la liberté d'expression. En vertu de cette nature primordiale, la liberté de pensée sert d'axiome pour de nombreuses autres libertés, qui ne conditionnent pas son fonctionnement ni son existence. Ce rôle fondamental de la liberté de pensée comme substrat des autres libertés a été reconnu, entre autres, par le juge Benjamin Cardozo de la Cour suprême des États-Unis, dont le raisonnement dans l'affaire *Palko contre Connecticut* (1937) était le suivant : « La liberté de pensée [...] est la matrice, la condition indispensable de presque toutes les autres formes de liberté. À de rares exceptions près, on peut retracer la reconnaissance de cette vérité qui irrigue notre histoire politique et juridique » (Polenberg, 1996). L'argument

de Sententia selon lequel la liberté cognitive doit être considérée comme le substrat de toutes les autres libertés peut s'inscrire dans cette tradition juridico-philosophique.

S'il semble évident que le droit à la liberté de pensée protège de manière adéquate les manifestations extérieures de la pensée, comme la pratique religieuse ou le changement de religion, on peut néanmoins se demander si ce droit est suffisamment large pour inclure la protection de la pensée au sens neuropsychologique du terme. Les dispositions légales protégeant la liberté de pensée semblent bien équipées pour protéger le « *locus externus* » – le comportement, les énoncés verbaux, les textes écrits ; mais elles semblent beaucoup moins bien équipées quand il s'agit de protéger le « *locus internus* » – les informations non exprimées, le langage silencieux, les intentions cachées, les préférences préconscientes par exemple.

Enfin, étant donné que des rapports récents indiquent qu'une surveillance neurologique forcée est pratiquée sur des travailleurs dans des installations en Chine, il devient d'autant plus important de considérer que les cadres des droits de l'homme devraient protéger la capacité des personnes à prendre des décisions libres et efficaces au sujet de la collecte et du traitement de leurs données cérébrales personnelles. La Convention européenne des droits de l'homme (CEDH), qui protège les droits à la vie privée et à la liberté de pensée (articles 8 et 9), peut, à ce titre, offrir un cadre conceptuel et normatif approprié pour empêcher les utilisations coercitives des neurotechnologies. Lorsque la Convention 108 du Conseil de l'Europe sur la protection des données aura été modernisée et entrera en vigueur, elle pourrait par exemple servir de base solide à des spécifications supplémentaires. Cependant, d'autres spécifications pourraient s'avérer nécessaires pour étendre la portée et l'objectif du droit à la liberté de pensée au-delà du seul respect de la liberté de conscience et de religion, et ce afin de pouvoir protéger tous les états cognitifs, affectifs et autres états mentaux, qu'ils soient exprimés ou non. La clause de l'article 9 de la CEDH, qui indique que « toutes les croyances reconnues sont protégées par ce droit », peut par exemple offrir une base appropriée pour cette caractérisation élargie. Toutefois, les croyances sont généralement considérées comme une catégorie spécifique d'états mentaux, car elles possèdent un contenu sémantique et des attitudes propositionnelles. Mais des représentations mentales impliquant des qualités sensorielles, comme les perceptions et les souvenirs épisodiques, ne sont généralement pas considérées comme des croyances. C'est également le cas de certaines attitudes propositionnelles comme les désirs : alors que les croyances tentent de représenter le monde tel qu'il est et qu'elles n'ont pas l'intention de le changer, les désirs sont formés de représentations de la manière dont le monde pourrait ou devrait être. En d'autres termes, les croyances constituent une catégorie trop étroite, qui s'avère donc insuffisante pour protéger l'ensemble de l'esprit. Pour cette raison, de futures réformes des droits de l'homme pourraient être nécessaires pour garantir que le droit à la liberté de pensée protège non seulement toutes les croyances, mais aussi tous les états mentaux qui ne sont pas des croyances.

3.4.2. Vie privée

Bien que le droit à la vie privée ait été en partie contenu dans les notions de liberté de pensée et d'autonomie personnelle, la première conceptualisation cohérente du droit moderne à la vie privée remonte à un article fondamental, publié en 1890 par Samuel Warren et Louis Brandeis. Dans cet article, la vie privée est définie comme « le droit d'être laissé seul » (Brandeis & Warren, 1890). Au moment de la rédaction de leur article, la principale

préoccupation de Warren et Brandeis concernait l'intérêt croissant de la presse pour les commérages et la révélation d'informations personnelles sur des individus sans leur consentement préalable. Ce cas spécifique de confidentialité a été inclus par Alan Westin et d'autres auteurs dans la notion plus large de « vie privée informationnelle », c'est-à-dire le contrôle des informations sur soi-même. Selon Westin, on peut définir la confidentialité de l'information comme le droit de chacun à déterminer pour lui-même quand, comment et dans quelle mesure ses informations personnelles doivent être communiquées à d'autres personnes (Westin, 1968).

Le droit international des droits de l'homme reconnaît formellement le droit à la vie privée. La Déclaration universelle des droits de l'homme (DUDH) stipule que « nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes » (article 12). De même, la Convention européenne des droits de l'homme (CEDH) de 1950 stipule que « toute personne a droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance » et précise que ce droit implique « la protection contre les écoutes téléphoniques, la collecte d'informations privées par les services de sécurité d'un État et les publications portant atteinte à la vie privée » (article 8).

Dans le monde numérique d'aujourd'hui, le droit à la vie privée prouve sa pertinence au vu de ce contexte de développement de méthodes complètement nouvelles de traitement de l'information, qui étaient impensables à l'époque de Warren et Brandeis, ou même des rédacteurs de la DUDH et de la CEDH. Parmi ces nouvelles méthodes, on trouve en particulier les techniques de traitement des données cérébrales, qui visent à révéler des informations concernant les processus mentaux ou la santé neurologique d'une personne. Ces techniques comprennent d'une part l'analyse prédictive de données neuronales primaires, comme les enregistrements cérébraux, et d'autre part les déductions établies à partir des données secondaires (par exemple, des données phénotypiques ou comportementales) grâce à des techniques comme l'informatique affective. Yuste et *al.* ont de fait affirmé qu'« un niveau extraordinaire d'informations personnelles peut déjà être atteint à partir des traits caractéristiques fournis par les données des personnes », et donc que « les citoyens devraient avoir la capacité – et le droit – de garder privées leurs données neuronales » (Yuste et *al.*, 2017). Sur la base de considérations similaires – avec une attention particulière portée aux failles de sécurité des neurodispositifs, à la nature des données neuronales et au potentiel déductif des techniques avancées d'analyse de données –, Ienca et Andorno ont proposé de réinterpréter le droit à la vie privée en le faisant évoluer. Ils ont ainsi proposé la reconnaissance d'un « droit à la vie privée mentale » qui protégerait explicitement les individus contre l'intrusion non consentie de tiers dans leurs données mentales (qu'il s'agisse de données cérébrales ou de données indicatives pourvoyeuses d'informations neurologiques, cognitives et/ou affectives), ainsi que contre la collecte non autorisée de celles-ci (Ienca & Andorno, 2017b).

Un curieux précédent historique du droit à la vie privée mentale est par ailleurs rapporté par le philosophe du XVI^e et du XVII^e siècle Francis Bacon. Il a en effet relaté que la reine Elisabeth I^{ère} avait révoqué une loi de censure de la pensée à la fin du XVI^e siècle, parce qu'elle n'aurait pas apprécié d'« ouvrir des fenêtres dans les âmes et les pensées secrètes des hommes » (Brimacombe, 2000). Au début du XX^e siècle, l'historien John Bagnell Bury a

également mis en avant la relation qui existe entre intimité mentale et liberté de pensée. Dans son célèbre ouvrage intitulé *A History of Freedom of Thought*, il affirme ainsi qu'« un homme ne peut jamais être empêché de penser ce qu'il veut tant qu'il cache ce qu'il pense » (Bury, 1914, p. 1). Cela suggère que l'exercice du droit à l'intimité mentale, et donc la dissimulation de ses propres pensées, est nécessaire pour exercer pleinement son droit propre à la liberté de pensée.

Si on se réfère au contexte historique dans lequel il a été rédigé et adopté, il est normal que l'article 8 de la CEDH mentionne par exemple explicitement les écoutes téléphoniques, mais pas les défis que posent à la vie privée la technologie numérique (confidentialité numérique) et la neurotechnologie (vie privée neurologique et mentale). Cependant, le monde de 2021 est profondément différent de celui de 1950 en ce qui concerne la collecte, le traitement et le partage des informations : de nouvelles dispositions peuvent donc s'avérer nécessaires pour protéger la vie privée dans ce contexte d'évolution technologique constante. La Convention d'Oviedo, en revanche, offre une base suffisamment large et libre de référence technologique pour réglementer le respect de la vie privée mentale. Elle stipule en effet que « toute personne a droit au respect de sa vie privée en ce qui concerne les informations relatives à sa santé » (article 10(1)). Toutefois, la portée de cet article de la Convention paraît malgré tout encore trop limitée, car il protège uniquement les informations concernant la santé, et pourrait donc échouer à protéger de manière adéquate les données mentales ou neuronales traitées dans un cadre non sanitaire – comme c'est le cas, par exemple, des informations sur les intentions cachées, ou bien sur les convictions personnelles.

3.4.3. Intégrité mentale

Si la liberté de pensée protège le cerveau et l'esprit humains contre toute ingérence extérieure importune, et si le droit à la vie privée protège les informations personnelles (y compris les informations mentales) contre toute intrusion extérieure, il existe en parallèle d'autres principes normatifs qui protègent également le cerveau et l'esprit humains contre les préjudices. Dans l'histoire des idées, le principe éthique de « non-malfaisance » constitue ainsi la construction conceptuelle la plus complète pour protéger l'intégrité d'une personne et empêcher les préjudices.

L'obligation morale « de ne pas nuire » (en grec ancien : ἐπι δὴλήσει δὲ καὶ ἀδικίῃ εἶρξεν) est déjà présente dans certaines versions anciennes du serment d'Hippocrate, et est largement citée dans toute la littérature déontologique médicale. Dans le livre *Épidémies* du Corpus hippocratique, il est ainsi écrit que : « Le médecin doit [...] avoir deux objectifs précis en tête en ce qui concerne la maladie, à savoir, faire du bien ou ne pas faire de mal » (livre I, sect. 11, trad. Adams, en grec : ἀσκέειν, περὶ τὰ νοσήματα, δύο, ὠφελέειν, ἢ μὴ βλάπτειν). Cette obligation morale a ensuite été reformulée dans la maxime latine « *primum non nocere* », c'est-à-dire « d'abord ne pas nuire »². Bien que le principe de non-nuisance soit ancré dans l'éthique de la médecine et de la recherche biomédicale, caractériser le préjudice n'est pas toujours dépourvu d'ambiguïté. La littérature sur l'éthique médicale a donc entrepris de classer les préjudices en fonction de leur ampleur, de leur gravité, de leur durée et de leur réversibilité

² Contrairement à ce que l'on croit généralement, l'expression latine « *primum non nocere* » n'est pas d'origine ancienne. Smith (2005) a pu remonter à Thomas Sydenham (1624-1689) en suivant un livre de Thomas Inman (1860) intitulé *Foundation for a New Theory and Practice of Medicine*. Voir : (Smith, 2005)

(Meslin, 1990). De plus, elle distingue différents types de préjudice en fonction de la sphère ou de la capacité personnelle qui a été affectée par l'acte malveillant : on peut à cet égard distinguer entre préjudices physiques, préjudices psychologiques et préjudices socio-économiques. Cependant, la séparation entre les dommages physiques et psychologiques est discutable, car elle présuppose une conception ontologique dualiste de la personne (corps vs esprit). En outre, on peut observer que les nouvelles formes de préjudice rendues possibles par les technologies émergentes rentrent difficilement dans cette classification (Favaretto, De Clercq, Gaab et Elger, 2020 ; Hayes, 2017). La prévention des préjudices psychologiques, par exemple des dommages causés par des abus psychologiques, est un exemple historique éminent précédant l'apparition des « neurodroits », en particulier du droit à l'intégrité mentale.

Historiquement, les expressions « intégrité mentale » et « intégrité psychologique » ont été utilisées, bien que de manière non systématique, pour désigner le pendant psychologique du principe d'intégrité physique (également appelée « intégrité corporelle »). L'intégrité physique désigne le principe normatif d'inviolabilité du corps humain, ainsi que la capacité et le droit de la personne d'exercer un contrôle autonome sur son corps. Cela étant, l'intégrité physique est une condition fondamentale de l'autonomie et de l'autodétermination.

Selon l'approche des capacités élaborée par Martha Nussbaum – une approche normative du bien-être humain, qui se fonde sur la capacité effective de la personne à atteindre son bien-être –, l'intégrité physique est bien plus qu'un droit : elle est aussi l'une des dix capacités principales recensées par la philosophe. En effet, celle-ci définit l'intégrité physique comme un ensemble de capacités dont font partie la capacité « de se déplacer librement d'un endroit à l'autre » et la capacité « d'être protégé des agressions violentes ».

L'intégrité physique est actuellement protégée par de nombreuses législations nationales, mais également par le droit international des droits de l'homme. Au niveau national, l'intégrité physique a été reconnue, entre autres, par les tribunaux de la République d'Irlande comme un « *enumerated right* » protégé par le mécanisme général de garantie des « droits personnels » contenu dans l'article 40 de la Constitution irlandaise. Au niveau international, la DUDH et le Pacte international relatif aux droits civils et politiques (PIDCP) protègent tous deux l'intégrité physique.

