



L'IA et les sciences humaines & sociales à Grenoble

Un état des lieux

Thierry Ménissier

Vice-Président en charge du développement des recherches en Sciences Humaines et Sociales

Thierry.menissier@univ-grenoble-alpes.fr

Sommaire

Présentation, p. 4

Liste des acronymes utilisés et des unités de recherche impliquées, p. 5

Axe 1 : Cognition, p. 6-9

Programme 1 « De l'Intelligence naturelle à artificielle, et vice-versa » (LPNC & GIPSA-lab)

Programme 2 : « Interface cognition naturelle et artificielle : innovations pour la cognition réparée, assistée ou étendue » et applications en neuroergonomie (LPNC & GIPSA-lab)

Programme 3 : « Philosophie de la cognition » (PPL)

Programme 4 : « Connaissances collectives & cognition individuelle : Conceptions, croisements disciplinaires, données » (LIDILEM)

Axe 2 : Appropriations de l'IA & pratiques sociales, p. 10-15

Programme 5 : « Emergence du lexique en contexte : apport des méthodes non supervisées sur grands corpus de données multimodales (LIDILEM)

Programme 6 : « Algorithmes et IA dans le rapport production-consommation » (Gresec)

Programme 7 « Le processus de naturalisation » (Gresec)

Programme 8 : « Le rôle des conseillers robotiques dans les marchés financiers » (GAEL)

Programme 9 : « Processus d'innovation dans les IA et impact sectoriel : énergie, agriculture, santé » (GAEL)

Programme 10 : « IA et innovation » (CERAG)

Programme 11 : « IA et gestion des risques » (CERAG)

Programme 12 : « L'Analyse Prédictive pour la production à la demande » (CERAG)

Programme 13 : « Projet ALGOLUX » (LARHRA)

Programme 14 : « IA et approches qualitatives autour des questions de la mémoire et du récit » (LARHRA)

Programme 15 : « Acceptabilité, usage et implications de l'usage des technologies incorporant l'IA » (LIP/PC2S)

Programme 16 : « Implications motivationnelles de l'usage des nouvelles technologies dans le domaine de la santé » (SENS)

Axe 3 : Ethique, p. 16-18

Programme 17 : « Éthique philosophique de l'IA » (PPL)

Programme 18 : « Ethique opérationnelle de l'IA : CER Grenoble Alpes » (LPNC)

Programme 19 : « Débat public et IA » (Gresec)

Programme 20 : « IA, comportements responsables et éthique » (CERAG)

Programme 21 : « Algorithmes, contrôle social technologique, justesse de décision et d'attribution » (PPL)

Axe 4 : Normes, p. 19-22

Programme 22 : « IA, justice prédictive et raisonnement probatoire » (CRJ / LJK)

Programme 23 : « Protection des créations issues de l'IA par droit d'auteur » (CRJ-CUERPI)

Programme 24 : « La protection des technologies d'IA » (CRJ)

Programme 25 : « IA, algorithmes et processus décisionnels. Vers une datacratie ? » (CESICE)

Programme 26 : « La réglementation des armes autonomes » (CESICE)

Programme 27 : “Adversarial examples in machine learning & AI security” (CESICE)

Programme 28 : « Des LegalTech à la Justice Prédictive : Solutions d'IA pour les métiers du droit » (CESICE)

Programme 29 : « La maîtrise du risque cyber/IA et son transfert vers l'assurance » (CESICE)

Présentation

Le rapport sur l'intelligence artificielle (IA) rédigé par le mathématicien et député Cédric Villani et rendu public le 28 mars 2018 met en valeur la nécessité d'accompagner le déploiement de la recherche en intelligence artificielle en France en lui donnant un « sens ». C'est d'ailleurs le titre même du rapport, qui comprend d'importants développements sur l'éthique (partie 5), terme entendu au sens large, ce qui accorde une place importante à la démarche réflexive dans les usages ou dans la conception même de l'IA. Parce qu'il oriente ainsi le questionnement sur les fins et les enjeux, les valeurs et les usages d'une société future souhaitable, le rapport Villani sollicite directement les disciplines SHS dans leur capacité à dialoguer avec les disciplines technologiques afin de construire et d'orienter la recherche sur l'intelligence artificielle.

Les unités de recherche en sciences humaines et sociales de l'Université Grenoble Alpes et plus généralement du site universitaire grenoblois sont conscientes que la recherche en IA ainsi conçue conduit leurs disciplines vers une étroite collaboration avec les disciplines technologiques, et, au-delà, que cette recherche les sollicite pleinement pour la compréhension de l'humain dans son monde à venir. Elles se positionnent pour une réponse possible sur quatre axes de recherche, intitulés : *cognition, appropriations de l'IA & pratiques sociales, éthique* et *normes*. Cette présentation ne vise nullement à organiser les axes en fonction d'une hiérarchie, elle n'induit aucune forme de priorité : les quatre axes apparaissent également importants et visent, dans le cadre de cet état des lieux, à faciliter l'exposition des projets possibles ou des programmes déjà en cours. De surcroît, dans ce document, ces termes sont entendus de manière générique et non restrictive : ils dessinent un panorama de la recherche en cours et possible, qui est pluridisciplinaire et qui participe déjà largement au questionnement des thématiques confirmées ou émergentes de la recherche actuelle en IA.

Ces axes sont nourri par ce que nous avons nommé des « programmes », portés par les équipes de recherche en SHS de l'UGA que nous avons sollicitées afin de donner aux travaux à la fois une visibilité et une forme de cohérence en dépit de leur variété. Chaque équipe impliquée a répondu en fonction de ses projets en cours, mais également en fonction des perspectives de travail qu'elle se donne dans un avenir proche. Chacune de ces propositions pourra être développée et gagnera en ampleur dans le cadre d'un institut de recherche pluridisciplinaire dédié à l'IA sur le site grenoblois. La description fournie pour chaque programme a ainsi pour vocation d'augmenter le niveau de connaissance mutuelle des chercheurs des différentes disciplines impliquées et de favoriser les rencontres interdisciplinaires en vue de nouveaux projets.

Thierry Ménissier,
Le 24 mai 2018

Liste des acronymes utilisés et des unités de recherche impliquées

- Centre d'études et de recherches appliquées à la gestion ([CERAG](#)) - UMR 5820
- Centre d'étude sur la sécurité internationale et les coopérations européennes ([CESICE](#)) - EA 2420
- Centre de recherches juridiques ([CRJ](#)) - EA1965
- Grenoble Applied Economics Lab ([GAEL](#)) - UMR 5313
- Institut des neurosciences de Grenoble ([GIN](#)) - UMR S836
- Grenoble images parole signal automatique ([GIPSA Lab](#)) - UMR 5216
- Centre de recherche sur les enjeux de la communication ([GRESEC](#)) - EA 608
- Institut des Langues et Cultures d'Europe, d'Amérique, d'Afrique, d'Asie et d'Australie ([ILCEA4](#)) - EA 7356
- Laboratoire de recherche historique Rhône Alpes ([LARHRA](#)) - UMR 5190
- Laboratoire de linguistique et didactique des langues étrangères et maternelles ([LIDILEM](#)) - EA 609
- Laboratoire d'informatique de Grenoble ([LIG](#)) - UMR 5217
- Laboratoire Interuniversitaire de psychologie / Personnalité, cognition, changement social ([LIP - PC2S](#)) - EA 4145
- Laboratoire Jean Kuntzmann ([LJK](#)) - UMR 5224
- Laboratoire de Psychologie et de NeuroCognition ([LPNC](#)) - UMR 5105
- Philosophie, Pratique, Langage ([PPL](#)) - EA 3699
- [Pôle Grenoble cognition](#) - FR 3381
- Techniques de l'imagerie, de la modélisation et de la cognition ([TIMC-IMAG](#)) - UMR 5525
- Laboratoire sport et environnement social ([SENS](#)) - EA 3742

1/ Cognition

Programme 1 « De l'Intelligence naturelle à artificielle, et vice-versa » (LPNC, Julien Diard, Martial Mermillod, Benoit Lemaire, Stéphane Rousset, Alan Chauvin, Christian Graff, David Alleysson / GIPSA-lab, environ 30 chercheurs distribués sur l'ensemble des 3 départements de cette UR).

