



Version provisoire

Commission des questions juridiques et des droits de l'homme

L'interface cerveau-machine : nouveaux droits ou nouveaux dangers pour les libertés fondamentales ?

Rapport*

Rapporteur : M. Olivier BECHT, France, Alliance des démocrates et des libéraux pour l'Europe

A. Projet de résolution

1. L'Assemblée note les progrès rapides réalisés ces dernières années dans les neurotechnologies, notamment la capacité à enregistrer et à stimuler directement l'activité neuronale, avec la possibilité de créer des interfaces cerveau-machine (ICM) de plus en plus efficaces. Ces progrès sont le fruit de la combinaison d'une meilleure compréhension du fonctionnement du cerveau, des avancées techniques et de la puissance croissante des systèmes d'intelligence artificielle. Le fait de pouvoir créer une connexion entièrement symbiotique entre le cerveau humain et des systèmes informatiques numériques, y compris l'internet et les systèmes d'intelligence artificielle, reste une aspiration lointaine. Néanmoins, c'est un objectif que les chercheurs et les entrepreneurs poursuivent déjà et que beaucoup pensent pouvoir atteindre à terme.

2. Plusieurs neurotechnologies, dont l'ICM, sont actuellement développées et appliquées en vue de diverses utilisations. Entre autres mesures prises, des sommes énormes sont investies dans la recherche pour créer de nouveaux traitements médicaux contre les troubles neurologiques et psychiatriques, notamment le contrôle direct de membres robotisés, la production de parole de synthèse ou le traitement de certains troubles de l'humeur insurmontables ou du stress post-traumatique. Les institutions militaires et de sécurité mènent des recherches dans les neurotechnologies en vue d'une application dans le renseignement, la propagande, les interrogatoires, la surveillance et l'amélioration des performances des combattants. Des entreprises privées étudient la possibilité d'utiliser des dispositifs grand public pour convertir directement les pensées en caractères dactylographiés ; offrent des services commerciaux de détection de mensonges basés sur des scanners cérébraux ; et vendent des dispositifs basés sur les neurotechnologies directement aux consommateurs, par exemple sous forme de jeux informatiques ou de produits dédiés au bien-être. Les chercheurs s'intéressent au développement de campagnes de « neuromarketing » qui exploiteraient les préférences subconscientes, et examinent si les modèles d'activité neuronale peuvent prédire la récidive criminelle.

3. Accéder aux processus neuronaux qui sous-tendent la pensée consciente, c'est accéder à un niveau du soi qui, par définition, ne peut être consciemment ni filtré ni dissimulé. Cela risque d'entraîner une profonde violation de la vie privée et de la dignité humaine et pourrait même subvertir le libre arbitre et porter atteinte au dernier refuge de la liberté : la pensée. L'amélioration des capacités cognitives et sensorielles par l'ICM pourrait créer deux catégories distinctes d'êtres humains, les « augmentés » et les autres, l'amélioration étant accessible uniquement à ceux qui disposent de la fortune et des privilèges nécessaires, ou utilisée à des fins de répression. L'identité individuelle, la liberté d'action et la responsabilité morale peuvent être diminuées par la fusion de l'expérience neurologique et sensorielle numérique et des processus décisionnels, ce qui pourrait changer la nature même de l'humanité et des sociétés humaines.

4. Si les applications hypothétiques les plus spectaculaires de l'ICM relèvent encore de la spéculation, les progrès déjà réalisés et les ressources consacrées à la poursuite de la recherche impliquent dès aujourd'hui un besoin urgent d'anticipation et de réglementation préventive. Les sociétés démocratiques doivent veiller à

* Projet de résolution et projet de recommandation adoptés à l'unanimité par la commission le 9 septembre 2020.

ce que les principes éthiques fondamentaux soient respectés. Les avantages potentiels immenses des neurotechnologies sont tels, en particulier dans le domaine médical, que le progrès et l'innovation ne doivent pas être étouffés. Toutefois, la recherche doit être tenue à l'écart des domaines vraisemblablement néfastes ou dangereux pour se tourner vers des applications positives qui ne menacent pas la dignité, l'égalité et la liberté des individus, sur lesquelles repose également la démocratie.

5. L'Assemblée estime qu'une approche sensible et calibrée de la réglementation des nouvelles neurotechnologies, y compris les technologies ICM, est nécessaire, qui englobe à la fois des cadres éthiques et une réglementation juridique contraignante. Elle constate les similitudes et les liens entre la « neuroéthique » et la bioéthique, ainsi que l'importance de l'intelligence artificielle dans le fonctionnement des technologies ICM. Elle se félicite donc du travail déjà entamé au sein du Conseil de l'Europe par le Comité de bioéthique (DH-BIO) et par le Comité ad hoc sur l'intelligence artificielle (CAHAI). Elle salue également les travaux menés par d'autres organisations internationales, notamment l'Organisation de coopération et de développement économiques (OCDE) qui a adopté récemment une recommandation sur l'innovation responsable dans le domaine des neurotechnologies. L'Assemblée note avec intérêt les avancées réalisées, comme au Chili, où l'on envisage de modifier la Constitution, d'adopter une législation et de prendre d'autres mesures pour protéger la société humaine des éventuelles retombées négatives des neurotechnologies.

6. L'Assemblée considère que les principes éthiques suivants doivent être appliqués au développement et à l'application des neurotechnologies en général, et des technologies ICM en particulier :

6.1. Bienfaisance et prévention de l'utilisation malveillante. Ces technologies ne doivent être développées et appliquées qu'à des fins compatibles avec le respect des droits de l'homme et de la dignité humaine. Toute recherche qui poursuit des objectifs incompatibles doit être interdite. Une attention particulière doit être portée aux technologies à double usage et à celles qui sont développées à des fins militaires ou de sécurité. Les nouvelles neurotechnologies doivent être soumises à une évaluation préalable de leurs effets sur les droits de l'homme avant d'être mises en œuvre.

6.2. Sécurité et précaution. Ces technologies et leurs conséquences prévues ou imprévues doivent être sûres, tant pour l'utilisateur que pour la société dans son ensemble. La sécurité doit être garantie avant la mise en service de toute nouvelle application.

6.3. Respect de la vie privée et confidentialité. Au minimum, les informations collectées par les dispositifs neurotechnologiques et d'ICM doivent être protégées conformément aux principes généraux de protection des données. La protection des « neurodonnées » en tant que catégorie spéciale doit également être envisagée, par exemple sur le modèle des interdictions relatives au commerce des organes humains.

6.4. Capacité et autonomie. Ces technologies ne doivent pas être utilisées contre la volonté d'un sujet ni d'une manière qui empêche le sujet de prendre librement d'autres décisions sur la poursuite de leur utilisation. Une attention particulière est nécessaire dans les cas où ces technologies sont utilisées pour soulager la douleur chronique, la toxicomanie ou d'autres affections pour lesquelles l'interruption du traitement pourrait entraîner un malaise ou une détresse.

6.5. Volonté et responsabilité humaine. Ces technologies ne doivent pas empêcher les individus d'agir librement et d'être responsables de leurs actes. Les êtres humains, agissant librement selon leur conscience naturelle (par opposition à une conscience augmentée ou symbiotique), doivent rester les seuls décideurs et les premiers acteurs de la société, surtout pour les questions qui peuvent avoir une incidence sur les droits de l'homme et les processus démocratiques.

6.6. Équité, intégrité et inclusivité. Ces technologies ne doivent pas créer de statut privilégié ou supérieur pour leurs utilisateurs ; elles doivent être mises en œuvre dans le respect de l'égalité et de la dignité humaine, y compris des membres des groupes marginalisés ou vulnérables ; et elles doivent être accessibles au plus grand nombre, en particulier dans la mesure où elles sont appliquées à des fins médicales.

6.7. Gagner la confiance du public par la transparence, la consultation et l'éducation/la sensibilisation. La mise en œuvre des technologies nouvelles comme les neurotechnologies destinées aux particuliers sera d'autant mieux favorisée et acceptée qu'elle se fera dans la confiance du public qui en connaîtra les bienfaits mais aussi les dangers potentiels.

7. Dans la mesure où les technologies ICM peuvent changer fondamentalement la relation entre le soi intérieur et le subconscient de l'individu, et le monde extérieur, elles représentent une menace sans équivalent et sans précédent pour les valeurs fondamentales des droits de l'homme et de la dignité humaine. L'Assemblée prend note avec un intérêt particulier des propositions visant à établir et à fournir une protection juridique pour les nouveaux droits de l'homme, parfois appelés « neurodroits ». Ces propositions sont censées combler les lacunes du cadre actuel des droits de l'homme par lesquelles les technologies ICM peuvent menacer la jouissance des droits actuellement protégés et, au-delà, le respect de la dignité humaine fondamentale. Les droits en question ont été désignés comme suit : la liberté cognitive, la vie privée sur le plan mental, l'intégrité mentale et la continuité psychologique.

8. L'Assemblée appelle par conséquent les États membres du Conseil de l'Europe à :

8.1. établir des cadres éthiques pour la recherche, le développement et l'application des neurotechnologies, y compris les technologies ICM, en tenant compte des principes énoncés au paragraphe 6 de la présente résolution ;

8.2. définir clairement les limites de la recherche, du développement et de l'application des neurotechnologies, y compris les technologies ICM, au moyen de cadres juridiques spécifiques qui garantissent le respect et la protection effectifs des droits de l'homme ;

8.3. veiller à ce qu'il existe des organes appropriés pour la surveillance et la réglementation de la recherche, du développement et de l'application des neurotechnologies, y compris les technologies ICM, de manière à garantir la mise en œuvre efficace des cadres juridiques et éthiques applicables ;

8.4. envisager la création et la protection juridique de nouveaux « neurodroits » afin d'offrir une protection particulièrement efficace contre les risques éventuels inhérents aux technologies ICM.

9. En ce qui concerne les travaux pertinents déjà menés au sein du Conseil de l'Europe, l'Assemblée :

9.1. encourage le DH-BIO à adopter une approche ouverte et constructive de la question des nouveaux « neurodroits », y compris la possibilité d'assurer leur protection en vertu du droit international par le biais d'un protocole additionnel à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STCE 005) ;

9.2. encourage le CAHAI à tenir compte des risques et des opportunités qui peuvent découler de l'application de l'intelligence artificielle dans le contexte des systèmes d'ICM et de ses effets particulièrement graves sur les droits de l'homme.