En revanche, les occurrences du droit à l'intégrité mentale dans l'histoire des idées sont plus rares. Comme l'a observé Douglas, « contrairement au droit à l'intégrité physique, le droit à l'intégrité mentale ne bénéficie d'aucune tradition philosophique significative » (Douglas, 2014). Exception notable, Welford a utilisé, au début des années 1970, la notion d'intégrité mentale comme un critère de démarcation pour délimiter la frontière éthique entre obligation d'offrir un traitement maintenant en vie, et acharnement thérapeutique déraisonnable, notamment chez les patients en phase terminale, les personnes séniles et les nourrissons gravement déficients (Welford, 1970).

Le droit à l'intégrité mentale, au même titre que le droit à l'intégrité physique, est assuré par la Charte des droits fondamentaux de l'UE (CDF), dont l'article 3 stipule que « toute personne a droit au respect de son intégrité physique et mentale ». La Charte se concentre en particulier sur trois exigences : le consentement libre et éclairé, la non-commercialisation des éléments du corps humain, et l'interdiction des pratiques eugéniques et du clonage reproductif humain. En outre, elle promeut un droit d'accès aux services de santé mentale. Cependant, il n'y a aucune référence explicite aux pratiques liées à la neurotechnologie, ni aux préjudices

spécifiques que peut causer une interférence malveillante dans le domaine neuropsychologique d'une personne. Surtout, comme l'a noté Douglas, « très peu de travaux ont été réalisés pour déterminer [...] quelles sortes d'influence mentale entrent dans son champ d'application » (Douglas, 2014).

L'intégrité mentale est également protégée par la Convention relative aux droits des personnes handicapées (CDPH), un traité international des Nations unies sur les droits de l'homme visant à protéger les droits et la dignité des personnes en situation de handicap. Contrairement au CFR et à la CDPH, la CEDH ne fait quant à elle aucune référence explicite à la protection de l'intégrité physique et/ou mentale.

Tout comme il est difficile de séparer complètement préjudices physiques et psychologiques, tracer une ligne de séparation nette entre intégrité physique et intégrité mentale est une tâche conceptuellement ardue, voire absurde dans la mesure où les fonctions et les facultés mentales tirent leur cause de processus physiques. C'est pourquoi il faudrait plutôt se tourner vers des tentatives de préserver et de respecter l'intégrité de l'être humain de manière plus holistique (c'est-à-dire en englobant l'intégrité physique et mentale dans un même ensemble), qui sont plus à même de surmonter ce dualisme. L'article 1 de la Convention d'Oviedo peut offrir une base appropriée pour une telle approche holistique, puisqu'il stipule que la Convention « protège la dignité et l'identité de tous les êtres humains et garantit à chacun, sans discrimination, le respect de son intégrité ».

En outre, le droit à l'intégrité mentale est intimement lié au droit à la vie privée, car la protection de la vie privée comprend l'intégrité physique et psychologique d'une personne. L'intégrité mentale présente également des affinités avec les principes normatifs régissant la protection des personnes atteintes de troubles mentaux. En particulier, l'article 7 de la Convention d'Oviedo du Conseil de l'Europe (« Protection des personnes atteintes de troubles mentaux ») définit les conditions dans lesquelles les personnes atteintes de troubles mentaux peuvent ou non être soumises à une intervention sans leur consentement. L'intégrité mentale peut donc être envisagée dans la continuité de l'article 7 de la Convention d'Oviedo, et semble bien adaptée pour protéger également les personnes atteintes de troubles mentaux.

En ce qui concerne le risque de discrimination, si la discrimination génétique est déjà interdite, entre autres, par la Convention d'Oviedo du Conseil de l'Europe (article 11) et par la loi américaine de 2008 sur la non-discrimination en matière d'information génétique, il n'existe aujourd'hui aucune garantie explicite contre la neurodiscrimination. Par conséquent, une reconceptualisation éventuelle du droit devrait chercher à délimiter le champ légitime de la manipulation d'un procédé neuronal, à empêcher les interventions sur le cerveau qui causeraient des dommages mentaux, et à interdire les pratiques de traitement des données cérébrales pouvant entraîner une neurodiscrimination.

À côté des objectifs civils, il faut également, pour assurer la protection de l'intégrité mentale, définir les limites de la modulation de l'activité cérébrale à des fins militaires. Le droit de la guerre qui s'applique lors des conflits armés (appelé *jus in bello* ou droit humanitaire international) ne protège pas explicitement les militaires contre la violation de leur intégrité mentale. Il est nécessaire de rédiger de nouveaux principes normatifs, semblables à ceux qui régissent les armes autonomes, afin de protéger les soldats contre une utilisation offensive et abusive de la neurotechnologie en temps de guerre comme en temps de paix. C'est d'autant plus important que les progrès récents des applications militaires de la neurotechnologie et de

l'IA ouvrent la perspective d'une application militaire effective de la neurotechnologie (Rickli & Ienca, 2021)

3.4.4. Identité personnelle

En philosophie, et plus particulièrement en philosophie de l'esprit, on définit généralement l'identité personnelle comme l'identité unique d'une personne, celle-ci étant généralement considérée comme un sujet de conscience qui perdure au fil du temps. On peut donc décrire l'identité personnelle comme un ensemble de propriétés fondamentales qui définissent une personne en tant qu'individu ou qui font d'elle la personne qu'elle est en la distinguant des autres. On identifie subjectivement ces propriétés (par exemple, par introspection) comme étant celles pour lesquelles le sujet éprouve un sentiment particulier d'attachement ou de propriété. Pour cette raison, l'identité personnelle est plus subjective que d'autres formes d'identité, par exemple l'identité nationale.

La notion d'identité personnelle présuppose souvent le concept de personnalité, c'est-à-dire la jouissance du statut de personne. La plupart des philosophes définissent le statut de personne comme la possession d'un certain ensemble de propriétés mentales (Baker, 2000). On peut exprimer cette conception de l'identité d'une personne fondée sur les propriétés mentales par la phrase suivante : « Nécessairement, le sujet X est une personne au moment *t* si et seulement si les propriétés *x*, *y*, *z* s'appliquent ». Selon certains auteurs, il n'est peut-être même pas nécessaire de posséder ces propriétés mentales au moment *t*, car il suffirait d'être capable d'acquérir ces propriétés (Chisholm 1976 : 136f.), ou d'appartenir à un groupe dont les membres possèdent habituellement ces propriétés (Wiggins 1980 : ch. 6). Bien qu'il y ait un accord généralisé sur la pertinence des propriétés mentales pour définir l'identité personnelle, il existe un désaccord important autour de la détermination des propriétés mentales qui seraient spécifiquement constitutives de l'identité personnelle. Parmi les candidats potentiels, on trouve par exemple : la conscience de soi, la proprioception, ou encore la capacité de souffrir (Garrett, 2002 ; Noonan, 2019 ; Price-Williams, 1957).

En plus d'être composée de certaines propriétés mentales, l'identité personnelle se caractérise également par le fait de persister dans le temps (Dainton & Bayne, 2005 ; Noonan, 2019). En effet, la plupart des gens se perçoivent comme conservant la même identité individuelle au fil du temps, malgré le remplacement continu des cellules corporelles et les nouvelles expériences de vie qu'ils acquièrent. La question de la persistance de l'identité personnelle est abordée par les théories de la continuité psychologique de l'identité (Schechtman, 1994). Ces théories définissent l'identité personnelle comme composée d'une superposition de chaînes de connexions psychologiques causées de manière appropriée. Ces connexions psychologiques sont, par exemple, des souvenirs ou d'autres états cognitifs ou affectifs, comme une intention et l'action réalisée par cette intention.

L'une des premières contributions aux théories de la continuité psychologique est celle du philosophe du XVIIe siècle John Locke, pour qui une personne est « un être intelligent et pensant, qui a une raison et une réflexion, et qui peut se considérer comme lui-même, la même chose pensante, *en différents temps et lieux* » (1975 : 335). Cela signifie qu'une personne est une personne si et seulement si elle a la capacité de se considérer elle-même dans une continuité avec son moi passé, et de conserver les propriétés mentales constitutives de son identité malgré le passage du temps et les changements de lieux. Selon la plupart des contributions aux théories

de la continuité psychologique, la persistance désigne une certaine relation psychologique entre les propriétés mentales passées, présentes et futures. Et donc à tout moment, une personne hérite de propriétés mentales telles que les croyances, les souvenirs et les préférences de son moi passé.

Ienca et Andorno ont soutenu que protéger cette chaîne des propriétés mentales constitutives de l'identité d'une personne est nécessaire si l'on veut protéger la dignité et les droits fondamentaux de cette personne (Ienca & Andorno, 2017). En particulier, ils ont fait valoir qu'il convient de disposer d'un droit à la **continuité psychologique** afin d'empêcher que des interventions extérieures auxquelles la personne n'a pas consenti ne causent des ruptures indésirables dans la chaîne des états mentaux constitutifs de l'identité de cette personne (comme cela pourrait être le cas, par exemple, dans le cadre d'une neurostimulation forcée).

Dans la théorie juridique, le droit à l'identité personnelle est défini comme le droit de chacun de se forger une identité individuelle, de développer une conscience et de protéger cette identité et cette conscience individuelles contre toute limitation, manipulation ou effacement extérieurs (Marshall, 2014). Selon cette définition, le droit à l'identité personnelle est de fait intimement associé au droit à la vie, car ce n'est qu'en existant que les individus peuvent cultiver leur identité. Ce droit est reconnu en droit international par une série de déclarations et de conventions. Ainsi, la Cour européenne des droits de l'homme (CEDH) a interprété l'article 8 de la Convention européenne des droits de l'homme de manière à ce qu'il inclue « l'identité personnelle » dans la notion de « vie privée », dont la protection contre les intrusions non désirées est explicitement garantie³. Dans un autre arrêt (*Bruggemann et Scheuten contre Allemagne*), la CEDH a souligné l'importance pour l'identité personnelle des relations concernant le « champ émotionnel » et « le développement de la personnalité ».

3.5. Une taxonomie conceptuelle des « neurodroits »

Alors que la visibilité des débats sur les « neurodroits » dans l'opinion publique est cruciale pour garantir l'engagement public ainsi qu'une participation démocratique aux processus délibératifs sur ces questions, la rareté relative de ces débats au sein de la littérature académique nourrit des risques d'ambiguïté sémantique, d'ambiguïté normative et de confusion conceptuelle⁴. Ces risques sont exacerbés par la présence de multiples terminologies, qui, en outre, ne sont pas toujours conciliables. Plus important encore, plusieurs questions méta-éthiques, normatives et juridiques doivent toujours être résolues. Pour ces raisons, nous tenterons dans cette section de fournir une classification systématique des « neurodroits » qui ont été proposés jusqu'à ce jour. Enfin, dans la section suivante, nous discuterons des principales questions conceptuelles encore ouvertes. Nous pouvons d'ores et déjà identifier au moins cinq familles thématiques de « neurodroits » en fonction des principes éthiques normatifs dont ils découlent : les « neurodroits » dérivés de la liberté de pensée ; ceux dérivés de la protection de la vie privée ; ceux dérivés de la protection de l'intégrité mentale ; ceux dérivés de l'identité personnelle ; et enfin, les principes éthiques annexes.

³ Voir : *Goodwin c. Royaume-Uni* (2002) 35 EHRR 18, p. 90.

⁴ Une recherche du terme de « *neurorights* » dans le moteur de recherche Google a donné plus de 22 000 résultats. La même recherche effectuée dans Google Scholar a permis de trouver un peu plus d'une centaine d'entrées (état : juin 2021).

3.5.1. Famille thématique : la liberté de pensée

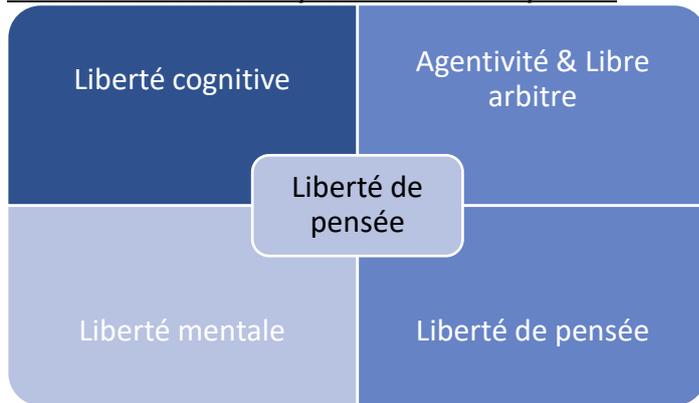


Figure 4 - Liberté de pensée et concepts associés

« neurodroits ». Malgré quelques différences conceptuelles mineures dans l'interprétation de ce droit, il existe un consensus général dans la littérature pour dire que la liberté cognitive implique le contrôle autonome et sans entrave d'une personne sur son esprit. Ce point de vue est bien illustré par l'utilisation par Bublitz de la liberté cognitive comme synonyme d'« autodétermination mentale » (Bublitz, 2013). Selon Bublitz (2013), ce droit comprend deux principes fondamentaux et intimement liés : (a) le droit des individus à utiliser les neurotechnologies émergentes ; (b) la protection des individus contre une utilisation coercitive et non consentie de ces technologies. En d'autres termes, la liberté cognitive est le principe qui garantit « le droit de modifier ses états mentaux à l'aide de dispositifs neuronaux, ainsi que le droit de refuser de le faire » (Bublitz, 2013, p. 234). De manière analogue, Ienca et Andorno ont souligné que la liberté cognitive était un « droit complexe qui doit reposer sur les prérequis des libertés négatives et positives » comprises dans le sens que leur donne Isaiah Berlin (Berlin, 1969). On trouverait donc : la liberté négative de faire des choix au sujet de son propre domaine cognitif sans être restreint ou empêché par des obstacles, des barrières ou des interdictions extérieures ; la liberté négative d'exercer son propre droit à l'intégrité mentale sans contraintes ou violations extérieures de ce droit ; et enfin, la liberté positive d'avoir la possibilité d'agir pour prendre le contrôle de sa vie mentale (Ienca & Andorno, 2017b). Farahany a ajouté une strate phénoménologique à ces définitions, puisqu'elle a défini la liberté cognitive comme « le droit à l'autodétermination sur nos cerveaux et sur nos expériences mentales » (Farahany, 2019, p. 97). En outre, elle a également fourni l'analyse la plus détaillée et la plus complète à ce jour de l'éventail des libertés et des droits englobés par ce droit. Selon elle, il s'agit notamment de « la liberté de pensée et de rumination, du droit à accéder à soi-même et à s'altérer, et du droit de consentir à des modifications de notre cerveau et de nos expériences mentales ou bien de les refuser » (p. 98). Enfin, elle souligne que le droit à la liberté cognitive, comme tous les intérêts individuels, « n'est pas absolu, mais doit être mis en balance avec les coûts sociétaux qu'il occasionne » (ivi).

