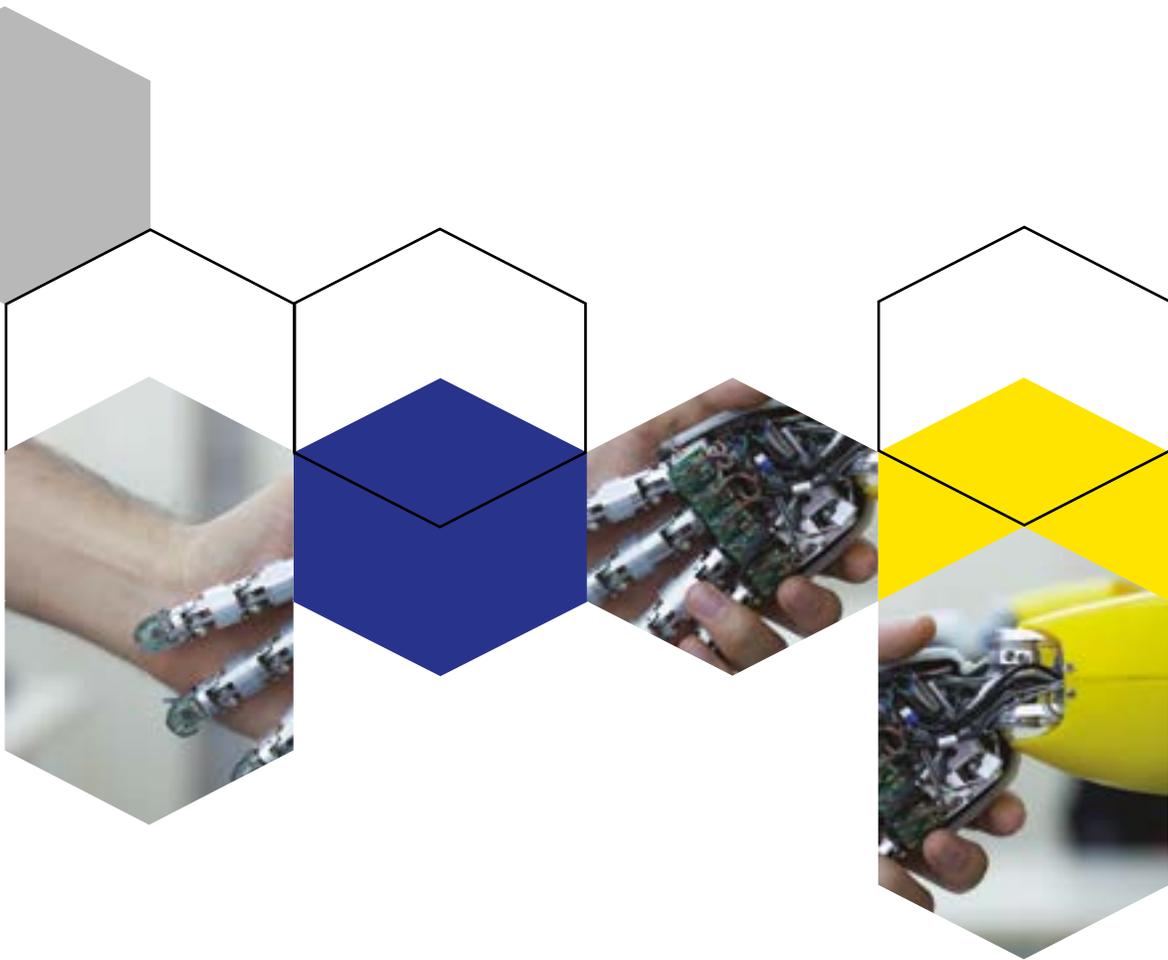


[Intelligences Artificielles]

Quelles promesses ? Quels défis ?

Les cahiers de veille de la Fondation Télécom // Cahier n°8





{ IA du 3^e type }

éditorial Juin 2016

L'intelligence peut se définir selon deux grandes catégories de facultés. La première intelligence, que nous partageons avec la plupart des espèces animales, est celle qui nous relie au monde extérieur et nous permet de percevoir, d'apprendre, de reconnaître, d'estimer et de décider. Elle est centrale à notre capacité d'adaptation et de survie dans un monde toujours fluctuant et son automatisation a été jusqu'ici le principal moteur de l'Intelligence Artificielle (IA). Grâce aux progrès constants de la microélectronique, de l'informatique, du traitement du signal, de l'analyse statistique et plus récemment de l'apprentissage profond opérant sur de vastes données, des résultats remarquables ont été obtenus dans l'automatisation des tâches de perception et de décision. Un exemple saisissant de cette IA du premier type en est la voiture autonome dont on ne mettra pas en doute la virtuosité dans son respect des règles de conduite et dans son attention pour les autres.

La seconde intelligence est ailleurs et est propre à chaque individu. Elle englobe les facultés de l'esprit, celles qui nous permettent d'imaginer, d'élaborer, d'inventer et d'espérer. Le seul modèle dont nous disposons pour essayer de reproduire dans une machine les propriétés de cette intelligence créatrice est notre cerveau, dont l'architecture est nettement différente de celle de l'ordinateur classique. Information et processus s'y entremêlent dans un même écheveau de connexions synaptiques, lesquelles se comptent en milliers de milliards. L'IA du deuxième type ne pourra pas faire l'économie de ce parallélisme massif qui pourra cependant et fort heureusement être décomposé à la façon des modules corticaux.

Quand les mystères de l'information mentale et de l'organisation corticale auront été complètement levés, dès lors aussi que la microélectronique saura offrir les moyens de contrôler un grand nombre de connexions (disons quelques centaines de millions par module, pas plus), rien n'empêchera de concevoir des cortex artificiels avec plus de modules que n'en contiennent nos cerveaux. Cette IA du troisième type, que certains appellent singularité technologique, sera l'aboutissement d'une alliance détonante entre les neurosciences, l'électronique, le calcul intensif, la datamasse et le principe de diversité.

Claude Berrou

Professeur à Télécom Bretagne

{ préface }

Avec ce huitième numéro des Cahiers de Veille de la Fondation Télécom, l'Institut Mines-Télécom a tenu à marquer plus encore son intérêt pour ces travaux en coproduisant son séminaire interne de début d'année, sa conférence large public après l'été, et bien sûr ce cahier.

Dans le cadre des missions et de la stratégie de l'IMT, les écoles de l'Institut, qui bénéficient d'une large autonomie scientifique, pédagogique et de recherche pour mener leurs activités en lien avec leurs écosystèmes, se rassemblent autour de certaines thématiques transversales. Ces grands sujets donnent l'occasion à la Fondation Télécom et ses partenaires de faire un point sur les actions menées dans les écoles qui font toute la valeur de notre réseau. Rappelons par exemple les cahiers précédents sur l'enseignement supérieur à l'ère du numérique, celui sur l'internet des objets ou celui sur les identités numériques.

Fortement transdisciplinaires, ces cahiers visent à présenter aussi bien les points techniques que les grands enjeux sociétaux, les questions de droit et les questions d'éthique, les fondements économiques comme les réflexions philosophiques. L'Intelligence Artificielle, dont on voit les développements fulgurants en quelques années, avait été abordée l'an passé dans le cahier sur l'humain augmenté. Plusieurs équipes dans nos écoles développent des recherches de pointe et apportent des contributions déterminantes à ce grand défi scientifique et sociétal. Certains de leurs travaux sont présentés ici. Je vous souhaite une excellente lecture à la découverte des IA.

Christian Roux

Directeur de la recherche et de l'innovation de l'Institut Mines-Télécom

(sommaire)

3

9

22

27

27



Naissance de l'Intelligence Artificielle

Naissance de l'Intelligence Artificielle

Il est temps de prendre l'Intelligence Artificielle au sérieux

Qu'est-ce que l'intelligence ?

*De nombreuses formes d'intelligence
Mieux évaluer l'intelligence
L'intelligence des artefacts*

Quelles directions pour l'Intelligence Artificielle ?

Un quotidien bercé d'Intelligences Artificielles

L'avènement de l'apprentissage machine

*Les types d'apprentissage
Les formes d'apprentissage*

Vers une informatique neuro-inspirée

Le système nerveux humain

Planche centrale: L'IA entre mythes & réalités

Les grands projets

*API et open source
Les architectures matérielles*

Considérations économiques

Intelligences Artificielles, que ferons-nous d'elles ?

Des IA créatives et conscientes ?

Que manque-t-il aux IA ?

Enjeux & défis modernes des IA

Apprendre à vivre avec les Intelligences Artificielles

Glossaire

Collaborations possibles avec l'Institut Mines-Télécom & la Fondation Télécom



Question : Vous semblez alors dire que les programmes IA seront quasi identiques aux humains. N'y aura-t-il aucune différence ?

Réponse : Les différences entre les programmes IA et les humains seront probablement plus grandes que les différences entre la plupart des gens. Il est impensable que le « corps » contenant un programme IA ne l'affecte pas profondément. C'est pourquoi, à moins que son corps ne soit une réplique étonnamment fidèle du corps humain, (et pourquoi le serait-il ?), il aurait probablement des vues extrêmement différentes de ce qui est important, de ce qui est intéressant, etc. [...] Je pense qu'un programme IA, même si nous pouvions le comprendre, nous paraîtrait assez étrange. C'est d'ailleurs pour cela que le moment où nous aurons affaire à un programme IA, et non pas à un programme simplement « bizarre », nous donnera beaucoup de fil à retordre.

Douglas Hofstadter, Gödel Escher Bach, 1985



Janvier 2015 : à l'initiative du britannique Stuart Russel, spécialiste en Intelligence Artificielle (IA), une dizaine de chercheurs signent une lettre ouverte appelant leurs collègues à aller au-delà du simple objectif historique de la performance des IA. « *Les progrès en Intelligence Artificielle sont tels qu'aujourd'hui il faut concentrer les travaux non seulement sur des IA plus performantes, mais également sur la poursuite d'une IA la plus bénéfique possible pour la Société. [...] Nous recommandons une large recherche visant à assurer que les IA soient de plus en plus robustes et bénéfiques ; et que ces systèmes fassent effectivement ce que nous voulons qu'ils fassent. [...] Cette recherche est nécessairement interdisciplinaire, car elle implique à la fois la Société et l'Intelligence Artificielle. Elle s'établit de l'économie au droit et à la philosophie, de la sécurité informatique aux méthodes formelles et, bien sûr, au sein des diverses branches de l'IA elle-même.* »

Le 37^e signataire, Elon Musk, s'était ému l'année précédente qu'une intelligence artificielle supérieure non bienveillante envers l'humanité puisse émerger d'ici quelques années, et

qu'il était peut-être déjà trop tard pour enrayer ce processus. Musk s'était convaincu de cette menace à la lecture du livre du philosophe Nick Bostrom, *SuperIntelligence*. Quelques chercheurs s'étaient alors réunis à l'issue de la [conférence NIPS 2014](#) pour réfléchir aux impacts que ces positions pouvaient avoir sur leur recherche et s'étaient donné rendez-vous fin 2015. C'est à cette même époque que Musk et quelques autres entrepreneurs créent Open AI, doté d'un fonds de un milliard de dollars US, pour promouvoir une IA à visage humain.

En deux mois les signataires représentent près de 300 groupes de recherche, parmi les principaux chercheurs en IA chez Google, Facebook, Microsoft et d'autres industriels, ainsi que les meilleurs informaticiens, physiciens et philosophes du monde entier. Ils sont plus de 8600 en juin 2016. Ils ne se contentent pas de souligner l'importance de leur constat et l'occasion à saisir, mais l'accompagnent d'une liste de pistes concrètes de recherche à mettre en œuvre aussitôt. 2016 s'est ainsi ouverte avec une approche et des objectifs nouveaux pour développer les Intelligences Artificielles.

Il est temps de prendre l'Intelligence Artificielle au sérieux

« Ce n'est plus une simple curiosité pour chercheurs, l'Intelligence Artificielle a maintenant un impact mesurable sur nos vies. » C'est par ces mots que le Wall Street Journal annonce à ses lecteurs fin août 2014 que l'Intelligence Artificielle n'est plus un simple sujet de prospective.

Les développements en IA ont été longtemps sous-estimés en raison d'un manque de clarté des définitions, entretenu par une confusion généralisée entre le *machine learning*, le *deep learning*, les *réseaux de neurones*, l'*analyse prédictive*, et l'*analyse et la fouille de données massives*. Le cinéma, la littérature et les médias ont souvent égaré la discussion sur l'IA en préférant des histoires fantastiques, de HAL 9000 dans l'Odyssée de l'espace à Terminator et son cortège de peurs. Mais, alors que les chercheurs prennent largement conscience de la nécessité de discuter de l'impact de l'IA sur la société, le grand public découvre dans la presse et les réseaux sociaux des avancées spectaculaires qui racontent une nouvelle histoire, celle d'une technologie qui est déjà parmi eux.

Trois ruptures majeures expliquent comment en quelques années un coup de fouet a été donné aux recherches en IA. Ces trois tendances ont fourni une plate-forme d'innovation accessible

et bon marché pour les développeurs, qui utilisent ces algorithmes comme des *commodités* de base pour opérer des transitions majeures dans de nombreux secteurs industriels :

- l'accès à des ressources de calcul parallèle à très bas coût,
- l'accès facilité à des données massives, pouvant servir d'ensemble d'apprentissage,
- des algorithmes nouveaux, profitant des deux ruptures précédentes.

Défi de l'IA s'il en est, la première victoire d'un programme de go (AlphaGo de Google DeepMind) en octobre 2015 sur un joueur professionnel avait déjà fait sensation. La manière dont le même programme, ayant appris de l'entraînement effectué par la suite avec l'humain perdant, battait ensuite un des meilleurs joueurs mondiaux et obtenait la 4^e place au classement de go mondial, achevait de frapper les esprits, d'autant plus qu'il semblait avoir fait montre de créativité pour ce faire.

Alors que des voitures autonomes circulent déjà sur les routes et que des logiciels traduisent de mieux en mieux des textes, en temps réel, serions-nous en train de créer ces intelligences qui dépassent l'homme ? Mais de quelle(s) intelligence(s) parlons-nous ?

60 ans d'Intelligence Artificielle

C'est en août 1956, lors de la conférence de Dartmouth, que l'expression "*Artificial Intelligence*" fait son apparition publique pour la première fois. Elle avait été utilisée par les pionniers John McCarthy, Marvin L. Minsky, Nathaniel Rochester et Claude E. Shannon l'été précédent pour proposer ce séminaire. Elle caractérise « *la possibilité de produire des programmes qui se conduiraient ou penseraient intelligemment* ». Ses ambitions d'alors, et le défi originel de l'Intelligence Artificielle, sont de « *chercher à produire, sur un ordinateur, un ensemble de sorties qui serait considéré comme intelligent s'il était produit par un être humain* ». L'IA peut se révéler à travers des simulations exactes des processus cognitifs humains, ou

bien via des programmes conduisant à des conséquences intelligentes. Elle a été traversée par de nombreuses dualités, entre l'inné et l'acquis, entre les symboles des *systèmes experts* et les sous-symboles des *réseaux de neurones formels*, entre la compétence et la performance, qui ont rythmé son histoire.

Technologie de la connaissance (nouvelle science de l'ingénieur) mais aussi *science générale du traitement de l'information* (par l'homme ou par la machine) ou encore *théorie de l'homme et des processus cognitifs*, cette discipline a eu tour à tour chacune de ces ambitions, ni incompatibles, ni indépendantes. Reliée intimement à un ensemble d'autres

disciplines au sein des *Sciences Cognitives*, elle a eu en 60 ans ses moments de gloire mais aussi ses moments de doute et de recul. Ce sont les *AI Winters* que Yann LeCun, aujourd'hui responsable des équipes IA au sein de Facebook, et dont les travaux depuis 30 ans ont mené aux avancées spectaculaires d'aujourd'hui, n'hésite jamais à rappeler pour modérer certains enthousiasmes actuels vis-à-vis du large public. Mais parmi les chercheurs, nombreux sont ceux qui estiment, comme John Giannandrea, vice-président de l'ingénierie chez Google et responsable de l'activité *machine learning*, que « *les choses sont en train de prendre une tournure incroyable* » et que nous assistons à un véritable *AI spring*.

