

SECTION 07
SCIENCES DE L'INFORMATION
SIGNAUX, IMAGES, LANGUES,
AUTOMATIQUE, ROBOTIQUE, INTERACTIONS,
SYSTEMES INTEGRES MATERIEL-LOGICIEL

Pierre-Olivier AMBLARD (président de section) ; Lucie BAUDOUIN (secrétaire scientifique) ; Patrice ABRY ; Nicolas ANDREFF ; Sébastien BRIOT ; Marie CHABERT ; Raphaëlle CHAINE ; Laurent CONDAT ; Christophe FONTE ; Gersende FORT ; Abdoulaye GAMATIE ; Michel GAY ; Nicolas HERAUD ; Denis LESUEUR ; Jean-Claude MARTIN ; Gilles MILLERIOUX ; Dimitri PEAUCELLE ; Maxime PELCAT ; Sophie ROSSET ; François ROUSSEAU ; Laure TOUGNE

Résumé La section 07 regroupe des thématiques de recherche dans les sciences de l'information, au coeur de l'intelligence artificielle moderne et de la société numérique. Les multiples disciplines de la section s'unissent pour relever des défis posés par les systèmes et les données en développant des méthodologies spécifiques. Parmi celles-ci, ce rapport donne des éléments sur l'apprentissage, la gestion des grandes dimensions, l'optimisation et les interactions comme lignes de force traversant la section. Dans un deuxième temps, les apports des recherches en science de l'information à quelques grands enjeux sociétaux sont présentés. Enfin, des éléments sur la place de la recherche française et quelques réflexions sur la pratique de nos disciplines terminent le rapport.

Introduction

La section 07 (S07) développe des recherches au coeur des sciences de l'information. Ses principaux thèmes concernent l'automatique, la science des données (images, signaux, géométries), les interactions homme machine, le traitement automatique des langues et de la parole, la robotique, les systèmes intégrés sur puce, et enfin les applications de toutes ces disciplines dans les sciences naturelles (physique, biologie, chimie), humaines (économie, sociologie, psychologie) et en santé. Les sciences de l'information développées en S07 sont héritières de la cybernétique de Norbert Wiener. Pour le paraphraser, « *le but de la cybernétique est de développer un langage et des techniques qui permettent d'attaquer le problème du contrôle et de la communication en général, mais également de trouver un répertoire propre d'idées et de techniques pour classifier leurs manifestations particulières dans certains concepts* » (dans *The Human Use of Human Beings* :

Cybernetics and Society, Houghton Mifflin, 1950). Cette phrase sous-tend les fondements de la S07 : le développement de méthodologies pour traiter ou mettre en forme l'information portée par des systèmes ou des données à des fins d'actions, d'interaction ou d'extraction de connaissances.

Les recherches de la S07 sont au coeur de l'intelligence artificielle moderne, ancrée dans les données réelles, tant par les questions méthodologiques fondamentales que par les recherches applicatives. Ces recherches sont en lien fort avec les sciences naturelles et humaines, et à ce titre sont très interdisciplinaires. Elles exploitent des interactions entre mathématiques, physique et informatique, manipulent des données et des systèmes issus de la physique, de la biologie, de la linguistique, de la psychologie ou encore de l'économie. A l'interface de multiples disciplines, les chercheurs relevant de la S07 échangent avec des collègues spécialistes de ces domaines variés pour développer de concert de nouvelles méthodologies dédiées.

La première partie de ce rapport décrit des défis méthodologiques actuels, de façon non exhaustive, qui définissent des lignes de force de la section. Quelques problèmes très actuels concernant les objets que manipulent les chercheurs relevant de la section, les systèmes, les données, leurs interactions, l'exploitation qui en est faite et pour quels types d'actions sont ensuite présentés.

La partie suivante illustre la place des recherches relevant de la S07 dans certains grands enjeux sociétaux. Le dernier paragraphe est dédié à la place de la recherche française dans le paysage mondial et précède quelques remarques sur la bonne pratique de nos disciplines.

I. Défis méthodologiques

Les chercheurs en sciences de l'information relevant de la S07 travaillent sur des objets complexes pour en extraire de l'information, les modéliser, les représenter, les observer, les contrôler ; ces objets présentent actuellement des défis importants, défis relevés à l'aide de méthodologies nouvelles développées en interaction avec les mathématiques, la physique, l'informatique et les disciplines dont ces défis sont issus.

A. Objets fondamentaux

1. Systèmes

De manière formelle, un système est un élément, ou un ensemble d'éléments qui interagissent et soumis à des lois de fonctionnement. Ces éléments peuvent être naturels issus des sciences du vivant ou de la terre (biologie, physique), artificiels (ordinateurs, robots, parcs éoliens) ou immatériels (économie, langue, parole). Ainsi, de nombreux systèmes sont des équipements aux technologies variées, associant mécanique, électronique, chimie ou physique, conçus par l'humain ou par le biais d'assistance par ordinateur. Ils doivent remplir des tâches spécifiques répondant à des contraintes de fonctionnement. Le système est souvent soumis à des perturbations issues de l'environnement dans lequel il évolue et ses propriétés de résilience jouent alors un rôle important.

Les systèmes sont décrits par des modèles théoriques. Ces derniers sont nécessaires pour comprendre des phénomènes, analyser des comportements, détecter des événements, contrôler ou superviser un procédé. Leur formalisation constitue souvent un défi et les méthodes utilisées pour le relever sont variées. Les modèles peuvent être obtenus à partir des principes physiques ou construits par observation sur des données produites par le système. Les modèles prennent des formes multiples : équations mathématiques, graphes, algorithmes. Généralement, les modèles ne décrivent les systèmes que de manière approchée. L'approximation dépend des échelles de temps ou d'espace considérées, des dynamiques retenues ou négligées ou encore de l'exactitude des mesures.

2. Données

Les données sont les vecteurs de l'information. Elles peuvent être observées et obtenues par des spécialistes de domaines scientifiques divers (par ex. images satellitaires, traces sismiques, matrices de contacts génomiques) et parfois mises en forme par les chercheurs pour être exploitées (par ex. signaux pour les communications, lois de commande pour l'automatique, planification de trajectoires pour des robots, données textuelles ou de parole collectées et annotées).

Les données sont représentées mathématiquement par des variables dans des espaces discrets (données textuelles, données catégorielles, données quantifiées) ou continus (mesures physiques), de dimension finie (signaux échantillonnés, images), voire infinie lorsque des représentations fonctionnelles sont appropriées (mouvements modélisés par des courbes, surfaces continues, phénomènes descriptibles par des équations aux dérivées partielles).