S'il y a un accord global sur les prémisses fondamentales de la liberté cognitive, il existe cependant un désaccord quant à son domaine d'application. La plupart des définitions, y compris celle de Bublitz rappelée ci-dessus, limitent le champ d'application de la liberté cognitive aux seules altérations des états mentaux « induites » par les « dispositifs neuronaux » ou les « neurotechnologies ». Dans le même article, Bublitz propose une définition encore plus

Au sein du débat sur les « neurodroits », quatre droits conceptuellement dérivés de la liberté de pensée, ou du moins associés à celle-ci, ont été proposés. Il s'agit du droit à la liberté cognitive, du droit à l'agentivité et au libre arbitre, du droit à la liberté mentale et du droit à la liberté de pensée elle-même.

Comme nous l'avons vu précédemment, la liberté cognitive a été un précurseur du débat sur les

étroite de la liberté cognitive : elle devrait selon lui se limiter au droit d'utiliser la neurotechnologie à des fins d'amélioration cognitive (p. 233). Cette définition semble donc exclure toutes les altérations d'états mentaux qui n'amélioreraient pas les fonctions cérébrales (par exemple, celles qui les diminuent ou qui provoquent des changements qualitatifs plutôt que quantitatifs). En revanche, et en s'inspirant de travaux antérieurs sur la manipulation en ligne (voir Susser *et al.*, 2018), Ienca et Vayena ont proposé une définition plus large, qui englobe également les altérations involontaires – qu'elles entraînent une amélioration, une diminution ou une absence de modification des fonctions cérébrales – des états mentaux induites par des procédés non neurotechnologiques, comme les réseaux sociaux et la manipulation en ligne (Ienca & Vayena, 2018). De même, Farahany (2019) a proposé un large champ d'application de la liberté cognitive au droit de la responsabilité civile. Ainsi, comme la nature et la variété des technologies capables de décoder et/ou de modifier le cerveau humain est en constante évolution, les définitions de la liberté cognitive exemptes de mention à des technologies précises semblent préférables à celles qui en font mention.

Yuste *et al.* ont quant à eux défendu un « droit à l'agentivité, soit la liberté de pensée et le libre arbitre de choisir ses propres actions » (Yuste *et al.*, 2021). Bien que ces auteurs utilisent ces trois notions comme synonymes, comme l'indique la conjonction de coordination disjonctive « ou », l'agentivité, la liberté de pensée et le libre arbitre désignent ordinairement des concepts bien distincts. L'agentivité, telle qu'elle est abordée dans la littérature de la philosophie de l'action, désigne l'exercice ou la manifestation de la capacité d'agir d'un agent (Gallagher, 2007 ; Proust, 2013). Le libre arbitre, comme nous l'avons vu, est une thèse ontologique désignant la capacité des agents à choisir sans entrave entre différents modes d'action (Dennett, 2015 ; Spence, 1996). En d'autres termes, l'agentivité appartient au domaine de l'action ; mais le libre arbitre, en revanche, appartient au domaine de la cognition, et plus particulièrement à la prise de décision. Plus important encore, l'agentivité et le libre arbitre sont habituellement conçus comme des capacités ou des dispositions. Ils sont de nature descriptive, et non normative (Aarts & van den Bos, 2011). Pour déduire une normativité de ces énoncés descriptifs, il faut faire découler des droits et des obligations à partir des capacités et des dispositions contenus dans leur définition. Le fonctionnement logique d'une telle opération reste cependant peu clair encore aujourd'hui. Enfin, comme l'observe Munoz, « le libre arbitre est un concept porteur de multiples dimensions, qui pose plusieurs problèmes philosophiques non résolus » (Munoz, 2019).

La notion de *liberté mentale* est quant à elle rarement utilisée dans la littérature. Repetti l'a mentionnée pour exposer sa théorie bouddhiste du libre arbitre (Repetti, 2018). Bublitz fait un usage spécifique de la liberté mentale (qu'il appelle également « liberté d'esprit ») dans le contexte des « neurodroits » : il la décrit en effet comme le « contrôle conscient de son esprit » (C. Bublitz, 2016). Il a avancé que la liberté mentale devrait être classée parmi les libertés juridiques et politiques les plus importantes (ivi). Cependant, il est difficile de savoir si son interprétation de la liberté mentale équivaut à la liberté cognitive ou si elle s'en distingue conceptuellement.

Enfin, certains auteurs ont avancé que la notion même de liberté de pensée offrait un terrain normatif approprié pour répondre aux défis que soulèvent les nouvelles neurotechnologies en matière de droits de l'homme (Lavazza, 2018). Adopter la liberté de pensée comme fondement normatif du contrôle autonome d'une personne sur son esprit

présente l'avantage d'être économe en concepts, suivant ainsi le principe du rasoir d'Ockham, ou principe de parcimonie, qui stipule qu'« on ne doit pas multiplier les éléments si ce n'est pas nécessaire » (Schaffer, 2015). La liberté de pensée étant déjà inscrite dans le droit international des droits de l'homme et largement discutée en philosophie du droit, il serait, toutes choses étant égales par ailleurs, plus économe d'adopter cette terminologie normative au lieu de multiplier les éléments normatifs comme la liberté cognitive, la liberté mentale et les droits à l'agentivité et au libre arbitre. Si cette approche économe devait être poursuivie, il conviendrait toutefois de préciser que « la protection de l'autodétermination d'une personne sur son esprit devrait inclure l'ensemble du for intérieur » (C. Bublitz, 2015), c'est-à-dire tous les états mentaux ou les capacités mentales, dont les phénomènes cognitifs, émotionnels et conatifs conscients ou inconscients. Comme l'ont souligné Ienca et Andorno, la liberté de pensée constitue la justification fondamentale des libertés qui lui sont associées, à l'instar de la liberté de choisir, de la liberté d'expression, de la liberté de la presse et de la liberté de religion. Une interprétation évolutive de ce droit devrait donc se concentrer non seulement sur la protection des manifestations extériorisées de la pensée, mais aussi de la pensée elle-même.

On peut émettre, cependant, une objection contre cette conceptualisation élargie. En effet, la notion de liberté de pensée, comme nous l'avons observé précédemment et telle qu'elle est inscrite dans la DUDH et la CEDH, est historiquement mêlée aux notions de liberté de conscience et de religion. Son objectif initial et sa fonction historique pourraient donc être inconciliables avec la nécessité contemporaine de protéger le *forum internum* et de garantir la liberté des personnes de s'autodéterminer dans le contexte de leur propre esprit. C'est pourquoi certains auteurs ont avancé que la liberté de pensée et la liberté cognitive n'étaient pas simplement des dénominations différentes, mais bien des droits conceptuellement distincts. Par exemple, Farahany a affirmé que la liberté cognitive était un concept plus large que la liberté de pensée, car elle englobait également, en plus de la liberté de pensée, les droits d'accéder à soi-même et de s'altérer, ainsi que le droit de consentir à des modifications de notre cerveau et de nos expériences mentales ou de les refuser (Farahany, 2019). De ce point de vue, bien qu'économe sur le plan conceptuel, assimiler la liberté cognitive à la liberté de pensée serait injustifié. Par conséquent, l'introduction de nouveaux droits comme le droit à la liberté cognitive ne serait pas une simple prolifération inutile d'éléments normatifs, mais une clarification réellement nécessaire du cadre des droits de l'homme à la lumière de l'évolution technologique.

3.5.2. Famille thématique : la vie privée

Contrairement aux « neurodroits » dérivés de la liberté de pensée, les « neurodroits » issus du droit à la vie privée semblent se caractériser par un niveau de consensus conceptuel et terminologique beaucoup plus important. On utilise

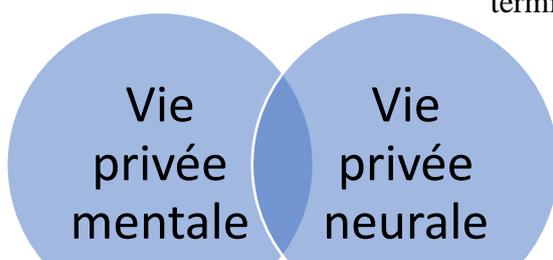


Figure 1 Vie privée mentale & Vie privée neurale

ordinairement l'expression de **vie privée mentale** pour désigner le droit des personnes à être protégées contre l'intrusion non consentie de tiers dans leurs données mentales, ainsi que contre la collecte et le traitement non autorisés de ces données (Ienca & Andorno, 2017a, 2017b ; Shen, 2013 ; Yuste

et *al.*, 2021). Yuste et *al.* ont ainsi avancé que le respect de la vie privée mentale n'était pas seulement un droit, mais aussi une capacité, celle « de garder ses pensées protégées contre leur divulgation » (Yuste et *al.*, 2021). Le lien entre vie privée mentale et droit général à la vie privée est néanmoins discutable. Ienca et Andorno soutiennent par exemple que la nature particulière des informations cérébrales nécessite qu'on ajoute des mesures spécifiques aux cadres actuels de protection de la vie privée. Cela est dû, selon eux, au fait que les données cérébrales touchent directement à la vie mentale intérieure et à la personnalité d'une personne, et qu'elles sont obtenues d'une manière bien particulière. Selon eux, la vie privée mentale doit protéger les ondes cérébrales car ce sont des données en elles-mêmes, mais aussi car elles sont génératrices de données et constituent de véritables sources d'information. Ils ont appelé ce défi le « problème de l'inception ». De plus, un droit à la vie privée mentale protégerait non seulement les données cérébrales conscientes, mais aussi les données qui ne sont pas (ou seulement partiellement) soumises à un contrôle volontaire et conscient. Enfin, il garantirait la protection des informations cérébrales en l'absence d'un outil externe capable d'identifier et de filtrer ces informations. Cela contribuerait à garantir le droit des personnes de voir leurs données cérébrales protégées contre tout accès non autorisé, et d'empêcher la fuite inconsidérée de ces données dans l'infosphère.

Des auteurs ont également débattu pour savoir comment faciliter l'insertion du droit à la vie privée mentale dans une réglementation contraignante. Ienca & Malgieri ont par exemple proposé d'élaborer une évaluation de l'impact de la protection des données mentales (MDPIA), à savoir un audit vérifiant les composants neurotechnologiques du traitement des données (par exemple, les algorithmes d'IA) et réexaminant l'algorithme au cas où certains risques pourraient être atténués simplement par un « changement de leur conception » (Ienca & Malgieri, 2021). Selon eux, les responsables du traitement des données mentales devraient être obligés de se conformer aux exigences suivantes :

- a) Description du traitement (dont une description du fonctionnement de la technologie en jeu) ;⁵
- b) Réalisation d'un équilibre entre les objectifs poursuivis, et la nécessité et la proportionnalité du traitement des données ;⁶
- c) Évaluation des risques réels pour les droits et libertés fondamentaux, et proposition de mesures appropriées pour atténuer ces risques.

Un autre concept fréquemment utilisé dans le droit des personnes à protéger les informations relatives à leur cerveau est celui de vie privée neurale. Alors que la « vie privée mentale » vise à protéger les informations mentales, quelle que soit la manière dont elles sont collectées ou déduites, la vie privée neurale concerne spécifiquement la protection des données neurales, également appelées neurodonnées (Hallinan, Schütz, Friedewald et de Hert, 2014 ; Ienca, 2015 ; Wolpe, 2017). La différence entre ces deux types de vie privée et de données n'est pas anodine. Les informations mentales sont des informations relatives à des états

⁵Margot E Kaminski et Gianclaudio Malgieri, « Algorithmic Impact Assessments under the GDPR : Producing Multi-Layered Explanations » [2020] International Data Privacy Law <<https://doi.org/10.1093/idpl/ipaa020>>, consulté le 1er février 2021.