Qu'est-ce que l'intelligence ?

Un modèle de l'architecture de l'esprit humain

Le *General Problem Solver*, créé en 1959 par Simon, Shaw et Newell, est une première tentative d'un système artificiel à l'heure de l'IA naissante, proposé pour résoudre n'importe quel problème, en confrontant les objectifs poursuivis et les moyens pour y parvenir. Ce système, qui a eu une grande influence sur l'évolution de l'Intelligence Artificielle et des Sciences Cognitives, résolvait très bien des problèmes simples, mais était limité dès lors que la combinatoire du problème augmentait. Ces travaux, et le livre de Newell & Simon, *Human Problem Solving* (1972), sont fondateurs du *paradigme cognitiviste*.

Dans son *Introduction aux Sciences Cognitives* en 1992, Michel Imbert préfère prudemment « *décrire ce qui fait tout être que l'on s'accorde à reconnaître comme intelligent* » (voir Cahier de Veille 2015 sur l'Humain augmenté, page 16). Sa définition a l'avantage de cerner les résultats observables qui caractérisent un processus intelligent, plutôt que d'en expliquer les mécanismes qui l'ont permis, dont beaucoup ne sont toujours pas compris.

Notre définition de l'intelligence ne cesse en effet d'évoluer. Nous nous interrogeons encore aujourd'hui sur la manière dont nous apprenons, sur la place de l'émotion dans le raisonnement, sur la place du rêve dans le renforcement des apprentissages. Pour comprendre ce que l'on peut attendre d'une intelligence artificielle, et pour mieux en éviter les nombreux mythes, il faut sans doute d'abord s'accorder sur différentes formes d'intelligence, plutôt que définir une unique notion d'intelligence.

De nombreuses formes d'intelligence

Mesurer et comparer l'intelligence des êtres humains est un exercice délicat. L'intelligence se développe au sein d'un environnement et d'une culture, elle se manifeste différemment selon le genre, l'âge, les expériences, les connaissances. Elle désigne tour à tour la faculté de comprendre, la faculté de raisonner

et de décider, la faculté de s'adapter, l'adresse et l'habileté, la somme des connaissances et celle des compétences, sans oublier la réussite à des examens.

L'intelligence est en réalité multiple et multiforme. La littérature en recense souvent plus d'une dizaine. Explorons-en quelques-unes en commençant par celles qui ont été très tôt reproduites par des *artefacts*. Première d'entre elles, *l'intelligence rationnelle* (la logique) est celle qui a été la plus mesurée depuis le fameux test de Binet, précurseur du Quotient Intellectuel (voir encart page 6). Elle rassemble les compétences de calcul, d'analyse, de logique, et de raisonnement par déduction ou induction, parfaites pour résoudre des problèmes mathématiques, des jeux, prendre des décisions. Cette *intelligence mathématique* a naturellement été la première à être implémentée dans des programmes, et notamment à travers des systèmes experts. On peut y ajouter *l'intelligence naturaliste* qui consiste à savoir classer des objets, et définir des catégories. Modéliser une situation, manipuler cette modélisation et y tester des hypothèses et leurs limites, est un ensemble de processus qui a également été rapidement mécanisé. Chez l'humain il s'agit de *l'intelligence systémique*. Couplées à *l'intelligence organisationnelle*, celle qui permet d'assembler des informations disparates, et à *l'intelligence stratégique*, celle qui comprend l'optimisation des ressources,

60 ans de liens entre jeux et IA

Dès 1956, l'IA s'intéresse aux jeux comme source de défis. Arthur Samuel développe un premier programme de jeu de dames qui se forme par *apprentissage par renforcement*, et bat en 1962 un amateur américain de bon niveau, sur un seul jeu, devenant ainsi la première *machine qui bat l'homme* de l'histoire. Le jeu de backgammon suit en 1979, avec les mêmes techniques qui font jouer le programme contre lui-même pendant son apprentissage. Il atteint ainsi des niveaux qui sont réputés ne pas pouvoir être enseignés par des humains, tandis que ces derniers apprennent beaucoup en observant les capacités de ces programmes. C'est ainsi que l'on prend conscience que les intelligences artificielles pourraient bien servir

à renforcer les intelligences humaines. Et c'est le jeu d'échecs, un jeu dont Claude Shannon avait dit dès 1950 qu'il consistait un bon défi pour la pensée mécanisée, qui tient en haleine chercheurs et joueurs jusqu'en 1997, quand l'ordinateur DeepBlue d'IBM bat le champion mondial Garry Kasparov. Et enfin le jeu de go en 2016, grâce à des techniques d'*apprentissage profond*, et un système ayant joué des milliers de parties contre lui-même, tombe à son tour.

Mais il y a encore beaucoup à apprendre du jeu. Jusqu'à présent, les joueurs ont en effet une connaissance parfaite de tous les éléments du jeu. Il faut maintenant s'intéresser à des jeux à connaissance incomplète, comme le

jeu de poker qui devient la prochaine étape à franchir. D'autres types de jeux, comme les robots footballeurs ou les voitures de course autonomes, présentent également de nombreux défis d'IA à relever, beaucoup plus généraux. Le jeu ultime reste ce que Alan Turing appelle en 1950 *The imitation game*. À l'origine une femme et un homme sont interrogés par une personne (homme ou femme) qui ne les voit pas, et qui doit deviner qui est la femme. L'homme caché doit donc imiter un "comportement féminin". Turing se demande si un programme pourrait prendre la place de l'homme caché et leurrer l'interrogateur. Aujourd'hui, la question est de savoir quelle proportion d'humains perdrait à ce jeu.

“ *l'intelligence est souvent renvoyée à la capacité d'apprendre* ”

des moyens, du temps et de l'espace et permet de prendre les décisions, toutes ces formes d'intelligence sont un premier groupe qui permet par exemple de planifier une série d'actions pour atteindre un but.

L'intelligence créative est le plus souvent ignorée par les tests. L'aptitude à la créativité est pourtant un des critères les plus souvent cités à notre époque pour déterminer le degré d'intelligence d'un artefact. **L'intelligence littéraire**, celle des mots, du sens, permet d'élaborer des raisonnements traduits dans des discours, de suivre et d'effectuer des conversations, de traduire et de manipuler des concepts abstraits.

L'intelligence émotionnelle, qui dispose de son propre quotient également, permet d'observer les émotions, en soi ou sur les autres, les interpréter, les canaliser pour les décharger de la conscience de l'individu. Que ce soit pour l'intelligence naturelle ou artificielle, le rôle et l'importance des émotions a été longtemps méconnu ou ignoré. On sait à présent depuis António R. Damásio qu'émotions et raisonnement sont fortement liés. Les émotions renseignent sur l'état du corps, qui est l'interface entre le lieu de l'intelligence et l'environnement qui la motive et où elle s'exerce. **L'intelligence kinesthésique** est également liée au corps : c'est elle qui permet la coordination des mouvements, leur force, leur précision. Avec

l'intelligence émotionnelle et **l'intelligence spatiale** (le sens de l'orientation), elle permet d'appréhender sa place dans le monde, ce qu'on peut y faire ou ne pas faire, et comment le faire. On y ajoutera **l'intelligence situationnelle** qui consiste à savoir s'adapter et survivre dans un environnement inconnu, voire hostile. Ce troisième groupe apporte les capacités à percevoir et agir à bon escient sur le monde.

Les caractères de l'intelligence sont le produit de l'évolution et de l'interaction avec le monde. Ils sont aussi activés par les interactions avec ses semblables, notamment pour certaines formes d'apprentissage, par imitation – l'intelligence est du reste souvent renvoyée à la capacité d'apprendre. **L'intelligence sociale** révèle des humains à l'aise au contact d'autrui, tandis que **l'intelligence au collectif** caractérise ceux qui mettent leur ego de côté pour un objectif commun supérieur. La dualité isolé / collectif a nourri un fort courant des sciences cognitives, celui du paradigme de la fourmière. Il s'agit de faire coopérer des **agents cognitifs** minimaux de manière à obtenir un comportement global qualifié d'intelligent, qu'aucun agent isolé n'aurait pu effectuer seul.

Des intelligences moins répandues existent également. Leur maîtrise apporte un plus à leurs bénéficiaires. Ainsi **l'intelligence pluri-sensorielle** fait utiliser simultanément l'ensemble de ses sens pour percevoir le monde d'une manière

Évaluer l'Intelligence

Corollaire de la définition de l'intelligence, la capacité à la mesurer et la quantifier, qu'elle soit naturelle ou artificielle, reste une question ouverte. Développé en 1905 pour détecter les élèves en difficulté, le test de Simon-Binet est une échelle métrique de l'intelligence qui est à l'origine du concept d'âge mental. Il ne doit pas être confondu avec le quotient intellectuel qui est un test psychométrique donnant une indication quantitative standardisée sur la performance intellectuelle, rang d'une personne relativement à la population. Il a lui aussi le défaut de ne pas prendre en compte l'ensemble des différentes formes d'intelligence. De plus, n'étant appliqué qu'aux êtres humains, et selon des principes de mesure qui leur sont propres, il a également contribué à ignorer l'étude des formes d'intelligence exercées par les animaux.

L'intelligence artificielle elle-même est encore mesurée en fonction de critères humains, et donc évaluée par comparaison avec les humains. Dans son article "*Computing Machinery and Intelligence*", Alan Turing pose en 1950 la question de la possibilité que des *machines* puissent *penser*, ces deux termes restant d'ailleurs à définir. Pour le savoir, il propose que si, lors d'une conversation, une machine arrive à passer pour un être humain auprès d'un interlocuteur, c'est qu'elle est réellement intelligente. Il existe en réalité plusieurs tests de Turing avec des protocoles plus fins. Leurs défauts sont de se limiter à des expériences de communication, et au seul jugement humain, alors que des comportements intelligents peuvent ne pas être associés au langage, ni même être humains.

Mieux évaluer l'intelligence

Des intelligences artificielles actuelles ont passé le test de Turing, mais des chercheurs s'accordent à dire que ce n'est pas tant la machine qui réussit le test, mais les humains qui échouent, car ils ont fini par ne plus être capable d'identifier les caractéristiques de comportement d'une machine, tant elles sont familières. De plus, de nombreuses expériences de passage du test ne résistent pas à un examen plus approfondi, et révèlent surtout des systèmes conçus avec ruse pour obtenir le label.

La communauté de recherche entre aujourd'hui dans une ère de l'après-Turing, enrichissant le test pour que la machine soit évaluée sur de plus nombreux critères et tâches à résoudre.

différente. On retrouve cette aptitude implémentée par exemple dans les objets connectés et notamment la fusion de capteurs. Enfin, **l'intelligence temporelle** offre pour sa part un sens aigu de l'axe du temps.

Les concepts de soi, de passé et de futur étant reliés au sein de la conscience primaire, et l'absence de conscience d'ordre supérieur empêchant de planifier le futur (en se servant de la mémoire à long terme), l'étude des personnes qui possèdent cette intelligence temporelle permettrait d'en savoir plus sur une autre notion qui pose question : **la conscience**. Car la conscience, et notamment la conscience de soi, est une condition nécessaire pour aller plus loin que les premiers stades du développement intellectuel d'un être vivant. Celle-ci n'est d'ailleurs pas limitée aux êtres humains, et de nombreux animaux passent le **test du miroir**, y compris des insectes. Les chercheurs prennent ainsi conscience que **la cognition animale** a été longtemps sous-estimée (pour leur absence supposée de langage) et doit être source d'inspiration pour nos artefacts. Comme l'est l'idée que le siège de l'intelligence ne réside pas uniquement dans le cerveau (humain), et que des processus cognitifs ont lieu également dans des neurones de la peau ou, ce qu'on appelle **l'intelligence incarnée**, dans la morphologie des êtres vivants.

La reconnaissance des chiffres manuscrits

Le **paradigme connexionniste** ne repose pas sur l'utilisation de symboles et de règles comme le paradigme cognitiviste, mais s'inspire des connexions entre les neurones biologiques, avec des réseaux connectant un espace d'entrée (par exemple les valeurs de blanc à noir des 16x16 pixels de l'image d'un chiffre) à un espace de sortie (10 **neurones artificiels** représentant les chiffres possibles) via un ensemble de neurones et relations intermédiaires, caractérisées par des paramètres, les **poinds synaptiques**, calculés lors d'une phase d'apprentissage. Dès les années 90, chèques bancaires et codes postaux étaient reconnus par de tels systèmes.

L'intelligence des artefacts

Le projet initial d'une **IA forte**, qui était de reconstruire la manière dont l'homme pensait, puis la dépasser, a fait place au fil des ans à un projet plus modeste d'**IA faible**, simulation spécialisée de comportements humains considérés comme intelligents, par des méthodes d'ingénieur, sans se soucier de similarité. Deux paradigmes principaux ont été utilisés pour s'attaquer aux défis de l'IA. L'approche classique, à base de manipulation de symboles et de règles, le **cognitivism**, a donné rapidement des résultats intéressants pour les problèmes relevant du premier groupe d'intelligences, celle des mathématiques, mais s'est heurtée à la complexité de la traduction automatique, malgré l'arrivée des **réseaux sémantiques**. L'approche neuro-inspirée, le **connexionnisme**, avec ses capacités d'apprentissage, s'est attaquée aux problèmes de perception visuelle ou auditive, limitée par la puissance des machines de l'époque. Des approches hybrides, sur des machines parallèles conçues pour, ont permis dans les années 90 des avancées avec l'émergence de comportements intelligents non programmés. Aujourd'hui, la multitude de formes d'intelligence existantes, humaines ou non, étant un fait acquis, ce n'est pas tant vers une **superintelligence** que vers une **nouvelle forme d'intelligence, propre aux artefacts**, ne se définissant ni par comparaison ni par extension, que les IA se dirigent.

Ai-je affaire à un être humain ?

La question de savoir si les dialogues qu'on exerce à travers une machine sont le fait d'êtres humains qui nous font face, ou bien de robots algorithmiques, ou d'une combinaison des deux, se fait de plus en plus prégnante avec le déploiement des services associés aux messageries instantannées comme Messenger de Facebook, et aux plates-formes de micro-blogging comme Twitter : les **bots**.

L'enjeu est important, car il ne s'agit pas moins de capter les utilisateurs qui passent de plus en plus de temps sur ces outils, et de leur proposer à travers des **dialogues intelligents** une nouvelle manière d'accéder à l'information et à la consommation. C'est en quelque sorte une architecture nouvelle pour un navigateur internet à l'heure de l'IA, avec comme objectif

de simplifier l'interaction avec les services en ligne. Le site d'information Quartz a ainsi déployé en février 2016 un **bot** avec lequel le lecteur dialogue, plutôt que consulter une longue liste d'articles, afin de se voir proposer des résumés des informations essentielles qui l'intéressent. Suivant Telegram et Kik, Facebook ouvre en avril 2016 son **botstore** sur Messenger. Pour la plupart, les 900 millions d'utilisateurs mensuels de Messenger pourront y croiser leurs premiers robots conversationnels, à condition que ceux-ci sachent ne pas être trop envahissants et rendent effectivement service.

Il faut en effet que l'expérience utilisateur soit parfaite pour que l'adoption soit au rendez-vous. Un **bot** qui réserverait des billets d'avion sans tenir compte des temps de débarquement

et d'embarquement et de contrôle lors des escales serait plus encombrant qu'utile, et desservirait la marque qui le diffuserait. Ces **bots**, développés pour assister les humains dans des tâches qui leur sont propres, doivent acquérir à terme des connaissances du niveau d'une IA forte, générale (voir ci-dessus), et une personnalité capable de s'adapter à chaque utilisateur, en prenant en compte ses habitudes, sa culture, ses croyances... C'est pourquoi pendant quelques temps encore les **bots** seront probablement toujours constitués d'équipes mixtes IA & humains.