B. Projet de recommandation

1. L'Assemblée renvoie à sa résolution ... (20...) intitulée « L'interface cerveau-machine : nouveaux droits ou nouveaux dangers pour les libertés fondamentales ? » Elle rappelle que cette résolution a été adoptée alors que des travaux pertinents étaient en cours au sein du Conseil de l'Europe, menés par le Comité de bioéthique (DH-BIO), pour les neurotechnologies, et par le Comité ad hoc sur l'intelligence artificielle (CAHAI).
2. En conséquence, l'Assemblée demande au Comité des Ministres de :
 - 2.1. soutenir les travaux du DH-BIO sur les droits de l'homme et les neurotechnologies, y compris en complétant son mandat existant afin de garantir que la possibilité de protéger les « neurodroits » par le biais d'un protocole additionnel à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STCE 005) soit prise en considération ;
 - 2.2. tenir compte des effets potentiellement sans équivalent et sans précédent de l'intelligence artificielle sur les droits de l'homme en relation avec les systèmes d'interface cerveau-machine (ICM) lors de l'évaluation de la faisabilité d'un cadre juridique pour l'intelligence artificielle.

C. Exposé des motifs par M. Olivier Becht, rapporteur

1. Introduction

1. La proposition de résolution que j'ai déposée le 23 janvier 2019 et sur laquelle se fonde le présent rapport ([Doc. 14814](#)) a été renvoyée en commission par le Bureau le 12 avril 2019. La commission m'a ensuite nommé rapporteur le 29 mai 2019.

2. Cette proposition rappelait que l'ordinateur avait toujours été présenté comme un progrès pour l'humanité, raison pour laquelle la connexion progressive entre le corps humain et des appareils utilisant les technologies numériques n'avait guère suscité de débat. Les appareils de ce type permettent aujourd'hui de connecter directement le cerveau humain à un ordinateur. D'un côté, une telle connexion peut avoir de précieuses applications médicales, rendre par exemple à des personnes handicapées la faculté de parler ou de manipuler des objets ; de l'autre, elle pourrait permettre de lire, d'ajouter ou de supprimer des informations dans le cerveau, mettant un terme au dernier refuge de la liberté : la pensée.

3. À la suite de l'examen de ma note introductive par la commission lors de sa réunion tenue à Berlin (Allemagne) les 14 et 15 novembre 2019, je me suis rendu en Californie (États-Unis d'Amérique) pour une visite d'information du 25 au 27 février 2020. À l'Université de Californie à Berkeley, j'ai rencontré le professeur Jack Gallant (avec qui j'ai discuté des dernières évolutions de la technologie d'imagerie cérébrale, notamment son utilisation pour reconstruire des images visuelles à partir de l'activité neuronale), ainsi que les professeurs Ehud Isacoff et Stuart Russel. À l'Université de Californie à San Francisco, j'ai rencontré Josh Chartier, avec lequel j'ai échangé sur l'utilisation de l'électrocorticographie et des algorithmes d'apprentissage automatique par « réseau neuronal » pour convertir l'activité neuronale décodée en parole de synthèse. Owen Phillips, PDG de la start-up BrainKey, m'a parlé de son projet de développement d'outils de diagnostic médical basés sur l'intelligence artificielle (IA) pour analyser les scanographies d'imagerie à résonance magnétique (IRM) d'une clientèle payante. José Carmena, PDG de la start-up Iota Biosciences, m'a présenté les dispositifs de diagnostic médical ultra-miniaturisés mis au point par son entreprise, parfois décrits comme de la « poussière neuronale ». À l'Université de Stanford, j'ai rencontré le professeur Hank Greely, Alix Roger et Daniel Palanker ; j'ai parlé robotique et prothèses avec le professeur Oussama Khatib et avec Shameek Ganguly ; j'ai discuté des implants rétiniens artificiels et des menaces technologiques avec le professeur E. J. Chichilnisky ; et j'ai échangé avec le professeur Nick Melosh sur les prothèses auditives. Le professeur Byron Yu de l'Université Carnegie Mellon m'a initié aux dernières découvertes en matière de technologie de stimulation et d'empreinte neuronale, dont l'imagerie optique. Alan Mardinly et Alex Feerst de Neuralink m'ont décrit leur technologie d'empreinte neuronale. J'ai discuté avec Luc Julia, vice-président de l'innovation chez Samsung, des applications qui permettent aux clients de se connecter à la technologie. Enfin, j'ai abordé avec Greg Corrado, directeur de recherche principal chez Google, la question de l'utilisation de la technologie et les dérives possibles, en général et dans le domaine de la santé en particulier. J'aimerais remercier toutes ces personnes pour leur temps et leurs contributions, sans oublier Emmanuel Lebrun-Damiens, consul général de France à San Francisco, et ses collègues pour leur aide inestimable dans l'organisation de ma visite.

4. J'avais également l'intention d'organiser une audition avec des experts lors d'une réunion de la commission, mais la pandémie de COVID-19 a malheureusement rendu cela impossible. Je tiens à remercier Marcello Lenca (président du Laboratoire de bioéthique, éthique et politique de la santé, D-HEST, École polytechnique fédérale de Zurich, Suisse), Timothy Constandinou (directeur adjoint, Centre for Bio-inspired Technology, Imperial College London, Royaume-Uni) et David Winickoff (analyste principal des politiques, OCDE) d'avoir accepté de participer à l'audition initialement prévue et de l'avoir remplacée par des contributions écrites au présent rapport.

2. Technologies

5. L'histoire des neurotechnologies est relativement récente. Le premier électroencéphalogramme (EEG) des signaux électriques dans le cerveau d'un animal date d'à peine plus d'un siècle ; le premier EEG sur un être humain a été effectué en 1924. La première stimulation électrique directe du système auditif humain a été réalisée en 1957. En 1965, un EEG a servi à composer de la musique ; en 1988, toujours par EEG, on a réussi à contrôler les mouvements d'un robot. En 1997, la Food and Drug Administration des États-Unis a autorisé le recours à la stimulation cérébrale profonde (SCP, une méthode invasive – voir ci-dessous) pour traiter le tremblement idiopathique et la maladie de Parkinson. Les progrès se sont accélérés ces 20 dernières années, parallèlement à l'augmentation exponentielle de la puissance des ordinateurs : en 2005, Matt Nagie est devenu la première personne à contrôler une main artificielle à l'aide d'une interface cerveau-machine (ICM) ; en 2013, un patient de l'équipe de recherche BrainGate a contrôlé un membre artificiel à l'aide de microélectrodes implantées dans son cerveau ; et en 2018, des chercheurs de Berkeley ont créé le plus petit

stimulateur nerveux du monde, un implant sans fil de la taille d'un grain de poussière (surnommé précisément *neural dust*, « poussière neuronale »). Les principales entreprises du domaine, comme Neuralink, fondée par Elon Musk, ou Facebook Reality Labs, cherchent maintenant à mettre au point des applications commerciales.

6. La technologie au cœur des interfaces cerveau-machine comporte deux éléments : un appareil qui enregistre ou stimule l'activité cérébrale, et un algorithme « décodeur » qui extrait les informations de l'activité enregistrée ou crée un signal pour stimuler une activité. Les technologies d'enregistrement et de stimulation peuvent être non invasives (extra-crâniennes) ou invasives (intracrâniennes). Outre l'EEG, les technologies d'enregistrement non invasives comprennent l'imagerie par résonance magnétique fonctionnelle (IRMf) et la spectrométrie fonctionnelle dans le proche infrarouge (NIRSf), qui enregistrent l'activité neuronale en mesurant la circulation sanguine dans le cerveau ; les technologies de stimulation non invasives comprennent la stimulation transcrânienne par courant continu (STCC) et la stimulation magnétique transcrânienne (SMT), qui créent un courant électrique dans le cerveau. Les technologies d'enregistrement invasives utilisées dans les ICM comprennent l'électrocorticographie (EcoG) et les implants corticaux, avec des électrodes placées directement sur le cortex cérébral. À un stade plus expérimental, on trouve des dispositifs plus perfectionnés encore : « poussière neuronale » (implants miniatures sans fil ni batterie, pourvus de capteurs et de stimulateurs et contrôlés par ultrasons), « dentelle neuronale » (minuscules électrodes disposées le long de « fils » en polymère insérés (« injectés ») dans le cerveau, développée par Neuralink) et « neuropixels » (autre type de dispositif à plusieurs électrodes capable d'accéder simultanément à de nombreuses régions du cerveau). Les technologies de stimulation invasives sont par exemple la stimulation cérébrale profonde (SCP) et la stimulation du nerf vague¹.

7. Des progrès impressionnants se poursuivent dans tous ces domaines. Par exemple, peu de temps avant mon arrivée en Californie, le scanner cérébral « NexGen 7T MRI » a été installé à l'Institut de neurosciences Helen Wills de l'Université de Berkeley, 7T faisant référence à la puissance des aimants utilisés (7 Tesla). Le NexGen 7T MRI est un projet extrêmement coûteux (13,4 millions de dollars) et véritablement international (l'appareil a été construit par Siemens, dont le siège est en Allemagne), comme la plupart des travaux de recherche dans ce domaine. Le NexGen 7T MRI aura une résolution d'imagerie d'environ 0,4 mm, ce qui correspond à l'échelle des structures des colonnes neuronales qui répondent aux caractéristiques spécifiques du monde sensoriel. Si la résolution spatiale devient ainsi moins limitée, la technologie IRM reste soumise à d'importantes limitations sur le plan de la résolution temporelle : alors que la fréquence de décharge neuronale est d'environ 100 par seconde, un IRM l'enregistre au rythme d'environ une fois par seconde.

8. Malgré les progrès rapides récemment accomplis par les ICM et la diversité des approches actuellement à l'étude, d'importants obstacles empêchent toujours le développement d'applications plus ambitieuses. Les dispositifs non invasifs sont incapables d'enregistrer des activités détaillées au niveau cellulaire et ne peuvent servir que pour des interfaces simples, à choix binaire. Aujourd'hui, toute recherche plus poussée requiert une lourde intervention chirurgicale. Les dispositifs invasifs présentent eux-mêmes de sérieuses limites, dont la dégradation progressive de la qualité des enregistrements issus des implants. Il faut également transmettre les données depuis l'intérieur du crâne, soit par un fil traversant la boîte crânienne, soit par un système sans fil, qui pose lui-même certains problèmes. À l'heure actuelle, les meilleurs dispositifs invasifs à plusieurs électrodes enregistrent via un millier de canaux au maximum et suivent des centaines de neurones dans une seule zone cérébrale ; une ICM plus généraliste exigerait des échantillons provenant de dizaines, voire de centaines de milliers de neurones, potentiellement répartis dans plusieurs régions du cerveau. Elle créerait donc de nouveaux défis en matière d'informatique et d'analyse des données. Il faut ajouter à cela les problèmes d'ingénierie et de chirurgie associés à la fabrication, et à l'implantation, de structures complexes en trois dimensions intégrant des dispositifs électroniques. Mais surtout, « nous ne comprenons pas assez bien la répartition des représentations d'informations et leur traitement par le néocortex pour pouvoir estimer, autrement que de façon rudimentaire, ce qu'une séquence d'activités particulière peut bien "vouloir dire"² ».