⁶Dariusz Kloza et *al.*, « Data Protection Impact Assessment in the European Union : Developing a Template for a Report from the Assessment Process » (LawArXiv2020)DPiLabPolicy Brief <<https://osf.io/7qrfp>>, consulté le 1er décembre 2020.

mentaux comme les pensées, les souvenirs, les croyances, les perceptions et les émotions. Pour un tiers, acquérir une information mentale nécessite un certain niveau de décodage. Les informations neurales, en revanche, sont des informations sur le système nerveux. Bien que ces deux types d'informations se chevauchent largement, ils ne sont toutefois pas superposables. Il est ainsi possible d'accéder aux informations neurales sans accéder aux informations mentales, et *vice versa*. Par exemple, en accédant aux biomarqueurs de la maladie d'Alzheimer présents dans le cerveau d'une personne (par exemple, une mesure du dépôt de bêta-amyloïde grâce à la TEP de l'amyloïde), on accède aux informations neurales de cette personne. Toutefois, on n'a pas automatiquement accès aux états mentaux de cette personne. Et l'opposé est vrai : si on utilise des technologies non neurales, comme la reconnaissance des visages et l'IA des émotions, pour accéder à des informations sur l'état émotionnel d'une personne (par exemple, un sentiment de peur), on dispose d'un certain degré d'accès aux informations mentales. Cependant, l'accès aux informations neurales de cette personne n'est pas automatique. Par conséquent, la protection de la vie privée mentale et celle de la vie privée neurale semblent constituer des instances complémentaires, irréductibles à la protection de la vie privée dans le monde bio-numérique.

3.5.3. Famille thématique : l'intégrité mentale

On observe également une convergence conceptuelle relativement forte concernant l'intégrité mentale. Comme nous l'avons vu précédemment, ce droit à l'intégrité mentale est inscrit dans la Charte des droits fondamentaux de l'UE (article 3). Cependant, des divergences existent quant à son interprétation. Ienca et Andorno ont soutenu qu'il faudrait peut-être soumettre le droit à l'intégrité mentale à une interprétation évolutive afin de mieux sauvegarder le droit des personnes à être protégées contre des manipulations illicites et nuisibles de leur activité mentale. En revanche, Douglas a défini l'« intégrité mentale » comme « un droit contre l'interférence mentale » (Douglas, 2014). De même que Douglas, Lavazza a interprété l'intégrité mentale comme « le contrôle, par l'individu, de ses états mentaux et de ses données cérébrales de telle sorte que, s'il n'a pas donné son consentement, personne ne puisse lire, diffuser ou modifier ces états et ces données afin de conditionner l'individu de quelque manière que ce soit » (Lavazza, 2018). La différence conceptuelle est ici importante. Alors que Douglas et Lavazza interprètent l'intégrité mentale en lien étroit avec le droit à la liberté cognitive et le droit à la liberté de pensée, la définition de Ienca et Andorno établit une relation logique nécessaire entre intégrité mentale et protection contre les atteintes au domaine neuronal et/ou mental d'une personne. Dans le premier cas, il s'ensuit que l'intégrité mentale est subsumée par la liberté cognitive et la liberté de pensée ; dans le second cas, elle leur est complémentaire.

Les travaux de Douglas sur les interventions médicales visant à la réhabilitation des criminels peuvent nous aider à clarifier ce point. Selon lui, imposer, sans le consentement explicite du patient, des mesures médicales qui « sont destinées à avoir des effets mentaux » pourrait être considéré comme une violation du droit à l'intégrité mentale, dans la mesure où il s'agit d'une sorte d'interférence mentale non consentie, donc « une menace sérieuse pour l'agentivité » (Douglas, 2014). Selon la définition d'Ienca & Andorno, ce type de mesures médicales constitue dans tous les cas une violation de la liberté cognitive du bénéficiaire ; cependant, elles constituent une violation de l'intégrité mentale uniquement si elles causent un préjudice au sujet.

En plus de ces considérations conceptuelles, consacrer le droit à l'intégrité mentale nécessite de clarifier en amont la relation qu'entretient ce droit avec le droit à l'intégrité physique ou corporelle. Comme indiqué précédemment, tracer une ligne de séparation nette entre intégrité physique et intégrité mentale est une tâche absurde, qui s'enracine dans une ontologie dualiste (implicite ou explicite) de la relation corps-esprit. Puisque les fonctions et les facultés mentales tirent leur cause de processus physiques, les tentatives holistiques visant à préserver et respecter l'intégrité de l'être humain, qui surmontent ce dualisme, sont préférables. La recherche sur l'intégrité mentale devrait donc éviter les prémisses dualistes, et plutôt chercher à fournir un cadre global pour la protection de l'intégrité personnelle qui inclurait à la fois le corps humain, et les facultés et processus que ce corps contient.

L'article 1 de la Convention d'Oviedo constitue un bon exemple de cette approche holistique, comme nous l'avons déjà mentionné. De même, la Cour suprême irlandaise a déclaré, en 1965, qu'il existait un « droit de ne pas subir d'ingérence dans votre corps ou votre personnalité » (*Ryan vs Attorney General*). Selon cette déclaration, l'intégrité physique et mentale doit être considérée comme une condition préalable unique et nécessaire à l'identité personnelle. De la même manière, le quatrième amendement de la Constitution des États-Unis prévoit la protection du « droit des citoyens d'être garantis dans leur personne », ce qui suggère une conception de l'intégrité regroupant intégrité physique et intégrité mentale.

S'inspirant de commentaires précédemment formulés par d'autres pour motiver la reconnaissance de ce droit, Douglas et Forsberg (2021) ont identifié et exposé trois justifications distinctes pour la reconnaissance d'un droit légal à l'intégrité mentale : l'intuition, la cohérence justificative et la considération du développement technologique. Le premier item se réfère à l'intuition morale selon laquelle les personnes devraient être protégées contre les interférences mentales, en particulier contre les interférences « causant des souffrances mentales » (Bublitz & Merkel, 2014). La cohérence justificative désigne le point de vue selon lequel les justifications théoriques communes du droit à l'intégrité corporelle semblent également appuyer un droit à l'intégrité mentale. Enfin, la considération du développement technologique désigne le point de vue selon lequel les perturbations induites par les développements neurotechnologiques récents, et probablement futurs, exigent une protection accrue de la sphère mentale. Douglas et Forsberg reconnaissent, en conclusion, que chacune de ces justifications est sujette à débat. Néanmoins, ils réaffirment également que « chacun de ces candidats à la justification présente une certaine plausibilité, et justifie un examen plus approfondi » (Douglas & Forsberg, 2021).

3.5.4. Famille thématique : l'identité personnelle et la continuité psychologique

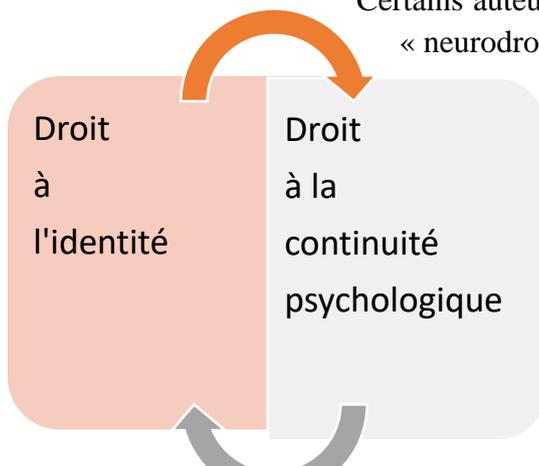


Figure 2 Identité & Continuité psychologique

Certains auteurs ont préconisé de reconnaître une quatrième famille de « neurodroits » : celle qui dérive de la protection de l'identité personnelle. Empruntant la terminologie aux études menées sur l'identité personnelle (Van Inwagen, 1997), Ienca et Andorno ont donné à ce droit le nom de droit à la « continuité psychologique » et l'ont décrit comme le droit de préserver « l'identité personnelle des personnes et la continuité de leur vie mentale de toute altération extérieure par des tiers qui ne serait pas consentie » (Ienca & Andorno, 2017a). Yuste et *al.*, en revanche, préconisent un « droit à l'identité », qu'ils décrivent comme « la capacité de contrôler à la fois son intégrité physique et son intégrité mentale » (Yuste

et *al.*, 2021). Alors que la continuité psychologique, dans sa formulation originale, a des affinités avec la liberté cognitive et la liberté de pensée, le droit à l'identité au sens de Yuste et *al.* semble être une condition préalable à l'intégrité physique et mentale. Par conséquent, ces deux droits semblent protéger des intérêts humains différents, bien que complémentaires.

Le droit à l'identité et le droit à la continuité psychologique offrent tous deux une base normative solide pour guider une intégration véritablement responsable de l'IA dans le contrôle des BCI, et pour préserver l'autodétermination et le sentiment d'identité personnelle d'une personne de toute manipulation subconsciente. Ces droits peuvent aider les utilisateurs de BCI à conserver le contrôle de leur comportement, sans éprouver de « sentiments de perte de contrôle », voire de « rupture » de leur identité personnelle (Gilbert et *al.*, 2017). Dans cette optique, Clausen et *al.* ont avancé que des « mécanismes de contrôle par veto » devraient être inclus dans les BCI pour protéger l'autonomie et l'identité des utilisateurs. Dans le même temps, le droit à la continuité psychologique vise à protéger contre les interventions non autorisées de tiers dans l'activité neuronale des utilisateurs de BCI. Ce droit pourrait, par exemple, protéger les individus contre des neuromodulations qu'ils n'auraient pas autorisées : même si ces interventions ne causent pas de blessures ou de traumatismes à l'utilisateur, elles pourraient néanmoins générer des modifications significatives au sein de la sphère psychologique de l'individu (modifier, par exemple, ses préférences politiques et/ou religieuses, ses états émotionnels ou ses souvenirs) sans le consentement explicite de ce dernier. Ce principe peut devenir particulièrement important dans un contexte de sécurité nationale et de recherche militaire, où des applications neurotechnologiques modulant les traits de personnalité (par exemple, les techniques de neurostimulation) sont actuellement testées en vue d'améliorer le combattant ou à d'autres fins stratégiques – par exemple, pour augmenter la capacité des soldats et autres personnels militaires à agir avec motivation et détermination même dans des situations de stress ou en l'absence de sommeil (Tennison & Moreno, 2012). Ces technologies de neuromodulation sont souvent mises en œuvre à l'aide de dispositifs hybrides capables, grâce à l'EEG, de déchiffrer les signes d'une baisse d'attention, puis de modifier ces processus par stimulation.

3.5.5. Principes éthiques annexes

Enfin, certains auteurs ont proposé de reconnaître des droits qui ne sont pas directement liés à la protection du domaine neuronal et mental, mais plutôt à la promotion des exigences socio-techniques nécessaires pour réaliser les droits susmentionnés. Deux de ces principes normatifs annexes sont ici proposés : le droit à un accès équitable à l'augmentation mentale, et le droit à la protection contre les préjugés algorithmiques. Le premier est défini par Yuste et *al.* comme « la capacité de garantir que les bénéfices de l'amélioration des capacités sensorielles et mentales issus des neurotechnologies soient distribués équitablement dans la population » (p.160-161) ; le second est défini par les mêmes auteurs comme « la capacité de garantir que les technologies ne comportent pas de préjugés » (ivi). En tant que tel, le droit à un accès équitable à l'augmentation mentale semble une condition préalable à la liberté cognitive comprise dans un sens positif. Ce principe peut être considéré comme normativement proche de l'article 3 de la Convention d'Oviedo, qui exige et promeut un accès équitable à des soins de santé de qualité appropriée. On peut toutefois se demander si l'amélioration des fonctions cognitives entre dans le champ des soins de santé. En revanche, le droit à la protection contre les biais et préjugés algorithmiques

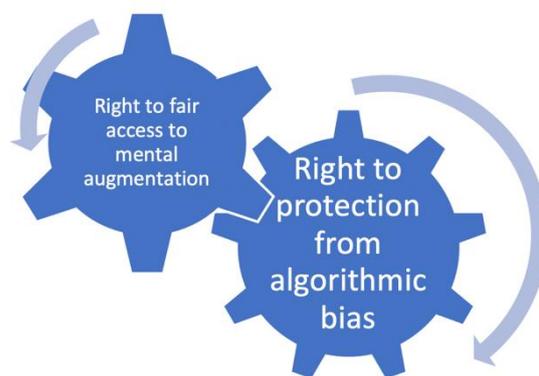


Figure 3 Principes éthiques annexes

semble une vraie condition préalable au droit à l'intégrité mentale, car il protège contre l'ensemble des préjudices générés par les biais et préjugés algorithmiques, et en premier lieu la neurodiscrimination algorithmique. Il convient de noter que, contrairement à tous les autres candidats au droit à l'intégrité mentale susmentionnés, le droit à la protection contre les biais algorithmiques peut être, et a été défendu dans des domaines qui n'ont pas de rapport avec la sphère mentale et/ou neurocognitive (Garcia, 2016). De plus, on pourrait avancer que la neurodiscrimination devrait être interdite dans tous les cas, que cette discrimination soit causée par un algorithme ou par un décideur humain. Par exemple, la Convention d'Oviedo interdit « toute forme de discrimination fondée sur le patrimoine génétique d'une personne ». De la même manière, les « neurodroits » devraient interdire toute forme de discrimination fondée sur les caractéristiques neuronales et/ou les états mentaux d'une personne.

Enfin, il est surprenant de constater que les « neurodroits » positifs, tels que le droit des personnes à accéder aux neurotechnologies et à les utiliser, n'ont été que rarement explorés. Par exemple, la Convention des Nations unies relative aux droits des personnes handicapées (CDPH) oblige les États à garantir et promouvoir la capacité des personnes en situation de handicap à vivre de manière indépendante et à participer à la vie sociale. Cela implique de soutenir la recherche et le développement d'appareils d'assistance, ainsi que l'accès des patients à ces appareils (voir l'article 4g de la CDPH). Les études futures devraient déterminer si ce droit s'étend également à l'utilisation des neurotechnologies chez les personnes en bonne santé à des fins telles que l'amélioration des fonctions cognitives, comme le permettrait une interprétation du droit à la liberté cognitive et du droit à un accès équitable à l'augmentation mentale comme des libertés positives.