L'implication du LPNC et de GIPSA-lab et, plus généralement, de la communauté des Sciences Cognitives de Grenoble dans le domaine de l'Intelligence Artificielle (IA) n'est pas récente, et ne fait pas que suivre la mode du jour, consécutive aux investissements massifs par les GAFAM (Google, Amazon, Facebook, Apple et Microsoft) dans le domaine. Depuis 2012 en effet, ces entreprises ont développé des centres de calculs permettant l'application à grande échelle de méthodes de réseaux de neurones artificiels profonds (Deep Learning) ; elles ont été suivies par des soutiens financiers forts des USA, de la Chine et, prochainement (cf. rapport Villani), de la France.

Cette IA « à la mode du jour », qui repose majoritairement sur les données, est à notre sens hémiplogique : elle oublie la modélisation et la compréhension de l'intelligence naturelle, et accouche de méthodes efficaces, mais non-explicables (c'est-à-dire dont le calcul n'est pas interprétable). Une autre approche, nécessairement complémentaire selon nous, milite pour l'étude bio-inspirée de l'intelligence artificielle, et l'étude conjointe de l'intelligence naturelle.

Ce point de vue a souvent été minoritaire dans l'histoire de l'IA et de ses précurseurs, mais a tout de même permis des avancées majeures, avec, pour ce qui est de Grenoble, des chercheurs prestigieux du GIPSA-Lab, du GIN, de TIMC, du LIG, du LJK, du LPNC, apportant des résultats reconnus internationalement dans les domaines de la séparation de sources (notamment de sources bio-inspirée), de la modélisation motrice (geste expert, plasticité, adaptation), des réseaux de neurones artificiels, de la modélisation bayésienne (programmation robotique probabiliste), l'apprentissage statistique, les simulations bio-inspirées des traitements visuels (allant du mosaïcage inspiré de la rétine humaine à la lecture de mots, de textes, l'exploration et la reconnaissance de scènes naturelles, la reconnaissance des visages et leurs émotions, etc.), la modélisation de la parole et des interactions de tous types et notamment communicationnelles (robot parlant, modélisation de l'apprentissage de la parole), la mémoire, etc.

Cette vue interdisciplinaire de l'IA, qui marie les STIC, SDV et SHS, est très active à l'UGA, en collaboration forte avec Grenoble INP, le CHU de Grenoble Alpes, le CEA et INRIA, et repose sur de nombreuses structures transverses (SFR Pôle Grenoble Cognition, CDP NeuroCog, Tremplin Carnot Cognition). Cette communauté publie dans des revues prestigieuses, organise les conférences et symposium internationaux du domaine (entre autres, en collaboration avec INRIA, le GDR BioComp et le CEA), et reçoit le soutien financier de nombreux opérateurs de recherche (ANR, ERC, financements privés). Rappelons enfin que Grenoble a été la première ville de France à proposer un diplôme de Sciences Cognitives de niveau M2, dès 1991 offrant par la même l'un des plus importants contingents d'étudiants en Intelligence Artificielle du site grenoblois ! Ces étudiants bénéficient d'une formation multidisciplinaire originale permettant de faire le lien entre cerveau, cognition naturelle et IA. Face à l'engouement des étudiants pour cette formation (le nombre de candidats dépassant les

capacités d'accueil de ce Master), le Hub 3IA pourrait bénéficier de ce volet enseignement pour la formation des chercheurs de demain en IA.

Cette communauté grenobloise est unique en France, de par son approche multidisciplinaire et sa taille critique. Nous affirmons qu'un va-et-vient entre l'IA et ses applications d'une part, et les Sciences Cognitives d'autre part, est l'approche la plus efficace pour l'avancée des deux domaines et leurs développements futurs, et espérons faire reconnaître la force grenobloise à ce sujet.

Programme 2 : « Interface cognition naturelle et artificielle : innovations pour la cognition réparée, assistée ou étendue » et applications en neuroergonomie, (LPNC, Hélène Loevenbruck, Chris Moulin, Aurélie Campagne, Monica Baciú / GIPSA-lab, (Département Parole et Cognition en totalité et plusieurs équipes (SAIGA, VIBS) et membres des deux autres départements soit plus de 20 personnes).

Ce programme, qui porte sur les interfaces entre cognition naturelle et cognition artificielle, permettra, d'une part, de mieux comprendre la cognition humaine et d'autre part de proposer de la réparer lorsqu'elle dysfonctionne, de l'assister lorsqu'elle est en apprentissage ou lorsqu'elle est empêchée, voire de l'étendre lorsqu'elle est limitée. Mieux décrire comment l'humain analyse, traite et interprète les informations issues de son environnement afin de produire des réponses adaptées permet en effet de proposer des interfaces permettant de remédier à des déficits, des handicaps ou des troubles. Cela permet également de développer des applications innovantes visant à assister l'humain et à étendre ses capacités, individuelles et collectives. Ce programme permettra également l'adaptation des technologies d'IA au fonctionnement cognitif humain (aux capacités et limitations cognitives humaines) et à ses capacités communicationnelles tout en améliorant la formation des futurs utilisateurs, sur la base des connaissances et données issues de la neuroergonomie et de la linguistique afin de garantir une efficacité plus importante du couple opérateur/système. Il permettra également de développer des systèmes plus efficaces, sûrs et éthiques en situation par une prise en considération en temps-réel des données humaines.

L'être humain a toujours cherché à s'équiper d'artefacts. Les recherches des dernières décennies sur les interfaces entre cognition artificielle et naturelle, en lien avec les importantes évolutions technologiques en électronique, informatique et biosciences, ont ouvert de nouvelles perspectives. Elles ont permis de décupler les capacités humaines de traitement de l'information et de communication. Les nouveaux artefacts que sont les smartphones permettent d'avoir sous le pouce une somme considérable de connaissances, d'y contribuer et de les partager. D'autres artefacts peuvent même être intégrés au corps ou au cerveau humain pour remédier à leurs déficits, handicaps, dysfonctionnements, vieillissements : implant cochléaire pour l'audition, prothèses pour la motricité, capteurs myo-électriques pour les prothèses, larynx artificiel pour la voix, substitution sensorielle pour la vision, stimulateur cérébral profond pour la maladie de Parkinson ou d'autres maladies et troubles mentaux.

Le programme que nous proposons revendique une démarche intégrative dans laquelle les recherches sur la cognition naturelle et les travaux sur la cognition artificielle seront associés et coordonnés, dans l'objectif de développer des dispositifs innovants, visant à réparer, assister ou étendre la cognition humaine et sociale.

L'objectif de ce programme est de mieux comprendre les grandes fonctions mentales (perception, action, mémoire, langage, raisonnement, fonctions exécutives, apprentissage, communication, motricité...) et de les mettre en relation avec l'étude des comportements en situation d'interaction notamment communicationnelle avec le monde environnant (naturel et artificiel). La démarche consiste à examiner ces grandes fonctions sous des angles

complémentaires : développemental, clinique, pathologique, en s'appuyant sur des théories et modèles computationnels et en confrontant données empiriques et simulations artificielles.