9. Sur le plan technologique du moins, certains chercheurs anticipent des progrès significatifs dans les années à venir, grâce à ce qu'ils nomment la « nanorobotique neuronale ». Introduits dans le système sanguin du sujet, des dispositifs microscopiques franchiraient si nécessaire la barrière hématoencéphalique pour localiser des types spécifiques de cellules cérébrales ou de structures cellulaires et s'y fixer. Ils enregistreraient et/ou stimuleraient alors l'activité neuronale pour « offrir un système *in vivo* non destructeur, en temps réel, sûr, de longue durée et pratiquement autonome, capable de réaliser la première ICM fonctionnelle » – en d'autres termes, de surmonter les obstacles évoqués ci-dessus. « De tels systèmes d'ICM humains pourraient

¹ Pour plus d'information sur les différents types de neurotechnologies, voir par exemple « iHuman: Blurring lines between mind and machine », Royal Society, septembre 2019.

² Mitrasinovic et autres, « Silicon Valley new focus on brain computer interface: hype or hope for new applications? », F1000Research 2018, 7:1327 (mis à jour le 21 janvier 2019).

révolutionner les communications homme-machine et ouvrir la voie à d'importantes améliorations cognitives chez l'être humain [...]»³ »

10. Ces propositions ne sont pas aussi farfelues qu'il y paraît. Sur des souris, des nanoparticules électromagnétiques ont déjà été utilisées pour contrôler des zones cérébrales prédéfinies, et des points quantiques de carbone fluorescents ont permis de cibler certaines cellules cérébrales pour en obtenir une image. Un cerveau humain a été relié à la moelle épinière d'un rat anesthésié ; un autre cerveau humain a guidé une blatte le long d'un circuit en forme de S via une stimulation électrique de ses antennes. Plusieurs cerveaux ont été connectés pour accomplir des tâches coopératives : on a constaté que quatre cerveaux de rats interconnectés réussissaient des opérations de calcul mieux qu'un seul cerveau. La connexion de trois cerveaux humains, appelée « BrainNet », a permis à trois sujets humains de prendre des décisions collaboratives via une communication non invasive de cerveau à cerveau (par EEG et SMT). Selon l'étude de la Royal Society, les neurotechnologies telles qu'on les imagine à l'avenir « sont encore très loin de nous, mais non totalement impossibles. N'oublions pas à quel point un alunissage ou l'internet auraient paru futuristes et lointains en 1950, alors que beaucoup de foyers n'avaient pas encore le téléphone [...] »

11. Les progrès des technologies ICM ont également été fortement stimulés par le développement rapide de l'intelligence artificielle (IA) observé ces dix dernières années. L'analyse des images du cerveau – en particulier celles obtenues par IRM – par les algorithmes d'apprentissage automatique a renforcé notre compréhension de la structure fonctionnelle du cerveau et notre faculté à décoder l'activité neuronale pour reconstruire les schémas de pensée qu'elle représente. Cela a permis aux chercheurs de reconstruire des images à partir de scanographies IRM de personnes regardant des bandes-annonces de films (ce que faisaient le professeur Gallant et son équipe dès 2011) ou de convertir des signaux cérébraux captés par ECoG en parole de synthèse (ce que fait M. Chartier – voir ci-dessous). Dans ce domaine comme dans d'autres, du diagnostic médical aux véhicules autonomes, les progrès de l'IA ont été déterminants. Pour aller plus loin, vous trouverez en annexe du présent rapport des informations générales sur l'IA, dont une description et un examen des principes éthiques applicables, ainsi que le texte de la recommandation pertinente du Comité des Ministres CM/Rec(2020)1.

3. Applications

12. Comme nous l'avons indiqué plus haut, les premières applications des ICM consistaient à contrôler des prothèses ou des robots simples à l'aide de signaux cérébraux⁴. Les progrès des capteurs et de la puissance informatique ont rendu possibles de nouvelles applications, plus complexes.

13. La recherche sur les troubles psychologiques et les troubles neurologiques gériatriques, par exemple, attire des financements considérables. Il n'est donc pas surprenant que la plupart des recherches actuelles sur les neurotechnologies concernent les applications médicales. Josh Chartier, de l'Université de Californie à San Francisco, m'a parlé du projet sur lequel il travaille pour détecter les schémas d'activité neuronale associés au contrôle du tractus vocal pendant la parole. Les signaux ainsi obtenus peuvent ensuite être décodés et utilisés pour produire une « parole de synthèse ». Cette approche permet de contourner l'immense difficulté qui consiste à identifier l'activité neuronale associée à des mots spécifiques (par rapport à l'activité associée à l'intention de prononcer ces mots). La recherche est menée sur des patients atteints d'épilepsie sévère, sur lesquels on a placé un « réseau neuronal » destiné à surveiller les crises et qui ont accepté que ce dispositif soit aussi utilisé dans le cadre de cette recherche. Très intrusive, cette neurotechnologie de synthèse vocale serait réservée aux personnes se trouvant dans des situations extrêmement difficiles, comme le syndrome d'enfermement. M. Chartier a admis la possibilité qu'une technique puisse fonctionner avec des personnes encore capables de parler, mais pas avec des personnes ne pouvant plus parler depuis longtemps et ne présentant peut-être plus les mêmes schémas détaillés et cohérents d'activité neuronale. Il a toutefois suggéré qu'avec de la pratique, les patients atteints du syndrome d'enfermement pourraient parvenir à « ré-entraîner » leur cerveau pour produire des schémas d'activité susceptibles d'être décodés et utilisés pour générer une « parole de synthèse ». L'idée première – obtenir un langage articulé à partir de l'activité neuronale – illustre à quel point la compréhension du fonctionnement du cerveau a progressé, grâce à l'amélioration de la technologie qui permet une lecture du cerveau. La nature intrusive de cette technologie et le recours à un effort mental conscient pour contrôler le tractus vocal peuvent laisser penser que certaines préoccupations éthiques – ayant trait à la vie privée et à l'intégrité mentale – sont ici moins prépondérantes.

³ Martins et autres (2019), « Human Brain/Cloud Interface », *Front. Neurosci.* 13:112.

⁴ Voir par ex. Chapin et autres, « Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex » (expérience réalisée sur un rat), *Nat. Neurosci.* juillet 1999 ; 2(7):664-670.

14. L'entreprise de José Carmena, Iota Biosciences, développe des dispositifs bioélectroniques sans batterie fonctionnant par ultrasons et ne mesurant que quelques millimètres – on les appelle « poussière neuronale » –, capables de se connecter directement au système nerveux central. Ces dispositifs peuvent collecter des données précises ou stimuler directement les nerfs et pourraient être utilisés pour diagnostiquer et traiter des affections allant de l'arthrite aux maladies cardiovasculaires. Iota cherche à réduire la taille de ses dispositifs à l'échelle inframillimétrique, ce qui pourrait à terme permettre leur utilisation dans la technologie d'interface cerveau-machine.

15. Certaines entreprises fournissent déjà des services ou des produits directement aux consommateurs. BrainKey, qui m'a été présenté par le PDG de l'entreprise, Owen Phillips, est en train de constituer une base de données de scanographies IRM obtenues principalement à partir de banques de données publiques (comme la Biobank au Royaume-Uni) et d'autres bases provenant de clients privés. Aujourd'hui, le service proposé par BrainKey est essentiellement descriptif/analytique et il contient des statistiques et un modèle 3D imprimé du cerveau de l'individu. L'objectif est de développer des outils de diagnostic qui fonctionnent grâce à l'IA. Si M. Phillips est opposé au fait d'autoriser les employeurs à imposer aux candidats à un poste qu'ils passent un IRM qui serait ensuite analysé pour tenter de prédire leurs performances professionnelles futures, il admet que la technologie pourrait être utilisée à cette fin.

16. L'entreprise canadienne InteraXon Inc. commercialise un multi-capteur EEG appelé Muse, destiné à aider les utilisateurs pendant la pratique de la méditation en enregistrant leur activité neuronale (ainsi que leur immobilité physique via un accéléromètre et leur rythme cardiaque via un moniteur optique de fréquence cardiaque). Ce dispositif s'accompagne d'une application pour smartphone qui fournit aux utilisateurs des informations sur leur pratique et leurs progrès. Les utilisateurs sont informés du fait qu'ils conservent le contrôle total des données EEG générées lors de l'utilisation de leur dispositif Muse, mais qu'ils peuvent participer à un programme de recherche qui partage les données EEG (et d'autres capteurs Muse) anonymisées avec des « tiers impliqués dans des recherches liées à l'amélioration de la compréhension scientifique du cerveau/corps, ou à l'amélioration des produits et/ou de l'offre de meilleurs services et expériences »⁵.

17. OpenBCI adopte une approche différente, en vendant une plateforme matérielle *open source* (qui comprend des EEG, des EMG pour la détection de l'activité musculaire, des ECG pour la fonction cardiaque et toutes sortes d'autres composantes associées) qui peut être utilisée avec des logiciels libres et gratuits pour divers projets. Comme l'indique le site internet d'OpenBCI, « nous travaillons à exploiter la puissance du mouvement *open source* pour accélérer l'innovation éthique des technologies d'interface homme-machine »⁶. Cela peut s'apparenter au modèle de « démocratisation » du développement neurotechnologique proposé par M. Ienca (voir ci-dessous).

18. Des recherches financées par Facebook ont utilisé l'ECOG pour comprendre, uniquement à partir de signaux neuronaux, à la fois ce que des sujets entendaient et ce qu'ils avaient l'intention de dire. D'après leurs auteurs, cela augure des « systèmes vocaux neuroprothétiques pour les personnes incapables de communiquer »⁷ ou, comme l'a annoncé sur Twitter Andrew Bosworth, vice-président de Facebook, « un appareil portable qui permet de taper du texte juste en y pensant », commentaire qui révèle peut-être mieux encore l'intérêt commercial de Facebook pour cette technologie. Si les recherches menées par M. Chartier, par exemple, ont déjà donné des résultats prometteurs grâce à la technologie invasive ECOG, il est difficile de voir comment des dispositifs « portables » non invasifs pourraient être aussi efficaces, du moins dans un futur proche.

19. Les enjeux commerciaux sont aussi présents dans ce qu'on appelle le « neuromarketing », « domaine récent et interdisciplinaire qui dépasse les frontières traditionnelles entre neurosciences, neuro-économie et recherches en marketing [...], dont le premier but est d'affiner les stratégies de marketing pour améliorer les ventes »⁸. Une étude a montré que chez les sujets qui préféraient la saveur du Pepsi à celle du Coca-Cola

⁵ Citation tirée du Contrat de licence utilisateur final Muse le 19 juillet 2020.