B. Défis actuels

1. Sur les systèmes

La nature des systèmes étudiés a connu un changement profond ces dernières années, principalement dû à l'avènement du numérique communément considéré comme la quatrième révolution industrielle. Durant

de nombreuses années, les systèmes étudiés étaient principalement des équipements isolés, possiblement complexes. La révolution numérique induit un changement de paradigme avec l'avènement des « grands systèmes » ou « systèmes de systèmes » ou « systèmes interconnectés ». Il est alors nécessaire d'étendre les modèles à des entités qui communiquent à des échelles sans commune mesure avec celles des années passées. Les nouvelles technologies de communication, filaires, sans fil ou en réseau, ont permis d'interconnecter et donc de faire interagir des équipements distants géographiquement dans des proportions inédites. Le nombre d'entités interconnectées et leur forte hétérogénéité nécessitent de revisiter profondément les outils de conception, d'analyse, de surveillance et de pilotage de ces équipements. Les formalismes classiques ne s'appliquent plus nécessairement et doivent être adaptés. De plus, le mode de communication doit être intégré dans le modèle du système.

L'Internet des Objets (IoT), les systèmes cyber-physiques, les systèmes multi-agents (flottes de drones, de robots, de véhicules), les nouveaux modes de production et de gestion de l'énergie (Smart-grids) entrent dans cette catégorie de systèmes modernes. Induits par l'ubiquité du numérique, de nouveaux modes de production sont apparus. On leur associe le concept d'industrie 4.0 ou industrie du futur. La notion de systèmes et les modèles correspondants nécessitent d'appréhender de façon nouvelle des problématiques telles que la fiabilité ou la sûreté de fonctionnement.

A contre-courant du tout numérique, de nouvelles préoccupations sont apparues prenant en compte « l'humain ». La cobotique, (collaboration homme-robot), l'IHM (Interaction Homme-Machine) et les dynamiques d'opinion dans les réseaux sociaux intègrent par essence cette dimension humaine et font l'objet d'une attention toute particulière. Les systèmes intégrant l'humain sont par nature plus difficiles à modéliser. Proposer de nouveaux formalismes pour tenir compte de la variabilité des comportements, des incertitudes et des interactions, constitue des défis actuels.

Par ailleurs, toujours dans le contexte de l'ère du numérique, le déploiement massif d'unités de calcul embarquées dans des

systèmes autonomes, improprement dits « intelligents », fait naître de nouvelles problématiques. Celles-ci doivent en effet réaliser de très nombreuses tâches au-delà de leur fonction principale, comme garantir la sûreté de fonctionnement, les communications avec les équipements extérieurs et la sécurité tout en consommant peu. Par ailleurs, la loi de Moore édictée depuis plusieurs décennies et stipulant que la puissance des calculateurs numériques doublait tous les deux ans a récemment été remise en question par des acteurs majeurs de l'industrie des semi-conducteurs. La conception même des systèmes doit intégrer ces nouvelles contraintes de limitation de la capacité de calcul et de ressources.

2. Sur les données

Le développement récent de ce que l'on appelle Intelligence Artificielle est dû à l'abondance des données de tous types d'une part, et à l'avènement d'architectures matérielles efficaces pour le calcul massivement parallèle (GPU) d'autre part.

Par exemple, les progrès considérables en traduction automatique utilisant de l'apprentissage profond viennent de l'analyse de grands corpus récemment disponibles. Il en est de même pour la reconnaissance de visages ou pour les premières versions de la machine à jouer au Go. Les chercheurs sont confrontés aujourd'hui aux données massives. Si cette abondance est bénéfique dans certains cas comme l'apprentissage supervisé, elle pose des difficultés quand il s'agit d'archivage, de transmission et de traitement. Par exemple l'IoT, qui repose sur la 5ème génération des communications numériques en cours de développement, nécessite des capacités de canaux grandissantes puisqu'il génère des masses de données gigantesques. Les ressources énergétiques étant limitées (en raison d'un enjeu sociétal majeur), des questions fondamentales de traitements de données massives avec des ressources limitées se posent.

Une des raisons de l'accroissement des données est également la multiplicité des modalités d'observation d'un même système ou d'un même phénomène. Par exemple, le cerveau humain peut aujourd'hui être observé à l'aide de multiples techniques d'imagerie

(activité électrique et IRM, activité électrique interne et imagerie calcique...); les interfaces homme machine intègrent plusieurs sens (vision, ouïe et/ou toucher); les flottes mixtes de drones et de véhicules autonomes s'échangent des données partielles et multimodales issues d'une zone d'opération étendue. Ces multiples modalités offrent des perspectives nouvelles dans les traitements, comme par exemple l'accroissement de la résolution des images, ou dans les actions/interactions. Toutefois, l'exploitation conjointe de plusieurs modalités n'est pas toujours aisée et nécessite souvent la mise en oeuvre de techniques *ad-hoc*.

Par ailleurs, la nécessité d'obtenir des corpus annotés dans certaines disciplines à des fins d'apprentissage supervisé, comme en TAL ou en vision, ne peut se faire qu'au prix de coûts prohibitifs, et se doublant de questions éthiques importantes (violation d'intimité, esclavage numérique, biais de genre, impact écologique, domination culturelle, fracture sociale). De nouvelles méthodologies dites semi-supervisées sont actuellement en plein essor (apprentissage par transfert, augmentation de données) et permettent de réduire les coûts.

Beaucoup de questions actuelles reposent sur des données qui peuvent exister dans des espaces de très grandes dimensions. Cette possibilité est ouverte par l'existence de systèmes de mesures permettant de sonder de nombreuses caractéristiques des phénomènes observés (comme l'imagerie hyperspectrale), ou simplement par l'existence de multiples capteurs délivrant chacun une information pertinente (réseaux de capteurs en surveillance sismique ou en génie civil). Il n'est pas rare aujourd'hui de traiter des données évoluant dans des milliers voire des millions de dimensions (par exemple en génomique), et les espaces de grandes dimensions posent des problèmes difficiles souvent dûs à des comportements contre-intuitifs des données (par ex. des points dans un espace de très grande dimension sont essentiellement équidistants les uns des autres).