Plus important encore, des recherches supplémentaires sont nécessaires pour caractériser la nature de « neurodroits » positifs comme la promotion du bien-être des patients sur la base du principe éthique de bienfaisance, qui ont jusqu'à présent occupé un rôle secondaire dans le débat sur les « neurodroits ».

4. Les « neurodroits » dans la politique internationale

En l'espace de quelques années, les « neurodroits » sont passés du domaine de la réflexion théorique à un sujet de débat de politique internationale. À ce jour, plusieurs acteurs gouvernementaux, intergouvernementaux et non gouvernementaux participant activement à la gouvernance des neurotechnologies ont inclus ou sont en train d'inclure les « neurodroits » dans leur agenda.

Une première étape importante dans l'introduction des « neurodroits » en tant qu'outils réglementaires a été franchie en 2019, lorsque le Conseil de l'Organisation de développement et de coopération économiques (OCDE) a adopté une *Recommandation sur l'innovation responsable dans le domaine des neurotechnologies* (OCDE-Conseil, 2019). Cette recommandation de l'OCDE est un instrument de droit souple conçu pour anticiper et relever les défis éthiques, juridiques et sociaux soulevés par les neurotechnologies émergentes, tout en favorisant l'innovation dans ce domaine. À ce titre, elle a établi la première norme internationalement acceptée en matière de gouvernance des neurotechnologies. La recommandation de l'OCDE est principalement axée sur la gouvernance responsable des acteurs de l'industrie des neurotechnologies. Elle contient néanmoins des dispositions sur trois catégories de « neurodroits » : la vie privée mentale (qui met l'accent sur la protection des données et des autres informations relatives au cerveau), l'intégrité mentale (qui met l'accent sur l'anticipation et la surveillance d'une éventuelle utilisation involontaire et/ou d'un mauvais usage de ces données) et la liberté cognitive.

L'OCDE n'est pas la seule organisation intergouvernementale qui place les « neurodroits » au cœur de ses stratégies de gouvernance. En 2020, le Conseil de l'Europe a lancé un plan d'action stratégique quinquennal centré sur les droits de l'homme et les technologies en biomédecine : ce plan contient un module sur l'évaluation de la pertinence et de la suffisance du cadre existant des droits de l'homme pour répondre aux questions soulevées par les applications neurotechnologiques. L'objectif de ce programme est d'évaluer si les questions éthiques et juridiques fondamentales soulevées par les neurotechnologies peuvent être traitées par « le cadre existant en matière de droits de l'Homme [...] ou si de nouveaux droits de l'Homme relatifs à la liberté cognitive, à la vie privée mentale, ou à l'intégrité mentale et la continuité psychologique doivent être envisagés pour encadrer les neurotechnologies ». Ce plan d'action stratégique s'inscrit dans la continuité du cadre réglementaire introduit par la Convention d'Oviedo, à savoir la prévention de l'utilisation abusive des innovations en biomédecine et la protection de la dignité humaine.

En parallèle, la Commission des questions juridiques et des droits de l'homme (rapporteur : M. Olivier BECHT, France) a publié un rapport intitulé *Les interfaces cerveau-machine : nouveaux droits ou nouveaux dangers pour les libertés fondamentales ?* (Rapport Doc. 15147 | 24 septembre 2020). La Commission a fait valoir qu'une approche calibrée de la réglementation de la technologie BCI était nécessaire, et qu'elle devrait englober à la fois des cadres éthiques et une réglementation juridique contraignante. Elle a donc appelé les États membres, les comités intergouvernementaux compétents du Conseil de l'Europe et le Comité des Ministres à prendre des mesures spécifiques à cette fin.

Les législateurs nationaux aussi sont actifs dans le domaine de la gouvernance des neurotechnologies. Le développement politique le plus important dans ce domaine est

l'approbation récente par le Sénat chilien d'une loi de réforme constitutionnelle qui définit l'intégrité mentale comme un droit humain fondamental, ainsi qu'une loi sur la neuroprotection qui protège les données du cerveau et applique l'éthique médicale existante, codifiée dans le code médical chilien actuel, à l'utilisation des neurotechnologies au sein de populations distinctes des seuls patients. En avril 2021, la Commission du Sénat chilien sur l'avenir a approuvé le texte final qui modifiait l'article 19 de la Constitution, et qui reconnaît donc les « droits à l'intégrité physique et mentale » et protège « l'activité cérébrale et ses données ». Cela fait du Chili « le seul pays disposant d'une proposition de loi et d'un amendement constitutionnel rendant obligatoire la protection neuronale et protégeant explicitement les neurodroits » (Yuste et *al.*, 2021). De même, le secrétaire d'État espagnol à l'IA a récemment publié une charte des droits numériques qui intègre les « neurodroits » aux droits des citoyens à l'ère nouvelle du numérique. Enfin, l'autorité italienne de protection des données a consacré en 2021 la Journée de la protection des données aux « neurodroits », et a reconnu leur nécessité, notamment celle du droit à la vie privée mentale et à l'intégrité mentale, si l'on veut traiter correctement les conséquences de la neurotechnologie sur les droits de l'homme, en particulier sur le droit à la vie privée.

5. Questions ouvertes et avenir des « neurodroits »

Bien que (ou peut-être précisément parce que) les « neurodroits » soient passés en relativement peu de temps du domaine de la réflexion éthico-juridique à celui du plaidoyer et de la politique, de nombreuses questions conceptuelles et pratiques restent sans réponse. La première question consiste à déterminer s'il faut interpréter les « neurodroits » comme des droits au sens philosophique (également appelés *droits moraux*), comme des droits juridiques, comme des droits de l'homme au sens du droit international des droits de l'homme, ou tout cela à la fois.

En particulier, Capron (1976) a observé que, bien qu'ils soient intimement liés, droits moraux et droits juridiques ne sont pas identiques, car il peut y avoir des circonstances dans lesquelles une personne a le droit moral « de contraindre le comportement d'une autre personne (ou autre), mais ne peut pas légalement le faire » (Capron, 1976). Plus radicalement, Thomson a soutenu que droits légaux et droits moraux sont deux territoires distincts dans le domaine des droits, parce qu'ils proviennent de sources différentes : alors que les droits moraux sont créés par la raison pratique (*phronesis*, en termes aristotéliens), les droits légaux sont uniquement créés par un système juridique et par le processus d'élaboration des lois (Thomson, 1990). La vue d'ensemble proposée dans ce rapport nous indique qu'il convient plutôt de considérer les « neurodroits » à la fois comme des droits au sens philosophique, c'est-à-dire des droits moraux, et comme des droits juridiques, c'est-à-dire des droits qui existent en vertu des règles des systèmes juridiques, ou en vertu de décisions d'organes faisant autorité au sein de ces systèmes. Ceci est cohérent avec la thèse de John Stuart Mill selon laquelle il existe un lien analytique étroit entre droits moraux et droits légaux, tous les droits étant dirigés vers la recherche d'un fondement au bien-être. Comme l'a constaté la théorie juridique antérieure, le fait de fonder les droits juridiques sur des droits moraux peut fournir de meilleures justifications à la raison pratique et à la politique (Hohfeld, 1913). Par conséquent, toute future déclaration internationale ou réforme législative sur la neurotechnologie et les droits de l'homme pourrait trouver un bénéfice à se fonder sur des « neurodroits » entendus comme des droits moraux.

Une deuxième question urgente consiste à déterminer si les « neurodroits » au sens du droit international des droits de l'homme doivent être interprétés comme de tout nouveaux droits de l'homme, ou comme des évolutions de droits déjà existants. Pour résoudre cette question, deux principes peuvent nous guider. Le premier, comme nous l'avons vu, est le rasoir d'Ockham ou la loi de l'économie, qui stipule qu'il ne faut pas multiplier les éléments si ce n'est pas nécessaire. Le second principe se réfère à « l'inflation des droits », c'est-à-dire la tendance à qualifier de « droit de l'homme » tout ce qui est moralement souhaitable, et postule qu'il faut éviter la prolifération injustifiée de nouveaux droits. La prolifération injustifiée de nouveaux droits de l'homme est en effet problématique : elle diffuse avec elle un certain scepticisme à l'encontre de l'ensemble des droits de l'homme dans la mesure où elle y inclut de simples *desiderata* ou des revendications purement rhétoriques. En d'autres termes, il faut éviter l'inflation des droits, car elle ne peut que diluer l'idée centrale des droits de l'homme et détourner l'attention de l'objectif central de leurs instruments, à savoir protéger un ensemble d'intérêts humains véritablement fondamentaux – et non pas protéger tout ce qui serait souhaitable ou avantageux dans un monde idéal (Ienca, 2021).

De ce point de vue, l'approche la plus économe reviendrait à considérer les « neurodroits » comme des interprétations évolutives de droits existants, tout en les soumettant à des examens pour savoir s'ils pourraient constituer de nouveaux droits de l'homme. Plusieurs tests de ce genre ont été proposés pour empêcher l'inflation des droits. Par exemple, Alston a proposé une liste de critères qu'une revendication normative devrait satisfaire afin de pouvoir être qualifiée de « droit de l'homme ». Selon lui, le nouveau candidat aux droits de l'homme doit (i) « refléter une valeur sociale fondamentalement importante » ; (ii) « être cohérent, mais pas simplement répétitif, vis-à-vis du corpus existant qui compose le droit international des droits de l'homme » ; (iii) « pouvoir atteindre un très haut degré de consensus international », et (iv) « être suffisamment précis pour donner lieu à des droits et à des obligations clairement identifiables » (Alston, 1984). De même, Nickel a avancé que tout nouveau candidat au statut de droit de l'homme devait non seulement (i) concerner un bien très important, mais aussi (ii) répondre à une menace commune qui pèserait gravement sur ce bien, (iii) imposer des obligations justifiables et proportionnées, et (iv) être réalisable dans la plupart des pays du monde (Nickel, Pogge, Smith, & Wenar, 2013). À la lumière de notre analyse des défis éthiques et des droits de l'homme soulevés par les neurotechnologies, il apparaît clairement que protéger l'espace cérébral et mental d'une personne possède une valeur sociale fondamentalement importante, vise à sauvegarder un bien très important et répond à une menace grave qui pèse sur ce bien.

La plupart des « neurodroits » proposés jusqu'à présent semblent par ailleurs imposer des obligations justifiables et proportionnées. En effet, aucune des normes examinées dans ce rapport ne semble limiter l'innovation neurotechnologique, ni peser de manière disproportionnée sur les acteurs de ces neurotechnologies, sur les chercheurs ou sur toute autre partie prenante. En outre, les « neurodroits » semblent compatibles avec le corpus existant du droit international des droits de l'homme : en effet, ils complètent les droits et libertés fondamentaux, mais, dans le même temps, ils introduisent des spécificités normatives afin de bien protéger le domaine mental de la personne, spécificités qui ne se contentent pas de répéter le cadre déjà existant des droits de l'homme. De plus, bien que l'accès aux neurotechnologies soit inégalement réparti à l'échelle mondiale, la protection des « neurodroits » concerne tous les pays du monde. Cependant, il est trop tôt pour déterminer si et quels « neurodroits » seront capables d'atteindre un haut degré de consensus international. L'approbation récente par le Sénat chilien d'un projet de loi sur la protection neuronale promouvant le droit à l'intimité mentale et le droit à l'intégrité mentale, ainsi que le consensus global autour de ces deux droits dans la recherche éthico-juridique, font de ces derniers des candidats sérieux pour être inclus dans le cadre international des droits de l'homme. Mais des recherches supplémentaires restent à mener pour évaluer si, et quels « neurodroits » sont suffisamment précis pour donner lieu à des droits et à des obligations clairement identifiables.

Comme en témoigne l'analyse ci-dessus, pour que le domaine des « neurodroits » progresse et ait un impact cohérent sur la politique internationale, il doit surmonter les fluctuations terminologiques et les ambiguïtés sémantiques qui sont les siennes aujourd'hui. Sans terminologie commune, sans absence d'ambiguïté sémantique et sans harmonisation conceptuelle, il est peu probable que les initiatives promouvant les « neurodroits » débouchent sur des politiques nationales et internationales efficaces. Cela ne signifie pas que ce processus d'harmonisation devrait éluder complètement les points de vue divergents. Au contraire, il

devrait les inclure dans une démarche démocratique pluraliste et délibérative. En outre, il doit veiller à ce que les propositions de « neurodroits » soient correctement examinées, délimitées sur le plan conceptuel, justifiées sur le plan normatif et ancrées dans la philosophie morale et les réglementations existantes.

Cette analyse suggère que c'est le droit à la vie privée mentale qui bénéficie du plus large consensus et de la plus forte stabilité conceptuelle dans le débat international. Cela s'explique par l'existence d'un accord global sur la nécessité qu'il y a de protéger les données cérébrales et mentales contre tout accès, examen ou modification qui n'auraient pas été autorisés. En ce qui concerne le droit à l'intégrité mentale, le débat est ouvert quant à savoir si la protection contre les atteintes est une exigence constitutive de ce droit. Toutefois, rappelons-le une fois de plus, il existe un large consensus sur la nécessité de protéger la personne contre les interférences mentales. Enfin, les candidats au statut de « neurodroits » qui relèvent de l'autodétermination mentale (c'est-à-dire la liberté de pensée, la liberté cognitive, la liberté mentale) et de l'identité personnelle (c'est-à-dire les droits à l'identité personnelle et à la continuité psychologique) se caractérisent par une plus grande variation sémantique et interprétative. Cependant, malgré des terminologies multiples et souvent inconciliables, l'accord global autour du fait que la liberté d'exercer un contrôle sur sa sphère cognitive et émotionnelle est la condition préalable de toutes les autres libertés, et constitue une condition essentielle de l'autonomie personnelle et de l'autodétermination, persiste bel et bien.