⁶ D'une certaine manière, les ingénieurs de Netflix ont réussi à combiner ces approches en piratant un bandeau Muse pour pouvoir l'utiliser comme une sorte de télécommande mains libres, c'est-à-dire faire défiler un menu (en utilisant l'accéléromètre pour détecter les mouvements de la tête) et sélectionner un programme (en utilisant l'EEG pour détecter une activité neuronale, en l'espèce le fait de se concentrer sur le mot « lecture »).

⁷ Moses et autres, « Real-time decoding of question-and-answer speech dialogue using human cortical activity », Nature Communications, 2019. D'autres chercheurs ont obtenu des résultats similaires à l'aide de la même technologie, par ex. Salar et autres, « Classification of Articular Movements and Movement Direction from Sensorimotor Cortex Activity », Scientific Reports / Nature Research, octobre 2019.

⁸ Ulman, Cakar et Yildiz, « Ethical Issues in Neuromarketing: "I Consume, Therefore I am!" », Sci Eng Ethics (2015) 21:1271-1284.

lors de tests en aveugle, une zone cérébrale, le putamen ventral, était particulièrement activée ; lorsque les sujets connaissaient la marque, presque tous préféraient le Coca, et c'est une autre partie du cerveau qui s'activait davantage (le cortex préfrontal, associé à la conscience de soi et qui, dans ce contexte, paraissait l'emporter sur les papilles de ceux qui préféraient auparavant le Pepsi). Une meilleure compréhension de ce mécanisme pourrait permettre aux publicitaires de développer, et de tester, des stratégies de marketing ciblant des mécanismes subconscients particulièrement sensibles aux images ou à la marque, et non aux qualités intrinsèques du produit.

20. Les neurotechnologies et les ICM pourraient présenter un intérêt particulier pour les procédures judiciaires et pénales. « Les neurosciences commencent à être utilisées pour tenter d'évaluer de manière objective des aspects par nature subjectifs. Distinction entre mensonge et vérité, capacités mentales, souffrance, fiabilité des souvenirs constituent des domaines d'étude intéressants pour la justice civile et pénale. Ces tentatives de lecture de l'esprit sont en train, théoriquement, de devenir possibles »⁹. Certaines de ces technologies sont même disponibles sur le marché, bien que leur application aux procédures judiciaires ne soit pas encore acceptée : par exemple, les entreprises Cephos et No Lie MRI commercialisent des produits utilisant l'IRMf pour estimer si une personne dit ou non la vérité. Par ailleurs, une étude suggère que l'activation du cortex cingulaire antérieur, associé au contrôle des impulsions, laisserait présager une nouvelle arrestation (récidive)¹⁰. De tels usages soulèveraient de sérieuses questions de respect des droits et des garanties procédurales.

21. Les recherches sur les neurotechnologies ICM visent des objectifs non seulement civils et commerciaux, mais aussi sécuritaires et militaires. La DARPA (Defense Advanced Research Projects Agency, États-Unis) est particulièrement active dans ce domaine. La plupart de ses travaux poursuivaient à l'origine des objectifs médicaux, l'accent étant mis sur les aspects les plus pertinents pour les forces armées. Les programmes « Revolutionising prosthetics » et « Reliable neural interface technology » (RE-NET), par exemple, visaient à accélérer le développement de prothèses contrôlées par ICM (neuroprothèses) ; le programme « Reorganisation and plasticity to accelerate injury recovery » (REPAIR) cherchait à rétablir les fonctions neuronales et comportementales après une atteinte du système nerveux ou une privation sensorielle.

22. D'autres recherches financées par la DARPA peuvent s'avérer plus délicates sur le plan éthique. Les programmes suivants en sont un exemple : « Restorative encoding memory integration neural device » (REMIND) a développé une technologie susceptible d'accroître, ou de réduire, la capacité d'un sujet à garder des événements en mémoire. « Accelerated learning » visait à « révolutionner l'apprentissage dans l'environnement militaire » (adresse au tir, par exemple). « Narrative networks » (N2) consistait à détecter l'activité cérébrale sous l'influence des récits, pour « accélérer et améliorer la communication des informations dans les opérations d'information à l'étranger », et à créer des technologies ICM « plaçant dans la même boucle celui qui écrit l'histoire et celui qui en prend connaissance, les réactions neuronales à un stimulus narratif devenant capables d'influer sur le récit ». Cela ouvrirait la voie à « des récits optimaux, adaptés à un individu ou à un groupe spécifique ». Le programme « Neurotechnology for intelligence analysts » (NIA) visait à développer de nouvelles ICM non invasives pour rendre plus efficaces et productives les analyses d'images, grâce à la détection des réactions neuronales lors de la perception d'un « point d'intérêt » sur une image. Via une approche similaire, « Cognitive technology threat warning system » (CT2WS) cherchait à renforcer la capacité, lors de la surveillance d'un site, à détecter les menaces et à y réagir. La DARPA travaille également à la mise au point d'un casque d'EEG à faible coût, avec pour but de « réunir les efforts de collecte des données au service de la neuro-imagerie ». Plus récemment, le programme « Systems-based neurotechnology for emerging therapies » (SUBNETS) a été lancé dans l'espoir de produire de nouveaux implants d'enregistrement et de stimulation de l'activité neuronale, en vue de traiter les troubles neuropsychiatriques et neurologiques, notamment parmi les anciens soldats souffrant de troubles mentaux. « Restoring active memory » (RAM) vise à rendre la mémoire à des patients souffrant de maladies incurables entraînant des défaillances à ce niveau. Signe peut-être d'une prise en compte des éventuels problèmes d'éthique, la DARPA aurait travaillé à ces deux derniers programmes en étroite collaboration avec la Food and Drug Administration¹¹. Les experts que j'ai rencontrés en Californie estiment que les projets financés par la DARPA sont encore loin d'atteindre les objectifs annoncés, lesquels ont souvent été délibérément exprimés en termes extrêmement ambitieux, avec l'assurance d'aller au-delà de ce qui est actuellement possible et dans l'espoir qu'au minimum quelques progrès utiles seront réalisés.

⁹ Goodnough et Tucker, « Law and Cognitive Neuroscience », *Annu. Rev. Law Soc. Sci.* 2010. 6:61–92.

¹⁰ Aharoni et autres, « Neuroprediction of future arrest », *PNAS Early Edition*, 2013.

¹¹ Miranda et autres, « DARPA-funded efforts in the development of novel brain-computer interface technologies », *Journal of Neuroscience Methods* 244 (2015) 52-67.

23. La mission de Neuralink, telle que la présente Elon Musk, est d'atteindre « une forme de symbiose avec l'intelligence artificielle » ; M. Musk est en effet convaincu que l'intelligence artificielle constitue « notre plus grande menace existentielle », « peut-être plus dangereuse que les armes nucléaires ». On trouve les mêmes préoccupations chez les tenants de la « nanorobotique neuronale » envisagée comme un moyen de connecter le cerveau humain au cloud : elle pourrait « être bénéfique pour l'humanité en contribuant à réduire les sérieux risques existentiels posés par l'émergence d'une intelligence artificielle généralisée. [La nanorobotique neuronale y parviendrait] en permettant la création de technologies d'augmentation de l'humain qui viendraient contrebalancer ces risques [...] »¹²

24. Alan Mardinly, directeur de recherche chez Neuralink, m'a expliqué que l'objectif de « symbiose » d'Elon Musk était loin de se concrétiser, de nombreuses inconnues techniques et scientifiques restant à régler avant d'y parvenir. L'objectif immédiat de l'entreprise consiste à produire des dispositifs médicaux, notamment pour aider les patients souffrant de lésions de la moelle épinière. Comme la technologie de « dentelle neuronale » intra-corticale de Neuralink implique des milliers d'électrodes, disposées le long de « fils » suspendus à une puce de 4x4 mm, elle offre une résolution suffisante pour contrôler les dispositifs. Pour M. Mardinly, la dentelle neuronale est la meilleure technologie disponible à l'heure actuelle pour l'interface cerveau-machine, même si sa nature invasive signifie que le spectre d'applications sera forcément limité. Il la juge néanmoins préférable au réseau d'électrodes standard Utah Array : la dentelle neuronale utilise des électrodes flexibles (par opposition aux électrodes rigides de l'Utah Array) et est donc à la fois moins perturbatrice pour le tissu cérébral et plus durable, car elle ne provoque pas de cicatrices localisées. Elle offre en outre une meilleure résolution, avec 8 à 16 électrodes par tige très étroite (contre une seule électrode sur une tige relativement épaisse pour l'Utah Array). À terme, la dentelle neuronale pourrait être utilisée pour implanter jusqu'à 100 000 électrodes, bien que cela implique de recourir à un système d'insertion intégré et automatisé. En résumé, Neuralink vise à créer une technologie sûre, stable et évolutive. Neuralink travaille déjà avec la Food and Drug Administration des États-Unis sur une « étude préliminaire de faisabilité » qui porte davantage sur les questions de sécurité – telles que la cybersécurité/le piratage, les implications en matière de données/confidentialité et le contrôle des entrées – que sur les problématiques éthiques.

25. Toujours en envisageant l'avenir (possible), les partisans de la nanorobotique neuronale imaginent de nombreuses applications à cette technologie, en conjonction avec l'infrastructure mondiale du « cloud » : réalité virtuelle totalement immersive et impossible à distinguer de la réalité ; réalité augmentée, des informations sur le monde réel étant directement surimposées sur la rétine ; interprétation en direct de langues étrangères, accès à de nombreuses formes d'informations en ligne, et même « expérience totalement immersive et en temps réel d'épisodes de la vie de tout volontaire humain sur la planète, via une "observation transparente" non intrusive ».

26. Dans sa contribution écrite au présent rapport, M. Winnickoff présente la taxonomie des neurotechnologies qui figure ci-dessous (N.B. j'ai légèrement modifié sa proposition à des fins de simplicité et de pertinence). Si « l'interface cerveau-machine » semble être en relation avec la fonction primaire d'une technologie particulière, elle englobe évidemment aussi les fonctions de lecture et d'intervention/modulation. Comme nous l'avons déjà indiqué dans ce rapport, les technologies ICM peuvent être utilisées dans tous les domaines cités.

- Domaines d'utilisation, entre autres :
 - o Clinique / médical (neurologie/neurochirurgie, psychiatrie, rééducation, traitement de la douleur)
 - o Activité professionnelle (formation, performance)
 - o Militaire (renseignements/interrogatoires, armes/perfectionnement des combattants)
 - o Grand public (utilisation directe par le consommateur/bricolage, éducation, bien-être/mode de vie/loisirs)
- Fonction primaire de la technologie, dont :
 - o Déchiffrement/lecture du cerveau (imagerie, modélisation/cartographie)
 - o Fonction d'intervention/modulation cérébrale
 - o « Construction » du cerveau (interface cerveau-machine, neuroprothèses)
 - o Fonction dérivée (inspirée par le cerveau, par exemple les réseaux neuronaux artificiels)
- Objectifs de santé – prévenir, réparer, remplacer, augmenter, etc.