La dimension de l'espace d'états peut devenir infinie, par exemple lorsque les données considérées sont des courbes (mouvement d'un membre), des surfaces, ou des fonctions d'une variable d'espace. Enfin, les données peuvent vivre dans des espaces

ayant des structures très particulières (variétés différentielles, ensembles de graphes par exemple). La géométrie de l'espace peut alors être la source de l'information comme dans le traitement des mesures de nuages de points, ou demande à être élucidée pour traiter efficacement les données (réduction de dimension).

Les espaces géométriques de faible dimension nécessitent une attention particulière. C'est le cas en traitement de la géométrie lorsqu'on s'intéresse à la reconstruction de la topologie d'une surface à partir de données 3D échantillonnées sur un objet, lorsque l'on cherche à reconstituer ou classifier le mouvement d'un personnage, ou suivre une route avec un véhicule autonome. Un autre objectif de l'analyse topologique de données est d'explorer, d'identifier et d'exploiter la structure sous-jacente d'un jeu de données. Les données se présentent dans ce cas sous forme de nuages de points équipés d'une métrique ou de mesures de dissimilarité, appartenant à des espaces de grandes dimensions mais concentrés autour de structures géométriques inconnues que l'on cherche à identifier.

La nature géométrique complexe de l'espace contenant les données peut ou doit également être prise en compte, comme en géométrie de l'information où les grandeurs vivent dans des espaces structurellement riemanniens (comme par exemple les matrices de rotation et/ou de déplacement en robotique ou les matrices définies positives en sciences des données).

C. Méthodologies

Un constat fort sur l'état actuel de la recherche en sciences de l'information S07 est l'omniprésence de l'apprentissage et de l'optimisation dans toutes les disciplines relevant de la S07. Par ailleurs, les problèmes liés à la gestion des grandes dimensions et des grandes masses de données sont également présents dans la plupart de nos disciplines. Enfin, les problèmes d'interaction sont très prégnants également. Ce paragraphe est consacré aux méthodologies dans ces domaines et développées par les acteurs des sciences de l'information S07.

1. Apprentissage

La théorie de l'apprentissage est au coeur de la S07. Si des recherches théoriques restent à mener, elle a atteint aujourd'hui une maturité technique telle que son utilisation conjuguée au flot de données devient ubiquitaire dans nos disciplines. L'outil incontournable à l'heure actuelle est l'apprentissage profond, qui est utilisé dans de multiples applications. Toutefois, un effort considérable reste à mener pour le comprendre et le rendre utilisable avec des garanties de performance, dans des applications critiques (transports, santé, emploi, éducation...). Nous donnons dans la suite quelques éléments très actuels sur les recherches menées dans nos disciplines sur l'apprentissage en général, profond en particulier, et leurs applications.

Comprendre. Si la théorie de l'apprentissage statistique automatique a eu des avancées significatives dans les dernières décennies, les progrès récents de l'apprentissage profond et de certaines machines à noyaux résistent à l'analyse explicative. Le bon comportement en généralisation d'algorithmes appris sur peu de données est non intuitif. Des efforts dans diverses directions tentent actuellement de comprendre les mécanismes sous-tendant ces comportements. Par ailleurs, la théorie des grandes matrices aléatoires est un outil puissant avec lequel des progrès ont été réalisés, à la fois pour des explications du comportement de réseaux de neurones simples, mais également de machines à noyaux. D'autres approches fondées sur des transformées itérées (comme la transformée en scattering) qui ne sont pas apprises mais ont une structure proche des réseaux profonds (itérations de structures linéaire/non-linéaire) offrent des angles d'attaque intéressants pour comprendre le fonctionnement des réseaux profonds. Par ailleurs, des approches issues de la physique statistique sont de plus en plus utilisées dans la compréhension des algorithmes d'apprentissage, qu'ils soient profonds ou plus classiques, comme en clustering spectral.

Garantir. Les résultats spectaculaires dans de nombreux domaines d'application de l'apprentissage profond poussent leur

utilisation pour des applications concrètes immédiates dans divers domaines. Si tel est déjà le cas par exemple pour des jeux ou de la traduction, ces applications sont loin d'être possibles dans des domaines où la conclusion prise par le réseau doit présenter des garanties (comme la santé, le transport autonome ou l'énergie). L'absence de compréhension du processus d'apprentissage automatique ne permet pas le développement actuel de garanties conduisant à l'acceptabilité, ou expliquant avec certitude la décision prise par ces réseaux. Développer ces garanties avec un niveau de précision dépendant des applications est un chantier majeur autour de l'apprentissage profond. A cet égard, des développements récents dans les inégalités de concentrations, en optimisation stochastique ou en analyse fonctionnelle par exemple, sont des pistes pour l'obtention de ces garanties.

Apprendre. L'extraction d'information reste un objectif prépondérant de l'apprentissage statistique dans les travaux relevant de la S07. La résolution de problèmes inverses vus comme des problèmes statistiques est toujours d'actualité, ainsi que les techniques d'apprentissage par renforcement (étude de bandits stochastiques, prototypes d'études de prises de décisions séquentielles). Enfin, la maturité de certaines branches de l'apprentissage, par exemple les méthodes bayésiennes, autorise aujourd'hui les chercheurs à développer des méthodes de co-conception au plus près des applications, comme en imagerie biologique par exemple.

Agir. L'apprentissage concerne également la commande de robots, via le développement de méthodes de perception d'environnement complexes, de compréhension de scènes ou encore de développement de stratégies sensori-motrices pour des tâches complexes ou des structures difficiles à modéliser. La concrétisation physique inhérente à la robotique rend toutefois encore plus critiques les questions de garantie de résultat de ces algorithmes, de stabilité de la commande et de coût d'acquisition des données.

Maîtriser les ressources. Le contrôle et l'optimisation des ressources dans les systèmes intégrés sur puces repose en grande partie sur l'observation et l'analyse de divers événements. Ces derniers concernent les

parties logicielles et matérielles d'un système. L'application des techniques d'apprentissage hors-ligne ou en-ligne aux échantillons de données collectées permet de prédire des métriques de performance (par exemple, le nombre d'instructions par cycle processeur, la fréquence d'accès à la mémoire), de consommation énergétique, de température. Le résultat d'une telle prédiction sert ensuite dans la prise de décisions concernant l'allocation des ressources matérielles afin d'optimiser les comportements fonctionnels et non fonctionnels du système. Plus en amont, on peut souligner l'exploitation possible des techniques d'apprentissage en conception (choix de paramètres les mieux adaptés) ou en test (par ex. détection de configurations ou de comportements fautifs) de systèmes intégrés sur puce. L'augmentation du nombre de transistors par processeur (des dizaines de milliards aujourd'hui) provoque une forte inflation de l'espace de conception qui correspond à la quantité de degrés de liberté lors de la création et de la programmation d'un système. L'intelligence artificielle est donc maintenant utilisée également pour prendre les décisions complexes de cartographie et d'ordonnement nécessaires au portage d'un système sur un substrat matériel.