Cette démarche intégrative requiert des plates-formes expérimentales de haut niveau :

- (neuro)imagerie cérébrale fonctionnelle et médicale (en particulier, imagerie fonctionnelle par Résonance Magnétique (IRMf), l'électroencéphalographie (EEG), neuro-stimulation magnétique transcrânienne (TMS), électrophysiologie, etc.).
- instruments de mesure mécaniques, biomécaniques (plateforme de force, systèmes d'analyse du mouvement MOCAP, électromyographie EMG) ou psychophysiques
- systèmes d'immersion cyber-physique (VR, AR, interface muscle-machine)

Elle requiert également des étudiants formés dans le domaine des sciences cognitives, des sciences du langage et des sciences de l'ingénieur.

Le site grenoblois présente l'originalité de rassembler les prérequis de ce programme :

- des laboratoires de recherche fondamentale et appliquée ainsi qu'un centre hospitalier universitaire à la pointe de ces questions de recherche,
- des plates-formes expérimentales de haute technologie
- et des formations académiques d'excellence, notamment les masters de psychologie, de sciences cognitives, de neurosciences, de sciences du langage, et les écoles d'ingénieurs de Grenoble INP.

La démarche intégrative revendiquée dans ce programme est mise en avant par le LPNC, qui se positionne en leader sur le site grenoblois dans le domaine de la psychologie cognitive, de la neuropsychologie et de la remédiation cognitive. Le LPNC a une expertise reconnue dans l'utilisation de ces outils expérimentaux, notamment l'imagerie cérébrale fonctionnelle et les mesures psychophysiques. Le Gipsa-lab a également une expertise sur les méthodologies, sur le développement des capacités communicationnelles humaines, la production et la perception de la parole humaine dans sa multimodalité (audio-visuelle et gestuelle), l'étude linguistique de terrain avec applications dans le cadre des systèmes d'information géographique (GIPSA-lab et LIG), et les interactions situées homme-homme et homme-machine. Le LPNC, le GIPSA-lab ont une longue tradition de recherche sur les questions fondamentales relevant des sciences cognitives et des sciences de la parole avec une forte composante sciences humaines et sociales, qui sont abordées dans le cadre de projets propres des laboratoires, mais aussi collaboratifs, avec d'autres équipes du site grenoblois au sein du CDP IDEX Neurocog « Cerveau et Cognition » et du Pôle Grenoble Cognition. La spécificité de ces projets est d'être étudiée à l'interface, c'est-à-dire dans le cadre du couplage cognition et communication naturelle / artificielle, grâce à des méthodes des sciences de l'ingénieur, des développements technologiques, méthodologiques, du traitement de l'information et de la parole, débouchant vers des problématiques d'interaction humain-machine, de robotique cognitive et sociale.

Programme 3 : « Philosophie de la cognition » (PPL, responsable Denis Perrin).

La recherche en philosophie à Grenoble conduit une réflexion sur la nature et les formes de la cognition, en particulier sous la forme d'une contribution forte au champ des sciences cognitives grenobloise et cette réflexion est bien sûr susceptible de contribuer à la question de l'IA. Elle prend la forme d'une interrogation relative aux grandes fonctions de l'esprit (la mémoire, la perception, les sentiments, les états mentaux comme l'intention, en particulier), d'un questionnement de la structure et des limites de l'esprit (métacognition, cognition sociale, en particulier), et de réflexions sur l'épistémologie des sciences cognitives.

Programme 4 : « Connaissances collectives & cognition individuelle : Conceptions, croisements disciplinaires, données » (Lidilem, Jean-Pierre Chevrot).

Ce projet de réflexion théorique (auquel un blog de recherche est consacré : <https://co3i.hypotheses.org/>) met en synergie des chercheurs en SHS (psychologie sociale, linguistique, sociologie), sciences des données, neurosciences, philosophie, physique des populations pour examiner les phénomènes de pensée collective, leur modélisation et leur description à partir de données massives ou de démarches expérimentales. Le but est de repenser les oppositions social/cognitif et collectif/individuel et d'en proposer une articulation plus satisfaisante. Il implique les membres de plusieurs laboratoires en France ou à l'étranger, dont plusieurs laboratoires de l'UGA (Lidilem, ILCEA4, LIP, PPL).

Il a été financé en 2017 par la mission pour l'interdisciplinarité du CNRS. Ce financement sera poursuivi par d'autres, en cours de finalisation.

2/ Appropriations de l'IA & pratiques sociales

Programme 5 : « Emergence du lexique en contexte : apport des méthodes non supervisées sur grands corpus de données multimodales » (Lidilem).

Ce programme financé pour trois ans par le CDP NeuroCog. Il est construit autour de la thèse de William Havard (LIDILEM, ED SHPT), et implique le LIG (Laurent Besacier) et Lidilem (Jean-Pierre Chevrot).

Il s'agit d'utiliser les ressources de l'apprentissage-machine ("machine learning") pour comprendre comment l'enfant découvre des unités lexicales nouvelles à partir d'énoncés accompagnés de scènes visuelles (décrites par les énoncés).

Le système d'apprentissage-machine apprend à apparier des scènes visuelles aux énoncés à partir d'une base de données constituée de signaux de parole (acoustiques) et de signaux visuels. Ce système est ensuite testé sur d'autres données : on lui présente des signaux de paroles contenant les mêmes objets dans d'autres contextes et on vérifie s'il est capable de reconnaître des images contenant ces objets.

Actuellement, trois étapes sont prévues pour le déroulement du projet.

1/ Interrogation sur les connaissances apprises par le système d'apprentissage-machine : simples appariements son-image ? Catégories sémantiques ? Cadres syntaxiques ?

2/ Interrogation sur les capacités du système à généraliser ce qu'il a appris

3/ Interrogation sur la sensibilité du système à l'input qu'on lui fournit : fréquences des mots, spécificité des contextes où apparaît chaque mot, variété des voix prononçant les différents énoncés où un mot donné apparaît, etc. Cette manipulation de l'input est facilitée par le fait que la base de données d'apprentissage est un corpus de parole synthétique que nous pouvons manipuler à notre guise.

Il s'agit d'un projet interdisciplinaire informatique-psycholinguistique développementale, utilisant les ressources de l'informatique pour modéliser le processus d'acquisition par lequel l'enfant découvre un lexique à partir de corrélations entre des indices sonores et des indices visuels.

Programme 6 : « Algorithmes et IA dans le rapport production-consommation » (Gresec).

Thème récurrent dans la tradition de recherche du laboratoire Gresec, le domaine de la culture, des réseaux sociaux numériques et des industries culturelles s'enrichit aujourd'hui de nouvelles interrogations importantes du fait de l'émergence de l'IA au cœur des pratiques sociales. Ainsi, dans le champ du design d'interface des services : éditorialisation des contenus, systèmes de recommandation, affichage et programmation, ou encore les modèles d'affaires).

Programme 7 « Le processus de naturalisation » (Gresec).

La machine (par exemple les Bots et leur dimension conversationnelle) est présentée puis perçue comme autonome, fonctionnant hors de tout déterminant socio-économique. Une approche critique de ce type de dispositif serait donc précisément de « ré-encadrer » ces outils dans leur contexte socio-économique de production. Mais plus que cela, en s'intéressant à l'automatisation de l'action, la recherche pourrait se déplacer vers les théories de l'intentionnalité, de la rationalité et les théories du choix. Un des aspects intéressants (et sur

lequel les sciences de l'information et de la communication peuvent intervenir) pourrait porter sur la compréhension et la mise à jour – « l'ouverture de la boîte noire » – de la performativité propre à ces dispositifs (personnalisation de l'énonciation éditoriale propre au dispositif et actions mises en œuvre dans le cas de commandes).

Programme 8 : « Le rôle des conseillers robotiques dans les marchés financiers » (GAEL, Paolo Crosetto).