Saura-t-on toujours distinguer la part du **bot** de celle de l'humain ? <http://botpoet.com/> invite à se tester dans un cas particulier : trouver qui d'un humain ou d'une IA a composé un poème.

Quelles directions pour l'Intelligence Artificielle ?

“ le prochain défi : apprendre

à partir de peu d'exemples ”

Après avoir exploré un grand nombre de problèmes et d'approches depuis sa création, la recherche en IA s'est intéressée depuis 20 ans à la construction d'agents intelligents, des systèmes percevant et agissant dans des environnements précis. Dans ce cadre, l'intelligence est définie selon des dimensions statistiques et la notion économique de rationalité : il s'agit de faire des déductions logiques, de planifier correctement et de prendre de bonnes décisions. Cette approche à base de représentations probabilistes et d'apprentissage statistique a eu pour effet une forte fertilisation croisée entre des disciplines comme l'Intelligence Artificielle, l'apprentissage automatique, les statistiques, les neurosciences... La création de cadres conceptuels et de théories communes, la disponibilité de grandes masses de données sur lesquelles apprendre, et la puissance de calcul des machines ont permis l'émergence soudaine de succès remarquables de l'IA. Tous les problèmes anciens ont simultanément subi des avancées majeures : la reconnaissance de la parole, la traduction automatique, le dialogue homme-machine, la classification des contenus d'une image, la marche des robots, les véhicules autonomes... Alors, que proposent les chercheurs pour la suite des événements ?

Trois sujets sont à traiter à court terme : **l'impact de l'IA sur l'économie, les questions d'éthique et de droit, la robustesse des artefacts**. Pour le premier, il s'agit de maximiser les effets bénéfiques sur l'économie tout en minimisant ses effets délétères. Quels vont être les effets sur le marché du travail, et sur la notion même de travail ? Comment des secteurs comme la

banque, les assurances, le marketing vont être modifiés par la connaissance extrêmement fine des comportements des clients ? Quelles politiques mener pour atténuer les effets négatifs et quelles nouvelles métriques utiliser pour prendre ces décisions ? Des questions d'éthique et de droit en découlent, notamment pour les véhicules autonomes : quelles décisions peuvent-ils prendre pour minimiser des accidents, et qui endosse les responsabilités ? Quel rôle jouent les informaticiens dans la construction des algorithmes et leurs conséquences, notamment dans le cadre de la surveillance ou de la gestion des données relevant de la vie privée ? Enfin, pour que la société accepte largement les artefacts intelligents, ils doivent être *vérifiés* (ils font ce qu'ils sont censés faire), *valides* (ils n'ont pas de comportement ayant des conséquences non souhaitées), *sûrs* (ils ne sont pas piratables) et *contrôlables* (ils peuvent être corrigés après leur déploiement).

À long terme, il s'agit de développer des systèmes qui peuvent apprendre de leur expérience – et notamment à partir de peu d'exemples, ce qui n'est pas le cas aujourd'hui avec l'apprentissage profond – d'une manière semblable aux humains, jusqu'à surpasser les performances humaines dans la majeure partie des tâches cognitives, ce qui aura un impact réellement majeur sur la société humaine.

La nature des IA est à présent claire. Ce sont des formes d'intelligence nouvelles, créées par l'humain et se créant elles-mêmes, accompagnant l'humanité dans son quotidien, de manière bienveillante et bénéfique.

Bibliographie & lectures complémentaires

An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence.
Janvier 2015 <http://goo.gl/u0EZNq>

Computing Machinery and Intelligence.
A. M. Turing (1950) *Mind* 49: 433-460.
<http://goo.gl/4sFMBw>

Quelle est votre intelligence dominante ?
Pierre Paperon. Mars 2016
<https://goo.gl/QhhO4W>

Superintelligence: Paths, Dangers, Strategies
Nick Bostrom (2014)

L'erreur de Descartes – La raison des émotions
Antonio R. Damasio (1995)

*What do games tell us about intelligence?
On measuring human intelligence from comparisons.* Johan Ugander. Mars 2016
<https://goo.gl/FmvaP1>

Are Insects Conscious?
Peter Singer, Mai 2016 <https://goo.gl/FRd0ya>



Un quotidien bercé d'Intelligences Artificielles

La reconnaissance du langage naturel

Le traitement du langage naturel est une discipline clé de l'IA, d'autant plus qu'elle est au cœur du problème de Turing, puisqu'elles doivent convaincre un être humain de leur humanité... Au-delà de la captation, puis la synthèse, de l'écrit ou de l'oral, c'est toute l'ambiguïté d'une langue, les jeux de mots, les intentions, la tonalité des locuteurs, sans oublier les sous-entendus, les savoir communs non explicités, la culture, qui doivent être pris en compte. Ces systèmes, de plus en plus performants grâce à des combinaisons d'algorithmes et l'accès à des corpus de textes et de conversations, s'apprentent à rendre tous nos objets causant.

« *Le modèle économique pour les 10000 prochaines startup est facile à prédire : prendre X, ajouter de l'IA. Tout ce que nous avons électrifié, nous allons le cognitiser* », pouvait-on lire fin 2014 dans Wired. Et effectivement 2015 a été l'année où les intelligences artificielles sont véritablement entrées dans notre quotidien, sous l'impulsion des grands groupes de l'Internet, et des startup se reposant sur les **commodités** qu'ils ouvraient (voir page 18). On peut même affirmer que la révolution numérique en cours est en train de libérer et d'augmenter nos capacités cognitives, comme la révolution industrielle a contribué à libérer et augmenter les capacités musculaires des humains, en leur apportant la puissance de la vapeur puis de l'électricité. Et comme pour ces révolutions précédentes, cela ne se fera pas sans une réorganisation du travail, voire de la notion de travail elle-même (voir page 20).

Alors que les IA originelles ont été développées pour émuler l'intelligence rationnelle, c'est principalement pour assister notre intelligence émotionnelle et notre intelligence sociale que les IA d'aujourd'hui sont conçues. Il s'agit de faciliter et d'augmenter tous nos processus de communication, qu'ils se fassent entre humains - outils de traduction en temps réel, reconnaissance des visages et des émotions qu'ils portent - ou entre humains et machines. "Take X, Add AI" peut s'entendre au sens littéral du terme : prendre n'importe quel objet, et l'augmenter de capacités intelligentes minimales pour que son utilisation se fasse le plus naturellement possible. Capacités intelligentes qui servent aussi à échanger entre objets connectés, par exemple quand le thermostat qui a appris des comportements des utilisateurs de la maison à quelle température régler les pièces

selon les heures et les personnes présentes, devient le cœur de la plate-forme de gestion domotique de la maison.

Peu à peu, des agents conversationnels généralistes, comme Siri d'Apple, deviennent familiers, même si la reconnaissance du langage naturel reste un vaste sujet de recherche, loin d'être résolu. Familiers, jusqu'à un certain point, car si un objet physique renseigne habituellement sur son usage par sa forme, ou un logiciel par son interface à base d'icônes et de menus, l'interaction avec un agent conversationnel nécessite encore un apprentissage de la part de l'humain, qui ne sait pas toujours quoi attendre de son compagnon virtuel, ce à quoi il saura répondre ou pas, et jusqu'où lui reprocher ses incapacités.

La maison est un des enjeux de déploiement de ces IA, un scénario qui nous est déjà connu grâce au cinéma : majordome virtuel ou présence artificielle. Lors de la conférence pour développeurs Google I/O en mai 2016 a ainsi été dévoilée la nouvelle génération de Google Assistant, un **bot** qui utilise le contexte des conversations en cours pour apporter une assistance pertinente à la volée, et qui enrichira également Google Home d'une « *expérience ambiante allant au-delà des appareils* ».

Pour proposer de tels services, ces assistants doivent écouter en permanence. Mais « *faciliter notre quotidien doit se faire tout en respectant notre vie privée* », rappelle Rand Hindi, CEO de Snips, un des *Innovators Under 35* en France en 2014, qui développe une IA pervasive, apprenant de nos habitudes et interagissant pour nous avec la technologie. Tout l'enjeu est de chiffrer les données personnelles collectées et apprendre à partir de ces données chiffrées.

L'avènement de l'apprentissage machine

Pour mettre en œuvre ces scénarii qui étaient encore futuristes il y a 5 ans, les machines doivent apprendre d'un environnement non prédictible, sur des données de tout type, arrivant massivement, dont la signification brute a été préalablement chiffrée. Algorithmes d'apprentissage et *big data*, deux des trois piliers de *la nouvelle Intelligence Artificielle*, sont donc aujourd'hui fortement liés.

Cependant, si on associe de plus en plus souvent l'intelligence naturelle comme artificielle aux capacités d'apprentissage et d'adaptation, cela n'a pas toujours été le cas. Pendant longtemps les problèmes ont été résolus sous forme de manipulations de règles – s'il pleut, prendre un parapluie – et d'arbres de décision où chaque branche est une option de règle. Mais de tels systèmes se sont avérés inefficaces face aux problèmes de reconnaissance de formes et de reconnaissance vocale, pour lesquels les données à comprendre sont complexes, variables, variées, bruitées. Il a fallu imaginer des systèmes capables de s'entraîner sur des bases d'exemples, d'en faire émerger des traits, de généraliser sur des exemples non encore rencontrés et de s'améliorer continuellement avec l'expérience : c'est le *machine learning*, ou *apprentissage automatique*, qui comporte aujourd'hui plusieurs dizaines d'algorithmes, que l'on peut classer de différentes manières, ou *types d'apprentissage*.

Les types d'apprentissage

L'approche la plus courante est celle de l'*apprentissage supervisé* : est présenté à un système algorithmique apprenant un (grand) ensemble d'entrées associées à des étiquettes, par exemple des objets avec leur catégorie, et le système ajuste ses paramètres internes – on parle de poids – jusqu'à pouvoir *classifier* des exemples non encore présentés, faisant

ainsi preuve de capacité de généralisation. Les réseaux de neurones sont de tels systèmes. Ils se construisent un modèle interne du monde qui leur est présenté, à base d'extracteurs de caractéristiques et de classifieurs simples.

Cet apprentissage nécessite une intervention extérieure pour que les règles d'apprentissage soient appliquées. Un enseignant fournit des cibles désirées pour chaque motif d'entrée et le système apprenant se modifie en conséquence. A contrario, *l'apprentissage non-supervisé* se fait sans enseignant guidant le système. Ce dernier doit se forger lui-même ses propres catégories de l'espace des motifs d'entrée, le plus souvent sur des bases statistiques. Entre les deux, *l'apprentissage semi-supervisé* travaille sur des données dont certaines sont étiquetées et d'autres non, ce qui peut améliorer les qualités de chacune des approches prises indépendamment. Le *co-apprentissage* est un sous-ensemble de cette classe, à base de classifieurs travaillant sur des caractéristiques différentes et idéalement indépendantes, et communiquant sur leurs résultats. *L'apprentissage hétéro-associatif* apprend à associer des formes de sortie à des motifs présentés en entrée. *L'apprentissage auto-associatif* permet de reconstituer un motif incomplet présenté en entrée et qui n'a été qu'évoqué.

L'apprentissage par renforcement est pour sa part un type spécial d'apprentissage supervisé. Très peu d'information est donné au système, et notamment aucune sortie désirée. Un jugement est émis sur les performances du système, à lui de se modifier pour que les jugements reçus soient de plus en plus positifs. Il ressemble en cela à des techniques d'apprentissage réservées aux animaux. Combiné à l'apprentissage profond, il a permis fin 2015 la victoire par AlphaGo contre le champion européen.

“ il reste beaucoup à faire en apprentissage non-supervisé... ”



Nearest Neighbors



Linear SVM



RBF SVM



Decision Tree



Random Forest



AdaBoost



Naive Bayes



Linear Discriminant Analysis



Quadratic Discriminant Analysis

Savoir choisir l'algorithme d'apprentissage, ou la combinaison d'algorithmes qui permet de résoudre un problème de classification, reste tout un art. Ce savoir-faire peut heureusement s'acquérir en partant d'exemples très visuels. L'image ci-dessus est extraite du site <http://scikit-learn.org/>, un ensemble d'outils en Python très prisé pour l'analyse et la fouille de données, au développement duquel Télécom ParisTech participe. Des sites web comme <http://playground.tensorflow.org/> offrent également la possibilité de s'exercer à choisir les architectures et les paramètres de départ de réseaux connexionnistes. <http://www.mln.io/resources/periodic-table/> présente plusieurs dizaines d'outils et de bibliothèques de machine learning.

Les succès du *deep learning*

La reconnaissance des formes, et notamment des images, a depuis longtemps été traitée par combinaison de différents extracteurs de caractéristiques et classifieurs apprenants. Ceux-ci sont souvent des réseaux neuronaux – leurs composants principaux modélisent de façon très simplifiée les neurones biologiques – multicouches – la topologie de ces réseaux permet d'identifier des neurones d'entrée, des neurones de sortie et une ou plusieurs couches intermédiaires de neurones, reliées par leurs synapses et modifiant leurs poids synaptiques. **L'apprentissage profond**, ou *deep learning*, implémente cette idée sur un très grand nombre de couches intermédiaires et un encore plus grand nombre de neurones. C'est ce gigantisme, permis par les capacités de calcul actuelle, qui est au cœur de la très grande majorité des avancées décisives de ces derniers mois. Le grand avantage de ces architectures profondes est « *leur capacité d'apprendre à représenter le monde de manière hiérarchique* », comme l'explique Yann LeCun, dont les travaux sur les *réseaux convolutifs*, « *une forme particulière de réseau neuronal multicouche dont l'architecture des connexions est inspirée de celle du cortex visuel des mammifères* », a permis dans les années 90 les premiers systèmes de reconnaissance de chèques. Tous les systèmes de reconnaissance vocale actuels mettent en œuvre ce type d'apprentissage aujourd'hui.

Les succès du *deep learning* ne doivent cependant pas donner à croire que toute l'IA s'y réduit aujourd'hui. Pour arriver à des *IA générales*, loin des IA actuelles qui ne font que savoir très bien jouer au Go, ou aux échecs, mais sont incapables sur d'autres tâches, il faut leur permettre de se remémorer, de prédire et de planifier. De nouvelles architectures de réseaux de neurones, les *réseaux récurrents à mémoire*, sont ainsi conçues, capables par exemple de légendier automatiquement des images.

Les êtres vivants comprennent, parce qu'ils savent comment fonctionne le monde, car ils y vivent. Cette aptitude à la **compréhension du**

sens commun, comme le souligne Yann LeCun, est la question fondamentale qu'il faut résoudre à présent. **L'apprentissage prédictif** pourrait être une piste. Il s'agit d'un type d'apprentissage très particulier : on montre au système apprenant une séquence d'événements (les images d'une vidéo, les mots d'une phrase, les comportements d'un utilisateur) et on lui demande de prédire les éléments suivants.

À Télécom ParisTech, Albert Biffet mène des recherches dans le cadre du *machine learning on real time*. Il s'agit de fouiller des données (*datastream mining*) arrivant à haute vélocité. Cela peut être par exemple des flux de données de télécommunications. Aucune donnée n'est stockée et il faut s'adapter aux changements en temps réel. Les enjeux actuels sont de permettre du *machine learning* de ce type dans le cadre de l'Internet des objets, notamment pour permettre à ces objets de prendre de meilleures décisions. Un équilibre est recherché entre avoir plus de données à disposition ou avoir de meilleurs algorithmes pour prendre ces décisions sous contrainte de temps réel. Dans certains cas les flux de données étudiées arrivent de manière distribuée. Quatre types d'algorithmes sont retenus : *classification, regression, clustering, et frequent pattern mining*.

Albert Biffet développe la prochaine génération de modèles et d'algorithmes pour fouiller les données en provenance de ces objets : des algorithmes qui ne cessent jamais d'apprendre, des algorithmes parallèles et distribués sur plusieurs machines, et enfin l'application de l'apprentissage profond à ces flux de données.