¹² Martins et autres, op. cit.

27. Comme le résume l'étude de la Royal Society :

« Des implants, casques, bandeaux ou autres appareils pourraient nous aider à mieux mémoriser, à apprendre plus vite, à parvenir plus vite à prendre de meilleures décisions, à résoudre des problèmes de manière objective [...]. Relier des cerveaux humains à des ordinateurs, en exploitant la puissance de l'intelligence artificielle, nous permettrait de fusionner l'intelligence émotionnelle des humains et leur capacité à prendre des décisions avec la puissance informatique de traitement des mégadonnées, créant une nouvelle forme d'intelligence collaborative. Un certain degré de télépathie deviendrait possible : nous pourrions parler non seulement sans parler, mais aussi sans mots, en accédant aux pensées de l'autre au niveau conceptuel [...]. On pourrait communiquer de cerveau à cerveau non seulement des pensées, mais aussi des expériences sensorielles [...]. Grâce à des améliorations physiques et mentales, les membres de la police ou de l'armée pourraient protéger la population en voyant mieux dans l'obscurité, en ressentant des présences et en réagissant plus vite [...] »

28. Il convient néanmoins de garder à l'esprit que la technologie actuelle est très loin d'atteindre la quasi-totalité de ces objectifs, du moins pas de manière concrètement utile ou inquiétante. Sans amélioration significative de notre compréhension de la structure et du fonctionnement du cerveau, les possibilités les plus spectaculaires de l'ICM resteront purement spéculatives. Certains experts doutent sérieusement du potentiel des neurotechnologies comme base d'interaction avec les ordinateurs. Greg Corrado (Google) fait valoir que le système nerveux humain a évolué pour interagir avec le monde extérieur en utilisant les sens naturels, qui resteraient la principale interface avec les ordinateurs : en particulier, les yeux et les oreilles pour recevoir des informations, et le bout des doigts et les voies vocales pour les transmettre. Les améliorations de l'interface homme-machine se produiraient donc du côté de l'ordinateur, comme avec le traitement de la parole naturelle (c'est-à-dire la capacité d'un ordinateur à « comprendre » la parole humaine). Selon lui, les possibilités les plus prometteuses pour l'interface bioélectronique résident dans la stimulation directe des nerfs musculaires et des implants cochléaires et rétinien pour entrer des informations (stimuler), et dans celle du cortex moteur pour lire des informations. D'après M. Corrado, l'état actuel des technologies ICM ne comporte aucun risque éthique – un point de vue partagé par un grand nombre de personnes que j'ai rencontrées en Californie. Cela dit, son utilisation abusive pourrait soulever des préoccupations éthiques, par exemple si la technologie IRMF était utilisée pour prévoir la personnalité, les propensions criminelles ou les capacités professionnelles des individus – ce que M. Corrado considère comme une sorte de phrénologie, une théorie discréditée selon laquelle les traits de personnalité et les capacités intellectuelles d'un individu peuvent être déterminés en mesurant les caractéristiques physiques de son crâne.

4. Préoccupations

29. L'état actuel des technologies ICM ne suscite peut-être pas de préoccupations immédiates qui dépassent manifestement la portée des cadres éthiques existants (notamment l'éthique médicale et les réglementations relatives à la protection de la vie privée et des données), mais le rythme de leur évolution reste supérieur aux préoccupations éthiques et le seuil de rupture qui pourrait amener à un décrochage du droit est proche. La marche du progrès technologique est inexorable et certains acteurs chercheront inévitablement à développer des applications qui posent sérieusement problème sur le plan éthique, que ce soit sur la base de technologies futures ou d'usages de technologies existantes. Il s'agit tout simplement de prendre des précautions pour anticiper ce qui pourrait arriver à l'avenir et pour fixer des limites qui détournent la recherche des domaines vraisemblablement néfastes ou dangereux et l'orientent vers des applications positives qui ne menacent ni les droits individuels ni les sociétés démocratiques. Certaines de ces questions éthiques n'entrent pas dans le champ du présent rapport, comme le fait qu'une connexion établie avec un cerveau « modifie inévitablement ce cerveau »¹³. Je privilégierai ici les préoccupations liées aux droits de l'homme, fondées sur la volonté d'assurer le respect de la dignité humaine et des caractéristiques essentielles de l'être humain, envisagé comme un agent moral autonome.

30. Accéder aux processus neuronaux qui sous-tendent la pensée consciente, c'est accéder à un niveau du soi qui, par définition, ne peut être consciemment ni filtré, ni dissimulé – la pire atteinte possible à la vie privée. Même les technologies actuelles de neuromarketing, associées à la diffusion massive de messages ciblés via la publicité et les réseaux sociaux, pourraient avoir un effet dévastateur sur la liberté de choix et les processus démocratiques. Les ICM pourraient servir à créer des détecteurs de mensonges très perfectionnés, jugés si fiables que leurs résultats pourraient être acceptés comme preuves dans les procédures pénales. Cela porterait cependant atteinte à la protection contre le fait de contribuer à sa propre mise en accusation, puisque les suspects concernés ne pourraient plus refuser de livrer des informations lors des interrogatoires.

¹³ Amy L. Osborn et Bijan Pesaran, « Parsing learning in networks using brain-machine interfaces », *Curr Opin Neurobiol.* octobre 2017 ; 46: 76-83.

Dans le même temps, les ICM pourraient servir à créer de faux souvenirs ou à modifier ou supprimer des souvenirs réels, si bien que les témoignages humains perdraient leur caractère de preuves fiables.

31. Les recherches « à double usage », comme celles que finance la DARPA (voir ci-dessus), inquiètent particulièrement. Ce terme désigne à la fois l'application de technologies civiles à des fins militaires, policières/judiciaires ou de sécurité nationale, et la possibilité que des acteurs, y compris non étatiques, détournent certaines technologies à des fins malveillantes. À cet égard, « le fait que ces technologies dépendent de l'informatique et du numérique les expose également aux cyberattaques »¹⁴, et les conséquences pour l'individu deviendront d'autant plus redoutables que les technologies ICM, qui affectent les processus neuronaux, gagneront en complexité, en puissance et en précision.

32. Se posent également des questions d'accès et d'équité. Les risques apparaissent aujourd'hui que l'accès à de puissantes améliorations cognitives à des fins non cliniques puisse dépendre du niveau de richesse – certains pourront se les offrir, d'autres non. Ainsi pourraient apparaître deux catégories d'êtres humains, les « augmentés » et les autres. Ces technologies seraient susceptibles également de laisser le choix des réalisations possibles et de bénéficiaires ciblés au marché non réglementé ou aux intérêts de gouvernements potentiellement autocratiques.

33. Enfin, « sur un plan plus philosophique, certains craignent que l'usage généralisé des interfaces neuronales ne place les décisions humaines sous la coupe d'un "neuro-essentialisme" qui réduirait chaque personne à son cerveau, et nos choix à de simples séries de processus neurobiologiques ; ce serait la fin de la liberté d'action individuelle et de la responsabilité morale [...] »¹⁵. La question se pose, en effet : à quoi sert de développer les ICM pour préparer l'humanité aux dangers d'une intelligence artificielle généralisée si, ce faisant, nous abolissons des qualités essentielles à chaque existence humaine ?

34. Le professeur Rafael Yuste de l'Université Columbia à New York a exprimé ces préoccupations au nom du « groupe Morningside », qui réunit 25 experts des neurotechnologies et des domaines connexes. Dans un discours prononcé en 2019 lors de la réunion de l'Union interparlementaire à Doha (Qatar), le professeur Yuste a décrit comment une éventuelle mauvaise utilisation et l'absence de réglementation pourraient entraîner des problèmes dans les cinq domaines suivants¹⁶.

- Identité personnelle : « plus nous sommes connectés au réseau par des dispositifs ou des interfaces cerveau-machine, plus nous diluons notre propre identité. La dépendance que nous constatons actuellement à l'égard de nos appareils est un avant-goût de ce qui nous attend : à mesure que nous augmentons la bande passante de notre connexion au réseau en utilisant des interfaces cerveau-machine non invasives, nous nous y dissolvons de plus en plus. »
- Libre arbitre : « si nous utilisons des informations et des algorithmes extérieurs pour prendre des décisions, nous renonçons à notre propre liberté d'action. Qui prend la décision ? [...] que se passera-t-il lorsque nous aurons un GPS de vie qui nous conseillera sur ce que nous devons faire à tout moment ? »
- Respect de la vie privée sur le plan mental : « si les données du cerveau sont accessibles et déchiffrables, alors nos processus mentaux et nos pensées seront accessibles à autrui. Même les pensées dont nous n'avons pas conscience – le subconscient – pourraient être décryptées. Nous pensons que les données du cerveau doivent être protégées avec la même rigueur législative que les organes. En fait, nos données cérébrales sont un organe – pas un organe physique, mais un organe mental – et il devrait être interdit d'en faire commerce, car elles représentent ce que nous sommes. »
- Augmentation cognitive, y compris l'apprentissage augmenté qui « peut permettre à certains groupes sociaux dans certains pays de renforcer leurs capacités mentales et physiques en leur donnant accès à des algorithmes externes et à la robotique pour la vie quotidienne. Nous pensons que la garantie du principe de justice dans le développement et le déploiement de ces technologies devrait assurer une égalité d'accès et que leur utilisation à des fins militaires devrait être strictement encadrée. »
- Protection contre les préjugés et la discrimination : « comme les algorithmes utilisés en IA fonctionnent souvent selon des préjugés implicites, ces technologies pourraient accidentellement implanter ces préjugés dans nos processus cérébraux. Il serait terrible de réduire à néant nos progrès vers l'égalité et la justice en diffusant des préjugés grâce aux nouvelles technologies. »

¹⁴ Ienca, Jotterand et Elger, « From Healthcare to Warfare and Reverse: How Should We Regulate Dual-Use Neurotechnology? », *Neuron* 97, 17 janvier 2018, 269-274.

¹⁵ The Royal Society, op. cit.

¹⁶ Voir https://www.ipu.org/sites/default/files/documents/yuste_speech_final_0.pdf.