Implanter efficacement. L'apprentissage est quasiment présent à toutes les échelles des systèmes, des gros serveurs centralisés jusqu'à des noeuds de calculs légers comme dans l'IoT. La réalisation efficace des algorithmes d'apprentissage prenant en compte à la fois la précision, la performance et la consommation énergétique pose un défi de conception conjointe logiciel/matériel. Si les performances des GPU ont grandement contribué à la réalisabilité d'algorithmes d'apprentissage profond complexes, l'efficacité énergétique reste une préoccupation d'actualité quant à l'implantation de ces algorithmes. Parmi les pistes pour y répondre, on peut mentionner les travaux sur les architectures *manycore*, les architectures neuromorphiques (spiking neural networks et technologies émergentes à base de memristors), les architectures sur puce basées sur le paradigme de calcul approximatif, les architectures basées sur le paradigme de calcul en mémoire, les architectures hétérogènes, etc. Cela couvre

divers aspects liés à la conception, à la vérification et au test entre autres.

Ethique et biais. Les développements spectaculaires de l'apprentissage profond ont mis en lumière des questions éthiques fondamentales que les chercheurs ne peuvent laisser de côté. Certaines causes, comme les bases d'apprentissage restreintes ou le besoin d'annotation de grandes masses de données, conduisent à des questions éthiques relevant non seulement de nos disciplines mais également de la sociologie, de la politique et de la philosophie.

2. Gestion des grandes dimensions

Les applications actuelles traitées en S07 concernent souvent des systèmes et des données vivant en grandes dimensions, voire en dimension infinie. Par ailleurs, la gestion et le traitement des données devraient représenter près du cinquième de la consommation électrique mondiale en 2025. Le traitement efficace de grandes masses de données est donc un enjeu majeur.

Compresser et calculer. Le développement des objets connectés a entraîné la génération de données massives (issues notamment de capteurs), accompagnées d'une quantité très importante de communications. L'impact sur les coûts en performance et en énergie des traitements et transmissions de données est fort. Pour y répondre, de nouveaux paradigmes voient le jour, à l'image de l'*edge computing*, ou il s'agit d'intégrer des infrastructures de calcul intermédiaires entre les supports périphériques tels que les capteurs et des unités de traitements centralisées (serveurs puissants) vers lesquels les données sont communiquées pour être traitées. Par exemple, en considérant une infrastructure formée par quelques machines légères proches des capteurs, des données collectées peuvent être pré-analysées au plus près des capteurs (sans solliciter les serveurs) pour éventuellement prendre des décisions (par exemple, actionnement ou contrôle). Les coûts de traitements et de transmission des données sont ainsi réduits. La conception de systèmes intégrés sur puce efficaces en performance et en énergie, adressant des paradigmes tels que le calcul à la périphérie représente ainsi un défi important pour la

gestion des données massives dans l'internet de ces objets connectés.

Dans ces approches décentralisées, pour des environnements contraints en énergie (où stockage et moyen de calculs sont limités), des stratégies de calculs efficaces nécessitent de ne conserver que les données pertinentes, ou des versions compressées des données. Les notions de *core-sets* développées initialement en traitement de la géométrie (sous-ensemble des données permettant des calculs à performances plus faibles mais garanties) ou de *sketching* permettent d'élaguer les données et de n'en conserver qu'un résumé très informatif. Ces techniques sont proches de la théorie de l'échantillonnage compressé. Une alternative à la conservation des données est le calcul en ligne qui permet d'extraire l'information sans conserver les données. Les algorithmes en ligne très en vogue dans la dernière décennie du 20ème siècle reviennent actuellement sur le devant de la scène comme moyen de calculs en ces situations contraintes.

Distribuer. Distribuer les calculs est inhérent à la gestion des données dans les réseaux de capteurs ou les flottes de robots. De nombreuses méthodologies (classification, commande, estimation, optimisation) ont des versions distribuées pour pouvoir être adaptées efficacement aux situations appelant la distribution.

Parcimonie. La grande dimensionnalité des données est souvent une difficulté mais parfois un leurre. Dans de nombreuses applications, la dimension effective de l'information est beaucoup plus petite que celle des données dans laquelle elle est cachée (par ex. variation d'une image sous l'effet du déplacement de la caméra, représentations éparses de signaux ou d'images dans des bases adéquates). Les techniques de réduction de dimension utilisent cet *a priori* et sont dessinées de façon *ad-hoc*.

Retour vers le continu. Le comble de la grande dimension est la dimension infinie qui appelle des approches particulières, relevant de l'analyse fonctionnelle. Par exemple, en automatique, de nombreux problèmes ne peuvent se modéliser qu'en envisageant des systèmes de dimension infinie, comme le

contrôle de phénomènes vibratoires ou flexibles, le contrôle aux frontières d'écoulements fluides, ou la prise en compte des retards dans les systèmes. Certaines applications en traitement du signal et des images nécessitent aussi des modélisations continues et de dimension infinie, comme en théorie du krigeage ou pour la modélisation de trajectoires complètes.

3. Optimisation

L'optimisation est mise en oeuvre pour l'estimation, la prise de décision et l'extraction d'informations des données. L'optimisation est une discipline qui va jusqu'aux mathématiques (analyse convexe, calcul des variations), et la recherche opérationnelle (planification industrielle, allocation de ressources) et qui imprègne aujourd'hui fortement les sciences de l'information S07.

Des méthodes modernes d'optimisation (éclatement proximal, activation aléatoire de variables et opérateurs) implémentées sur des architectures séquentielles, parallèles ou distribuées, sont développées par les chercheurs en sciences de l'information en lien avec les applications. L'exploitation par ces méthodes des propriétés de parcimonie (ou de rang faible) des solutions recherchées joue un rôle majeur dans leur efficacité. Les approches par factorisation sous contraintes de matrices ou tenseurs, ou encore par représentations parcimonieuses dans des dictionnaires, ont un vaste champ d'applications, de la génomique aux neurosciences, en passant par la géophysique, la finance, la télédétection ou l'étude du climat.