Ces recherches bénéficient de l'existence de la chaire **Machine Learning for Big Data à Télécom ParisTech**, pilotée par **Stephan Cléménçon**, et dont les travaux ont déjà été traités dans le cahier de veille Big Data. Créée en septembre 2013 avec le soutien de la Fondation Télécom, et financée par quatre entreprises partenaires (Safran, PSA Peugeot Citroën, Criteo et BNP Paribas), la chaire produit une recherche méthodologique répondant au défi que constitue l'analyse statistique des données massives.

Les IA au service du marketing

Les interfaces homme machine permises par les capacités conversationnelles des Intelligences Artificielles, et les apprentissages prédictifs sur des grandes masses de comportements, offrent une interaction d'un nouveau type avec les consommateurs, plus naturelle. « *Le but ultime est d'être pertinent dans les interactions avec les consommateurs* », rappelait Samir Amellal (Publicis ETO) lors d'un séminaire interne de la Fondation Télécom. « *Il s'agit plus de faire des découvertes d'atypisme, de nouveaux marchés et de nouveaux usages, que de reconnaître dans des masses / flux de données des éléments déjà connus.* »

“...et dans la compréhension
du sens commun”

Les formes d'apprentissage

Certains mécanismes d'apprentissage, comme ceux liés à la motricité des robots, ne peuvent pas se reposer uniquement sur le *big data*, comme la plupart des algorithmes d'IA en vogue, car l'adaptation d'un robot à son environnement dépend des robots dans toutes leurs variétés et leurs états, et des environnements dans toutes leurs déclinaisons et variabilités. Il faut alors s'intéresser à des *formes d'apprentissage* qui mettent en situation dans son environnement le système apprenant. Quand les *types d'apprentissage* mettent en exergue les processus en jeu, les *formes d'apprentissage* s'intéressent au contexte, comportemental et environnemental, dans lesquels ils agissent. Elles sont multiples et dépendent du développement cognitif de l'être – ou de l'artefact – observé.

L'empreinte et l'habitation sont des apprentissages primaires et irréversibles. La première est acquise brutalement, réalisation de comportements innés. La seconde permet de ne pas considérer comme nouveaux tous les événements qui se présentent. Elle rend possible l'expérience et le raisonnement. **L'apprentissage par essai-erreur** repose sur les systèmes de valeur, et les récompenses / punitions. Il permet l'acquisition d'automatismes. **L'apprentissage par l'action**, fortement dépendant de l'environnement, regroupe l'acquisition du savoir-faire par l'entraînement, la recherche active d'informations, la manipulation (d'objets, de modèles...). Il peut avoir comme objet la pensée elle-même (échafaudage d'hypothèses et vérifications) et permettre ainsi la puissance créatrice. **L'apprentissage par observation / imitation** décrit des situations où l'apprenant copie le comportement d'un modèle sans que celui-ci ne l'instruise ni ne le corrige. C'est un type d'apprentissage courant dans les premières périodes du développement intellectuel, et plus tard dans les phénomènes de mimétisme. **L'apprentissage coactif** s'effectue entre des êtres ayant un niveau d'expertise comparable. Toute sa force réside dans la dynamique d'échange entre les deux, à base d'hypothèses, de tests et de critiques. **L'apprentissage par instruction**, a contrario, s'effectue entre deux êtres aux niveaux d'expertise différents. Complétant les apprentissages par action et par observation / imitation, l'expert dirige l'avancée de l'élève en communiquant avec lui.

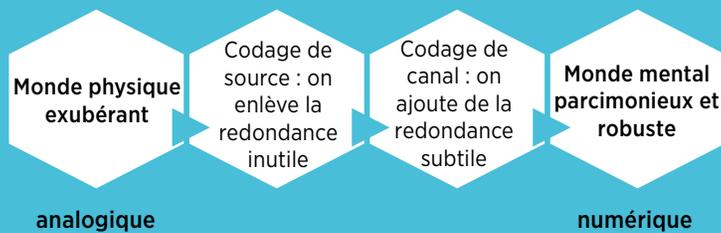
Enfin, une dernière forme d'apprentissage est celle qui consiste à apprendre à apprendre. **Le méta-apprentissage** permet, en observant ses propres processus d'apprentissage, d'activer les bons mécanismes d'apprentissage pour chaque situation nouvelle, et créer éventuellement de nouveaux mécanismes d'apprentissage.

À Télécom Bretagne, Mai Nguyen élabore des robots adaptatifs, dans le cadre des services d'aide à la personne. L'apprentissage étudié est par exemple l'apprentissage actif par curiosité artificielle. C'est ce que fait un enfant entre 0 et 6 ans qui ne va pas à l'école, et pourtant apprend à se développer. Il joue, en autonomie. C'est l'enfant qui décide, parmi tous les jeux possibles. Les psychologues disent que l'enfant va choisir de manière intuitive la tâche qui va lui permettre d'apprendre le plus. Chaque enfant a envie de faire des progrès, à sa juste mesure, et choisit des choses à apprendre pour le bien qu'il en tire lui-même. Il dispose pour cela de différentes stratégies d'apprentissage à disposition, par guidage social, ou en autonomie. C'est en s'inspirant de ces observations que la chercheuse imagine des robots capables d'apprendre de leur environnement physique.

L'objectif est celui d'une robotique de service qui se propose de créer des robots intelligents grâce à l'agrégation des meilleurs algorithmes d'IA. Un premier défi consiste à agréger ces algorithmes en un système cohérent et temps réel, sur un système embarqué. Un deuxième défi est de construire des algorithmes pour que le robot puisse interagir physiquement et affectivement avec son environnement physique et ses utilisateurs. Un troisième défi est de permettre au robot d'apprendre des nouvelles tâches qui n'avaient pas été préprogrammées. Il s'agit donc d'aller au-delà des défis les plus connus, où la tâche (unique) ainsi que les cas de test sont définis à l'avance et avec précision. La robotique intelligente bénéficie indéniablement des récentes avancées en IA, mais doit également continuer à trouver de nouveaux paradigmes pour avancer dans les défis qui lui sont propres. Par exemple, quand ces robots sortent d'usine, ils ne sont pas encore terminés, pour pouvoir s'adapter par la suite à leur environnement. Ils nécessitent donc une structure cognitive leur permettant d'apprendre.

À Télécom Bretagne, Claude Berrou et son équipe cherchent au sein du projet européen Neucod (*Neural Coding*) à modéliser le cerveau humain par une approche jusque-là inédite : la théorie de l'information. Ils partent du constat que le néocortex, ce « milieu de propagation qui permet à des processus biologiques de passer d'îlots de connaissances à d'autres », possède une structure très proche de celle des **décodageurs** modernes. Leurs travaux ont permis de développer des codes de représentation et de mémorisation de l'information expliquant sa robustesse et sa pérennité.

Modèle de Shannon appliqué au cerveau



Tout démarre d'une analogie entre la théorie de l'information de Shannon et le cerveau humain. Cette analogie a permis le développement de mémoires associatives robustes et à grande diversité d'apprentissage, grâce aux **travaux de Vincent Gripon**. Elle prend en compte un monde qui est réel, riche, complexe, et surtout analogique. La phase de perception revient à du codage de source (enlever la redondance de l'environnement), et celle de mémorisation à du codage de canal (rajouter de la redondance pour rendre le signal plus robuste).



Vers une informatique neuro-inspirée

Le calcul informatique a connu des développements considérables ces quarante dernières années, grâce aux avancées de la physique et de la microélectronique. L'algorithmique a également progressé, notamment depuis les travaux pionniers de Donald Knuth, mais à un rythme nettement plus lent que celui de la microélectronique. Nos ordinateurs ultrarapides, multicœurs et dotés de vastes mémoires sont restés de rudimentaires machines à programmer, non apprenantes, non adaptatives, indifférentes, en un mot inintelligentes.

Deux raisons peuvent être données pour expliquer l'échec, provisoire, de l'IA. Tout d'abord, comme l'a expliqué Yann LeCun, trop de scientifiques ont manqué de pragmatisme dans l'élaboration des méthodes qui auraient pu servir aux progrès de l'IA : « *il fallait avoir une théorie à toute épreuve, et les résultats empiriques ne comptaient pas. C'est une très mauvaise façon d'aborder un problème d'ingénierie* ». Et s'il y a bien un domaine où, en l'état actuel des connaissances, aucune théorie unificatrice ne s'est imposée, c'est celui de l'intelligence, qu'elle soit naturelle ou artificielle. Trop souvent également l'approche descendante – *top-down*, de la cognition vers les modèles – a été privilégiée au détriment de l'approche *bottom-up* – des circuits vers la cognition – moins propice à la théorisation.

Ensuite et surtout, l'idée d'une Intelligence Artificielle n'a été portée, jusqu'à ce jour, que par des experts de l'informatique (au sens le plus large des sciences de l'information), lesquels ont cru pouvoir se passer d'imiter les seuls modèles disponibles et en ordre de marche : les circuits de notre cerveau. Pourtant, si l'on veut obtenir des machines électroniques des propriétés cognitives supérieures, telles que l'imagination, la création, le jugement, on doit cesser de suivre à la lettre le modèle de la machine informatique conventionnelle ou plutôt essayer de lui trouver un modèle compagnon, le premier se chargeant des gros calculs (ce que le cerveau humain ne sait pas faire) et le second de l'intelligence. Et nous n'avons qu'une seule source d'inspiration pour obtenir cette Intelligence Artificielle : notre cortex.

Il semble donc impératif, dans un nouvel élan vers l'IA, qu'on appellera plutôt **informatique neuro-inspirée**, de bien connaître le cerveau dans ses aspects **mésoscopiques**, c'est-à-dire le codage neural et les circuits neuraux. Cette approche a également l'avantage de répondre à certains verrous actuels du domaine, comme la consommation énergétique ou d'autres problématiques liées aux systèmes embarqués.

Dans le cortex humain, l'intelligence s'appuie sur une structure dans laquelle les opérations

(suite page 16...)

L'IA entre mythes & réalités

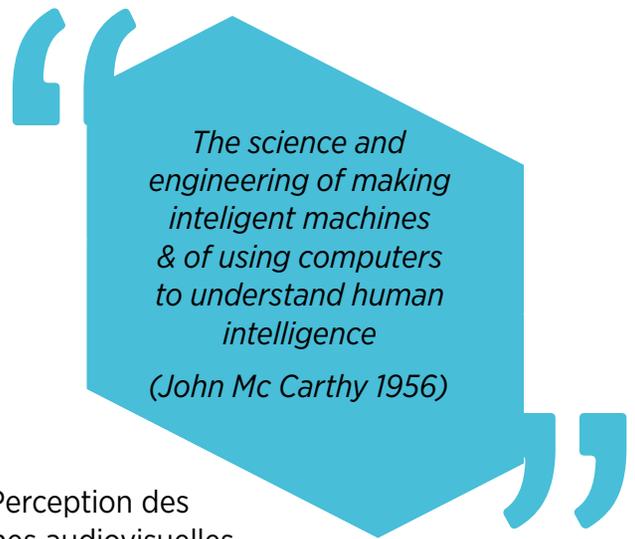
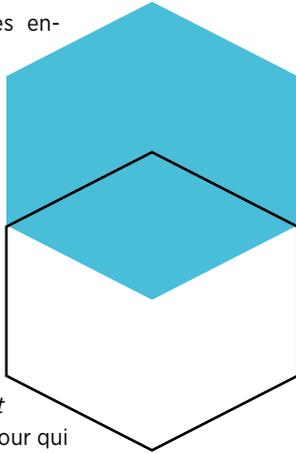
Une science qui (s')interroge

En 1972 paraît " *What Computers Can't Do : The Limits of Artificial Intelligence* ", un essai de Hubert Dreyfus, qui propose une critique philosophique de l'Intelligence Artificielle. Il y relève que les travaux réalisés dans les années 60 ont eu des résultats très en-deçà des espérances et des annonces faites. Il montre les limitations d'une approche de la représentation des connaissances sur la seule base de manipulation des symboles et de règles. Jetant un froid sur les Sciences Cognitives de l'époque, un *AI winter*, il propose néanmoins de se reconcentrer sur trois secteurs négligés : le corps, la situation et les intentions. Paru en France sous le titre " *L'intelligence artificielle : mythes et limites* ", ces réflexions, et les mises à jour qui ont suivi, ont marqué les sciences cognitives naissantes, ont permis d'explorer de nouveaux paradigmes, et sont allés puiser des idées dans des disciplines variées, tout en gardant une vision critique des pistes explorées.

Aujourd'hui encore, l'Intelligence Artificielle ne cesse d'être questionnée, dans ses objectifs, dans ses réalisations, dans l'organisation de sa recherche, dans sa définition même. Elle alimente et inspire en retour les disciplines qui la composent, et grâce aux avancées importantes récentes, permises

par la combinaison de nouveaux algorithmes, de grands corpus de données à explorer et de capacités de calcul décuplées pour ce faire, elle est entrée de plein pied dans notre quotidien, et devient un objet de débat public.

Prendre conscience de leurs limites tout en admirant leurs réalisations, et dépasser les craintes et les peurs qu'elles peuvent susciter, c'est se donner toutes les chances de préparer un monde d'intelligences artificielles bienveillantes avec lesquelles évoluer.

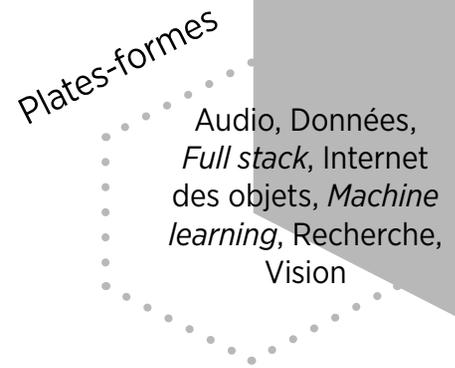
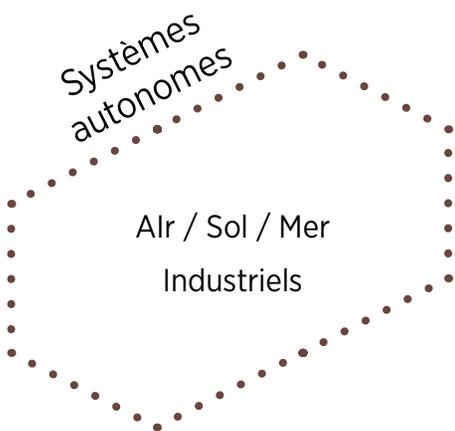
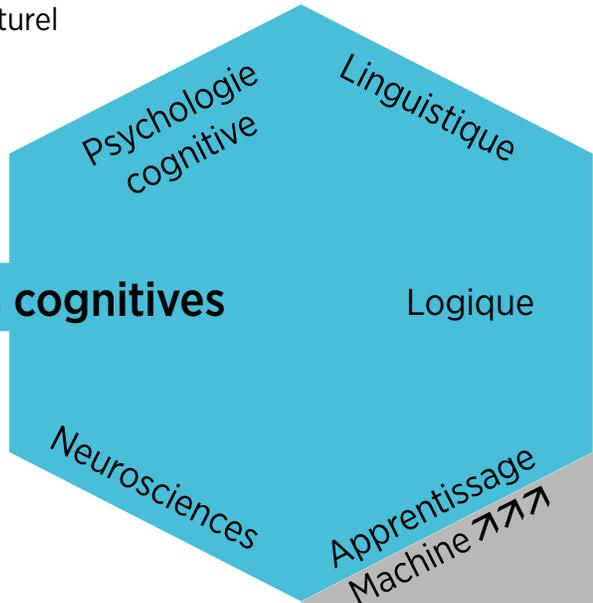


The science and engineering of making intelligent machines & of using computers to understand human intelligence
(John Mc Carthy 1956)

- Perception des formes audiovisuelles
- Interactions naturelles
- Indexation des contenus
- Extraction de connaissances
- Compréhension du langage naturel

- Représentation des connaissances &
- Recherche s
- Systèmes à base de co

Sciences cognitives





Légende

Au centre : vue synthétique de l'Intelligence Artificielle.
D'après Henri Sanson, Orange.