5. Réactions

35. Le monde de la recherche s'accorde largement sur le besoin d'anticiper la réglementation des neurotechnologies émergentes, dont les ICM, à travers une approche « précoce et fréquente », comme le préconise l'étude de la Royal Society. Les acteurs industriels soutiennent également cette approche. Pour citer Mark Chevillet, directeur de recherche aux Facebook Reality Labs, « nous ne pourrions ni anticiper, ni résoudre seuls toutes les questions éthiques soulevées par ces technologies. Ce que nous *pouvons* faire, c'est surveiller leurs progrès et informer la population lorsqu'elles dépassent ce qu'on croyait possible jusqu'alors. La conception neuroéthique est l'un des piliers de notre programme : nous voulons être transparents sur nos domaines de travail, pour que les gens puissent nous dire ce que ces technologies ont d'inquiétant pour eux »¹⁷. D'autres plaident pour une coordination entre les recherches en neuroéthique et dans les domaines voisins, comme l'IA – ces disciplines, ayant en commun certaines technologies, devraient se voir appliquer une approche commune des principes éthiques¹⁸. Certains avancent des arguments en faveur d'un « cadre de *neurosécurité* [...] conçu et mis en œuvre pour assurer une sécurité maximale sur tout le continuum allant de la recherche scientifique à la société (et inversement) [...], particulièrement axé sur l'anticipation et la détection rapide des menaces propres aux neurotechnologies, surtout celles qui concernent l'aspect mental [...]. [Ce cadre] devrait inclure, au minimum, trois grands niveaux de garanties : interventions réglementaires adaptées, codes de déontologie et activités de sensibilisation »¹⁹.

36. En 2013, le Conseil Nuffield sur la bioéthique a publié un rapport détaillé intitulé « Novel neurotechnologies: intervening in the brain ». Ce rapport, qui traite principalement (mais pas exclusivement) des applications médicales et des neurotechnologies, note que « le cerveau a un statut spécial dans la vie humaine qui le distingue des autres organes. Son fonctionnement sain joue un rôle central dans le fonctionnement de notre corps, dans nos capacités d'agir librement et de façon autonome, dans la conception que nous avons de nous-mêmes et de nos relations avec les autres – et donc dans notre aptitude à mener une vie épanouie. Cela signifie que les nouvelles neurotechnologies que nous examinons dans le présent rapport, qui toutes interviennent dans le cerveau, suscitent des préoccupations éthiques et sociales que ne font pas naître dans la même mesure d'autres technologies biomédicales ». Le rapport propose un cadre éthique établi en trois étapes :

- Les principes fondamentaux de bienfaisance et de précaution découlant « d'une tension entre le besoin et l'incertitude ». Alors que les troubles cérébraux graves génèrent une grande souffrance, il n'existe pas d'(autres) interventions efficaces. Dans le même temps, les avantages et les risques des neurotechnologies ne sont pas encore pleinement compris du fait de leur nouveauté et d'une compréhension incomplète du fonctionnement du cerveau lui-même. « Le statut particulier du cerveau fournit donc à la fois une raison de faire preuve de bienfaisance en intervenant lorsqu'une blessure ou une maladie provoque des troubles cérébraux, et une raison de nous montrer prudents lorsque nous ne sommes pas certains des effets d'une telle intervention. »
- Les implications des principes de bienfaisance et de précaution doivent être examinées à la lumière d'une série de cinq « intérêts » : la sécurité, les conséquences (involontaires) sur la vie privée, l'autonomie (tant dans les décisions relatives au traitement que dans le contexte plus large de la vie des patients), l'équité de l'accès aux nouveaux traitements et la confiance dans les nouvelles technologies.
- Le rapport propose par ailleurs trois « vertus » pour encadrer les pratiques des acteurs dans ce domaine, à savoir l'inventivité (que ce soit dans l'innovation ou dans l'élargissement de l'accès), l'humilité (admettre les limites de la connaissance et de nos capacités à utiliser la technologie pour soulager les troubles cérébraux) et la responsabilité (par le recours à des pratiques de recherche solides et en s'abstenant de toute allégation exagérée ou prématurée à propos des neurotechnologies).

37. Des approches similaires se retrouvent dans d'autres propositions de cadres éthiques visant à réglementer l'ICM et les autres neurotechnologies. Une équipe de chercheurs, par exemple, a proposé une liste de « Principes directeurs de la neuroéthique » :

- i. D'abord et avant tout, évaluer les aspects de sécurité.
- ii. Anticiper les enjeux liés à la capacité, à l'autonomie et à la liberté d'action.
- iii. Protéger le caractère privé et confidentiel des données neuronales.

¹⁷ « Imagining a new interface: Hands-free communication without saying a word », Tech@facebook, 30 juillet 2019.

¹⁸ Marcello Lenca, « Neuroethics Meets Artificial Intelligence », theneuroethicsblog.com, 1^{er} octobre 2019.

¹⁹ Lenca, Jotterand et Elger, op. cit.

- iv. S'attaquer aux éventuels usages malveillants des neurotechnologies et des outils neuroscientifiques.
- v. Agir avec prudence au moment d'appliquer les neurotechnologies et les outils neuroscientifiques à des usages médicaux ou non médicaux.
- vi. Recenser et traiter les inquiétudes spécifiques de la population au sujet du cerveau.
- vii. Encourager l'éducation du public et le dialogue.
- viii. Agir avec équité et partager les avantages tirés de la recherche en neurosciences et des technologies qui en résultent²⁰.

38. La question du « double usage des technologies cognitives » (dont l'ICM) a été abordée par Marcello Lenca dans un article de 2018²¹. Notant que les « technologies cognitives ont le potentiel d'accélérer l'innovation technologique et d'apporter des avantages considérables aux individus et aux sociétés », M. Lenca considère que « du fait de leur double usage possible, elles peuvent être cooptées par des acteurs étatiques et non étatiques à des fins qui ne sont pas bénignes, notamment la cybercriminalité, le cyberterrorisme, la cyberguerre et la surveillance de masse. À la lumière de la crise démocratique mondiale récente, de la militarisation accrue de l'infosphère numérique et de la potentialisation simultanée des technologies cognitives (TC), il importe de concevoir de manière proactive des stratégies susceptibles d'atténuer les risques émergents et de conformer l'avenir des TC aux principes de base de la démocratie libérale dans des sociétés libres et ouvertes ».

39. M. Lenca propose donc une « démocratisation des TC », en adoptant des éléments des deux extrêmes de la gouvernance et de la réglementation, à savoir le pur laisser-faire et la réglementation stricte. La « démocratisation » partage avec l'approche réglementaire stricte l'idée que la nouveauté des TC en fait un domaine relativement récent et encore immature, dépourvu de consensus sur les concepts ou les politiques de base qui seraient nécessaires pour maximiser les avantages tout en minimisant les risques. Elle reconnaît également l'ampleur des possibilités et des risques qu'elles présentent – « la perspective d'influencer les capacités cognitives humaines, déterminant ainsi un effet non négligeable sur l'évolution culturelle humaine et les équilibres mondiaux » – et le fait que la nouveauté des TC signifie que « les sociétés humaines se trouvent aujourd'hui à un tournant crucial, où elles peuvent prendre des décisions proactives sur le type de coexistence qu'elles souhaitent établir avec ces technologies. Privilégier les approches du laisser-faire à ce stade du développement reviendrait à reporter les interventions de gestion des risques à un moment où les TC seront largement développées et utilisées, donc hermétiques à toute modification ». La démocratisation pourrait cependant admettre le point de vue défendu par les partisans du laisser-faire : « l'excès de réglementation peut (a) anéantir les avantages des technologies cognitives pour la société au sens large et, si elle est gérée par des gouvernements non démocratiques, (b) engendrer une concentration indésirable de pouvoir et de contrôle ».

40. Cette « démocratisation » des TC s'appuierait sur six principes normatifs – qui sont comparables aux composantes des cadres éthiques proposés pour la réglementation des neurotechnologies de façon plus générale :

- i. Prévention du contrôle centralisé – « principe selon lequel il est moralement préférable d'éviter tout contrôle centralisé sur les TC pour prévenir les risques associés à l'accumulation illimitée de capital, de pouvoir et de contrôle sur les technologies au sein de groupes organisés tels que les grandes entreprises ou les gouvernements [...]. Les interventions normatives visant à limiter ce risque de centralisation peuvent être perçues comme le pendant cyberéthique des lois antitrust ».
- ii. Ouverture – « principe consistant à promouvoir l'accès universel à (des éléments de) la conception ou au schéma directeur des technologies cognitives, et la redistribution universelle de cette conception ou de ce schéma directeur par un processus ouvert et collaboratif de production entre pairs ». La prévention du contrôle centralisé (principe I ci-dessus) et l'ouverture « sont des exigences essentielles pour rendre les capacités qui seront enregistrées par ou intégrées dans les technologies cognitives [...] accessibles à tous ». « Dans un sens plus abstrait, l'ouverture dans le contexte des TC consiste à enrichir chaque application à laquelle nous sommes connectés, sur n'importe quel appareil, à n'importe quel moment, avec les (composantes des) technologies cognitives. »
- iii. Transparence – « principe consistant à permettre au grand public de comprendre les processus internes des technologies cognitives ».
- iv. Inclusivité – « principe consistant à s'assurer qu'aucun groupe d'individus ou minorité n'est marginalisé ou laissé pour compte pendant le processus d'infiltration des technologies cognitives dans notre société [...]. Le principe d'inclusivité [s'applique à tout] préjugé social éthiquement

²⁰ Greely et autres, « Neuroethics Guiding Principles », *J.Neurosci.*, 12 décembre 2018, 38(50) :10586 – 10588.

²¹ « Democratizing cognitive technology: a proactive approach », Marcello Lenca, 19 juin 2018.

- pertinent qui peut émerger, intentionnellement ou non, pendant le développement des TC. Il peut s'agir de préjugés culturels, politiques, linguistiques, etc. »
- v. Approche centrée sur l'utilisateur – « les technologies cognitives émergentes doivent être conçues, développées et mises en œuvre en fonction des besoins et des choix personnels des utilisateurs [...]. Les utilisateurs finaux (caractérisés aussi largement que possible, conformément aux principes d'ouverture et d'inclusivité) [doivent être] associés à la conception, au développement et à la mise en œuvre des technologies cognitives sur un pied d'égalité. »
 - vi. Convergence – « au sens strict, la convergence est le principe d'interopérabilité, d'intercommunication et de facilité d'intégration entre toutes les composantes des technologies cognitives [...] [même si] une interopérabilité excessive peut entraîner une insécurité accrue des données [...]. Au sens plus large et plus abstrait, c'est aussi le principe de convergence des différents types de technologies cognitives – les neurotechnologies, en particulier – d'une part, et des systèmes d'intelligence artificielle, d'autre part ».

41. Concernant les neurotechnologies conçues spécifiquement pour des applications militaires, on note ce commentaire : « Bien que l'éthique ne semble pas justifier, pour l'heure, une interdiction totale des neurotechnologies militaires ou un moratoire à leur sujet, des interventions réglementaires plus souples et plus calibrées pourraient être nécessaires pour atténuer les risques *d'usage disproportionné des neurosciences en guise d'armes* ». Là encore, cela supposerait « d'assurer au plus vite un suivi et une évaluation approfondie des risques. Même si pour l'heure, les avantages semblent l'emporter sur les risques, des mécanismes préventifs devraient être mis en place pour détecter rapidement les évolutions de ce ratio »²². Il serait néanmoins nécessaire de ne pas se priver de la capacité de se défendre dans l'hypothèse d'un adversaire utilisant ces technologies.