L'optimisation stochastique, résultant du couplage de procédures d'optimisation déterministes et de l'aléatoire, revient très fortement sur le devant de la scène de par ses applications en apprentissage et en intelligence artificielle, notamment pour l'entraînement des réseaux de neurones profonds ou comme modélisation de l'apprentissage séquentiel. L'aléa résulte souvent soit de la procédure d'acquisition des données ou observations (dans le cas d'apprentissage séquentiel), soit de procédures d'échantillonnage Monte Carlo introduites comme outil numérique pour approcher des quantités incalculables (par

exemple, dans la sélection d'un petit nombre d'échantillons dans une grande masse trop coûteuse à manier dans sa globalité). Il s'agit ainsi d'estimer des paramètres minimisant une certaine fonction de perte en moyenne ou au pire cas dans une classe d'intérêt, à partir de données où les incertitudes règnent.

Un terrain fertile pour l'optimisation, à la fois comme champ d'application et comme source de nouvelles approches, est la résolution de problèmes inverses régularisés de restauration et reconstruction d'images. Dans les problèmes de maximisation d'un critère d'attache aux données avec régularisation, les dernières avancées numériques mêlent des notions issues de l'optimisation à des concepts de simulation stochastique de type Monte-Carlo.

Par ailleurs, l'optimisation joue un rôle majeur dans le champ de l'automatique. Il est par exemple devenu standard de considérer une question comme résolue si elle se formule comme la solution d'un problème d'optimisation convexe, et presque résolue si elle est associée à des outils d'optimisation (éventuellement sur des fonctions non lisses, non convexes) efficaces en pratique. Cette évolution est marquée par une utilisation très importante en automatique de la programmation semi-définie (formulations par les inégalités matricielles linéaires) pour l'analyse des propriétés des systèmes dynamiques, et par le développement de logiciels pour la synthèse de correcteurs structurés.

A côté du progrès des algorithmes, sur le volet de la modélisation, le développement de relaxations convexes de problèmes non convexes de complexité combinatoire constitue un champ d'étude dynamique et les hiérarchies de Lasserre sont une avancée remarquable du domaine.

On observe aussi un regain d'intérêt pour l'application de méthodes de l'automatique à l'étude des algorithmes d'optimisation en eux-même (par exemple obtention récente des vitesses de convergence des méthodes de gradient accéléré de Nesterov par les outils théoriques de la dissipativité). C'est un retournement étonnant que les inégalités matricielles linéaires trouvent les vitesses de convergence des algorithmes utilisés pour les résoudre ! Mieux, ces outils peuvent permettre la synthèse de nouveaux

algorithmes d'optimisation aussi simplement que l'on fait la synthèse de lois de commande pour les systèmes dynamiques.

Enfin, le transport optimal connaît un intérêt croissant ces dernières années. Alors que les méthodes numériques disponibles ont longtemps cantonné cette théorie à la manipulation d'ensembles de milliers d'éléments au plus, des avancées majeures récentes, par exemple avec la régularisation entropique des problèmes, permet le passage à la grande échelle. Ainsi, il est maintenant possible d'appareiller rapidement deux nuages de millions de points, en minimisant le coût de transport de l'un à l'autre. Le transport optimal permet également de calculer des moyennes, d'interpoler, ou de classifier des formes géométriques, des visages dans des images, des textes avec des représentations en sacs de mots.

4. Interactions

Les interactions et leur étude ont une large place dans l'ensemble des disciplines de la S07, mais sont au coeur de quelques thématiques.

En IHM. L'Interaction Homme-Machine (IHM) est la discipline qui se consacre à la conception, l'évaluation et la mise en œuvre de systèmes informatiques interactifs pour l'usage humain ainsi qu'à l'étude des phénomènes majeurs qui les entourent.

Un des objectifs des recherches en IHM est donc de créer des systèmes interactifs adaptés aux capacités humaines. Les chercheurs étudient, développent et évaluent différentes formes de partenariats dynamiques humain-machine, et notamment de partenariats humain-IA, adaptées aux utilisateurs grâce à leurs capacités de personnalisation, d'adaptation dynamique et d'apprentissage. L'humain et la machine sont donc considérés ici comme un système complexe que l'on cherche à modéliser, à comprendre et pour lequel il est nécessaire de mener des recherches à de multiples niveaux : méthodes et cadres de conception, techniques d'interaction, protocoles d'évaluation. Ces recherches concernent aussi les données produites par ces interactions qu'il convient d'analyser et qui peuvent être complexes puisqu'elles concernent des contextes très variés : plusieurs humains interagissant avec

plusieurs machines dans divers environnements et lors d'interactions répétées et continues au cours du temps (par exemple permettre à un utilisateur d'interagir avec ses données personnelles de santé issues de multiples objets connectés).

Parmi les thèmes de recherche actuels, on peut par exemple mentionner la conception de techniques d'interaction sans contact, l'intégration de multiples modalités de communication, la continuité tangible virtuel augmenté, les capacités d'interactions sociales et émotionnelles de la machine, les technologies motivationnelles, ou encore la visualisation et les interactions immersives grand écran.

En TALP. Le traitement automatique de la langue et de la parole occupe une place intéressante au sein des travaux en interaction humain-robot ou plus largement IHM. En effet, la parole et la langue sont des vecteurs de communication pour l'humain et les utiliser lors d'une interaction est naturel, rendant efficace la communication. Toutefois, les interactions en langue soulèvent des défis importants comme le traitement temps réel, l'adaptation à de multiples locuteurs ou utilisateurs, à la situation en cours, ainsi que la prise en compte dans l'analyse des énoncés des utilisateurs de l'ensemble de l'interaction (temporalité mais aussi modalité). De tels systèmes d'interaction vont mobiliser notamment des travaux sur les architectures hybrides permettant de tenir compte des différents niveaux de communication mis en jeu dans un modèle (alignement de modalité, modèle utilisateur et modèles statistiques). La multimodalité nécessaire à l'interaction humain-robot en langue implique également des modèles de génération et des modèles de comportements qui ont un impact fort sur l'interaction elle-même. L'ensemble de ces modèles doit répondre à la contrainte forte du temps réel.