Dans ce projet s'appuie sur la méthode de l'économie expérimentale, et notamment les expériences de laboratoire et de terrain, pour étudier comment les humains et les algorithmes d'IA interagissent dans le domaine des décisions financières. Un focus particulier est porté au cas extrême où les décisions financières sont déléguées à des *robo-advisors*, une pratique de plus en plus répandue dans le secteur financier. L'objectif est d'étudier le rôle que les *robo-advisors* jouent dans la correction de deux biais bien connus de la littérature en finance comportementale, l'*aversion aux pertes* et l'étude et suivi de la *tendance* du marché. Les *robo-advisors* seraient par construction immunisés de ces et d'autres biais qui impactent les décisions humaines ; d'autre part, ils pourraient être utilisés comme *commitment devices* pour limiter les problèmes d'incohérence inter-temporelle des êtres humains.

Différents types d'utilisateurs pour ces *robo-advisors* peuvent être envisagés. Des investisseurs sophistiqués, conscients de leurs biais et limites cognitives, feraient un usage optimal de ces algorithmes, en leur déléguant les décisions de façon continue et en leur octroyant toute confiance. Des investisseurs *naïfs*, en revanche, pas conscients de leurs biais, seraient ceux qui bénéficieraient le plus des aides numériques, mais ceux qui moins en demanderaient.

Les expériences de laboratoire permettent d'éliciter l'attitude face aux *robo-advisors*, le profil des préférences inter-temporelles et leur cohérence, et leur aversion aux pertes avec le but d'évaluer quels impacts une diffusion d'algorithmes d'aide à la décision dans un contexte financier pourrait bénéficier les différents types de sujets.

Partenaires : ce programme est en partenariat avec Alexia Gaudeul (Georg-August Universität Göttingen, Allemagne) et Caterina Giannetti (Università degli Studi di Pisa, Italie)

Programme 9 : « Processus d'innovation dans les IA et impact sectoriel : énergie, agriculture, santé » (GAEL, axe Innovation, Nadine Massard, Stéphane Lemarié).

Ce programme qui est aujourd'hui prospectif a pour objectif d'étudier les processus d'innovation en IA et leur impact sur différents secteurs sur lesquels GAEL a une expertise (énergie et agriculture) ou envisage de s'investir dans l'avenir (santé). Pour ce qui concerne le processus d'innovation, l'enjeu est d'identifier les acteurs économiques qui prennent part au processus d'innovation, d'étudier leurs stratégies et les politiques publiques de type STI (Science Technologie, Innovation) qui peuvent favoriser de telles innovations. Pour ce qui concerne l'impact sectoriel, l'objectif est d'étudier comment l'IA modifie le partage de la valeur et la concurrence dans différents secteurs. L'IA est souvent associée à l'émergence de plateformes, si bien que ces questions peuvent être étudiées dans le cadre des travaux sur les marchés à plusieurs faces. Enfin, l'IA est souvent associée au développement des nouveaux

équipements (capteurs, robots, etc.) et il nous semble particulièrement intéressant d'étudier comment se partage la valeur entre ces innovations complémentaires.

L'IA présente d'importantes perspectives dans les différents secteurs présentés plus haut : les réseaux intelligents dans le domaine de l'énergie, l'agriculture de précision dans le domaine agricole, l'accompagnement des malades et les relations entre médecins et patients, etc.

Programme 10 : « IA et innovation » (CERAG, axe « Innovation et Complexité Organisationnelle »).

Recherches actuelles sur les capacités organisationnelles d'innovation de business modèles digitaux, dans lesquels intervient naturellement l'IA. Thèse encadrée à ce sujet. Toutefois actuellement dans les terrains explorés par la doctorante, les entreprises étudiées ne mettent pas en place explicitement des technologies d'IA dans leur projet.

Projet de monter une thèse sur créativité et l'IA. Discussions avec des collègues luxembourgeois, pour le moment au point mort.

Liens possibles entre IA et recherches en innovation

- Les pratiques managériales dans les organisations innovantes
- L'innovation organisationnelle : digitalisation, compétences, GRH
- Les capacités d'absorption d'innovation, les capacités dynamiques
- Le management des industries créatives

Liens possibles entre IA et complexité

1. Multi : niveau ; espace ; acteur ; relation ; enjeux, intérêt ; représentation
2. Processus : évolutif ; non linéaire ; non planification ; collaboratif ; réseaux
3. Objets : dispositifs ; instruments ; comportement ; organisation ; technologie ; internet des objets ; big data ; distribution, etc.

Programme 11 : « IA et gestion des risques » (CERAG).

En finance

En présence de différences d'information entre les acteurs financiers sur la valeur fondamentale des entreprises, l'Intelligence artificielle (IA) pourrait orienter les acteurs financiers dans une même direction et de ce fait amplifier les bulles spéculatives. Le trading haute fréquence implique des robots, et non pas des humains... De plus, les systèmes basés sur l'IA ont été conçus pendant des périodes de volatilité faibles sur les marchés ; risque donc de ne pas prendre les décisions optimales en cas de retournement économique majeur.

De nombreuses recherches au CERAG sont basées sur le postulat de présence d'asymétries informationnelles (AI) entre investisseurs sur les marchés financiers, ou entre les investisseurs et les entreprises. Ces expertises peuvent être mobilisées sur le champ de l'IA. Autres expertises :

- Analyse de la performance : l'IA rend-t-elle les entreprises plus performantes ?
- Trading haute fréquence et IA : expertises en microstructure des marchés financiers.
- Expertises sur comment se créent les bulles spéculatives, risque systémique, ...

IA et asymétrie d'information

Comment l'IA peut réduire l'asymétrie d'information entre acteurs et ainsi améliorer les prises de décisions au sein des organisations ? Cela pourrait représenter un champ de recherche porteur est intéressant. La notion d'asymétrie informationnelle est présente dans des recherches du CERAG en marketing (réduction d'AI entre l'entreprise et le client), logistique (informer toutes les parties prenantes pour optimiser les flux), systèmes d'information (étude de « signaux faibles » pour mieux anticiper, il s'agit là d'une « niche » de recherche pour le CERAG).

IA et résilience

Expertises en résilience collective ou d'un système. Elle se met en œuvre dans des situations de crises (aléa*risque) pour permettre à un collectif de faire face à la situation (choc brutal, traumatisme, rupture) et rebondir. Liens possibles avec l'IA :

- Source de résilience : Collaboration, confiance, capital social, routines organisationnelles, flexibilité, etc.
- Processus de résilience : dynamic capabilities, renaissance, reconstruction, aptitude à se réinventer (exemple business model)
- Apprentissage et intelligence collective : amélioration et évolution (des routines, de la confiance, de la flexibilité), capitalisation.

Programme 12 : « L'Analyse Prédictive pour la production à la demande » (CERAG / Paul Ready).

Dans ce projet de recherche, nous étudions les processus de prévisions déployés dans les industries appliquant le mode « production à la demande » et qui sont confrontés à une incertitude liée à la variabilité de la demande. L'intérêt d'associer les nouvelles technologies comme l'Internet des Objets et l'analyse prédictive, pour la production à la demande constitue le véritable enjeu de ce projet. Il se décline actuellement de deux manières :

(1) sous la forme d'un programme du projet régional de l'ARC 7 (2014-2016) intitulé « L'entrepôt collaboratif » et qui consiste à mettre en œuvre une approche « Bottom-Up » à base d'agents intelligents pour la gestion des commandes des entrepôts mutualisés dans un environnement ambiant et collaboratif (laboratoires partenaires : CERAG (UGA) et DISP (INSA de Lyon), enseignants-chercheurs impliqués Karine Samuel (Grenoble INP), Paul Ready (Grenoble IAE), Valérie Botta-Genoulaz (INSA de Lyon), Thibaud Monteiro (INSA de Lyon)) ;

(2) sous la forme d'un projet de recherche doctorale qui débutera en 2019 (laboratoire partenaire : CERAG(UGA), enseignants-chercheurs responsables : Paul Ready (Grenoble IAE) et Karine Evrard Samuel (Grenoble INP), entreprise partenaire : Probayes, région Rhône-Alpes).