En pointillé : écosystème du marché de l'Intelligence Artificielle.
Par Shivon Zilis, Bloomberg Beta.
Analyse et détail des entreprises & projets à découvrir sur www.shivonzilis.com



ation des
& raisonnement

sémantique

s experts
nnaissances

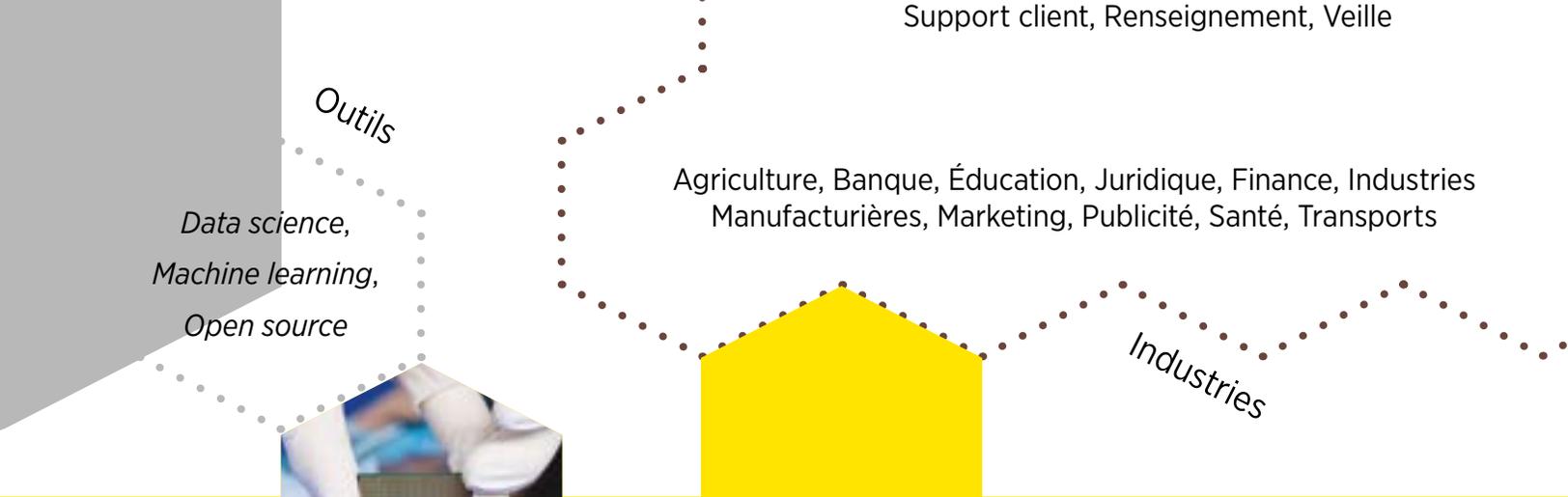
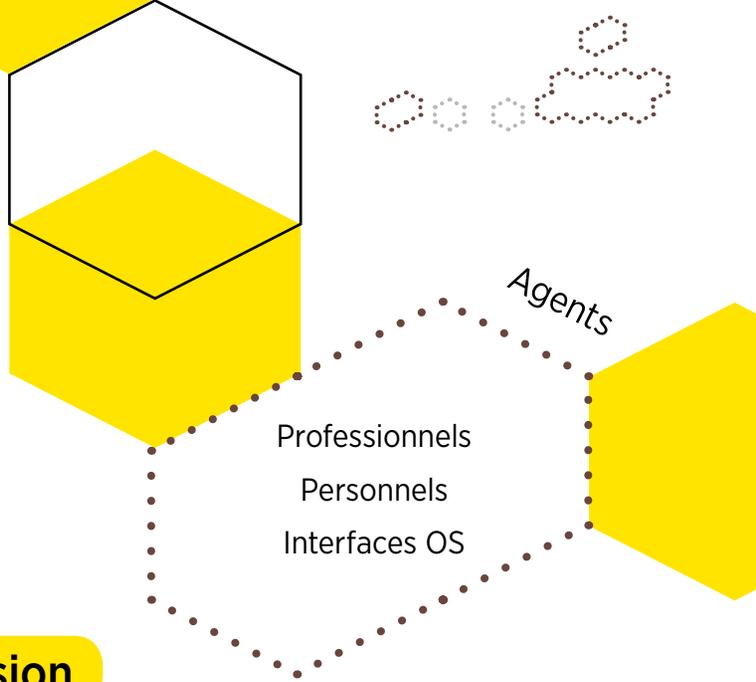
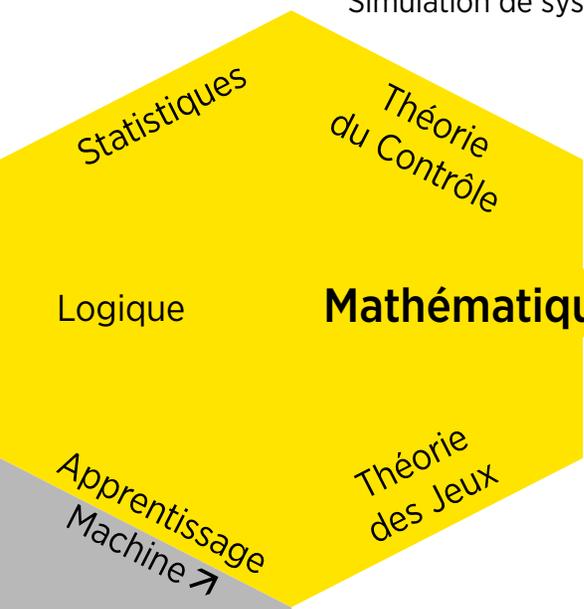
Systèmes multi-agents

Processus décisionnels

Pilotage de processus

Stratégie & tactique

Simulation de systèmes complexes



Outils

Data science,
Machine learning,
Open source

et la mémoire s'intriquent intimement. L'équipe Neucod a fait le pari que c'est de la compréhension de la mémoire, ses circuits, ses principes – la règle de Hebb par exemple, voir encart ci-dessous –, que découlera la compréhension des opérations, et non l'inverse. Un premier sujet d'étude a donc été la mémoire à long terme, et en particulier la manière dont elle s'organise, emmagasine et restitue. Des travaux menés depuis 2009 ont conduit à proposer des modèles de cette mémoire à grande efficacité, biologiquement plausibles, et reposant sur l'hypothèse d'une information mentale numérique. Ces travaux s'appuient sur la confrontation de modèles artificiels avec la réalité observée de la connectivité de réseaux de neurones réels.

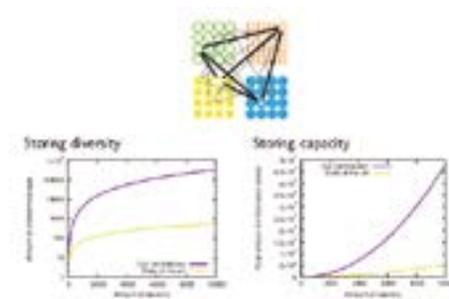
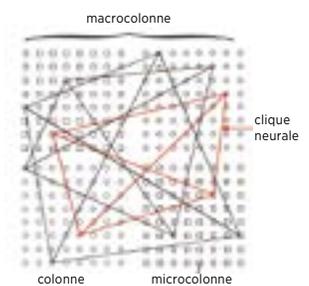


Figure ci-dessus. L'objet mental est identifié à l'état physique créé par l'entrée en activité (électrique et chimique), corrélée et transitoire, d'une large population ou "assemblée" de neurones distribués au niveau de plusieurs aires corticales définies, expliquait JP Changeux en 1983 (voir encart ci-dessous).

Les systèmes *deep learning* nécessitent des millions d'exemples et soulèvent des problèmes de robustesse : si on modifie l'architecture et la topologie choisies, tout s'écroule. Ils présentent également des problèmes d'intégration dans des petits circuits, ce qui limite les capacités des IA embarquées. **Nicolas Farrugia** les implémente, au sein de l'équipe Neucod, dans des *FPGA*, tandis que des recherches sont faites pour « passer en binaire » ces réseaux *deep learning*, c'est-à-dire les compresser. De nombreux travaux estiment en effet que le cerveau « stocke en binaire » : ce qui compte sont les connexions, pas leur force. Les résultats ont montré une bonne résistance des taux d'erreur à des compressions fortes.



Cette assemblée qui se décrit mathématiquement par un graphe, est "discrète", close et autonome, mais n'est pas homogène. C'est le pattern de connectivité qui contient l'information que l'on cherche à stocker, en introduisant une redondance spatiale. On obtient alors une capacité de stockage bien supérieure à celle de l'état de l'art.

Le système nerveux humain

100 milliards de cellules nerveuses, 30 milliards de neurones dans le cortex, 200 000 nouveaux neurones par seconde entre les 3^e et 5^e mois de la vie fœtale, le cerveau humain est une machine (!) complexe, fascinante, qui n'a pas encore révélé tous ses secrets. Quelques notions élémentaires sur sa topologie, la dynamique de sa création et certaines règles d'activation des influx nerveux sont utiles à connaître.

Le développement du système nerveux commence dès l'embryogénèse et se poursuit bien après la naissance. Il est constitué de deux types de cellules, les *neurones*, et les *cellules gliales* qui en forment l'environnement et dont la part dans les processus cognitifs a trop

longtemps été ignorée et reste mal comprise. En 5 semaines chez l'embryon humain un *tube neural* est formé, matrice initiale qui va produire l'ensemble des cellules du cerveau, par division cellulaire. Celles-ci vont *migrer* vers leur position finale, se différencient pendant leur migration. L'observation en coupe en profondeur d'un cortex final montre distinctement 6 zones, numérotées I à VI à partir de la surface. La couche IV contient un type de cellules particulier, les cellules étoilées, intercalées entre les cellules dites pyramidales des couches inférieures et supérieures. Ceci se retrouve dans toutes les aires du cerveau, avec cependant des variations en densité.

Un aspect essentiel du développement de ces six couches corticales est qu'il s'effectue

à l'inverse de ce que l'on pourrait imaginer. La couche la plus profonde, la VI, se crée en premier, puis les neurones de la couche V doivent la traverser pour se positionner, et ainsi de suite sur toute la hauteur jusqu'à la couche superficielle. La poussée des axones et des dendrites des neurones, et la formation des synapses, s'effectuent au fur et à mesure de la constitution de ces couches. Le thalamus et le cortex sont ainsi reliés, *bouclés*, avant que les couches III et II ne se forment. Les connexions s'établissent pendant cette épigénèse au fur et à mesure de la mise à disposition du matériel neuronal. Toute une dynamique de croissance, de redondance transitoire des liens entre cellules neuronales puis de stabilisation sélective se met en place, bien décrite par J.P. Changeux dans son ouvrage *L'homme neuronal* en 1983.

Il s'agit à présent de lier ces modèles, d'une part en amont aux procédés connus ou à découvrir de l'apprentissage machine, d'autre part en aval aux mécanismes de la communication cérébrale. L'apprentissage machine sur réseaux de neurones, sous l'impulsion d'acteurs industriels majeurs, fait aujourd'hui l'objet d'une Recherche & Développement de grande ampleur et ambitieuse. Cependant, les méthodes mises en œuvre requièrent d'énormes moyens de calcul, qui ne correspondent pas aux aptitudes du cerveau humain ni à sa consommation d'énergie. Là encore, une meilleure compréhension des mécanismes sensoriels naturels permet d'orienter la recherche vers des solutions algorithmiques bien plus efficaces parce que véritablement neuro-inspirées. Quant à la communication cérébrale, sous ses aspects purement informationnels, peu d'équipes à travers le monde s'attellent à la modéliser et à en exploiter les principes.

L'objectif principal d'une telle thématique de recherche est la spécification, la conception et le test de machines informatiques connexionnistes cortico-inspirées capables d'emmagasiner de grandes quantités d'information et de les croiser pour pouvoir répondre à des questions inédites. Ces machines, appelées à devenir les auxiliaires intellectuels dont l'humanité aura besoin dans les années à venir pour l'aider à répondre à des défis de plus en plus nombreux et prégnants,

fonctionneront à la vitesse électronique, c'est-à-dire un million de fois plus rapidement que le cerveau humain, auront des capacités d'apprentissage des centaines de fois plus étendues que ce dernier, mais devront être conçues à la façon du cortex et non pas selon le modèle informatique classique. Seule une machine connexionniste neuro-inspirée pourra offrir cette propriété de l'intelligence, avec les avantages supplémentaires de la vitesse électronique et de capacités mnésiques accrues.

Pour avancer dans cette direction, il faut réunir des scientifiques issus de disciplines aujourd'hui aussi distantes que peuvent l'être, par exemple, la neuropsychologie et la théorie de l'information. Il faut rassembler des chercheurs et des ingénieurs issus des disciplines des neurosciences traditionnelles (neurobiologie, neuropsychologie, imagerie...), des sciences de l'information (informatique, théorie de l'information, apprentissage...), des mathématiques (théorie des graphes, probabilités...) et de l'électronique (architecture et circuits). La thématique dont l'équipe de recherche Neucod à Télécom Bretagne préconise la consolidation sera la première réponse française à ce besoin d'interdisciplinarité, voire de transdisciplinarité, lesquelles devront allier excellence de la recherche et souci de l'application et du transfert industriel.

Parmi les principaux axes de travail d'un grand projet tel que celui-ci, on peut citer :

- l'élaboration de modèles connexionnistes de représentation et de mémorisation de l'information mentale, ouvrant la voie à de nouvelles techniques de classification et de généralisation,
- la modélisation de la transmission de l'information dans le cerveau,
- la modélisation de l'intentionnalité, de la pertinence, de la motivation, de la préférence, de la curiosité et de quelques autres concepts fondamentaux des sciences cognitives,
- la recherche d'éléments de validation de ces modèles par l'imagerie cérébrale,
- l'étude d'architectures de circuits neuromorphiques, analogiques et/ou numériques,
- la diversification des domaines d'applications de ces modèles connexionnistes : mémoires associatives, classificateurs, moteurs de recherche, traitement du langage naturel, traduction, ... et bien sûr auxiliaires intellectuels généralistes.

S'inspirer du cerveau

On peut constater que les neurones connectés aux sens n'ont pas un accès direct au cortex. Un relais s'effectue dans les centres sous-corticaux comme le thalamus. Il en est de même des autres entrées du cortex, en provenance des aires motrices ou associatives. Un grand nombre d'entrées au cortex provient enfin du cortex lui-même, d'un même hémisphère ou de l'hémisphère opposé. La couche IV, celle des cellules étoilées, dont les axones s'étendent horizontalement dans cette couche, constitue la porte d'entrée principale du cortex. Ici s'établissent des connexions entre fibres montantes (les entrées) qui s'éclatent en une arborisation sphérique, avec d'autres fibres verticales et les cellules étoilées, une influence latérale étant possible. On arrive ensuite aux cellules pyramidales qui sont les

portes de sortie du cortex, qui ne s'éloignent pas sans établir des boucles de réentrée via des connexions collatérales, avant de rejoindre les centres sous-corticaux comme le thalamus et y établir une seconde forme de réentrée. Finalement les influx nerveux sont transmis par exemple aux muscles par l'intermédiaire des neurones moteurs.

La topologie et la connectique du cerveau sont deux éléments qui se renforcent. L'influence latérale entre neurones proches est souvent de nature inhibitrice. Ceci permet, par exemple, de bien délimiter dans le cas de neurones liés à la peau le point d'impact sur les sens, et ne pas le véhiculer bruyamment jusqu'au cortex. On identifie par ailleurs dans le cortex des colonnes fonctionnelles. C'est le cas du cortex

visuel où des colonnes de matériel neuronal sont spécialisées dans la reconnaissance des couleurs, d'autres dans l'orientation de l'objet observé, et d'autres encore sont spécifiques à tel ou tel œil. Les voisinages au niveau des sens, ainsi que la topologie, sont conservés au sein de cartes d'axones puis dans le cortex au sein de cartes de neurones. L'étude des échanges entre ces groupes de neurones a suggéré des algorithmes d'apprentissage, comme ceux construits sur la règle de Hebb, qui énonce que « des neurones qui stimulent en même temps, sont des neurones qui se lient ensemble », hypothèse vérifiée plus tard dans les synapses d'un gastropode marin. Épigénèse, connectique, topologie, règles d'activation... sont autant de sources d'inspiration pour des futures IA neuro-inspirées.