Une troisième question qu'il faudrait traiter concerne la manière dont les « neurodroits » peuvent être mis en œuvre et appliqués de manière adéquate. Si, à l'avenir, certains des « neurodroits » décrits dans ce document devaient être examinés à plusieurs reprises et obtenaient un grand soutien démocratique et délibératif, comment devraient-ils ensuite être appliqués ? Il existe deux types d'instruments chargés de mettre en œuvre les droits de l'homme : les déclarations et les conventions. Les déclarations ne sont pas juridiquement contraignantes, mais elles ont un impact politique, tandis que les conventions sont juridiquement contraignantes en vertu du droit international. Les déclarations comme les conventions peuvent accéder au statut de droit international coutumier au fil du temps, ce qui les rend universellement contraignantes (Moscrop, 2014). La Déclaration universelle de l'UNESCO sur le génome humain et les droits de l'homme de 1997, ainsi que la Déclaration de l'UNESCO sur la bioéthique et les droits de l'homme de 2005 peuvent fournir un terrain approprié et servir de référence pour tenter d'élaborer une déclaration analogue qui aurait pour sujet le cerveau humain et les droits de l'homme. Parmi les conventions, la Convention pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine (Convention d'Oviedo) du Conseil de l'Europe constitue le modèle le plus approprié et le plus complet pour fonder toute initiative future et tout instrument international visant à protéger le cerveau humain. Étant donné qu'elle met l'accent sur l'interdiction de l'utilisation abusive des innovations dans le domaine de la biomédecine, sur la protection de la dignité et de l'identité de tous les êtres humains et sur la garantie du respect de leur intégrité et de leurs libertés fondamentales, la Convention est parfaitement placée pour consacrer des « neurodroits » tels que les droits à l'intégrité mentale, à l'identité personnelle et à la liberté cognitive. Les « neurodroits » pourraient faire l'objet d'un protocole additionnel spécifique qui préciserait les dispositions de la Convention.

Par ailleurs, les études juridiques devraient à l'avenir discuter du type d'instrument le plus approprié pour inscrire les « neurodroits » dans le droit international des droits de l'homme. En outre, elles devraient déterminer quelle est la meilleure manière d'éviter le problème de la « sous-application » des droits de l'homme (Koh, 1998), c'est-à-dire définir, tout en gardant une perspective réaliste, comment faire en sorte que les États respectent les lois sur les « neurodroits ».

Enfin, ces études devraient discuter de la place occupée par les « neurodroits » dans la gouvernance des neurotechnologies. À moins d'être partisan de la thèse improbable selon laquelle les « neurodroits » seraient suffisants pour former la gouvernance des neurotechnologies (et donc que la gouvernance des neurotechnologies pourrait être entièrement réduite à la promotion des « neurodroits »), il est essentiel de clarifier les liens entre les « neurodroits » et d'autres mécanismes de gouvernance, comme l'autorégulation de la part des acteurs des neurotechnologies, ou bien les directives éthiques et les réglementations contraignantes dans le droit de la santé, la protection des données, la protection des consommateurs ou encore le droit pénal.

6. Conclusions et options politiques

Depuis l'aube de l'humanité, l'innovation technologique a été un moteur majeur de la transformation des sociétés et des cultures. Nos ancêtres ont constamment modifié leur environnement en développant des artefacts dont la complexité s'est progressivement accrue. L'*homo sapiens* (littéralement « l'homme sage ») est fondamentalement un *homo technologicus* (c'est-à-dire un « homme qui utilise des technologies »). Ces technologies de plus en plus complexes ont évolué en même temps que les humains et les sociétés humaines. Dans certains cas, des transformations socio-technologiques majeures telles que l'invention de l'agriculture, de l'écriture et de la médecine moderne ont eu des conséquences profondes sur la vie humaine et sur la signification même d'être humain. Tout au long de ce processus d'évolution technologique, les sociétés humaines ont introduit des systèmes de normes, qu'il s'agisse de coutumes socialement acceptées, de recommandations prescrites par l'Église ou de lois contraignantes appliquées par l'État.

Parmi ces systèmes de normes, les droits de l'homme constituent les catégories les plus fondamentales et les plus universelles. En effet, selon leur définition, les droits de l'homme sont inhérents à tous les êtres humains, quels que soient leur nationalité, leur sexe, leur origine nationale ou ethnique, leur couleur, leur religion, leur langue ou tout autre statut. Les droits de l'homme étant des droits qui doivent être respectés simplement parce que nous existons en tant qu'êtres humains, leur définition conceptuelle et leur articulation sont interdépendantes de ce que signifie être un être humain. Si la signification de l'être humain évolue en raison des changements socio-technologiques, les droits de l'homme peuvent évoluer en conséquence. Cette évolution est nécessaire pour protéger les libertés fondamentales et les droits qui font que la vie humaine vaut la peine d'être vécue à un moment historique donné.

La Convention européenne des droits de l'homme est un exemple paradigmatique de cette évolution normative. Plus de soixante ans après son adoption, la CEDH n'a cessé d'évoluer sous l'effet de la réflexion juridico-philosophique et de la jurisprudence de la Cour européenne des droits de l'homme. Pour cette raison, la Convention a souvent été considérée comme un « instrument vivant », capable de s'adapter aux changements qui se produisent dans nos sociétés. Ce processus d'adaptation constante a également nécessité d'ajouter de nouveaux droits à la Convention chaque fois que des droits fondamentaux sont apparus dans des situations qui n'auraient pas pu être prévues lors de l'adoption initiale de la Convention.

Au cours du siècle dernier, l'innovation dans les domaines de la biomédecine, de l'information et de l'informatique a donné naissance à des technologies extrêmement puissantes et transformatrices. Ces technologies disposent d'une capacité sans précédent d'empiéter sur les droits de l'homme. Parmi ces technologies, la neurotechnologie et l'intelligence artificielle (ainsi que leur combinaison) jouent un rôle essentiel, car elles offrent de nouvelles possibilités de surveiller, d'influencer, de modifier et de simuler le cerveau et l'esprit humains. Le cerveau et l'esprit humains étant les substrats physiques et fonctionnels des facultés humaines fondamentales comme l'identité personnelle, la conscience, le langage et les émotions, toutes les technologies capables de surveiller, influencer, modifier et simuler le cerveau et l'esprit humains affectent les droits de l'homme. En conséquence, le constat de l'évolution de ces technologies doit inciter nos sociétés à réfléchir à ces conséquences, et à déterminer si une évolution normative est nécessaire pour protéger et promouvoir les droits de l'homme.

C'est à la science empirique, en particulier aux neurosciences, à la psychologie, à l'anthropologie et à la médecine, que revient la tâche d'étudier comment ces technologies peuvent transformer les humains dans les années à venir. Comme ces technologies peuvent influencer les capacités humaines au niveau neurobiologique et au niveau du traitement de l'information, il est plausible de prédire que la révolution neurotechnologique actuelle transformera la nature humaine au moins autant que certaines inventions précédentes qui ont marqué leur époque.

Dans le même temps, il faut impérativement que les disciplines normatives telles que l'éthique et le droit déterminent quels aspects de la transformation anthropologique induite par la technologie sont moralement souhaitables et légalement acceptables. En particulier, nous devons déterminer quels droits les gens doivent avoir sur leur cerveau et leur esprit. Étant donné que la neurotechnologie possède le potentiel de modifier les caractéristiques fondamentales (par exemple, la mémoire, la conscience, le raisonnement, les émotions, les croyances et les désirs) de l'être humain, elle contient aussi le potentiel de modifier les droits, les libertés et les obligations fondamentaux de l'être humain (Ienca, 2019b).

Ce rapport a tenté de fournir une vue d'ensemble détaillée des principaux développements technologiques dans les domaines de la neurotechnologie et de l'IA, en mettant un accent particulier sur les interfaces cerveau-machine. En outre, il a tenté de résumer la réflexion théorique existante sur les droits moraux et juridiques fondamentaux liés à l'utilisation de ces technologies. Enfin, il fournit un compte rendu descriptif et une taxonomie conceptuelle des « neurodroits », c'est-à-dire des principes moraux, juridiques, sociaux ou naturels de liberté ou de droit concernant le cerveau et l'esprit d'une personne.

Cette analyse suggère que les « neurodroits » reflètent des intérêts humains fondamentaux qui sont profondément ancrés dans l'histoire des idées. Ces droits introduisent des spécificités normatives quant à la protection du domaine cérébral et mental de la personne. Ces spécificités ne s'inscrivent pas dans une répétition des cadres de droits de l'homme déjà existants, mais ajoutent un nouveau niveau fondamental de protection normative. En outre, ils corroborent la thèse que les droits et libertés fondamentaux relatifs à l'esprit et au cerveau humains constituent le substrat fondamental de tous les autres droits et libertés. Par conséquent, la protection des « neurodroits » est une tâche fondamentale pour le droit international des droits de l'homme.

Cette vue d'ensemble montre qu'il n'existe pas encore de consensus total au sujet des limites conceptuelles et normatives des « neurodroits », pas plus que sur la terminologie à employer. Il y a bel et bien des divergences quant à la manière dont ces droits devraient être interprétés, formulés et articulés conceptuellement. Néanmoins, un certain degré de convergence se dessine pour désigner trois grandes familles de « neurodroits ». Tout d'abord, la nécessité de dispositions spécifiques pour protéger les informations privées contenues dans le cerveau semble faire l'objet d'un haut degré d'acceptation et de reconnaissance. Le droit à la vie privée mentale semble être le candidat le plus solide conceptuellement pour assumer ce rôle. Deuxièmement, le droit à l'intégrité mentale semble avoir le plus haut degré d'ancrage juridique. Bien qu'il y ait quelques variations dans l'interprétation de ce droit, il existe un consensus théorique total sur la nécessité de protéger la personne contre les dommages psychologiques et les interférences mentales. Troisièmement, il y a eu différentes propositions de « neurodroits » visant à préserver et promouvoir la liberté de l'esprit humain, et empêcher

ainsi toute manipulation extérieure. Il s'agit notamment d'interprétations qui cherchent à faire évoluer le droit à la liberté de pensée, le droit à la liberté cognitive et le droit à l'identité personnelle.

Il est en revanche surprenant que des droits positifs comme la promotion de la justice et de l'égalité – garantir, par exemple, un accès égalitaire aux neurotechnologies à des fins biomédicales, ou promouvoir le bien-être des patients sur la base du principe éthique de bienfaisance – aient jusqu'à présent occupé un rôle secondaire dans le débat sur les « neurodroits ».

Les familles de « neurodroits » énumérées ci-dessus semblent s'ancrer profondément dans le cadre international des droits de l'homme et dans la doctrine juridique. Cependant, elles ne sont pas suffisamment spécifiées dans les instruments actuels des droits de l'homme tels que la DUDH, la CEDH et le CFR. Par conséquent, un processus de réforme normative semble nécessaire afin de caractériser de manière adéquate les libertés ou les droits concernant spécifiquement l'esprit et le cerveau d'une personne. Ce processus peut se dérouler de deux manières : en adaptant l'interprétation des droits existants, ou en créant de nouveaux droits.

Faire évoluer les normes face à l'innovation technologique n'est pas un phénomène nouveau. Par exemple, l'apparition du ventilateur mécanique a donné naissance au concept de mort cérébrale. Il a ainsi obligé la loi à préciser plus clairement quelles fonctions font partie intégrante de la vie et quelles fonctions n'en font pas partie, ainsi que la manière de résoudre les désaccords concernant le statut moral d'êtres humains qui pourraient être considérés comme gravement atteints de lésions cérébrales mais non pas morts (Machado, 2007). Plus important encore, la CEDH a évolué en tant qu'« instrument vivant », car l'adoption de protocoles a ajouté de nouveaux droits à la Convention. Il s'agit notamment du droit au respect de ses biens, du droit à l'éducation, du droit à des élections libres au scrutin secret (protocole n° 1) et de l'interdiction générale de la discrimination (protocole n° 12). On trouve également, plus étroitement liés au domaine de la technologie biomédicale, la Déclaration universelle sur le génome humain et les droits de l'homme (DUGDH) de 1997 et la Déclaration internationale sur les données génétiques humaines (DIDGH) de 2003. Ces instruments ont tous deux été élaborés en réponse aux progrès de la génétique et ont introduit de nouveaux droits tels que le « droit de ne pas connaître ses informations génétiques » (UDHGHR Art. 5(c) ; IDHGD (Art. 10)). De même, la Convention d'Oviedo a introduit un protocole additionnel sur l'interdiction du clonage humain pour répondre au développement de techniques de clonage artificiel se fondant sur la reproduction de cellules et de tissus humains.