Programme 13 : « Projet ALGOLUX » (LARHRA, Lucie Goujard et Sophie Raux).

À la croisée de l'informatique et des sciences historiques, le projet Algolux vise à renouveler l'étude de la représentation de la lumière dans les arts visuels, par l'élaboration d'outils numériques spécifiques. À partir de la reconstitution en 3 dimensions d'une sélection de tableaux emblématiques, il s'agira d'étudier, à l'aide d'algorithmes de simulations d'éclairage, les effets de lumière dans la peinture occidentale, en privilégiant les dimensions pratiques et techniques de leur mise en œuvre. En mettant en évidence de manière tangible et inédite l'écart entretenu par l'imaginaire des peintres et la réalité optique et physique du

travail d'atelier, le projet Algolux permettra de découvrir de fascinants décalages ouvrant la voie à de nouvelles interprétations. Il vise donc à apporter une contribution majeure à l'un des principes les plus méconnus de l'histoire des arts visuels : la fabrique de l'éclairage.

Programme 14 : « IA et approches qualitatives autour des questions de la mémoire et du récit » (LARHRA, Anne Dalmasso, Anne-Marie Granet Abisset, Pascale Ancel, Nasima Moujoud).

Cette thématique peut s'inscrire comme une des pistes de prolongement du GDR Mémoires.

Programme 15 : « Acceptabilité, usage et implications de l'usage des technologies incorporant l'IA » (LIP/PC2S, Michel Dubois, Marc Gandit, Emmanuel Monfort, Caroline Poulet, Anna Tcherkassof).

Un certain nombre de travaux du LIP/PC2S porte sur l'analyse des processus sociocognitifs dans le travail et des réflexions autour de la dialectique entre conception d'objets intermédiaires et usage de ces objets, notamment des objets donnant une part centrale l'IA. Le LIP/PC2S effectue depuis de nombreuses années des travaux portant spécifiquement sur les objets techniques (notamment les NTIC) et sur leurs modalités d'appropriation. Le dénominateur commun des travaux de ces recherches vise à la fois à produire une *meilleure compréhension* des processus cognitifs et socio-cognitifs mis en œuvre lors de l'introduction de NTIC et une meilleure compréhension des problèmes liés à l'acceptabilité de ces technologies. Plusieurs thématiques de recherche ont été entreprises dans ce sens :

- Des travaux portant sur « *l'utilisabilité* » des objets techniques. Le champ d'études abordé est celui de l'IHM (Interface Homme Machine). L'utilisateur est appréhendé sous l'angle de son interaction avec le dispositif. Ces questions sont abordées sous un angle cognitif, « fondamental » et expérimental avec des protocoles de laboratoire. Elles visent à faire avancer les modèles et les hypothèses de recherche dans la théorie du traitement de l'information, notamment sur les effets des présentations multimodales et des modalités d'interaction.

- Des travaux portant sur « *l'usage* » des objets techniques et de leur impact sur la santé. Il s'agit d'une approche réalisée en contexte réel. La question de recherche porte sur la manière dont les utilisateurs confèrent de nouvelles significations et de nouveaux usages aux dispositifs techniques mis à leur disposition. On s'intéresse notamment au rôle de l'usager – et des groupes auxquels il se rattache – dans le processus d'assimilation de connaissances, mais aussi de détournement, de contournement et d'innovation des fonctionnalités et des dispositifs techniques.

- Enfin des travaux portant sur le « *développement* » de logiciels et d'interfaces de plus en plus « intelligents », de plus en plus conviviaux et davantage centrés sur l'utilisateur. Dans ce cadre des thématiques de recherche innovantes sur les interfaces émotionnelles, sur « Informatique et émotion », « Informatique et persuasion » et sur des conceptions de dispositifs symbiotiques.

Sur le plan théorique, l'objectif général est de montrer que l'utilisation et l'usage sont déterminés par une série de variables relevant des fonctionnalités du système, de l'utilisabilité de la technologie et des conditions organisationnelles de son déploiement. Le modèle général d'analyse s'organise autour des processus psychologiques d'acceptabilité et d'acceptation de produits, démarches, dispositifs technologiques, etc. dans des contextes de changement et/ou d'innovation. La question qui se pose actuellement est de resituer ces variables (d'acceptabilité a priori) avec les variables d'acceptation en situation réelle. Comment les

utilisateurs passent-ils d'un stade à l'autre ? Quelles en sont les conditions suffisantes ? Quelle est la place de l'émotion dans ce passage ?

Programme 16 : « Implications motivationnelles de l'usage des nouvelles technologies dans le domaine de la santé » (SENS, Aïna Chalabaev, Jean-Philippe Heuzé, Sandrine Isoard-Gauthier, Philippe Sarrazin, Damien Tessier, David Trouilloud).

Les études sur les effets de l'activité physique sur la santé se sont accumulées ces dernières années. D'un côté, il a été observé que l'inactivité physique est un facteur déclenchant ou aggravant de nombreuses pathologies chroniques (e.g., obésité, diabète de type 2), et constitue un des principaux facteurs de risque de mortalité dans le monde. D'un autre côté, une pratique minimale d'activité physique diminue la mortalité et augmente la qualité de vie à tous les âges de la vie. Cependant, malgré ces preuves scientifiques, la population reste en majorité insuffisamment active. En effet, si beaucoup d'individus sont conscients de l'importance de l'activité physique pour la santé, peu changent durablement leurs habitudes de vie. Il est donc crucial d'identifier les freins et leviers motivationnels à l'adoption et au maintien d'une activité physique régulière. Le Laboratoire SENS travaille à cet objectif en s'appuyant sur des modèles théoriques contemporains issus de la psychologie sociale. Dans cette optique, les innovations technologiques actuelles, comme les objets connectés (e.g., bracelets connectés), représentent a priori des outils à fort potentiel pour favoriser l'adoption et le maintien d'une activité physique régulière, car ils peuvent potentiellement, et de multiples manières, jouer sur les freins et leviers motivationnels à la pratique. Mais qu'en est-il réellement ? La possibilité que ces outils offrent d'obtenir des mesures objectives de son activité physique ou physiologique (*quantified self*) favorise-elle un engagement durable dans une pratique physique régulière ? La possibilité, par l'intermédiaire d'Internet, d'interagir d'une certaine manière avec autrui, en rendant son activité publique par exemple, ou en favorisant la pratique en collaboration ou compétition avec d'autres pratiquants, agit-elle comme un frein ou un levier motivationnel ? Le laboratoire SENS mène des projets visant à apporter des éléments de réponse à ces questionnements.

3/ Ethique

Programme 17 : « Éthique philosophique de l'IA » (PPL, Denis Perrin).

Il s'agit de développer, sous différentes formes et notamment collaboratives-interdisciplinaires, des travaux de recherche sur les normes éthiques et politiques par lesquelles il convient de réguler les développements actuels de l'intelligence artificielle. Cette réflexion est conduite par les chercheurs de PPL en particulier dans les domaines de la neurotechnologie, de l'innovation technologique, de la médecine et plus largement des sciences cognitives.

Programme 18 : « Ethique opérationnelle de l'IA : CER Grenoble Alpes » (LPNC, Carole Peyrin).

L'éthique représente un champ disciplinaire qui est de plus en plus sollicité. En effet, son développement est relatif aux inquiétudes liées aux avancées de la science et des technologies. De plus, l'application et la valorisation de certaines découvertes amènent les scientifiques à se poser des questions sur leur impact sociétal et environnemental. Enfin, l'éthique s'interroge sur le caractère invasif des expériences réalisées qui ne doivent pas porter atteinte à l'intégrité de l'individu. L'intérêt est donc grand de former un axe Ethique pour favoriser et organiser ce questionnement au sein de la communauté scientifique.