Les grands projets

S'inspirer du cerveau humain, le reproduire ou bien le comprendre, telles sont les motivations qui animent actuellement des équipes de recherches réunies à travers le monde pour faire des percées décisives dans les sciences cognitives. Deux grands projets internationaux concurrents, aux approches différentes, existent notamment pour tenter de comprendre les mécanismes du cerveau humain. Le premier par ordre d'apparition est d'origine européenne, *Human Brain Project*, le second d'origine nord-américaine, *Brain - Brain Research through Advancing Innovative Neurotechnologies - Initiative*. Une polémique a entaché les débuts du projet européen, sur des questions de méthodologies et de réalisme des objectifs poursuivis.

Human Brain Project (démarré fin 2013), issu de travaux antérieurs effectués en Suisse à l'EPFL (*Blue Brain project* en 2005, qui avait comme objectif de simuler une colonne corticale de neurones du cerveau du rat), cherche à bâtir une simulation informatique du cerveau humain, en partant de toutes les connaissances compilées à ce jour. 90 instituts de recherche européens et internationaux, répartis dans 22 pays différents, sont réunis dans le projet. Premier résultat important, en octobre 2015, l'équipe du Blue Brain Project publie un article qui décrit une simulation d'un cerveau de rat : 31 000 neurones et 40 millions de synapses, soit à un volume du néocortex d'environ 0.29 mm³.

L'initiative américaine, débutée en avril 2013, est plus ambitieuse encore, et mieux dotée financièrement. Elle s'est fixée pour objectif de « cartographier les circuits du cerveau, mesurer les fluctuations d'activité chimique et électronique se produisant dans ces circuits ; et comprendre comment leur interaction donne naissance à nos capacités uniques, tant dans les domaines cognitifs que comportementaux ». À ce type de grand projet on peut également ajouter la recherche du connectome par Sébastien Seung, l'équivalent de la cartographie du génome humain, pour les connexions neuronales, et le site de sciences participatives *Eye Wire* où chacun peut tenter d'identifier dans des images en coupe d'un cerveau les prolongations d'un neurone au sein du matériel cortical.

Tous les grands acteurs industriels d'Internet ont mis en place leurs équipes de recherche. Chez IBM la stratégie se construit autour des capacités de calcul de Watson, conçu pour répondre à des questions formulées en langage naturel, et qui s'est distingué en 2011 en remportant le jeu télévisé Jeopardy. Rob High, directeur technique de Watson trouve cependant qu'il est trop tôt pour employer le terme d'intelligence artificielle. Il préfère parler d'outils capables d'élargir les capacités cognitives humaines. Un des enjeux est d'apprendre aux machines que l'intonation a une influence sur le sens des paroles prononcées, un autre est de les aider à faire le tri au sein des informations dont elles se nourrissent sur Internet.

Enjeux de l'IA en entreprise

L'IA constitue un enjeu de recherche fort chez Orange, avec pour objectifs l'optimisation des performances de ses infrastructures, l'amélioration de la relation et de l'expérience client, ainsi que l'innovation de services.

Orange est ainsi impliquée dans les projets coopératifs ANR *Netlearn* et *H2020 PPP 5G CogNet* sur l'utilisation des techniques d'apprentissage machine pour, respectivement, optimiser l'allocation des ressources radio des réseaux d'accès mobiles et garantir les niveaux de service dans les réseaux virtualisés SDN/NFV (*Software Defined Network/Network Function Virtualization*). Orange participe

également au projet ANR *DATCHA* sur l'extraction de connaissances dans les conversations écrites temps réel (*chatbot*) entre clients et téléconseillers par compréhension automatique du langage naturel.

À découvrir sur <https://recherche.orange.com/dialoguer-avec-des-machines-entre-mythes-et-realite/>

Pour cela, Orange peut compter sur son expertise élevée dans le domaine du *Machine learning* en général, et du *Deep learning* en particulier, dans lequel elle mène depuis plus de 10 ans des travaux de recherche, appliqués notamment à la reconnaissance visuelle (reconnaissance

de textes, reconnaissance faciale, analyse de personnes dans les images) et à la reconnaissance de la parole. En particulier, Orange vient de remporter la tâche 1 du défi technologique *ChaLearn* organisé dans le cadre de la Conférence CVPR 2016 (*Computer Vision and Pattern recognition*), portant sur la reconnaissance de l'âge apparent de visages dans les images.

Chez Airbus *Defence and Space*, l'utilisation des IA se fait notamment dans des produits à longue durée de vie et maintenance difficile. Les IA vont pouvoir leur donner de l'autonomie et réduire la charge de travail en conséquence. Dans l'air, les enjeux sont beaucoup plus

Microsoft a lancé plusieurs projets d'agents évoluant dans des plates-formes logicielles, comme le *bot* Tay qui a maheureusement fait des premiers pas trébuchants sur Twitter, ou le projet Malmo qui consiste à entraîner des agents à se déplacer dans le monde Minecraft. Enfin, quatre catégories d'algorithmes sont disponibles sur son *Azure Machine Learning* : clustering, régression, classification et détection d'anomalies.

Baidu, sous l'impulsion d'Andrew Ng, se concentre sur le développement de véhicules autonomes, un enjeu majeur en Chine, et c'est l'occasion de lire régulièrement sous sa plume des réflexions techniques, juridiques et éthiques très abouties sur ces technologies.

Facebook a déployé en 2015 son premier centre de recherche dédié à l'IA hors des USA à Paris, et développe des recherches tous azimut pour améliorer l'expérience utilisateur de sa plate-forme, en reconnaissance et traitement des images et de leur contenu, et en traitement du langage, notamment pour les *bots* de son application Messenger.

API et open source

En plus des avancées scientifiques majeures, l'engouement autour des IA s'est accéléré grâce à la mise à disposition de capacités IA sous forme d'*API* d'une part (dans les domaines notamment de la vision, ou du prédictif), et du code source des plates-formes de *machine*

learning libéré par des grands opérateurs de l'Internet, d'autre part.

IBM a ainsi transformé Watson en une *IA as a service*, interrogée actuellement par plusieurs centaines de startup qui peuvent ajouter à leurs applications, très facilement et à des coûts réduits, les mêmes comportements intelligents que les plus grands industriels. *Amazon Machine Learning* est un service qui fournit des outils et des assistants de visualisation pour créer des modèles de *machine learning* sans en maîtriser nécessairement les fondements techniques. Les résultats, orientés apprentissage prédictif, sont disponibles pour les applications via de simples API. En mai 2016, Amazon passe en *open source* sa librairie de *deep learning* DSSTNE (*Deep Scalable Sparse Tensor Network Engine*).

L'"*IA as a service*" et l'"*IA for everyone*" font de nombreux émules. Citons par exemple dans le domaine des agents conversationnels l'offre Wit.ai rachetée par Facebook, ou bien encore orbit.ai, assistant.ai ou indicio.io.

C'est ici que Google est le plus inventif, avec un grand nombre d'outils d'apprentissage profond libérés au fil des mois, et expliqués dans des billets de leur blog de la recherche. Ils s'articulent autour de leur deuxième génération de plate-forme d'intelligence artificielle, *TensorFlow*, dont le code a été ouvert fin 2015. Et pas seulement ouvert, mais accompagné d'outils de visualisation des algorithmes, directement sur un navigateur internet.

très prometteurs. Un des outils développés est un assistant spécialisé en reconnaissance visuelle, qui observe et analyse ce qui se passe dans un *workflow*, et le documente. Les process peuvent ainsi être suivis précisément, les procédures rejouées si nécessaire, pour y apporter des précisions ou les optimiser, et les équipes sont mieux accompagnées, notamment en cas de passage de témoin. Des humains restent nécessaires pour entraîner et superviser l'assistant, qui est ainsi intégré dans un groupe mixte, dit "*bots & crowds*", et qui permet à ses collègues humains de se concentrer sur des tâches plus valorisantes.

importants qu'au sol. Il faut pouvoir déterminer en cas de défaillance qui est responsable : les algorithmes, les senseurs, l'humain qui est à côté ? Il faut notamment envisager toutes ces questions avant de s'adresser aux autorités pour obtenir les certifications. D'autres usages chez *Airbus Defence and Space* relèvent de l'apprentissage prédictif et rejoignent les problématiques des professionnels du marketing : détecter des comportements qui se démarquent de l'habitude.

Chez **Accenture**, l'IA est utilisée à la fois en interne dans sa division Accenture Operations, et à destination de ses clients, avec des ROI

Les architectures matérielles

Plus de vingt ans après les premières architectures dédiées à l'hybridation des systèmes experts et des réseaux de neurones, comme la machine ArMenX créée à Télécom Bretagne, les machines hautement parallèles, distribuées, et les processeurs simulant les neurones animent toujours les équipes de recherche, les fondateurs de puces et les industriels. Citons, parmi d'autres, le projet Spinnaker, dédié à la fois aux neurosciences, à la robotique, et à la recherche d'une science de la computation modernisée. Du côté des grands acteurs de l'Internet, *Big Sur open compute project*, initiative de Facebook, ouvre les spécifications de ses équipements matériels dédiés, sur base des *GPU* de Nvidia, et lance un partenariat avec des universités européennes qu'elle équipera de ces machines.

Mais la nouvelle la plus surprenante a été donnée par Google mi-mai 2016, qui a annoncé lors de Google I/O avoir équipé depuis un an ses centres de données d'unités de traitements IA spécifiquement développées en interne. Conçus sur la technologie TensorFlow, ces *TPU* sont déjà utilisées par plus de 100 équipes en interne, mais également par tout le monde, sans le savoir : « *si vous utilisez de la reconnaissance vocale sur le cloud, cela passe par nos TPU. Si vous utilisez la reconnaissance vocale Android, cela passe par nos TPU.* » Ces TPUs auraient un ratio performance / consommation énergétique treize fois supérieur à ceux d'un GPU traditionnel, et auraient procuré 7 ans d'avance à Google.

Un autre projet, Amelia, en partenariat avec l'éditeur américain IPsoft, est pour Accenture l'agent virtuel le plus abouti à l'heure actuelle. Combinant traitement du langage naturel – en anglais –, compréhension des documents et recherche, il est destiné aux actions de conseil en vente et de support client, à la fois pour les interlocuteurs internes et externes de l'entreprise, comme des banques, des sociétés d'assurance et des agences de voyages, qui sont les trois secteurs visés actuellement.

Ces projets s'ajoutent à d'autres hébergés par un laboratoire spécialisé en IA, que le cabinet de conseil a ouvert à Dublin fin 2015.

Considérations économiques

Les grands acteurs de l'Internet nous ont désormais habitués aux annonces surprenantes, mais celle des TPU de Google viendrait jusqu'à bouleverser la Loi de Moore, et remettre en cause les courbes de prévision de développement technologique. Et ce n'est pas la seule certitude qui est ainsi bousculée. La singularité technologique, supposée advenir d'ici vingt ans, portée par la révolution numérique, pourrait coïncider avec une encore plus grande révolution, sociale cette-fois. Une des craintes les plus entendues concernant les IA est celle du grand remplacement par les robots, dans presque toutes les activités humaines – y compris les plus intimes. Les études se multiplient avec des chiffres les plus divers, difficiles à vérifier, mais estimant tous qu'à court et moyen terme jusqu'à 50% des emplois connus à ce jour disparaîtraient. Cette peur du remplacement par les machines ne date pas d'hier, elle peut entraîner des rejets violents de la technologie, le *luddisme* dont nous avons parlé dans le précédent cahier de veille sur l'Humain augmenté.

À cet égard, le cas du déploiement massif des véhicules autonomes, et l'impact prévisible que cet événement aura sur l'économie mondiale, est riche d'enseignement à plus d'un titre : de la disparition de millions d'emplois à la redéfinition du concept des assurances, en passant par la nécessité d'un revenu universel. On estime ainsi qu'aux États-Unis l'arrivée massive prochaine de camions automatisés sur les routes va entraîner la disparition de l'économie de milliers de villes moyennes, une perturbation aussi importante que celle qui avait résulté de la construction du réseau d'autoroutes contournant des villes. Regardons les chiffres.

Selon l'*American Trucker Association*, il y a 3,5 millions de chauffeurs routiers professionnels aux États-Unis, et 5,2 millions de personnes supplémentaires employées dans l'industrie du transport routier. Les revenus de ces 8,7 millions de personnes créent les emplois des autres. Les 3,5 millions de conducteurs de camions s'arrêtent régulièrement pour manger, boire, se reposer et dormir. Des entreprises entières ont été construites pour assurer leurs besoins, comme les restaurants et les hôtels. Des millions d'autres emplois sont donc à nouveau impactés directement et indirectement, si d'aventure les chauffeurs routiers étaient voués à disparaître.

Et ils le seront, dans tous les pays, à un horizon 2020-2030, s'accordent à dire la majorité des études, car les routes seraient ainsi beaucoup plus sûres, les camions autonomes pourront rouler continuellement, ils n'auront besoin ni de salaire, ni d'assurance, ni de retraite...

Un phénomène supplémentaire est à prendre en compte. Les chauffeurs de camion sont beaucoup mieux payés que la moyenne aux États-Unis. Ils forment une partie de la classe moyenne qui n'a pas eu besoin d'une forte éducation pour s'y trouver. Leur disparition, et celle des autres emplois qu'elle entraînerait inévitablement, reviendrait à accroître les inégalités entre quelques propriétaires du capital, et une multitude de travailleurs sous-qualifiés et sous-payés.

Cette non-redistribution supplémentaire, brutale et définitive de l'argent contribuerait à bloquer encore plus la croissance économique. Au fur et à mesure que les robots remplaceront les emplois actuels, ne se contentant pas d'en améliorer les conditions, et sur un horizon probable de 10 ans, le travail humain sera de plus en plus inutile et même économiquement non viable par rapport au travail de la machine. Nous devons donc bientôt trouver un autre moyen de nous fournir l'argent nécessaire pour acheter ce que les machines produisent, c'est-à-dire la mise en place généralisée d'un revenu universel et inconditionnel.

Les effets de l'automatisation par les IA commencent à se faire sentir dans tous les secteurs économiques. Dans le commerce électronique et le marketing, elles permettent l'avènement de l'hyper-personnalisation de l'offre, en sachant anticiper nos besoins. Dans le domaine médical, leurs capacités de diagnostic des maladies, ou de prédiction des effets médicamenteux, les objets connectés de suivi de bien-être et les robots d'assistance vont faire que la santé va littéralement changer de forme. Dans l'industrie, les robots passent à l'étape suivante, élaborer leurs gestes par simple imitation d'un instructeur, et se programmer eux-même. Dans la sécurité, et notamment la cybersécurité, les capacités prédictives des IA et la détection des fraudes font envisager la conception de systèmes de défense uniquement à base d'IA. Avec la possibilité de prévoir des délits, le juri-

dique n'est pas en reste. Il est concerné également par l'apparition d'agents intelligents capables de rédiger des contrats et d'absorber l'immense jurisprudence. Dans l'éducation à l'ère numérique, des étudiants ont été surpris de découvrir que le professeur assistant de leur MOOC était en réalité une IA. Même dans les industries créatives, ces dernières parviennent à faire du montage vidéo automatique ou créer des bandes son.