42. La discipline baptisée « neuroéthique » présente de nombreux points communs avec la bioéthique. Dans ce dernier domaine, le Conseil de l'Europe a en 1997 adopté la « Convention d'Oviedo », sur les droits de l'homme et la biomédecine (STE n° 164), dont l'objectif est de « [protéger] l'être humain dans sa dignité et son identité et [de garantir] à toute personne, sans discrimination, le respect de son intégrité et de ses autres droits et libertés fondamentales à l'égard des applications de la biologie et de la médecine ». La Convention d'Oviedo affirme notamment que « l'intérêt et le bien de l'être humain doivent prévaloir sur le seul intérêt de la société ou de la science », oblige les Parties à prendre « les mesures appropriées en vue d'assurer, dans leur sphère de juridiction, un accès équitable à des soins de santé de qualité appropriée », et établit que « toute intervention dans le domaine de la santé, y compris la recherche, doit être effectuée dans le respect des normes et obligations professionnelles ». Elle comporte aussi des dispositions détaillées sur le consentement, les usages acceptables des tests prédictifs et les objectifs acceptables des interventions sur le génome humain.

43. En novembre 2019, peu après la présentation de ma note introductive à la commission, le Comité intergouvernemental de bioéthique « DH-BIO » du Conseil de l'Europe a adopté un plan d'action stratégique sur les droits de l'homme et les technologies en biomédecine (2020-2025). Ce document énonce que « [l']application de technologies émergentes et convergentes en biomédecine a pour conséquence l'effacement des frontières entre les sciences physiques et biologiques, entre ce qui relève de la thérapie et ce qui fait partie de la recherche, de même qu'entre les finalités médicales et non médicales. Si elles offrent des possibilités intéressantes dans le domaine de la biomédecine et au-delà, les technologies émergentes et convergentes posent aussi de nouveaux problèmes éthiques liés entre autres à l'identité, à l'autonomie, à la vie privée et à la non-discrimination ». S'appuyant sur la coopération (entre les organes du Conseil de l'Europe et avec les autres instances intergouvernementales pertinentes) et la communication (avec les parties prenantes externes), ce plan d'action s'articule autour de trois piliers : la gouvernance (basée sur les droits de l'homme, le débat public, la gouvernance démocratique et la transparence), l'équité (dans l'accès aux traitements et technologies innovants, pour lutter contre les disparités sanitaires dues à l'évolution sociale et démographique) et l'intégrité (y compris le renforcement de la participation des enfants, la protection des droits des enfants et la protection des droits des personnes atteintes de troubles mentaux).

44. En ce qui concerne l'ICM et les technologies connexes, le plan d'action note que « [l]es avancées dans le domaine des neurotechnologies, telles que la stimulation cérébrale profonde, les interfaces cerveau-ordinateur et les réseaux de neurones artificiels, ouvrent la perspective d'une compréhension, d'une surveillance et d'un contrôle accrus de l'esprit humain, mais posent aussi des questions de vie privée, de personnalité et de discrimination [...] Il est donc nécessaire d'évaluer si le cadre existant en matière de droits de l'Homme est suffisant pour y répondre 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

²² Ienca, Jotterand et Elger, op. cit.

encadrer les neurotechnologies. D'autres formes flexibles de bonne gouvernance peuvent constituer une alternative plus adaptée pour réglementer les neurotechnologies ». Sans surprise, le DH-BIO partage mes préoccupations et les grandes lignes de ma perception des possibilités et des risques que présentent les technologies ICM. Je salue et je soutiens ses travaux futurs sur les réponses les plus pertinentes, à savoir un renforcement ciblé du cadre des droits de l'homme (voir ci-dessous) et la mise en place d'un régime réglementaire souple qui puisse soutenir et orienter la recherche et le développement vers des fins positives et constructives.

45. D'autres organisations internationales s'intéressent également aux neurotechnologies émergentes. En décembre 2019, le Conseil de l'OCDE a adopté une recommandation sur l'innovation responsable dans le domaine des neurotechnologies. Cette recommandation « met l'accent sur l'importance (1) de valeurs fondamentales telles que la gestion responsable, la confiance, la sécurité et le respect de la vie privée dans ce contexte technologique, (2) du développement des capacités d'institutions phares comme les organismes de prospective et de surveillance et les organes consultatifs, et (3) des processus en matière de débats sociétaux, d'innovation inclusive et de collaboration ». Elle appelle ensuite les États membres et non membres et « tous les acteurs » à promouvoir et mettre en œuvre un ensemble de neuf « principes pour l'innovation responsable dans le domaine des neurotechnologies », qui sont décrits avec leurs actions spécifiques. Ces principes sont les suivants :

- i. Promouvoir une innovation responsable
- ii. Donner la priorité à l'évaluation de la sécurité
- iii. Promouvoir l'inclusivité
- iv. Encourager la collaboration scientifique
- v. Favoriser les débats sociétaux
- vi. Développer les capacités des organismes de surveillance et des organes consultatifs
- vii. Protéger les données cérébrales personnelles et autres informations
- viii. Promouvoir une culture de la gestion responsable et de la confiance dans les secteurs public et privé
- ix. Anticiper et surveiller les éventuels usages non intentionnels et/ou abusifs

46. Les neurotechnologies en général, et les ICM en particulier, pourraient modifier fondamentalement les relations entre for intérieur et monde extérieur. Certains chercheurs (dont M. Ienca) appellent donc de leurs vœux des réponses juridiques innovantes, y compris la création (ou une définition adaptée) de quatre « nouveaux » droits de l'homme : le droit à la liberté cognitive, le droit au respect de la vie privée sur le plan mental, le droit à l'intégrité mentale et le droit à la continuité psychologique. La « liberté cognitive » peut être définie comme « le droit et la liberté de contrôler sa propre conscience et les processus électrochimiques qui président à sa propre pensée, [...] condition préalable nécessaire à toutes les autres libertés ». Elle est proche en cela de la liberté de pensée, qu'on peut considérer comme indispensable aux libertés de religion, d'expression et d'association. Proche, mais elle est fondamentalement différente et distincte, et va plus loin : si la liberté de penser est le droit de penser ce qu'on veut, la liberté cognitive y est la condition préalable – la liberté du cerveau à générer des pensées sans interférences technologiques (ou autres) dans ce processus. Le droit au respect de la vie privée sur le plan mental protégerait les individus contre l'observation non consentie de leurs processus mentaux conscients. Le droit à l'intégrité mentale offrirait une protection contre des usages néfastes de type « piratage malveillant du cerveau » permettant une prise de contrôle des pensées et des actions de la personne. Le droit à la continuité psychologique protégerait l'individu contre des actes susceptibles d'avoir une incidence sur « la perception que chacun a de sa propre identité [...], [qui] consiste à se voir comme une même personne dans la durée »²³ – de rester psychologiquement soi-même. L'article dans lequel apparaissent ces propositions, ambitieux et détaillé, suscite la réflexion. Les auteurs se demandent entre autres si ces « nouveaux » droits sont déjà présents ou non dans les droits existants et si leur « création » constituerait une « inflation des droits ».

47. Un pays, le Chili, travaille déjà sur la protection juridique des « neurodroits », en collaboration avec le professeur Yuste et l'initiative NeuroRights de l'Université Columbia (voir ci-dessus). Une proposition de modification de l'article 19 de la Constitution chilienne définit l'identité mentale comme un droit fondamental, qui ne peut être modifié que par application des lois futures. Le projet de loi sur la « neuroprotection » qui l'accompagne contient des définitions juridiques des neurotechnologies, des interfaces cerveau-machine et des neurodroits. Toutes les données obtenues à partir du cerveau seraient définies comme des « neurodonnées » et intégrées dans le champ d'application de la législation existante sur les dons d'organes, interdisant ainsi leur commerce. Toute utilisation et tout développement futurs des neurotechnologies seraient soumis à la législation médicale. Parallèlement à ces initiatives, l'Université catholique au Chili travaille sur

²³ Ienca et Andorno, « Towards new human rights in the age of neuroscience and neurotechnology », *Life Sciences, Society and Policy*, (2017) 13:5.

des directives éthiques pour les secteurs de l'informatique, de l'IA et de la neuro-ingénierie. Toutes ces activités sont associées à une campagne de sensibilisation du grand public soutenue par le président chilien, les ministres du gouvernement et les parlementaires²⁴. Ce nouveau cadre juridique ferait du Chili le premier pays au monde à réglementer et à protéger les données qui pourraient être extraites du cerveau humain, en vue de réserver leur utilisation à des fins altruistes²⁵. On dit par ailleurs qu'il pourrait rendre illégales des technologies telles que celles du projet « taper du texte par la pensée » de Facebook²⁶.

48. Les appels des neuroéthiciens en faveur d'une protection juridique des neurodroits se heurtent à un certain scepticisme. Alan Mardinly de Neuralink, par exemple, a émis l'hypothèse que le droit à la liberté cognitive pourrait être violé par la publicité classique, qui est souvent délibérément conçue et ciblée de façon à exploiter des préférences subconscientes. De même, le traitement des addictions pourrait également être considéré comme une ingérence extérieure dans la liberté de choix de consommation d'un individu.

49. En ce qui concerne l'aspect IA des technologies ICM, le Comité des Ministres a mis sur pied en septembre 2019 le Comité ad hoc sur l'intelligence artificielle (CAHAI). Le CAHAI a pour mission d'étudier la faisabilité et les éléments potentiels d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle. Son travail se fonde sur les normes du Conseil de l'Europe dans le domaine de la démocratie, des droits de l'homme et de l'État de droit, ainsi que sur d'autres instruments juridiques internationaux pertinents et sur les travaux menés au sein d'autres organisations internationales et régionales. Avec les participants habituels représentant les États membres et observateurs du Conseil de l'Europe et d'autres organes du Conseil de l'Europe (dont l'Assemblée), le CAHAI bénéficie d'un niveau d'implication exceptionnellement élevé des représentants des organismes du secteur privé, de la société civile et du monde universitaire et de la recherche.

50. Le CAHAI a tenu sa première réunion du 18 au 20 novembre 2019. Il a décidé, entre autres, qu'un des éléments clés de la future étude de faisabilité serait une « cartographie des risques et opportunités découlant du développement, de la conception et de l'application de l'intelligence artificielle, notamment l'impact de cette dernière sur les droits de l'homme, l'État de droit et la démocratie ». À ce jour, le CAHAI prévoit d'adopter l'étude de faisabilité lors de sa troisième réunion, qui doit avoir lieu en décembre 2020.