Au sein des établissements d'enseignement supérieur et des organismes de recherche, les questions éthiques peuvent être traitées à trois niveaux au moins. Le premier niveau concerne nos formations universitaires : par exemple, nos étudiants et doctorants reçoivent-ils une formation à l'éthique ? Le second niveau concerne nos pratiques de chercheurs : par exemple, comment mettons-nous en œuvre des préoccupations éthiques dans notre travail quotidien ? Le troisième niveau concerne l'environnement de notre recherche : par exemple, comment répondre aux revues, qui exigent l'approbation d'un comité d'éthique pour publier les travaux ? Plus largement, comment ces trois niveaux rencontrent-ils les questions éthiques plus globales qui se posent au niveau de la société ?

Pour répondre au troisième niveau, la COMUE UGA s'est dotée d'un comité d'éthique opérationnel qui procède à une évaluation des protocoles de recherche soumis par les équipes du site grenoblois. Ce comité a été créé le 13 septembre 2011 au sein du Pôle Grenoble Cognition (Fédération de Recherche 3381 du CNRS et SFR du site grenoblois) sous le nom de CERNI (Comité d'Ethique pour les Recherches Non Interventionnelles). Ce comité est rattaché depuis le 11 juillet 2016 à la COMUE Université Grenoble Alpes et porte désormais le nom de CER Grenoble Alpes (Comité d'Ethique pour les Recherches de Grenoble Alpes), mais il reste indépendant dans son organisation et ses décisions. Le CER Grenoble Alpes est composé de membres volontaires, dont le nombre et la qualité sont représentatifs de la diversité des disciplines concernées. Il est accompagné par des juristes du Groupe de Recherche Droit et Sciences de l'Université Grenoble Alpes et par des Correspondants Informatiques et Libertés (CIL) et relais CIL.

Le CER Grenoble Alpes est sollicité pour examiner et fournir un avis sur les aspects éthiques concernant les protocoles de recherches impliquant la personne humaine qui ne relèvent pas de la loi n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (dite loi Jardé). Entrent dans le champ d'application de la loi Jardé, les recherches organisées et pratiquées sur l'être humain en vue du développement des

connaissances biologiques ou médicales. Il en résulte par conséquent que les recherches scientifiques impliquant la personne humaine en vue du développement des connaissances en intelligence artificielle ne relèvent pas, en général, des dispositions de cette loi et seront alors amenées à solliciter les compétences du CER Grenoble Alpes.

Ces avis éthiques pouvant être demandés dans le cadre d'une publication, lors de la réponse à un appel à projets (European Research Council), ou lors de projets de recherches internationaux, le CER Grenoble Alpes est devenu une composante indispensable au site grenoblois. Pour renforcer sa visibilité et sa reconnaissance internationale, le CER Grenoble Alpes a été enregistré auprès de l'Office of Human Research Protection (OHRP) du Département de la Santé des Etats-Unis comme un Institutional Review Board (IRB00010290 COMUE Université Grenoble Alpes IRB #1). Enfin, le comité apporte depuis plusieurs années son expertise à plusieurs sites institutionnels français souhaitant créer un Comité d'Éthique analogue (SFR Neuroimagerie Fonctionnelle Tours-Poitiers, COMUE Université de Toulouse, Université de Picardie Amiens, COMUE Bourgogne-Franche-Comté, etc.).

Programme 19 : « Débat public et IA » (Gresec).

Ce programme entend prendre en compte les modalités de mise en public d'une politique publique et la manière dont s'organisent (ou pas le débat) et les questionnements sur l'IA (rôle des médias, place des experts, communication scientifique et technique, les registres utilisés...). Cette recherche s'assortit d'une réflexion à mener sur les relations entre usages de l'IA et pratiques sociales (en particulier dans les organisations sociales dans le domaine de la santé, de l'éducation et de l'environnement (et sur ce dernier point la question des risques). Une autre perspective est ouverte par-là, qui concerne la relation entre IA et traitement des informations (big data). Il s'agit alors de traiter l'éditorialisation des données (leur écriture, mise en forme, mise à disposition de publics différenciés, gestion des connaissances).

Programme 20 : Programme 20 : « IA, comportements responsables et éthique » (CERAG, axe « Comportements responsables et enjeux sociétaux »).

Les chercheurs de cet axe ont un intérêt commun pour la compréhension des conditions et mécanismes conduisant à la mise en place de comportements responsables et pour l'étude de leurs effets (image, performance, sociale, bien-être, etc.). Ils partagent un référentiel de valeurs éthique, altruisme, respect, équité et responsabilité. Les champs d'application de leurs recherches sont la santé, le travail, l'éducation, l'écologie et les transports-mobilité (pour certains décrits comme prioritaires par le rapport Villani).

Les chercheurs de cet axe mènent ou possèdent une expertise forte pour répondre aux défis majeurs posés par l'IA

- **L'éthique de la collecte et de l'utilisation des données.** Les questions éthiques relatives à la collecte, au stockage et à l'utilisation des données sont présentes dans les travaux des membres de l'axe. Citons le projet transdisciplinaire DUTY (E. Hoareau, et K. Samuel), lauréat IRS 2018. Ce projet a pour objectif d'étudier le concept de transparence dans le cadre de l'éthique des données anonymes ou anonymisées. Il réunit des chercheurs en droit (CECISE), Sciences de gestion (Laboratoire CERAG) et informatique (LIG) afin de proposer un programme de mise en conformité comprenant des règles normatives, un outil ethics-by-design et un code déontologique associé.
- **La (co)-conception d'un dispositif s'appuyant sur l'IA.** La conception implique la compréhension du fonctionnement des acteurs. Les postulats, limites et enrichissements (avec par exemple les modèles basés sur des hypothèses de rationalité des acteurs vs. des hypothèses d'irrationalité des acteurs, l'existence de biais comportementaux,...) pourrait fonder des méthodes de conception. De plus, il est nécessaire dès la phase de conception,

comme le souligne le rapport Villani, de savoir « penser les modes de complémentarité entre l'humain et les systèmes intelligents ». L'extension des travaux actuels sur la collaboration, les interactions au sein des organisations, la coopération interentreprises mais aussi avec les utilisateurs / citoyens, les relations avec les parties prenantes, pourraient apporter un éclairage sur un enjeu de société majeur.

- **Le partage de la valeur** entre l'ensemble des parties prenantes, en lien par exemple avec l'automatisation de certaines tâches dessine de nouvelles relations sociales mais aussi économiques. De même, l'IA modifie le processus de décision. La (création) de valeur et le processus de décision sont des notions centrales dans les travaux de l'équipe.
- **La perception et l'acceptation de l'IA par les cibles.** C'est un enjeu important qui conditionne le développement d'un dispositif d'IA. Elle est met en jeu des mécanismes bien étudiés par les chercheurs de l'axe : l'expérience anticipée ou vécue et notamment l'imaginaire, l'imagerie mentale, les émotions mais aussi la résistance / réactance, etc.
- **Les effets de l'IA sur la prise de la prise de décision, la performance ou la mise en place de comportements durables** sont des enjeux de recherche et de société majeurs qui rentrent dans les champs d'expertise de l'équipe. A cet égard, citons le projet transdisciplinaire, Mov2Health (A. Helme-Guizon), lauréat IRS 2018-2021, qui réunit des chercheurs en Sciences de gestion (CERAG), en information & communication et un Pr en pneumologie (LBFA-CHU). Il porte sur les effets des objets connectés du *Quantified-Self*, en lien avec un ensemble de dispositifs, sur l'engagement durable des personnes à risques et de patients BPCO dans une activité physique.