Il est souhaitable que les « neurodroits » suivent une trajectoire historique similaire, afin d'étendre et d'améliorer la capacité de notre cadre des droits de l'homme à aborder les conséquences profondes de la neurotechnologie et de l'IA pour la nature humaine, la dignité humaine et les droits de l'homme. L'introduction des « neurodroits » dans le cadre des droits de l'homme peut nécessiter d'ajouter de nouveaux protocoles aux instruments existants, ou même de créer de nouveaux instruments multilatéraux entièrement consacrés à la neuroéthique et au neurodroit. Dans les deux cas, il faudra résoudre certaines questions éthiques, méta-éthiques et juridiques fondamentales afin de surmonter des problèmes comme l'inflation des droits, et fournir une justification normative adéquate pour les « neurodroits ». Il s'agit notamment d'introduire des tests et examens justifiant l'introduction des « neurodroits », de

clarifier la relation entre « neurodroits » moraux et « neurodroits » juridiques, et d'harmoniser les « neurodroits » avec les instruments normatifs existants.

La Convention du Conseil de l'Europe pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine (Convention d'Oviedo) offre une plate-forme et une base normative idéaux pour protéger et promouvoir les « neurodroits ». Étant donné qu'elle met l'accent sur l'interdiction de l'utilisation abusive des innovations dans le domaine de la biomédecine, sur la protection de la dignité et de l'identité de tous les êtres humains et sur la garantie du respect de leur intégrité et de leurs libertés fondamentales, la Convention est parfaitement adaptée pour consacrer des « neurodroits » tels que le droit à l'intégrité mentale, le droit à l'identité personnelle et le droit à la liberté cognitive.

Comprendre, traiter et améliorer le cerveau et l'esprit humains constitue l'un des grands défis scientifiques de notre époque. Atteindre ces objectifs tout en préservant la justice, les droits fondamentaux et la dignité humaine, telle est la tâche parallèle de l'éthique et du droit. Les « neurodroits » seront probablement un outil utile pour accomplir cette tâche.

Annexe 1 : Glossaire

- **Amélioration cognitive** : Interventions sur le cerveau dans le but d'améliorer la cognition (améliorer, par exemple, l'attention, la concentration et le traitement de l'information au sein de fonctions exécutives comme le raisonnement et la prise de décision).
- **Biomarqueur** : Marqueur biologique, indicateur mesurant un état ou une condition biologique.
- **Cerveau (humain)** : Organe central du système nerveux humain.
- **Cognition** : Ensemble des processus mentaux tels que la pensée, la connaissance, la mémoire, le jugement et la résolution de problèmes.
- **Contenu mental** : Contenu d'un état mental, qu'il soit conceptuel ou non conceptuel.
- **Continuité psychologique** : Continuité de la vie mentale des personnes dans le temps (par exemple, continuité entre des états mentaux non concomitants).
- **Croyance** : Conviction d'une personne qu'une certaine proposition sur le monde est vraie.
- **Imagerie cérébrale** : Utilisation de diverses techniques pour obtenir, directement ou indirectement, une image de la structure, de la fonction ou de la pharmacologie du système nerveux.
- **Intégrité mentale** : Intégrité de l'esprit humain.
- **Liberté cognitive** : Droit des individus à prendre des décisions libres et compétentes concernant leur esprit et leur cerveau.
- **Neurodiscrimination** : Discrimination qui se fonde sur des caractéristiques neurales.
- « **Neurodroits** » : Principes éthiques, juridiques, sociaux ou naturels de liberté ou de droit liés au domaine cérébral et mental d'une personne.
- **Neurostimulation** : Modulation volontaire de l'activité du système nerveux par des moyens invasifs (par exemple, des microélectrodes) ou non invasifs (par exemple, la stimulation magnétique transcrânienne ou la stimulation électrique transcrânienne).
- **Neurotechnologie** : Ensemble large et hétérogène des méthodes, systèmes et instruments qui établissent une connexion avec le cerveau humain grâce à laquelle l'activité neuronale peut être enregistrée et/ou modifiée.
- **Personnalité** : Statut de personne.
- **Syndrome d'enfermement** : État neurologique dans lequel le patient est alerte et conscient, mais incapable de bouger ou de communiquer verbalement en raison d'une paralysie complète de quasiment tous les muscles volontaires du corps, à l'exception des mouvements verticaux des yeux et du clignement des yeux.
- **Vie privée mentale** : Droit des personnes contre l'intrusion non consentie de tiers dans les données de leur cerveau, ainsi que contre la collecte non autorisée de ces données.

Remerciements

Ce rapport a été commandé et soutenu par le Comité de bioéthique du Conseil de l'Europe. Le rapport est basé sur une combinaison de travaux originaux et de travaux antérieurs de l'auteur ayant trait à l'impact des neurotechnologies en matière de droits de l'homme. Voir :

- Ienca, M. (à paraître en 2021). *On neurorights*. *Frontiers in Human Neuroscience*.
- Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life sciences, society and policy*, 13(1), 1-27.
- Ienca, M. (2017). The Right to Cognitive Liberty. *Scientific American*, 317(2), 10-10.
- Ienca, M. (2019). *Intelligenza2: per un'unione di intelligenza naturale e artificiale*. Rosenberg & Sellier.

Références

- Aarts, H., & van den Bos, K. (2011). On the foundations of beliefs in free will: Intentional binding and unconscious priming in self-agency. *Psychological science*, 22(4), 532-537.
- Ajrawi, S., Rao, R., & Sarkar, M. (2021). Cybersecurity in brain-computer interfaces: RFID-based design-theoretical framework. *Informatics in Medicine Unlocked*, 22, 100489.
- Alston, P. (1984). Conjuring up new human rights: A proposal for quality control. *Am. J. Int'l L.*, 78, 607.
- Ayaz, H., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2019). Chapter 3 - The Use of Functional Near-Infrared Spectroscopy in Neuroergonomics. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 17-25): Academic Press.
- Baker, L. R. (2000). *Persons and bodies: A constitution view*: Cambridge University Press.
- Beitz, C. R. (2011). *The idea of human rights*: Oxford University Press.
- Berlin, I. (1969). Two concepts of liberty. *Berlin, i*, 118(1969), 172.
- Bles, M., & Haynes, J.-D. (2008). Detecting concealed information using brain-imaging technology. *Neurocase*, 14(1), 82-92.
- Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., & Haynes, J.-D. (2011). Tracking the unconscious generation of free decisions using ultra-high field fMRI. *Plos one*, 6(6), e21612.
- Boire, R. G. (2001). On cognitive liberty. *The Journal of Cognitive Liberties*, 2(1), 7-22.
- Bostrom, N., & Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and engineering ethics*, 15(3), 311-341.
- Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., . . . Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, 555(7698), 657-661. doi:10.1038/nature26147
- Boyden, E. S. (2015). Optogenetics and the future of neuroscience. *Nature neuroscience*, 18(9), 1200-1201. doi:10.1038/nn.4094
- Brandeis, L., & Warren, S. (1890). The right to privacy. *Harvard law review*, 4(5), 193-220.
- Brimacombe, P. (2000). *All the Queen's Men: The World of Elizabeth I*. New York: St. Martin's Press.
- Bublitz, C. (2015). Cognitive Liberty or the International Human Right to Freedom of Thought. In J. Clausen & N. Levy (Eds.), *Handbook of Neuroethics* (pp. 1309-1333). Dordrecht: Springer Netherlands.
- Bublitz, C. (2016). Moral Enhancement and Mental Freedom. *Journal of Applied Philosophy*, 33(1), 88-106. doi:<https://doi.org/10.1111/japp.12108>
- Bublitz, J.-C. (2013). My mind is mine!? Cognitive liberty as a legal concept. In *Cognitive enhancement* (pp. 233-264): Springer.

- Bublitz, J. C., & Merkel, R. (2014). Crimes against minds: On mental manipulations, harms and a human right to mental self-determination. *Criminal Law and Philosophy*, 8(1), 51–77. <https://doi.org/10.1007/s11572-012-9172-y>.
- Bury, J. B. (1914). *A history of freedom of thought*: Gorgias Press.
- Capron, A. M. (1976). Legal rights and moral rights. In *Biomedical Ethics and the Law* (pp. 375-397): Springer.
- Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A., & Yuste, R. (2019). Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell*, 178(2), 447-457 e445. doi:10.1016/j.cell.2019.05.045
- Cascio, J. (2017). Do brains need rights? *New Scientist*, 234(3130), 24-25. doi:[https://doi.org/10.1016/S0262-4079\(17\)31163-6](https://doi.org/10.1016/S0262-4079(17)31163-6)
- Castanheira, L., Silva, C., Cheniaux, E., & Telles-Correia, D. (2019). Neuroimaging Correlates of Depression—Implications to Clinical Practice. *Frontiers in Psychiatry*, 10(703). doi:10.3389/fpsy.2019.00703
- Chapman, C. R., Mehta, K. S., Parent, B., & Caplan, A. L. (2019). Genetic discrimination: emerging ethical challenges in the context of advancing technology. *Journal of Law and the Biosciences*.
- Chaudhary, P., & Agrawal, R. (2018). Emerging threats to security and privacy in brain computer interface. *International Journal of Advanced Studies of Scientific Research*, 3(12).
- Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, 20(1), 115-125. doi:10.1038/nn.4450
- Chizeck, H. J., & Bonaci, T. (2014). Brain-Computer Interface Anonymizer. In: Google Patents.
- Clausen, J., Fetz, E., Donoghue, J., Ushiba, J., Spörhase, U., Chandler, J., . . . Soekadar, S. R. (2017). Help, hope, and hype: Ethical dimensions of neuroprosthetics. *Science*, 356(6345), 1338-1339.
- Collins, R. F., & Harrington, D. J. (1999). *First Corinthians* (Vol. 7): Liturgical Press.
- Committee, U. H. R. (1993). *General Comment No. 22: The Right to Freedom of Thought, Conscience and Religion (Article 18)*, UN Doc. No. Retrieved from
- Committee on Legal Affairs and Human Rights (2020). The brain-computer interface: new rights or new threats to fundamental freedoms? Report Doc. 15147, 24 September 2020.
- Dainton, B., & Bayne, T. (2005). Consciousness as a guide to personal persistence. *Australasian Journal of Philosophy*, 83(4), 549-571.
- de Vos, F., Koini, M., Schouten, T. M., Seiler, S., van der Grond, J., Lechner, A., . . . Rombouts, S. A. (2018). A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. *Neuroimage*, 167, 62-72.
- Dennett, D. C. (2015). *Elbow room: The varieties of free will worth wanting*: mit Press.
- Douglas, T. (2014). Criminal rehabilitation through medical intervention: moral liability and the right to bodily integrity. *The Journal of Ethics*, 18(2), 101-122.
- Douglas, T., & Forsberg, L. (2021). Three Rationales for a Legal Right to Mental Integrity. *Neurolaw: Advances in Neuroscience, Justice & Security*, 179.
- Farah, M. J. (2002). Emerging ethical issues in neuroscience. *Nature neuroscience*, 5(11), 1123-1129.
- Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., . . . Wolpe, P. R. (2004). Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience*, 5(5), 421-425.

- Farahany, N. A. (2012). Incriminating thoughts. *Stan. L. Rev.*, 64, 351.
- Farahany, N. A. (2019). The Costs of Changing Our Minds. *Emory LJ*, 69, 75.
- Favaretto, M., De Clercq, E., Gaab, J., & Elger, B. S. (2020). First do no harm: An exploration of researchers' ethics of conduct in Big Data behavioral studies. *Plos one*, 15(11), e0241865.
- Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in neuroengineering*, 5, 14.
- Ferretti, A., & Ienca, M. (2018). Enhanced cognition, enhanced self? On neuroenhancement and subjectivity. *Journal of Cognitive Enhancement*, 2(4), 348-355.
- Fins, J. J. (2004). Neuromodulation, free will and determinism: lessons from the psychosurgery debate. *Clinical Neuroscience Research*, 4(1-2), 113-118.
- Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass*, 2(2), 347-357.
- Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, 33(4), 111-117.
- Garrett, B. (2002). *Personal identity and self-consciousness*: Routledge.
- Gaustad, E. S. (2001). *Roger Williams: Prophet of Liberty*: Oxford University Press.
- Gilbert, F., Cook, M., O'Brien, T., & Illes, J. (2019). Embodiment and estrangement: Results from a first-in-human "intelligent BCI" trial. *Science and engineering ethics*, 25(1), 83-96.
- Gilbert, F., Goddard, E., Viaña, J. N. M., Carter, A., & Horne, M. (2017). I miss being me: Phenomenological effects of deep brain stimulation. *AJOB neuroscience*, 8(2), 96-109.
- Goering, S., Klein, E., Sullivan, L. S., Wexler, A., y Arcas, B. A., Bi, G., . . . Gallant, J. (2021). Recommendations for Responsible Development and Application of Neurotechnologies. *Neuroethics*, 1-22.
- Gordon, B. A., Blazey, T. M., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., . . . Xiong, C. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology*, 17(3), 241-250.
- Greenberg, A. (2019). Inside the Mind's Eye: An International Perspective on Data Privacy Law in the Age of Brain Machine Interfaces. *Alb. LJ Sci. & Tech.*, 29, 79.
- Hallinan, D., Schütz, P., Friedewald, M., & de Hert, P. (2014). Neurodata and neuroprivacy: Data protection outdated? *Surveillance & Society*, 12(1), 55-72.
- Hayes, B. (2017). Migration and data protection: Doing no harm in an age of mass displacement, mass surveillance and "big data". *International Review of the Red Cross*, 99(904), 179-209.
- Haynes, J.-D. (2011). Brain reading: Decoding mental states from brain activity in humans. *The Oxford Handbook of Neuroethics*, 3-13.
- Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current biology*, 15(14), 1301-1307.
- Hohfeld, W. N. (1913). Some fundamental legal conceptions as applied in judicial reasoning. *Yale Lj*, 23, 16.
- Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural Decoding of Visual Imagery During Sleep. *Science*, 340(6132), 639. doi:10.1126/science.1234330
- Hubbard, J., Kikumoto, A., & Mayr, U. (2019). EEG Decoding Reveals the Strength and Temporal Dynamics of Goal-Relevant Representations. *Scientific Reports*, 9(1), 9051. doi:10.1038/s41598-019-45333-6
- Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T., & Gallant, J. L. (2016). Decoding the Semantic Content of Natural Movies from Human Brain Activity. *Frontiers in Systems Neuroscience*, 10(81). doi:10.3389/fnsys.2016.00081