En robotique. On peut distinguer trois formes principales d'interaction concernant les systèmes que sont les robots. Leur interaction avec un environnement dynamique est un sujet historique (boucle perception-action) mais qui reste très actuel, que ce soit pour la préhension d'objets déformables, le positionnement précis

d'instruments médicaux, le micro-assemblage 3D sur fibres optiques, la navigation de véhicules autonomes urbains ou de drones de maintenance. En particulier, l'interaction avec le vivant (santé, biologie, agriculture) pose le problème fondamental d'une variabilité (temporelle, spatiale, structurelle) difficilement modélisable. La deuxième forme est l'interaction humain-robot, déjà évoquée plus haut, dans lequel le partage du contrôle, les garanties de stabilité, le respect de l'intégrité et de l'éthique sont des problèmes ouverts. La troisième forme est l'interaction entre robots, qui mobilise communication et coopération. L'interaction multi-robots, par passage à l'échelle, implique de surcroît stratégie de coordination, multi-modalité de perception, hétérogénéité des flottes, répartition et efficacité des calculs, modération énergétique, ou encore robustesse aux erreurs (capteurs, communication, calculs).

Ces interactions nécessitent l'utilisation ou le développement de capteurs et d'observateurs en lien étroit avec la conception mécanique, la commande, les communications et les architectures.

En informatique graphique et réalité virtuelle. L'activité de recherche autour du rendu d'une scène 3D peut être inscrite en amont de l'activité d'interaction, au sens où elle construit et donne à voir des contenus visuels à partir de modèles géométriques (maillages) en prenant en compte des modèles parfois complexes d'illumination. Ces algorithmes de rendu sont développés en s'appuyant sur les capacités des dernières générations de cartes graphiques, en exploitant des stratégies d'échantillonnage des scènes ou de compression ad-hoc, ainsi qu'une modélisation de la lumière et de son interaction avec la matière. Pour permettre l'interaction d'un ou plusieurs utilisateurs avec de telles scènes virtuelles (ou augmentées), il est nécessaire de satisfaire des contraintes temps réel, associées ou non à des contraintes de photo-réalisme. Ces contraintes sont des plus ardues à satisfaire lorsque les scènes sont en mouvement. Une des grandes tendances de ce domaine, notamment dans l'industrie du jeu vidéo et de l'animation, concerne la génération procédurale de scènes complexes.

Par ailleurs, le domaine de la visualisation scientifique est en plein essor. L'idée est de coupler l'analyse de données pouvant appartenir à des espaces de grande dimension avec le développement de stratégies de génération d'images ou de projections, pour appréhender et gérer des masses de données, avec lesquelles le scientifique peut alors interagir.

II. Enjeux sociétaux

Les sciences de l'information développées en S07 sont au cœur de la société numérique. Elles contribuent intensément à de nombreux enjeux sociétaux. Cette section illustre des enjeux actuels pour lesquels les disciplines de la S07 apportent des avancées et des solutions.

A. Avancée des connaissances

Les succès récents et retentissants de l'astrophysique (imagerie du fond diffus cosmologique, mesure des ondes gravitationnelles, première image d'un trou noir) n'ont pu être obtenus sans des apports fondamentaux de traitement du signal et des images et l'utilisation de techniques d'apprentissage et d'optimisation puissantes et développées au plus près des expériences. La France est d'ailleurs partie prenante dans les deux premiers exemples avec des contributions majeures de membres de la S07. Par ailleurs, la théorie du contrôle a permis des exploits lors de la mission Rosetta, avec la première mise en orbite d'un satellite artificiel autour d'une comète. Le problème de la re-calibration rapide en vue de la synthèse efficace de contrôleur sur un modèle identifié au moment de l'approche du lander chargé d'expériences était un réel enjeu.

B. Société numérique

Près de la moitié des habitants de la planète possèdent aujourd'hui un smartphone et les réseaux sociaux comptent plus de trois milliards d'utilisateurs. En conséquence des évolutions récentes des technologies de l'information et de la communication, les frontières entre monde numérique et monde

réel se font moins évidentes et plus complexes. Les aides à la mobilité changent notre rapport à l'espace urbain. Les réseaux sociaux influent sur nos relations inter-individus, les systèmes de notation changent nos habitudes d'achat, la compréhension du langage naturel par les machines rend la communication avec les systèmes plus aisée et plus efficace et les robots font de plus en plus partie de notre quotidien.

Ces nouveaux usages des technologies numériques ajoutent de nouveaux défis aux problèmes traditionnels liés au traitement de très larges masses de données :

- l'utilisation de la plupart des systèmes numériques nécessite aujourd'hui de transmettre ses données à un tiers et de lui faire confiance sur leur non-divulgaration, posant la question fondamentale de la sécurité informatique ;
- l'accès permanent au réseau et au monde numérique est considéré comme un acquis depuis le déploiement de la 4G en 2010. La 5G devrait pousser plus l'ubiquité du numérique, regroupant en un standard tous les types d'interactions entre systèmes : à latence contrôlée, par exemple pour coordonner des véhicules, à faible débit et basse consommation pour interagir avec des objets connectés, et à haut débit pour échanger des données de grandes tailles telle la vidéo ;
- l'impact écologique croissant des systèmes se décline selon deux grandes modalités : l'énergie consommée par le calcul, le transport et le stockage des données (en embarqué ou dans le cloud) ; l'obsolescence rapide des systèmes associée au coût environnemental de leur recyclage ;
- l'accès aux données, briques nécessaires à l'apprentissage des modèles, soulève des questions d'éthiques, par exemple celle du respect de la vie privée ;
- l'interaction désormais physique avec l'humain amplifie les risques mais réveille aussi quelque peu les consciences quant à l'usage des technologies.

Les évolutions des systèmes annoncées pour les prochaines décennies incluent une multiplication des capteurs et actionneurs déployés dans l'IoT, une immersion possible dans un monde numérique ou de réalité augmentée, et une intelligence accrue des systèmes automatisés sous la forme d'une adaptation aux besoins des utilisateurs.

Toutes ces évolutions renforcent les défis cités précédemment.

Les disciplines de la S07 sont au cœur de l'évolution de nos comportements vis à vis des technologies numériques. Le traitement en-ligne de quantités de données, la compréhension et la traduction automatique des langues, les systèmes sur puce à basse consommation d'énergie, ainsi que les dispositifs d'interaction homme-machine participent à l'évolution des pratiques dans la vie privée et professionnelle, dans la pratique des sciences, des arts (l'art numérique est devenu un domaine à part entière), ou de l'enseignement.