Programme 21 : « Algorithmes, contrôle social technologique, justesse de décision et d'attribution » (PPL).

Ce programme de recherche philosophique, qui se situe à l'intersection des axes « pratiques sociales », « normes » et « éthique », vise d'abord à souligner la tension qui existe aujourd'hui entre les deux attentes dominantes en matière d'usages de l'IA : d'une part le contrôle social qu'elle exerce déjà (et qu'on attend qu'elle exerce) dans les organisations pour lesquelles elle s'applique, et de l'autre sa valeur normative en termes d'intelligence sociale, notamment en matière de justesse dans les décisions – qu'il s'agisse de détermination de procédure ou de choix d'affectation ou de distribution de biens.

Prospective, cette recherche vise à associer aux philosophes (éthiques et politiques) des chercheurs ingénieurs informaticiens, politistes et juristes, urbanistes, aménageurs et architectes – autant de spécialités de recherche qui rencontrent le problème de l'organisation sociale dans sa dimension politique. Normative, elle s'attachera à énoncer des règles rationnellement désirables pour encadrer les usages de l'IA, dans une double dimension : (1) dans une approche critique de ses pouvoirs de contrôle en fonction de la valeur de liberté, valeur cardinale pour les sociétés démocratiques, (2) dans une approche visant à instaurer les principes d'équité fondamentaux pour tout calcul d'attribution de ressources ou d'affectation de biens de base. Une « théorie de la justice » appropriée à l'IA apparaît souhaitable : ce programme de recherche entend y contribuer.

Cette recherche s'appuie sur les travaux engagés au sein de l'axe « philosophie pratique » de PPL, et plus particulièrement sur le séminaire « philosophie des techniques » proposé depuis 2015 par cette équipe. Elle peut se décliner par un séminaire pluridisciplinaire, par la réponse à des AAP (financement ANR et d'allocations doctorales). Elle intéressera de plus des partenaires extérieurs à l'académie comme les collectivités et les organisations concernées par des problèmes de décision publique ou d'affectation justes (à la fois en termes de justesse et de justice distributive) de ressources (par exemple les mutuelles d'assurance).

4/ Normes

Programme 22 : « IA, justice prédictive et raisonnement probatoire » (CRJ, Etienne Vergès / LJK, Caroline Bazzoli).

Ce programme vise à mobiliser les compétences en IA afin de déterminer si des algorithmes peuvent être utilisés pour analyser et prédire des raisonnements probatoires dans les décisions de justice. Ce projet interroge donc le thème dit de la « justice prédictive ».

Ce projet s'intègre dans un contexte très récent de recherche sur la prédiction des décisions de justice. Les progrès dans l'analyse de données textuelles et l'apprentissage automatique laissent à disposition des outils pour construire des modèles prédictifs (Sebastiani 2005, Korde et al. 2012) dans un contexte d'open data. L'apprentissage automatique vise à construire des systèmes qui vont apprendre à classer les décisions de justice. Les travaux les plus innovants ont recours à des modèles de classification (Support Vector Machine Vapnik, 1998) utilisant un algorithme de type N-grammes pour extraire des séquences de mots contiguës des décisions et les utiliser comme facteurs de prédiction (Aletras et al. 2016). Des méthodes de type forêts aléatoires sont aussi utilisées (Katz et al. 2017). Leur particularité consiste à réaliser des prédictions sur le résultat de la décision de justice sans connaître ou comprendre le raisonnement du juge. Ces prédictions reposent seulement sur l'analyse des faits et des éléments de preuve émanant de la décision de justice. L'enjeu de ce type de méthodes est de fournir aux praticiens (ex. avocats) des outils d'anticipation d'une éventuelle décision de justice à partir des éléments de preuve d'un dossier. Les travaux les plus innovants portent sur des décisions rendues par des cours suprêmes (États-Unis, Europe), ce qui entraîne certains biais dans les résultats. Le projet prévoit de tester ces méthodes d'analyse en utilisant des décisions de la justice quotidienne (arrêts de cours d'appel) qui présentent une plus grande variabilité. Cette variabilité devrait opérer comme un obstacle à la prédictibilité des raisonnements probatoires des juges.

Programme 23 : « Protection des créations issues de l'IA par droit d'auteur » (CRJ-CUERPI, Jean-Michel Bruguière).

HDR en cours de Bérengère Gleize (MCF, Université d'Avignon et des Pays du Vaucluse)

Les créations des machines intelligentes sont aujourd'hui de plus en plus nombreuses et la Commission juridique du Parlement européen a invité le 12 janvier 2017 la Commission Européenne à soumettre une directive définissant « des critères de création intellectuelle propre applicables aux œuvres protégeables par droit d'auteur créés par des ordinateurs ou des robots »¹. Les défenseurs de cette position soulignent qu'une telle protection stimulerait ainsi la création dans ces domaines et conférerait à l'Europe un avantage concurrentiel. Il s'agira ici, sans dogme, de savoir, d'une part, si notre système favorise bien le processus de création de ces machines intelligentes (notamment avec le *deep learning*) et, d'autre part, si notre droit de la propriété intellectuelle (droit d'auteur *et copyright*) est bien à même de protéger ces créations.

¹ Projet de rapport contenant des recommandations à la Commission concernant les règles de droit civil sur la robotique, 2015/2103 (INL)

Programme 24 : « La protection des technologies d'IA » (CRJ, Amélie Favreau).

Il y a un lien presque naturel à questionner la propriété « intellectuelle » sur les questions d'« intelligence » artificielle. Apte à appréhender l'incorporalité des objets, la propriété intellectuelle est en premier lieu sollicitée pour la protection des créations issues de l'IA, particulièrement pour le droit d'auteur (programme 17). En second lieu, il est possible d'envisager la propriété intellectuelle comme un outil de protection des technologies d'IA. Cette seconde démarche implique de saisir la technologie dans sa complexité. En effet, l'intelligence artificielle est le résultat d'une combinaison d'algorithme et de données.

La protection de l'algorithme s'inscrit dans le prolongement des réflexions sur la protection du logiciel (droit d'auteur) ou des inventions mises en œuvre par un ordinateur (brevet). Plus largement, elle interroge sur la titularité des droits (titularité de lignes de codes générées par l'IA), sur les contours des prérogatives des titulaires et les possibles transferts de technologies.

La protection des données qui nourrissent l'algorithme est également délicate. Elle implique de déterminer si nous sommes en présence de données personnelles ou non-personnelles, données publiques ou données privées, pour adapter le régime des droits. Ce sont par exemple, les questions relatives à la titularité des données et à une titularité partagée avec le propriétaire de l'algorithme qui seront au cœur de cette seconde problématique. La valeur de la donnée et de la valeur de l'entraînement de l'algorithme seront au centre de cet arbitrage sur la titularité.

La réflexion sur l'IA a démarré par une conférence de M. Nicolas Bronzo, Maître de conférences à Aix-Marseille Université, lors d'un colloque organisé par Mme Amélie Favreau à Bangkok (Thaïlande) le 26 février 2018 sur « l'impact des nouvelles technologies sur la propriété intellectuelle ». Elle s'est poursuivie par l'organisation d'une conférence le 27 avril 2018, à l'IUT2 de Grenoble, ayant réuni : Mme Carole ADAM, Maître de conférences en informatique, LIG, UGA ; Maître Jérémy Bensoussan, avocat au barreau de Paris, Cabinet Lexing Bensoussan Avocat et M. Clément Marc, juge auprès de la Cour administrative d'appel de Lyon. Et elle sera prolongée par l'inscription en thèse de Mme Mathilda Moynet à l'EDSJ (UGA).