Que faire alors avec des millions d'êtres humains inutiles ? Les humains n'ont que deux capacités de base, physiques et cognitives. Quand les machines nous ont remplacés dans les emplois physiques, nous sommes passés à des emplois qui exigeaient des capacités cognitives... Si les IA nous surpassent dans les capacités cognitives, il n'y a pas de troisième voie vers laquelle se diriger. À moins de repenser le travail, non pas comme un moyen pour pouvoir vivre, mais comme une activité de vie facilitée par les machines, et vivre des vies d'humains, ce qui pourrait bien être notre troisième capacité différenciante. Sur ces questions, le juriste Alain Supiot, professeur au Collège de France, rappelle qu'à la révolution numérique en cours correspond un changement d'imaginaire, qu'il faut identifier, comprendre, et prendre en compte. Cet imaginaire imprègne toutes nos façons de penser : les institutions, les arts, les sciences et les techniques. Du modèle de l'horloge qui rythmait le monde occidental depuis la fin du Moyen âge (et a donné le taylorisme), nous sommes passés au modèle de l'ordinateur, une représentation qui influence nos rapports sociaux, notre conception du droit, et l'ensemble des règles qui rendent la vie en société possible. Il est intéressant de noter que dans les cas précédents, « *le changement d'imaginaire a commencé dans l'ordre juridique avant de s'exprimer au plan scientifique et technique.* » Le fantasme aujourd'hui poursuivi est celui d'une mise en pilotage automatique des affaires humaines, sous *la gouvernance par les nombres*. Cela crée une forme de déshumanisation, des travailleurs asservis à la satisfaction d'indicateurs de performance chiffrés, évalués indépendamment des effets réels de leur travail. Cette déconnexion du travail de la réalité de ses produits met en péril, non plus la santé physique, mais la santé mentale. « *Restaurer un travail réellement humain est, sur le long terme, la clé du succès économique.* »

Une IA-FrenchTech en création

Dans une série d'articles consacrés à l'IA publiée au deuxième trimestre 2016, Olivier Ezratty, un des fins observateurs de l'écosystème numérique, identifie une petite cinquantaine de startup françaises – toutes ne se montrent pas encore – utilisant ou développant des techniques d'Intelligence Artificielle, dans des domaines aussi divers que la santé, le commerce et le marketing, le traitement des données et l'intelligence économique, et bien sûr les objets connectés.

C'est dans le domaine des applications métiers que la FrenchTech de l'IA, comme aux États-Unis, est la plus inventive : agents conversationnels, outils de prédiction de places libres, saisie automatique de pièces juridiques ou comptables, outils d'aide à la gestion de son emploi... ces startup bénéficient du résultat des travaux d'équipes de recherche, ou sont le fruit de développements propres, souvent aidés en cela par les *API* et les sources ouvertes par les grands groupes de l'Internet.

À lire à partir de <http://goo.gl/8MLIFz>

Lectures complémentaires

Take X and add AI – Entering an era of intelligent automation. Avril 2015
<https://goo.gl/qRQXPa>

The Three Breakthroughs That Have Finally Unleashed AI on the World. Octobre 2014
<http://goo.gl/OWshVl>

Why artificial intelligence always seems so far away. Janvier 2015
<http://goo.gl/lAvdPp>

Qu'est-ce que l'intelligence artificielle ? Yann LeCun au Collège de France. 2016
<http://goo.gl/Mpl1y4>

Self-Driving Trucks Are Going to Hit Us Like a Human-Driven Truck – The imminent need for basic income in recognition of our machine-driven future. Scott Santens, Mai 2015
<https://goo.gl/xkzrY7>

The Human Connectome Project
<http://www.humanconnectome.org/>

Siri-creator shows off first public demo of Viv, 'the intelligent interface for everything'. Mai 2016
<http://goo.gl/LmdGwY>

Google supercharges machine learning tasks with TPU custom chip. Mai 2016
<https://goo.gl/y8EvYQ>

Restaurer un travail réellement humain est, sur le long terme, la clé du succès économique. Alain Supiot. Avril 2015
<http://goo.gl/s0oMAq>



Intelligences Artificielles, que ferons-nous d'elles ?

Des IA créatives et conscientes ?



« La nuit dernière, j'ai rêvé, dit calmement LVX-1 »

- Et qu'est-ce que tu as rêvé ? »

- Je fais toujours à peu près le même rêve, docteur Calvin. De petits détails sont différents, mais il me semble que je vois un vaste panorama où travaillent des robots »

- Des robots, Elvex ? Et aussi des êtres humains ?

- Je ne vois pas d'être humains, dans le rêve. Pas au début. Seulement des robots, docteur Calvin.

- Que font-ils, Elvex ?

- Ils travaillent. J'en vois qui sont mineurs dans les profondeurs de la terre, et d'autres qui travaillent dans la chaleur et la radiation. J'en vois dans des usines et sous la mer.

Isaac Asimov, Le robot qui rêvait, novembre 1986



« Les robots peuvent bien prendre notre travail. Heureusement il restera toujours aux humains l'imagination, la créativité, l'art... » Mais est-ce si sûr ? Et si les IA pouvaient elles aussi faire preuve de créativité ? Et si les robots devaient éprouver de l'empathie et l'exprimer pour être utiles aux humains ? Et si c'était nécessaire pour que les IA atteignent ce niveau d'intelligence, le leur, certes, qui les rendent finalement comparables aux êtres humains ?

En juin 2015 des ingénieurs de Google détaillent dans un article le fonctionnement d'un réseau de neurones possédant entre 10 et 30 couches successives, spécialisé pour identifier des objets sur des images. Méthode classique à présent, les caractéristiques de l'image sont analysées de couche en couche : couleurs, formes, orientations... Ils ont alors l'idée de renverser le réseau : non pas lui présenter en entrée des images à interpréter, mais au contraire des images bruitées, et demander au réseau de les améliorer en ayant comme objectif de sortie une interprétation fixée à l'avance (par exemple, une fleur). L'objectif des ingénieurs était de

comprendre le contenu des couches cachées des réseaux, et voir comment peu à peu le réseau passait de l'image à son interprétation. Mais l'exercice se transforme rapidement en un jeu d'exploration du savoir emmagasiné dans les poids synaptiques. Et c'est finalement l'humain qui interprète les images produites par ces couches cachées. Psychédéliques, oniriques, il n'en fallait pas plus pour que la presse titre sur l'arrivée d'intelligences artificielles pouvant rêver, et peut-être bien capables de créativité. Ce serait en tout cas un premier pas vers l'apprentissage de la sensibilité et l'apparition de l'inspiration créative.

Cependant, qui a déjà travaillé depuis les années 90 avec des réseaux de neurones multicouches sait que les couches internes contiennent des motifs intermédiaires plus ou moins étranges. Ce qui a été produit à l'occasion de DeepDream relève plutôt du traitement de filtres sur des images, les poids synaptiques du réseau jouant le rôle de paramètres de la fonction de filtre, plutôt que d'une véritable créativité émergente, et encore moins du rêve (qui suppose un

Les illustrations en tête des trois parties de ce cahier ont été réalisées par DeepArt.io, une IA créative, elle-même développée sur un réseau d'apprentissage profond à 19 couches. Il suffit de fournir une image de départ et de choisir un style à la manière de tel artiste, pour se voir produire une image unique, que l'on peut acheter en haute définition et exposer chez soi...

<http://deepart.io/>

<http://deepdreamgenerator.com/>

détachement de l'environnement physique). Ceci dit, les expérimentations continuent, avec d'autres types d'architecture et de réseaux, notamment pour de l'écriture automatique, et les résultats sont suffisamment intéressants pour créer des poésies ou des romans qu'il est envisageable de juger.

Plus réaliste, et possible preuve d'un embryon de créativité, est le phénomène d'émergence de *comportements non programmés, non prévus par l'humain*. C'est le fameux 37^e coup de la deuxième manche AlphaGo contre Lee Sedol. C'est l'effet wow vécu par de plus en plus de chercheurs, fascinés par les capacités exploratoires des récents algorithmes. « L'IA a fait des choses qu'une personne n'aurait pu deviner, comme modifier la puissance d'un laser d'un côté ou d'un autre, et compenser avec un autre. Je n'aurais jamais imaginé qu'une machine puisse apprendre à faire l'expérience elle-même, partant de rien, en moins d'une heure », s'exasiait en mai un physicien ayant fait reproduire à une IA l'expérience du condensat de Bose-Einstein, expérience qui avait ouvert le prix Nobel à trois chercheurs en 2001.

Rendre les IA réellement créatives est devenu un enjeu. Fin mai, Google annonçait mettre en place une équipe de recherche, dénommée Magenta, pour rendre l'IA créative, en commençant par la création musicale. Sur la base de Google TensorFlow, les réalisations seront rendues publiques pour être reprises par d'autres chercheurs.

Que manque-t-il aux IA ?

Que ce soit pour l'intelligence naturelle ou artificielle, **le rôle des émotions** dans les processus cognitifs a longtemps été ignoré. Les travaux

d'Antonio R. Damasio dans les années 90, et notamment son *hypothèse des marqueurs somatiques*, corroborée par de nombreuses observations sur des personnes souffrant de lésions dans les lobes préfrontaux, ont changé la donne. Nous l'avons vu pages 16-17, le cerveau humain est issu d'une évolution, et comporte plusieurs étages. Des parties anciennes du cerveau, situées en-dessous des parties plus récentes, s'occupent du contrôle des mouvements, des comportements fondamentaux comme la faim, le sommeil, la recherche de sécurité ou de partenaires, et des régulations hormonales. Ces bas-étages hébergent émotions et interactions corporelles, tandis que dans le néocortex, particulièrement développé chez l'humain, règnent les intelligences rationnelles et notamment la faculté de raisonner. Damasio et ses équipes ont montré que les premiers ne font pas que la maintenance logistique du corps. « *Les mécanismes neuraux sous-tendant la faculté de raisonnement ne semblent pas fonctionner sans ceux qui sous-tendent la régulation biologique. La nature semble avoir construit les mécanismes sous-tendant la faculté de raisonnement, non pas seulement au-dessus des mécanismes neuraux sous-tendant la régulation biologique, mais aussi à partir d'eux et avec eux.* »

Les émotions, on le comprend, sont liées à **la nécessité d'une intelligence incarnée**. Elles impliquent le corps. Elles impliquent le soi. L'implication de soi – des échecs sont ressentis difficilement, des réussites sont fortement appréciées – est un caractère important du chemin d'un apprenant vers la maîtrise d'un savoir-faire puis vers l'expertise. Deux types d'émotions existent : les émotions primaires d'abord, innées, qui découlent des systèmes limbiques et sont reliées au **système de valeur**. Elles permettent une réponse modulable en fonction de l'histoire individuelle de l'interaction de l'être vivant avec son environnement. Les émotions secondaires établissent quant à elles des rapports entre les émotions primaires et les situations en cours. L'hypothèse de Damasio a été que lors d'un raisonnement, si un résultat néfaste peut sembler survenir après une action, un signal d'alarme est envoyé par les réseaux neuronaux qui se chargent des émotions, pour rejeter immédiatement cette action, restreindre le champ des possibles, et focaliser l'attention sur ce qui est essentiel.

L'attention est un autre aspect de la cognition qui n'a pas encore été suffisamment mécanisé. C'est une capacité de l'éveil qui ajoute une composante directionnelle aux comportements. Il s'agit d'un processus qui dirige le faisceau de la conscience sur un nombre restreint d'éléments. Parmi les éléments qui font partie du champ de la conscience, ce sont les plus excités, correspondant souvent à un petit groupe de neurones actifs entourés de neurones passifs, par inhibition latérale.

À l'inverse, l'esprit incarné a besoin également de relâcher son attention vers l'environnement externe, dans des phases dites de sommeil paradoxal, où **le rêve** se joue. Lors de ces moments un système interne prend possession du corps pour inhiber les mouvements, et du cerveau pour activer des processus cognitifs sans conséquences observables. Plusieurs fonctions connues du rêve pourraient être nécessaires aux machines pour parfaire leurs apprentissages : le rêve oublié qui permet le désapprentissage des éléments erronés, le rêve aide-mémoire qui transfère les souvenirs vers la mémoire à long terme, le rêve préparateur d'émergence qui fait tester des hypothèses, des embryons d'idées, sans effet néfaste sur l'environnement et le soi.

Aujourd'hui les IA conversationnelles comprennent à peu près ce qu'on leur demande, savent s'exprimer correctement et répondre à des questions simples. Elles ne savent pas encore comprendre le sens des conversations ni vraiment s'adapter au contexte. Il reste encore beaucoup de progrès à faire, et cela passera par une expérimentation du monde pour comprendre le langage qui le décrit, ou l'invention de leur propre langage. Nous avons vu que l'IA doit être acceptée comme étant différente – et non pas simulant – des intelligences humaines et animales. Les voies pour atteindre les capacités de créativité, et la conscience de soi, doivent être également radicalement nouvelles. Vicarious, une startup qui est restée sous le radar quelques années, s'est donnée ainsi pour objectif d'apporter l'imagination aux machines. Son credo est d'apprendre aux machines avec peu d'exemples. Ses architectures neuronales font la part belle aux boucles de réentrée. Après avoir vaincu les captchas en 2015, la startup compte surprendre en 2016 dans le secteur de la kinesthésie robotique.

Enjeux & défis modernes des IA

Olivier Boissier, à Mines Saint-Étienne, est secrétaire de l'Affia, l'Association française pour l'Intelligence Artificielle. L'objectif de cette association est de désenfourner l'IA. La liste des bulletins de l'Affia montre d'ailleurs à quel point l'IA est un liant entre différentes disciplines. Finalement, l'IA est d'abord une science.

Pour le chercheur, il n'existe pas une seule IA, mais plusieurs types d'IA. Elle est présente partout mais on ne la voit pas ou plus. Elle n'est pas uniquement là où les médias et l'actualité la positionnent le plus facilement : robots, voitures autonomes, jeu de go... On la trouve pour commencer *dans les systèmes autonomes*, les ateliers de production robotisés, les systèmes de traitement de données et les systèmes d'aide à la décision... On la trouve ensuite *dans les systèmes qui interagissent* avec l'être humain ou les être humains, dans les systèmes en interaction avec d'autres systèmes autonomes, dans les systèmes décentralisés... Enfin, l'Intelligence Artificielle est au cœur des systèmes socio-techniques, comme support à l'interaction et à la coopération entre humains, machines et logiciels, et y soulève des questions d'éthique, de droit, de régulation...

Olivier Boissier travaille dans une équipe qui s'intéresse aux intelligences connectées, celles qui lient des algorithmes, des **agents**

et des humains, ainsi qu'aux problématiques de prise de décision distribuée. Ce qui les intéresse particulièrement est d'interagir sur ces 3 dimensions : les systèmes autonomes, les systèmes interactifs (en interaction avec les humains) et, parce que s'ouvrent alors des questions de vie privée et d'éthique, les systèmes socio-techniques (là où algorithmes, agents et humains interagissent). Ces trois dimensions sont un cadre conceptuel qui permet de mieux présenter les enjeux et défis contemporains de l'Intelligence Artificielle : en premier lieu les apprentissages, les raisonnements, la connaissance... puis les émotions, les interactions, le partage d'autorité, l'autonomie... et pour finir les normes, les valeurs, la place de la culture, l'éthique...

Les questions qui se posent alors vont beaucoup plus loin que des choix d'algorithmes : comment représenter explicitement la délégation d'une tâche d'un homme à un agent, ou le contraire ? Comment donner l'impression de l'empathie ? Le partage d'autorité : jusqu'où un robot social peut pousser quelqu'un à faire quelque chose ? Comment détecter qu'un agent ne respecte pas une norme ? Et comment le sanctionner ? Comment prendre en compte dans les IA des valeurs différentes selon la culture de l'humain rencontré ? Et derrière toutes ces interrogations, il y a des techniques à imaginer.



Par exemple, pour intégrer des valeurs et des règles morales dans les IA, il faut avec les philosophes, comprendre leurs concepts, les règles morales, et les transmettre aux algorithmes. Il s'agit notamment des approches de *value sensitive design*. On ne sait hélas pas encore faire de la reconnaissance de situations. Pourrions-nous faire de l'apprentissage de concepts moraux en lisant aux IA des contes pour enfants ? Ce qui est sûr c'est qu'il faut mettre ces règles éthiques dès la conception, et savoir vérifier après si elles sont bien prises en compte, sans oublier le fait que des systèmes IA ouverts collaborent entre eux. Et peut-être que les IA pourraient nous proposer des règles d'éthique auxquelles nous n'avons pas pensé, comme AlphaGo a su explorer des tactiques de jeu qu'aucun humain n'avait encore vues.