Le développement extrêmement rapide et varié de l'apprentissage, déploiement dû à la conjonction de méthodes de calcul matériel-logiciel efficaces, d'exploitation de larges bases de données annotées, et du développement itératif d'algorithmes neuro-inspirés, rend difficilement prévisibles les impacts sociétaux à court terme des sciences de l'information. Cependant, il est incontestable que ces impacts seront importants et que les problèmes posés nécessiteront une recherche agile, performante et réfléchie.

C. Santé

La santé est une application transverse pour tous les thèmes méthodologiques de la S07. Les travaux actuels répondent à des enjeux sociétaux importants comme la gestion du vieillissement de la population (par exemple, l'étude de la maladie d'Alzheimer, le développement de prothèses ou l'assistance domotique), les recherches sur le cancer, ou encore la gestion de données médicales, soulevant actuellement des questions d'éthique et de sécurité. Les méthodes développées au sein de la S07, notamment liées à l'intelligence artificielle, ouvrent de nombreuses perspectives pour l'amélioration de la qualité des soins, la réduction de leur coût à travers une prise en charge plus personnalisée et prédictive, mais également une meilleure traçabilité et une aide à la décision médicale améliorée. Dans ce contexte, le patient et son environnement (médecins, soignants, hôpital, famille) forment un écosystème complexe devenant le cadre de recherches en santé. On peut aussi souligner la variété des données à étudier : les

données cliniques, le dossier patient, les données d'imagerie ou encore les données génomiques.

Un défi concerne l'analyse et la compréhension du vivant dans le but d'améliorer la prise en charge des patients. Il s'agit par exemple de tenter de modéliser un phénomène physiologique par un modèle numérique construit à partir de données *in-vivo* permettant d'effectuer des simulations par ordinateur afin de confirmer ou infirmer une hypothèse sur le fonctionnement du corps humain. Cela repose sur le développement de systèmes d'acquisition de données *in-vivo* rendu possible par un rapprochement entre des disciplines comme la physique et les sciences du numérique. On assiste ainsi à l'amélioration de techniques comme l'imagerie par résonance magnétique, la tomographie spectrale ou la microscopie super-résolue et l'émergence de nouvelles approches comme l'imagerie ultrasonore ultrarapide ou à trois photons.

Les méthodes numériques développées permettent des avancées majeures dans l'aide au diagnostic et au geste thérapeutique (chirurgie et rééducation). Des outils d'aide au diagnostic performants sont développés grâce aux méthodes d'apprentissage statistique et la constitution de grandes bases de données, dans des domaines variés comme l'ophtalmologie, la dermatologie, la mammographie ou encore la cardiologie. Concernant le geste thérapeutique, dans le cas de la chirurgie, on parle alors de « cockpit chirurgical » où l'assistance peut intervenir en amont par un planning opératoire guidé par les données 3D du patient, et en per-opératoire par la chirurgie mini-invasive, la chirurgie robotique ou la réalité augmentée. En rééducation, de nouvelles approches se développent pour proposer une thérapie personnalisée au patient prenant en compte ses progrès fonctionnels grâce à un suivi contextualisé. La robotique a également des apports importants pour la rééducation fonctionnelle de certains patients paraplégiques.

Les avancées en TAL ou en IHM notamment ont permis le développement d'outils pour un meilleur suivi des patients par le biais d'agents conversationnels dotés de capacités langagières grandissantes. Enfin, la formation médicale est un point crucial et les méthodes de simulations et de

développements de patients virtuels (permettant par exemple à l'étudiant en médecine d'acquérir les bonnes techniques d'anamnèse) ne cessent de progresser.

D. Climat/environnement

La robotique se met actuellement au service de l'environnement, de sa préservation et de sa gestion. L'impulsion lancée autour de la robotique agricole au profit de l'environnement et de l'individu est notable pour :

- soulager l'agriculteur en créant des robots d'assistance logistique (cobots "mules" porteurs-suiveurs) ou de désherbage automatique afin d'éviter l'utilisation de produits néfastes à l'environnement.

- aider l'agriculteur à prendre des décisions en collectant des données (cartographie des cultures afin de définir le volume de végétation à traiter ou le type d'adventices avant pulvérisation afin de diminuer la quantité de pesticides/fongicides à utiliser ; mesurer la qualité et la quantité d'herbe dans les parcelles pour aider l'exploitant à ajuster et déplacer le bétail). Les outils de traitement d'images automatiques sont dans ce cadre de plus en plus utilisés avec des types d'imagerie variés allant de la "simple" photo RGB prise par smartphone aux images issues de caméras hyperspectrales embarquées sur le tracteur ou sur des drones.

La robotique au service de l'environnement ne se limite pas à l'agriculture. On retrouve des drones dans le cadre de la surveillance des incendies de forêt, afin de détecter au plus vite les dépôts de feu dans les régions à risque et aider les sapeurs-pompiers dans leur tâche. Les drones sont également utilisés pour effectuer la surveillance routière. Combinés à du traitement vidéo, ils permettent d'envisager une compréhension fine de la circulation pour éviter les congestions et par conséquent la pollution. Il est aussi envisagé d'utiliser des robots pour aider au nettoyage des environnements pollués par l'homme (collecte des déchets du continent de plastique dans le Pacifique par exemple). Enfin, des robots ont été récemment utilisés afin de protéger la grande barrière de corail : le sous-marin autonome LarvalBot a pour but de repeupler les récifs coralliens en larguant des larves de coraux.

E. Transports

L'utilisation du véhicule autonome, bardé de capteurs et capable de percevoir son environnement, de l'interpréter et d'y évoluer, se développe. Si la plupart des véhicules pleinement autonomes évoluent actuellement à basse vitesse, les tests associés aux navettes autonomes pour les transports en commun se multiplient dans les grandes villes (Lyon, Paris, Nantes). A Paris, la navette circulant dans le quartier de la Défense a déjà déplacé près de 40000 usagers.

Si tous les grands constructeurs automobiles se sont emparés du sujet, il en reste néanmoins de nombreux enjeux en matière de recherche amont. Malgré des avancées spectaculaires, il y a encore de nombreux progrès à faire en matière de garanties de sécurité et les enjeux sont considérables au regard des contraintes de multimodalité et de quantité d'informations variées à traiter en temps réel. Il faut également noter que les véhicules autonomes ne seront pas utilisés que dans le contexte routier, mais également en contexte industriel, agricole ou militaire.