51. C'est dans ce contexte institutionnel que l'Assemblée débattre du présent rapport et de divers autres rapports relatifs à l'IA actuellement en préparation dans diverses commissions. L'Assemblée a choisi d'aborder le sujet sous un angle contextuel, en examinant les effets de l'IA dans différents domaines. Au sein de la commission des questions juridiques et des droits de l'homme, par exemple, il existe aussi des rapports sur les effets de l'IA sur les services de police et le système de justice pénale, sur les véhicules autonomes et (dans les premières phases de préparation) sur les armes meurtrières autonomes. Les recommandations que l'Assemblée peut adopter sur la base de ces rapports fourniront donc au CAHAI des orientations importantes lorsqu'il s'agira de cartographier les risques et les opportunités de l'IA et ses effets sur les droits de l'homme, l'État de droit et la démocratie, et de déterminer ensuite la nécessité d'un cadre juridique international contraignant.

6. Conclusions et recommandations

52. Comme beaucoup de technologies, les interfaces cerveau-machine ouvrent de nouvelles possibilités, mais créent aussi des risques. Elles pourraient rendre à des personnes la capacité de se mouvoir et de communiquer, permettre un accomplissement coopératif de certaines tâches ou rendre nos actes plus efficaces ; augmenter nos capacités cognitives en nous donnant directement accès aux données et en nous permettant d'exploiter davantage la puissance de l'informatique, ou encore nous ouvrir de nouvelles expériences sensorielles, voire émotionnelles. D'un autre côté, elles pourraient ne tenir aucun compte de la liberté d'expression et des droits au respect de la vie privée, à l'intégrité et à la protection contre le fait de contribuer à sa propre mise en accusation ; influencer les choix, les comportements et même la personnalité dans la durée ; ou saper des aspects essentiels de l'égalité et de la dignité humaines.

53. Bien que ni les découvertes scientifiques, ni les technologies ne soient suffisamment avancées à l'heure actuelle pour produire simultanément tous ces dangers, certains d'entre elles sont déjà opérationnelles. Les exemples évoqués ci-dessus représentent tous des conséquences réalistes et prévisibles des avancées rapides qui sont en train de se produire. Comme l'ont conclu de nombreux observateurs dont les points de

²⁴ Pour en savoir plus, voir <https://nri.ntc.columbia.edu/projects>.

²⁵ « Computers Accessing Human Brain: Colombia Neuroscientist Calls for Regulation », Colombia Global Centers, 11 octobre 2019.

²⁶ « Neurorights and Why We Need Them », Elena Blanco-Suarez, Psychology Today, 25 juin 2020.

vue différent, il y a urgence : il faut anticiper dès aujourd'hui les risques potentiels et adopter des mesures réglementaires pour les atténuer ou les éviter. Comme pour l'IA, discipline voisine, cela pourrait passer par des codes de déontologie, des réglementations contraignantes, voire de nouveaux droits, ou, plus probablement, par une combinaison des trois. Si les technologies ICM s'appuient de plus en plus sur l'IA, elles soulèvent un ensemble de préoccupations distinctes qui leur sont propres. Les principes éthiques et la réponse réglementaire requise sont donc, d'une certaine manière, plus complexes, reflétant l'importance de ce que signifie l'intrusion de ces technologies au cœur même de l'être humain.

54. Mes propositions concrètes et politiques sont exposées dans les projets de résolution et de recommandation ci-joints.

Annexe

Intelligence artificielle – Description et principes éthiques

On a tenté à plusieurs reprises de définir le terme « intelligence artificielle » depuis sa première utilisation en 1955. Ces initiatives s'intensifient aujourd'hui, car les organes normatifs, notamment le Conseil de l'Europe, réagissent aux capacités et à l'omniprésence croissantes de l'intelligence artificielle en œuvrant en faveur de son encadrement juridique. Il n'existe cependant toujours pas de définition « technique » ou « juridique » unique qui soit universellement admise²⁷. Aux fins du présent rapport, il est toutefois indispensable de décrire cette notion.

À l'heure actuelle, le terme « intelligence artificielle » désigne en général les systèmes informatiques capables de percevoir et d'extraire des données de leur environnement, puis d'utiliser des algorithmes statistiques pour traiter ces données, afin d'obtenir des résultats qui correspondent à des objectifs prédéterminés. Les algorithmes se composent de règles définies par l'homme ou par l'ordinateur lui-même, qui « forme » l'algorithme en analysant des ensembles de données considérables et continue à affiner ces règles à mesure qu'il reçoit de nouvelles données. Cette méthode, connue sous le nom « d'apprentissage automatique » ou « d'apprentissage statistique », est actuellement la technique la plus utilisée pour les applications complexes ; elle est uniquement devenue possible ces dernières années grâce à l'augmentation de la puissance de traitement des ordinateurs et à la disponibilité de données suffisantes. « L'apprentissage en profondeur » représente une forme particulièrement avancée d'apprentissage automatique, qui utilise plusieurs couches de « réseaux neuronaux artificiels » pour traiter les données. L'analyse ou la compréhension intégrale par l'homme des algorithmes développés par ces systèmes n'est pas toujours possible ; aussi sont-ils parfois qualifiés de « boîtes noires » (il arrive que ce terme désigne également, mais pour une raison différente, les systèmes d'intelligence artificielle propriétaires protégés par les droits de propriété intellectuelle).

Les formes actuelles d'intelligence artificielle sont toutes « restreintes », c'est-à-dire affectées à une tâche unique définie. L'intelligence artificielle « restreinte » est également qualifiée parfois de « faible », même si les systèmes modernes de reconnaissance faciale, de traitement du langage naturel, de conduite autonome et de diagnostic médical, par exemple, sont incroyablement sophistiqués et effectuent certaines tâches complexes avec une rapidité et une précision étonnantes. « L'intelligence artificielle générale », parfois qualifiée d'intelligence artificielle « forte », qui est capable d'exécuter toutes les fonctions du cerveau humain, est encore à réaliser. La « super-intelligence artificielle » désigne un système dont les capacités dépassent celles du cerveau humain.

Comme le nombre de domaines dans lesquels les systèmes d'intelligence artificielle sont appliqués est en augmentation, puisqu'ils se propagent dans des domaines qui peuvent avoir un impact important sur les droits individuels et les libertés, ainsi que sur les systèmes démocratiques et l'État de droit, la dimension éthique de ce phénomène a fait l'objet d'une attention croissante et de plus en plus urgente.

Un large éventail d'acteurs a formulé de nombreuses propositions d'ensemble de principes éthiques qui devraient être appliqués aux systèmes d'intelligence artificielle. Ces propositions sont rarement identiques et diffèrent à la fois dans les principes qu'elles énoncent et par la manière dont elles définissent ces principes. Les études montrent que la teneur essentielle des principes éthiques qui devraient être appliqués aux systèmes d'intelligence artificielle fait néanmoins l'objet d'un large consensus ; c'est notamment le cas des principes suivants²⁸ :

- **Transparence.** Le principe de transparence peut faire l'objet d'une interprétation élargie, de manière à englober l'accessibilité et l'explicabilité d'un système d'intelligence artificielle, en d'autres termes la possibilité donnée à un individu de comprendre le fonctionnement de ce système et le mode de production de ses résultats.
- **Justice et équité.** Ce principe englobe la non-discrimination, l'impartialité, la cohérence et le respect de la diversité et du pluralisme. Il suppose également que la personne à laquelle est appliqué un système d'intelligence artificielle puisse en contester les résultats, disposer d'une voie de recours et obtenir réparation.

²⁷ Pour une vue d'ensemble élargie des tentatives de définition de « l'intelligence artificielle », voir *AI Watch: Defining Artificial Intelligence – Towards an operational definition and taxonomy of artificial intelligence*, Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F. et Delipetrev, B., European Commission Joint Research Centre, 2020.

²⁸ Voir *Lignes directrices sur l'éthique en matière d'IA : situation en Europe et dans le monde*, rapport provisoire commandé par le Comité ad hoc sur l'intelligence artificielle (CAHA) du Conseil de l'Europe, M. Ienca et E. Vayena, mars 2020.

- *Responsabilité.* Ce principe englobe le fait d'exiger qu'un être humain soit responsable de toute décision qui a des conséquences sur les droits et libertés individuels, et que l'obligation de rendre des comptes et la responsabilité juridique de ces décisions soient définies. Il est donc étroitement lié au principe de justice et d'équité.
- *Sûreté et sécurité.* Ce principe suppose que les systèmes d'intelligence artificielle fassent preuve de solidité, de sécurité contre toute ingérence extérieure et de sûreté contre la commission d'actes involontaires, conformément au principe de précaution.
- *Respect de la vie privée.* Si le respect des droits de l'homme en général peut être considéré comme inhérent aux principes de justice et d'équité, de sûreté et de sécurité, le droit au respect de la vie privée est particulièrement important chaque fois qu'un système d'intelligence artificielle procède au traitement de données à caractère personnel ou privé. Les systèmes d'intelligence artificielle doivent par conséquent respecter les normes contraignantes du règlement général sur la protection des données (RGPD) de l'UE et de la Convention 108 du Conseil de l'Europe sur la protection des données (et de sa version actualisée, la Convention 108+), le cas échéant.

La mise en œuvre effective des principes éthiques applicables aux systèmes d'intelligence artificielle exige une approche intégrée de l'éthique, notamment une évaluation de leur impact sur les droits de l'homme, de manière à garantir le respect des normes établies. Il ne suffit pas que ces systèmes soient conçus uniquement sur la base de normes techniques et que des éléments soient ajoutés à un stade ultérieur pour tenter de faire respecter les principes éthiques.

Dans quelle mesure le respect de ces principes doit-il être intégré dans des systèmes particuliers d'intelligence artificielle ? Cela dépend des utilisations prévues et prévisibles de ces systèmes : plus leur impact sur l'intérêt général et les droits et libertés individuels est important, plus les garanties doivent être strictes. La réglementation éthique peut donc être mise en œuvre de différentes manières, depuis les chartes internes volontaires pour les domaines les moins sensibles jusqu'aux normes juridiques contraignantes pour les plus délicats. Dans tous les cas, il importe qu'elle prévienne des mécanismes de contrôle indépendants selon le niveau de réglementation.

Ces principes essentiels portent sur les systèmes d'intelligence artificielle et leur environnement immédiat. Ils n'ont pas vocation à être exhaustifs ni à exclure des préoccupations éthiques plus générales, telles que la démocratie (participation pluraliste des citoyens à l'élaboration de normes éthiques et réglementaires), la solidarité (qui admet les différences de point de vue des divers groupes) ou la durabilité (préservation de l'environnement de la planète).