À Télécom École de Management, Ritta Baddoura, docteur en psychologie, explore ainsi le potentiel thérapeutique de l'interaction avec des robots sociaux, particulièrement pour des patients présentant la maladie d'Alzheimer ou des troubles du spectre autistique. De tels robots sont des systèmes artificiellement intelligents, autonomes, ou semi-autonomes s'ils sont téléopérés, avec des compétences sociales, affectives, cognitives et kinesthésiques, et capables d'interagir avec un humain sur le plan verbal ou non verbal. Dotés d'une corporéité, le plus souvent humanoïde ou animaloïde, parfois mécanique ou mixte, ils sont qualifiés d'émotionnels ou d'affectifs lorsqu'ils sont susceptibles de manifester certains états affectifs et des émotions par des expressions faciales, des gestes, une posture ou des paroles, et d'en susciter chez les humains, voire d'adapter leurs réponses et leur comportement à ces derniers.

Il existe diverses modalités d'utilisation d'un robot social dans le cadre d'un programme thérapeutique. Dans le cas de triades adulte / patient / robot, le praticien peut proposer au patient d'interagir avec le robot, en privilégiant, selon les cas, les instructions explicites ou le jeu libre, ou placer le robot à sa disposition dans l'espace de manière à lui laisser le choix de l'utiliser ou pas. Ce type de configurations favorise entre thérapeute et patient l'émergence d'une attention partagée qui peut être un atout précieux. Cette attention portée à un objet commun s'apparente à une co-pensée.

Le rôle de médiateur du robot social est celui qui est le mieux démontré actuellement. Ce rôle favorise l'émergence de situations et d'interactions à fortes valences sociales et affectives. Ces dernières contribuent à une amélioration des troubles de la communication, de l'interaction sociale, du comportement et de l'humeur. Il reste toutefois nécessaire de conduire des études complémentaires pour évaluer les impacts et les effets indésirables encore inconnus de cet usage, ainsi que les risques psychopathologiques. Par exemple, les interactions avec un robot pourraient maintenir certains patients dans un cycle répétitif, figé et fermé au changement parce que trop sécurisant. Ou bien exercer sur certaines personnes un attachement poussé qui les enfermerait dans une relation inadaptée et aliénante. Le cas de militaires américains risquant leur vie pour sauver un robot combattant, ou effectuant une cérémonie en vue de l'enterrer, constitue un tel exemple d'attachement a priori inadapté. Reste que l'usage des robots sociaux dans le cadre thérapeutique donne au praticien les moyens de modifier et d'enrichir sa pratique et lui permet d'en repenser le cadre et les enjeux éthiques.

Dans le domaine médical et ailleurs, les questions de droit relatives aux IA, ou en présence des IA, posent également de nombreuses questions. Citons par exemple l'évolution du droit de la preuve, ainsi que les possibilités ouvertes par la « *justice prédictive* »... Certains appellent dès maintenant à une réflexion sur les droits des robots et des objets de l'Internet mus par des IA. En complément de ces questions de droit, les robots mécaniques ou les robots algorithmiques entrent également dans le champ de la régulation.

Lors d'un récent colloque rapporté dans la *Harvard Business Review*, le spécialiste en analyse prédictive Vasant Dhar a proposé un outil conceptuel nouveau, conçu pour aider chacun d'entre nous, et donc les politiques et les régulateurs, à estimer quand faire ou non confiance à un robot. La *Decision Automation Map* permet d'évaluer le coût des erreurs des systèmes automatisés prédictifs, comme les véhicules autonomes, les drones tueurs, les systèmes d'anti-fraude à la carte bancaire ou les outils de diagnostic médical, et de jauger s'il faut plus de technique, plus de droit ou plus de régulation.

Lorsqu'on parle d'éthique, l'exemple robotique vient rapidement à l'esprit, et notamment le dilemme de la voiture autonome en cas d'accident prévisible.

Va-t-elle sauver son occupant, ou le sacrifier pour éviter un bus d'enfants ? Mais l'éthique peut être nécessaire sur de nombreux autres sujets, comme celui de la gestion des données privées.

C'est l'objectif du **projet ETHICAA**, auquel participent plusieurs chercheurs de l'Institut Mines-Télécom, qui est de définir ce que devrait être un système composé d'un ou plusieurs agents, y compris des humains, capable de gérer des conflits éthiques, aussi bien au niveau individuel qu'au niveau collectif.

<http://ethicaa.org/>

Apprendre à vivre avec les Intelligences Artificielles

Loin des craintes éprouvées devant des IA mal-faisantes, plusieurs interventions sur les IA bienveillantes se tiendront pendant la grande conférence *IJCAI*, à présent annuelle, qui se déroule à New-York en juillet 2016. Mais si les IA sont bienveillantes, le sommes-nous avec elles ? Les commentaires sur une vidéo (ci-contre) d'un robot de Boston Dynamics maltraité par son propriétaire (pour le bien de l'expérience), l'acharnement contre l'agent conversationnel Tay de Microsoft, poussé à proférer des propos racistes sur Twitter, peuvent en faire douter. Le choix des termes dans la presse, « *dresser des IA* », pour décrire la fin proche de la programmation des machines via les instructions d'un langage informatique, remplacée par des instructions données en langage naturel, ou par imitation, est également préoccupant. Une vision paternaliste et très masculine des IA, qui préfère doter les agents conversationnels des GPS de voix féminine, au motif qu'« *ainsi elles ne semblent pas vous contrôler* », fait frémir. L'appel des chercheuses à féminiser beaucoup plus les équipes de recherche en IA est également le constat que nous devons revoir notre manière de considérer et concevoir nos artefacts intelligents.

Comme le rappelle le docteur en philosophie **Pierre-Antoine Chardel**, à **Télécom École de Management**, le défi éthique est entier dans une époque où nous sommes, d'une part, dans

une relation plutôt consumériste aux nouveautés technologiques et où, d'autre part, la majorité des usagers est loin de disposer de clés de compréhension nécessaires au déchiffrement de leurs environnements hyper-technologiques (objets connectés, agents autonomes, puces RFID, biométrie...) qui apportent souvent dans la vie de tous les jours un plus grand confort. Or, comme l'exprimait déjà Herbert Marcuse dans une conférence de 1941, plus une certaine *commodité* nous est technologiquement garantie, et plus elle se voit induite par la normalisation de nos pratiques technologiques, moins le champ des interrogations que nous serions en droit de formuler vis-à-vis de toute innovation technoscientifique (en termes de sens, de valeurs, de qualité du lien social...) est intense et profond. Se donner pour tâche « *d'orienter le présent vers un avenir durable* » suppose donc de travailler sans relâche à l'éveil d'une démarche critique constructive à l'ère hypermoderne, ainsi qu'à l'élaboration d'une explication et d'une interprétation de notre agir technologique.

Il est donc grand temps de s'interroger tous ensemble sur nos pratiques avec les IA, algorithmiques ou robotiques, bien au-delà du cercle des scientifiques, et de nous interroger sur nous-même. Politiques et citoyens doivent s'emparer de ces sujets avec sérénité, et les éprouver dans des *living labs* partagés avec nos compagnons artificiels, pour élaborer tous ensemble notre futur commun. ■



Références & lectures complémentaires

Inceptionism: Going Deeper into Neural Networks.
Juin 2015. <http://goo.gl/qgv5R7>

Google's AI writes spooky post-modern poetry after reading thousand of romantic books.
Mai 2016. <http://goo.gl/MVE8cu>

Here's why we should thank Microsoft for its AI bot that turned into a foul-mouthed racist.
Mars 2016. <http://goo.gl/DOA6tn>

Soon We Won't Program Computers. We'll Train Them Like Dogs.
Mai 2016. <http://goo.gl/kiZN4K>

Le potentiel thérapeutique de l'interaction homme-robot. La Lettre du Psychiatre Ritta Baddoura. Janvier 2016.

Comment créer une intelligence artificielle malveillante ? Manuel à l'usage des apprentis.
Mai 2016. <http://goo.gl/RKgJA4>

Glossaire

Agents cognitifs : logiciel qui agit de façon autonome, et intelligente

AI winters : moments de l'Histoire de l'IA où les doutes ont remis en question les enthousiasmes précédents

Analyse et fouille de données : extraction de connaissances à partir de données. **Data mining**

Analyse prédictive : techniques issues des statistiques, de *data mining* et de la théorie des jeux pour échafauder des hypothèses

API : *Application Programming Interface*, un ensemble normalisé de méthodes par lequel un logiciel offre des services à d'autres logiciels

Apprentissage machine : techniques et algorithmes fournissant des capacités d'apprentissage aux ordinateurs. Voir pages 10 et 11

Apprentissage par renforcement : présenté page 10

Apprentissage profond : présenté page 11

Artefact : objet façonné par l'Humain

Big data : données massives. Elles ont fait l'objet d'un cahier de veille à retrouver dans l'espace documentation de la Fondation Télécom.

<https://www.fondation-telecom.org/documentation/>

Bots : robots algorithmiques. Les **chatbots** sont des bots conversationnels.

Cognitivism : paradigme des **sciences cognitives** s'intéressant aux symboles et aux règles

Commodités : néologisme pour désigner l'anglais *commodity*, produit de base, de consommation courante

Conférence NIPS : Conférence annuelle, *Neural Information Processing Systems*, <https://nips.cc/>

Connexionnisme : paradigme des **sciences cognitives** s'inspirant des **réseaux de neurones**

Décodeur : élément de la chaîne de traitement du signal chargé de le récupérer après passage dans un canal bruité

Deep learning : voir **Apprentissage profond**

FPGA : *Field-Programmable Gate Array*, circuit intégré qui peut être reprogrammé après sa fabrication

GPU : *Graphics Processing Unit*, processeur spécialisé en traitement du signal et bien adapté aux calculs de **réseaux de neurones**

IA faible // IA forte : une IA faible est spécialiste du jeu d'échec mais nulle en cuisine. Une IA forte a des capacités étendues dans tous les domaines où les humains en ont.

IJCAI : Conférence annuelle, *International Joint Conference on Artificial Intelligence*, <http://www.ijcai.org/>

Machine learning : voir **Apprentissage machine**

Mésoscopique : échelle intermédiaire entre le macroscopique et le microscopique

Réseaux de neurones formels : représentations mathématiques et informatiques des neurones biologiques et de leurs connexions

Réseaux sémantiques : graphes modélisant la représentation des connaissances

Sciences cognitives : ensemble de disciplines scientifiques regroupant notamment les neurosciences, l'Intelligence Artificielle, la psychologie, la philosophie, la linguistique, l'anthropologie... Champ transdisciplinaire très vaste s'intéressant à la pensée humaine, animale et artificielle

Superintelligence : au-delà de l'**IA forte**, elle surpasse tous les humains dans tous les domaines

Systèmes experts : systèmes prenant des décisions à partir de règles et de faits

Test du miroir : test permettant d'évaluer si un animal reconnaît son propre reflet comme étant une image de son corps, et donc est conscient de son propre corps

TPU : *TensorFlow Processing Unit*

Value sensitive design : Conception de technologies prenant en compte les valeurs des Humains

Collaborations possibles avec l'Institut Mines-Télécom & la Fondation Télécom

L'Institut Mines-Télécom est un établissement public dédié à l'enseignement supérieur et la recherche pour l'innovation, dans les domaines de l'ingénierie et du numérique. Il est placé sous la tutelle des ministères en charge de l'Industrie et du numérique. À l'écoute permanente du monde économique, il conjugue une forte légitimité académique, une proximité concrète avec les entreprises et un positionnement unique sur 3 transitions majeures du XXI^e siècle : Numérique, Énergétique et Écologique, Industrielle. Ses activités se déploient au sein des dix grandes écoles Mines et Télécom.

La Fondation Télécom soutient l'Institut Mines-Télécom dans son ambition d'être acteur de la transition numérique de notre société. Créée en 2008 et reconnue d'utilité publique depuis 2012, elle contribue au développement et au rayonnement des écoles Télécom de l'Institut Mines-Télécom dans 4 domaines d'intervention : Formation, Recherche, Innovation et Prospective.

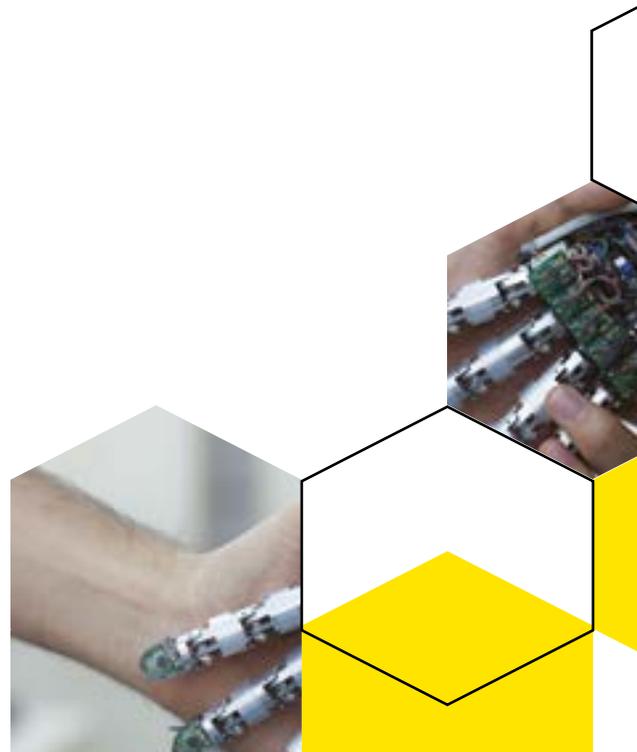
De multiples formes de collaborations sont possibles avec l'Institut Mines-Télécom et la Fondation Télécom. Ces partenariats donnent l'opportunité aux entreprises de participer dans un cadre mutualisé d'échanges et de partage, à d'ambitieux programmes de recherche sur des thématiques stratégiques pour le monde industriel, de partager l'expertise des enseignants-chercheurs, d'accéder au potentiel d'innovation des laboratoires et des incubateurs et d'identifier des talents dans les écoles.

Quelques-unes des équipes et thèmes relatifs à l'Intelligence Artificielle dans les écoles de l'Institut-Mines Télécom

Réseaux thématiques de l'Institut : Contenus, Connaissances et Interactions // TIC et société // Informatique Fondamentale et Mathématiques Appliquées /// **Télécom ParisTech** : Machine Learning au sein de la filière Big Data (Stephan Cléménçon) // AxSE Design-Interaction-Perception /// **Télécom Bretagne** : Neucod (Neural Coding) (Claude Berrou) // Département Lussi, notamment psychologie cognitive // Physique computationnelle (Francesco Andriulli) // Robotique, IA et sciences cognitives (Mai Nguyen) /// **Télécom école de Management** : Etos (Pierre-Antoine Chardel) /// **Télécom SudParis** : Intermedia (Bernadette Dorizzi) // Groupe AlgoRès /// **Mines Nantes** : Intelligence incarnée (Frédéric Boyer) /// **Mines Saint-Etienne** : Intelligence connectée (Olivier Boissier) ///

Les cahiers de veille de la **Fondation Télécom**

Les cahiers de veille de la Fondation Télécom sont le résultat d'études menées conjointement par des enseignants-chercheurs de l'Institut Mines-Télécom et des experts industriels. Chaque cahier, qui traite d'un sujet spécifique, est confié à des chercheurs de l'Institut qui réunissent autour d'eux des experts reconnus. Tout à la fois complet et concis, le cahier de veille propose un état de l'art technologique, et une analyse tant du marché que des aspects économiques, sociologiques, juridiques et éthiques, en mettant l'accent sur les points les plus cruciaux. Il se conclut sur des perspectives qui sont autant de pistes possibles de travail en commun entre les partenaires de la Fondation Télécom et les équipes de l'Institut Mines-Télécom.



Fondation Télécom

37-39 rue Dareau - 75014 Paris - France
Tel.: + 33 (0) 1 45 81 77 46

Directrice des opérations & des programmes
audrey.loridan-baudrier@fondation-telecom.org
www.fondation-telecom.org

Avec le soutien de :

BNP Paribas, Nokia et Orange,
partenaires fondateurs de la Fondation Télécom

Et Accenture, Airbus Defence & Space, SFR,
Solucom, Sopra Steria, Thales et Streamwide