Programme 25 : « IA, algorithmes et processus décisionnels. Vers une datacratie ? » (CESICE, Serge Slama, Théodore Christakis et d'autres collègues dans le cadre du Grenoble Alpes Data Institute).

Admission post bac, Parcoursup, voiture « intelligente », gestion du trafic, calcul des impôts, détection des plaques d'immatriculation pour verbaliser les contrevenants, guidage par internet, les algorithmes, et plus largement les procédés reposant sur des systèmes d'IA, ont de plus en plus d'impact dans notre vie quotidienne et surtout orientent de plus en plus la prise de décision aussi bien de la collectivité publique que par des personnes privées.

Or, cette influence grandissante des algorithmes sur la vie des citoyens à travers la prise de décision est appréhendée par les normes de manière tout à fait insatisfaisante. On peut certes relever l'article L. 311-3-1 du Code des relations entre le public et l'administration, créé par la loi 2016-1321 du 7 octobre 2016 pour une République numérique. Cet article impose que l'existence d'un algorithme soit mentionnée lorsque celui-ci fonde la prise de décision administrative. Un droit à se faire communiquer les détails de cet algorithme est alors ouvert. Le Règlement européen sur la protection des données (RGPD), entré en vigueur mai 2018, énonce pour sa part à son article 22 un principe d'interdiction de fonder « exclusivement » une décision individuelle sur un « traitement automatisé », moyennant

certaines exceptions comme le consentement de la personne intéressée. La CNIL veille à ce caractère individuel du traitement des demandes.

Dans un récent rapport, la CNIL a formulé « 6 recommandations opérationnelles » à destination tant des pouvoirs publics que des diverses composantes de la société civile (entreprises, grand public, etc.) visant notamment à « former à l'éthique tous les acteurs-maillons de la « chaîne algorithmique » » ou encore à « rendre les systèmes algorithmiques compréhensibles en renforçant les droits existants et en organisant la médiation avec les utilisateurs ». Mais cet encadrement sera-t-il suffisant eu égard au rôle de plus en plus important donné aux algorithmes dans le processus décisionnel ?

Quel est l'impact réel des algorithmes dans la prise de décisions des acteurs publics et privés ? Quelle est l'influence de ceux-ci sur les comportements et sur nos vies ? Est-on dans une situation où les personnes sont, dans leurs décisions, de plus en plus enfermés par ce que déterminent les algorithmes ? Quelles garanties faut-il apporter pour assurer le respect effectif des droits fondamentaux ? Quelles seraient les règles « éthiques » qui pourraient être dégagées pour assurer ce respect ? Quelles évolutions ou mutations de ces droits fondamentaux seraient nécessaires afin de garantir les droits ? Comment éviter que les algorithmes contribuent à entretenir les discriminations et stéréotypes ? Comment encadrer l'usage des algorithmes afin qu'ils n'envahissent pas toutes les sphères de la vie privée (data-panoptisme) et que les données personnelles soient également protégées et respectent le droit à l'autodétermination informationnelle ?

L'encadrement de l'usage des algorithmes dans le processus de prise de décision est aussi un impératif démocratique afin d'éviter que, suivant un processus purement rationnel et instrumentalisé, les machines décident à notre place développant au mieux une « Datacratie » et ou pire une « Datacature ».

La recherche pourrait aussi porter sur l'impact potentiel des processus décisionnels reposant sur l'IA, en particulier sur le deep learning, dans lesquels la machine entre dans des processus de perfectionnement, d'autoapprentissage voire d'autodétermination échappant en partie à l'être humain ce qui n'est pas sans poser des questions juridiques en termes de responsabilité, de titularité de la personnalité juridique et d'imputabilité du processus décisionnel, mais aussi de compréhension des critères présidant à la prise de décision.

Programme 26 : « La réglementation des armes autonomes » (CESICE, Théodore Christakis, Karine Bannelier, Thierry Garcia, Serge Slama et d'autres collègues dans le cadre du Grenoble Alpes Data Institute et du Grenoble Alpes Cybersecurity Institute).

Dans la continuité de ses travaux sur le droit de la sécurité internationale le CESICE travaille sur la question de la « course aux armements » fondés sur des systèmes d'intelligence artificielle. Les armes autonomes posent des questions juridiques et éthiques redoutables et ont fait l'objet de vives réactions et de plusieurs propositions de réglementation. La question de la capacité du droit international existant de répondre au défi des armes autonomes se pose avec force – tout comme la question de savoir si et comment un traité international pourrait être conclu dans ce domaine.

Programme 27 : « Adversarial examples in machine learning & AI security » (CESICE, Théodore Christakis, Karine Bannelier, Thierry Garcia, Serge Slama et d'autres collègues dans le cadre du Grenoble Alpes Data Institute et du Grenoble Alpes Cybersecurity Institute).

Security often relies on machine learning to detect malicious behaviors using binary classification. It was shown, however, that using standard learning algorithms for security tasks leads to poor performance because attackers are adaptive and are able to avoid detection by adjusting or manipulating the data that they generate while crafting their attacks. In this project, we argue that the core origin of the problem is that standard learning algorithms were developed and studied under the fundamental assumption that the data follows a distribution independent from the learning algorithm, which is not valid in security scenarios. Therefore, we need to find new data models adapted to security applications and use them to derive better performing algorithms for this type of data. To that end, we propose to use game theory to model the objective and adaptiveness of the attacker generating data, and the objective of the defender. To successfully address these ambitious objectives, we have put together, under the leadership of Patrick Loiseau (LIG) a team that gathers all the necessary competences on the Grenoble site: experts in game theory and learning and their applications to security from the POLARIS team of LIG; experts on practical security from the PRIVATICS team of Inria; and experts on (cyber)security laws from the CESICE.

Programme 28 : « Des LegalTech à la Justice Prédicative : Solutions d'IA pour les métiers du droit » (CESICE, Théodore Christakis, Karine Bannelier, Thierry Garcia, Serge Slama et d'autres collègues dans le cadre du Grenoble Alpes Data Institute et du Grenoble Alpes Cybersecurity Institute).

Dans le cadre du Grenoble Alpes Data Institute nous menons une réflexion sur les différents domaines où l'IA intervient déjà dans les différents métiers du droit. Nous avons dressé une liste avec les LegalTech existants, leurs différents domaines d'intervention et les nouvelles technologies utilisées (intelligence artificielle, machine learning, prédictif...). De façon plus générale nous travaillons sur le thème « Artificial Intelligence and Legal Analytics » afin d'identifier comment les nouvelles applications de l'AI changent la pratique du droit, quels sont les arguments pour et contre, et quelles sont les solutions à imaginer afin que le domaine du droit puisse pleinement bénéficier de l'IA.

Programme 29 : « La maîtrise du risque cyber/IA et son transfert vers l'assurance » (CESICE, Théodore Christakis, Karine Bannelier, Thierry Garcia, Serge Slama et d'autres collègues dans le cadre du Grenoble Alpes Data Institute et du Grenoble Alpes Cybersecurity Institute).

Dans le cadre de la participation du CESICE au sein du Grenoble Alpes Cybersecurity Institute nous avons (avec des collègues d'autres disciplines, y compris le CERAG) un grand projet qui consiste à évaluer, gérer et maîtriser le risque cyber pour ensuite le transférer éventuellement vers l'assurance. Nous souhaitons comprendre le risque cyber et essayer de le qualifier et le quantifier financièrement et juridiquement. Nous souhaitons aussi étudier la réponse du marché de l'assurance en termes de couverture du risque cyber/AI. Ainsi, pour ne donner qu'un exemple, nous essayons d'imaginer comment l'assurance peut/doit gérer les risques liés à des accidents de voitures autonomes et quelles sont les solutions que le droit peut proposer en termes de répartition de la responsabilité.