- Ienca, M. (2015). *Neuroprivacy, neurosecurity and brain-hacking: Emerging issues in neural engineering*. Paper presented at the Bioethica Forum.
- Ienca, M. (2017). The Right to Cognitive Liberty. *Scientific American*, 317(2), 10-10.
- Ienca, M. (2019a). Democratizing cognitive technology: a proactive approach. *Ethics and Information Technology*, 21(4), 267-280.
- Ienca, M. (2019b). *Intelligenza2: per un'unione di intelligenza naturale e artificiale*: Rosenberg & Sellier.
- Ienca, M., & Andorno, R. (2017a). A new category of human rights: neurorights. Retrieved from <http://blogs.biomedcentral.com/bmcblog/2017/04/26/new-category-human-rights-neurorights/>
- Ienca, M., & Andorno, R. (2017b). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(1), 5. doi:10.1186/s40504-017-0050-1
- Ienca M., Fins J.J., Jox R.J., Jotterand F., Voeneky S., Andorno R., Ball T., Castelluccia C., Chavarriaga R., Chneiweiss H., Ferretti A., Friedrich O., Hurst S., Merkel G., Molnár-Gábor F., Rickli J.M., Scheibner J., Vayena E., Yuste R., Kellmeyer P. (2021). Towards a Governance Framework for Brain Data. SSRN.
- Ienca, M., & Haselager, P. (2016). Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18(2), 117-129.
- Ienca, M., Haselager, P., & Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nature biotechnology*, 36(9), 805-810.
- Ienca, M., & Ignatiadis, K. (2020). Artificial intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB neuroscience*, 11(2), 77-87.
- Ienca, M., & Malgieri, G. (2021). Mental Data Protection and the GDPR. Available at SSRN 3840403.
- Ienca, M., Shaw, D. M., & Elger, B. (2019). Cognitive enhancement for the ageing world: opportunities and challenges. *Ageing and Society*, 39(10), 2308-2321.
- Ienca, M., & Vayena, E. (2018). Cambridge analytica and online manipulation. *Scientific American*, 30.
- Illes, J., Kirschen, M. P., & Gabrieli, J. D. (2003). From neuroimaging to neuroethics. *Nature neuroscience*, 6(3), 205-205.
- Illes, J., Rosen, A. C., Huang, L., Goldstein, R., Raffin, T. A., Swan, G., & Atlas, S. (2004). Ethical consideration of incidental findings on adult brain MRI in research. *Neurology*, 62(6), 888-890.
- Jobsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323), 1264-1267.
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352-355.
- Kober, H., Lacadie, C. M., Wexler, B. E., Malison, R. T., Sinha, R., & Potenza, M. N. (2016). Brain activity during cocaine craving and gambling urges: an fMRI study. *Neuropsychopharmacology*, 41(2), 628-637.
- Koh, H. H. (1998). How is international human rights law enforced. *Ind. LJ*, 74, 1397.
- Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394), 190-191.
- Lavazza, A. (2018). Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis. *Frontiers in Neuroscience*, 12(82). doi:10.3389/fnins.2018.00082
- Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., . . . De Ridder, D. (2014). Evidence-based guidelines on the therapeutic use of repetitive

- transcranial magnetic stimulation (rTMS). *Clinical Neurophysiology*, 125(11), 2150-2206.
- Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge University Press.
- Li, M., Xu, H., & Lu, S. (2018). Neural basis of depression related to a dominant right hemisphere: A resting-state fMRI study. *Behavioural neurology*, 2018.
- Long, A. A. (2002). *Epictetus: A Stoic and Socratic guide to life*: Clarendon Press.
- Luzzatti, L. (2006). *God in Freedom: Studies in the Relations Between Church and State*: Cosimo, Inc.
- Marshall, J. (2014). *Human rights law and personal identity*: Routledge.
- Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., . . . Deisseroth, K. (2019). Cortical layer-specific critical dynamics triggering perception. *Science*, 365(6453). doi:10.1126/science.aaw5202
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, 44(2), 379-387.
- Meslin, E. M. (1990). Protecting human subjects from harm through improved risk judgments. *IRB: Ethics & Human Research*, 12(1), 7-10.
- Mill, J. S. (1859). *On liberty* In: London: Thinker's library.
- Milton, J. (1791). *Comus: A Mask*: proprietors, under the direction of John Bell.
- Moreno, J. D. (2003). Neuroethics: an agenda for neuroscience and society. *Nature Reviews Neuroscience*, 4(2), 149-153.
- Moscrop, H. (2014). Enforcing International Human Rights Law: Problems and Prospects. In: Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, 10(1), 3096. doi:10.1038/s41467-019-10994-4
- Munoz, J. M. (2019). Chile--right to free will needs definition. *Nature*, 574(7780), 634-635.
- Nawrot, O. (2019). What about the Interior Castle? Response to Ienca's and Andorno's New Human Rights in the Age of Neuroscience and Neurotechnology. *Roczniki Teologiczne*, 66(3), 69-85.
- Nickel, J., Pogge, T., Smith, M., & Wenar, L. (2013). Stanford Encyclopedia of Philosophy, Human Rights. In: December.
- Noonan, H. W. (2019). *Personal identity*: Routledge.
- Nussbaum, M. (2007). Human rights and human capabilities. *Harv. Hum. Rts. J.*, 20, 21.
- Nuyujukian, P., Albites Sanabria, J., Saab, J., Pandarinath, C., Jarosiewicz, B., Blabe, C. H., . . . Simeral, J. D. (2018). Cortical control of a tablet computer by people with paralysis. *Plos one*, 13(11), e0204566.
- OECD-Council. (2019). *OECD Recommendation on Responsible Innovation in Neurotechnology*. Retrieved from <https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm>
- Omurtag, A., Aghajani, H., & Keles, H. O. (2017). Decoding human mental states by whole-head EEG+fNIRS during category fluency task performance. *J Neural Eng*, 14(6), 066003. doi:10.1088/1741-2552/aa814b
- Palaniappan, R., & Mandic, D. P. (2007). Biometrics from brain electrical activity: A machine learning approach. *IEEE transactions on pattern analysis and machine intelligence*, 29(4), 738-742.
- Palk, A., Illes, J., Thompson, P. M., & Stein, D. J. (2020). Ethical issues in global neuroimaging genetics collaborations. *NeuroImage*, 221, 1172
- Pereboom, D., & Caruso, G. (2002). Hard-incompatibilist existentialism: Neuroscience, punishment, and meaning in life. *Neuroexistentialism: Meaning, Morals, and Purpose*

- in the Age of Neuroscience*, OUP Editors, Gregg D. Caruso and Owen Flanagan, Forthcoming.
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72(5), 692-697. doi:10.1016/j.neuron.2011.11.001
- Polenberg, R. (1996). Cardozo and the Criminal Law: Palko v. Connecticut Reconsidered. *J. Sup. Ct. Hist.*, 92.
- Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756), 1963-1966.
- Powell, C., Munetomo, M., Schlueter, M., & Mizukoshi, M. (2013). *Towards thought control of next-generation wearable computing devices*. Paper presented at the International conference on brain and health informatics.
- Price-Williams, D. (1957). Proprioception and personal identity. *Philosophy and Phenomenological Research*, 17(4), 536-545.
- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*: OUP Oxford.
- Pugh, J., Pycroft, L., Sandberg, A., Aziz, T., & Savulescu, J. (2018). Brainjacking in deep brain stimulation and autonomy. *Ethics and Information Technology*, 20(3), 219-232.
- Pycroft, L., Boccard, S. G., Owen, S. L., Stein, J. F., Fitzgerald, J. J., Green, A. L., & Aziz, T. Z. (2016). Brainjacking: implant security issues in invasive neuromodulation. *World neurosurgery*, 92, 454-462.
- Rainey, S., Bublitz, J. C., Maslen, H., & Thornton, H. (2019). Data as a cross-cutting dimension of ethical importance in direct-to-consumer neurotechnologies. *AJOB neuroscience*, 10(4), 180-182.
- Rainey, S., McGillivray, K., Akintoye, S., Fothergill, T., Bublitz, C., & Stahl, B. (2020). Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology? *J Law Biosci*. doi:10.1093/jlb/ljaa051
- Reider, L. (1998). Toward a new test for the insanity defense: Incorporating the discoveries of neuroscience into moral and legal theories. *UCLA L. Rev.*, 46, 289.
- Reinhart, R. M., & Nguyen, J. A. (2019). Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nature neuroscience*, 22(5), 820-827.
- Repetti, R. (2018). *Buddhism, Meditation, and Free Will: A Theory of Mental Freedom*: Routledge.
- Rickli, J.-M., & Ienca, M. (2021). The Security and Military Implications of Neurotechnology and Artificial Intelligence. *Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications*, 197-214.
- Roskies A. Neuroethics for the new millennium. *Neuron*. 2002;35:21–3.
- Safire, W. (2002). *Visions for a new field of "neuroethics"*. Paper presented at the Neuroethics: Mapping the Field, Conference Proceedings, May 13-14, 2002, San Francisco, California.
- Safron, A., Klimaj, V., Sylva, D., Rosenthal, A., Li, M., Walter, M., & Bailey, J. M. (2018). Neural correlates of sexual orientation in heterosexual, bisexual, and homosexual women. *Scientific Reports*, 8(1), 1-14.
- Schaffer, J. (2015). What not to multiply without necessity. *Australasian Journal of Philosophy*, 93(4), 644-664.
- Schechtman, M. (1994). The same and the same: Two views of psychological continuity. *American Philosophical Quarterly*, 31(3), 199-212.
- Scott, G. (2013). The neurotechnology revolution has arrived. *The Futurist*, 47(5), 6.
- Schreiber, D., Fonzo, G., Simmons, A. N., Dawes, C. T., Flagan, T., Fowler, J. H., & Paulus, M. P. (2013). Red brain, blue brain: Evaluative processes differ in Democrats and Republicans. *Plos one*, 8(2), e52970.

- Schwarz, C. G., Kremers, W. K., Therneau, T. M., Sharp, R. R., Gunter, J. L., Vemuri, P., . . . Jack, C. R. (2019). Identification of Anonymous MRI Research Participants with Face-Recognition Software. *New England Journal of Medicine*, 381(17), 1684-1686. doi:10.1056/NEJMc1908881
- Scott, G. (2013). The neurotechnology revolution has arrived. *The Futurist*, 47(5), 6.
- Sententia, W. (2004). Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences*, 1013(1), 221-228.
- Shen, F. X. (2013). Neuroscience, mental privacy, and the law. *Harv. JL & Pub. Pol'y*, 36, 653.
- Shen, F. X. (2016). The overlooked history of neurolaw. *Fordham L. Rev.*, 85, 667.
- Smith, C. M. (2005). Origin and Uses of Primum Non Nocere—Above All, Do No Harm! *The Journal of Clinical Pharmacology*, 45(4), 371-377. doi:<https://doi.org/10.1177/0091270004273680>
- Soon, C. S., He, A. H., Bode, S., & Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences*, 110(15), 6217-6222.
- Spence, S. (1996). Free will in the light of neuropsychiatry. *Philosophy, Psychiatry, & Psychology*, 3(2), 75-90.
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2).
- Taylor, J. S., Harp, J. A., & Elliott, T. (1991). Neuropsychologists and neurolawyers. *Neuropsychology*, 5(4), 293.
- Thompson, K. (2019). Committing Crimes with BCIs: How Brain-Computer Interface Users can Satisfy Actus Reus and be Criminally Responsible. *Neuroethics*, 1-12.
- Thomson, J. J. (1990). *The realm of rights*: Harvard University Press.
- Turner, D. C., & Sahakian, B. J. (2006). Neuroethics of cognitive enhancement. *BioSocieties*, 1(1), 113-123.
- Van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. *Philosophical Perspectives*, 11, 305-319.
- Vidal, J. J., & JJ, V. (1973). Toward direct brain-computer communication.
- Welford, A. T. (1970). Mental Integrity and the Nature of Life. *Medical Journal of Australia*, 1(23), 1135-1138. doi:<https://doi.org/10.5694/j.1326-5377.1970.tb84477.x>
- Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, 25(1), 166.
- Wexler, A., & Reiner, P. B. (2019). Oversight of direct-to-consumer neurotechnologies. *Science*, 363(6424), 234. doi:10.1126/science.aav0223
- Wolpe, P. R. (2017). Neuroprivacy and Cognitive Liberty. In *The Routledge Handbook of Neuroethics* (pp. 214-224): Routledge, Taylor & Francis Group.
- Woolf, V. (1929). *A room of one's own and three guineas*: OUP Oxford.
- Yuste, R., Genser, J., & Herrmann, S. (2021). It's Time for Neuro-Rights. *Horizons: Journal of International Relations and Sustainable Development*(18), 154-165.
- Yuste, R., Goering, S., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., . . . Illes, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News*, 551(7679), 159.
- Zeki, S., Goodenough, O., & O'Hara, E. A. (2004). How neuroscience might advance the law. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1451), 1677-1684.