F. Energie

La question énergétique occupe aujourd'hui une place primordiale pour les sociétés modernes aux besoins croissants. Au centre des préoccupations majeures, on retrouve la réduction de l'impact environnemental lié à la production de l'énergie (notamment l'émission de gaz à effets de serre), de sa distribution et de sa consommation sous diverses formes (électricité, gaz, essence). La transition énergétique vise à apporter une réponse en promouvant les énergies renouvelables à la place des énergies fossiles avec pour objectif une meilleure indépendance énergétique réduisant les coûts de consommation. Néanmoins, l'exploitation efficace sources variées d'énergies renouvelables est un réel défi qui nécessite des technologies dédiées au stockage optimal de l'énergie et à sa gestion intelligente, en tenant compte de la variété des modes de consommation. Au coeur des enjeux des villes intelligentes figure la maîtrise de la demande et de la planification énergétiques des territoires (à travers les smart-grids notamment). Les technologies

numériques jouent un rôle crucial dans ce contexte : avec d'une part l'intégration de systèmes embarqués mis en réseaux (IoT), permettant ainsi de collecter des données du terrain pour prendre des décisions, économiser de l'énergie ou réduire les émissions de gaz à effet de serre, et d'autre part l'enjeu complexe du contrôle de réseaux électriques très hétérogènes et dispersés avec la multiplication de sources d'énergie variées.

Par ailleurs, il est important de garder à l'esprit l'impact délicat des technologies numériques au travers de la transition numérique. Un rapport¹ du groupe de travail « Lean ICT » dresse entre autres les constats suivants : la transition numérique augmente de 9% par an l'empreinte énergétique directe du numérique (comprenant la fabrication et l'utilisation des équipements numériques) ; au niveau mondial, l'intensité énergétique de l'industrie du numérique augmente de 4% par an alors que celle du PIB diminue de 1,8% par an. Ce constat appelle à questionner d'une part les modes de fabrication des équipements numériques et d'autre part les usages au sein de nos sociétés.

G. Défense/Sécurité

La mise en place de traitements vidéo presque temps réel et avec des performances satisfaisantes permet désormais des applications en vidéosurveillance. Par exemple en surveillance de l'environnement pour la prévention des crues grâce au comptage des bois flottants sur les rivières ou le monitoring de la circulation grâce à la détection et au suivi des véhicules, la protection des personnes par le biais de leur détection et de leur suivi, éventuellement dans des foules, ou encore la gestion et l'estimation du temps d'attente dans des files dans des supermarchés ou des aéroports.

L'analyse faciale, qui joue un rôle important dans la communication humaine et qui est une des meilleures biométries pour l'identification, fait l'objet d'énormes efforts de recherche. Plus récemment, les recherches dans ce domaine sont également axées sur l'analyse de gestes et de comportements avec

des applications notamment dans le milieu médical, par exemple pour le maintien à domicile des personnes âgées ou pour le suivi de traitements en lien avec des maladies cognitives. L'avènement des méthodes d'apprentissage profond et des bases de données publiques de plus en plus nombreuses dans ces domaines permettent d'envisager que l'on pourra prochainement comprendre et prévoir le comportement d'une personne par analyse vidéo, levant à nouveau des questions éthiques importantes.

H. Industrie du Futur

Les robots sont utilisés de longue date dans l'industrie, mais sont généralement cantonnés à des zones spécifiques que les humains ne peuvent pas pénétrer. Une nouvelle génération de robots dits collaboratifs, ou cobots, intègre le monde industriel depuis le début des années 2010. Le chiffre d'affaire associé aux ventes de ces robots est estimé à 1,23 milliard de dollars pour 2025. Ces robots sont conçus pour être intrinsèquement sûrs pour les humains, afin de ne pas blesser un opérateur en cas de contact accidentel, et doivent donc être capables d'interagir et de collaborer avec eux en toute sécurité.

Les exosquelettes sont d'autres exemples de systèmes mécaniques automatisés que l'on rencontre de plus en plus en contexte industriel. Ces mécanismes se greffent sur certaines parties du corps de l'utilisateur afin de lui conférer une force accrue ou de lui permettre de supporter des charges trop lourdes pour lui, évitant ainsi l'apparition précoce de troubles musculo-squelettiques. Le besoin industriel de reconfigurabilité des chaînes de production amène à l'utilisation de robots industriels montés sur bases mobiles ou sur rails, robots capables de changer rapidement de zones opérationnelles en fonction des besoins. Un exemple marquant est le robot du projet Batiprint3D / Yhnova qui est capable d'imprimer des maisons en 3D, et a construit la première maison imprimée en 3D à Nantes en 2018 (95 m², 5 pièces).

Les interfaces homme-machine impactent aussi l'industrie actuelle. Elles permettent

¹ "Pour une sobriété Numérique", <https://theshiftproject.org/wp-content/uploads/2018/11/Rapport-final-v8-WEB.pdf>.

d'interagir avec les machines et d'optimiser des processus industriels complexes. Les chercheurs visent à concevoir des interactions avec des dispositifs industriels en prenant en compte l'humain, dans le but de surveiller les processus, diagnostiquer les problèmes, ou visualiser des données industrielles.

La réalité augmentée, les interactions tangibles et les robots sociaux se développent aussi dans des contextes industriels à des fins de contrôle qualité, d'apprentissage pour l'opérateur, de gains de productivité et de sécurité.

III. Place de la recherche française

La recherche française en sciences de l'information est située au meilleur niveau européen et international. En témoignent la participation dans toutes les disciplines de nombreux acteurs français aux activités de sociétés savantes prestigieuses (plusieurs chapitres des IEEE, IFAC, ISCA, ACMI, ...) et de nombreuses reconnaissances de type *fellow*. Les acteurs français sont très actifs dans l'organisation, sans discontinuer et toutes disciplines confondues, d'évènements internationaux majeurs tels que DATE, INTERSPEECH, ICML, IFAC World Congress 2017, ICRA 2020, RoboCup 2020, ou de *workshops/challenges* plus spécialisés. Enfin au niveau européen, l'obtention de nombreuses bourses ERC par des chercheurs du périmètre S07 atteste de la dynamique de notre recherche.

CONCLUSION

La visibilité et la proximité des disciplines couvertes par notre section scientifique avec les enjeux sociétaux modernes nous impose une réflexion forte sur les orientations et dérives possibles des applications de nos recherches. Cependant, une fois ce constat effectué, la question de la mise en pratique de cette réflexion, en particulier éthique, par les chercheurs eux-mêmes autant que par leurs institutions, reste une gageure. Nous n'allons pas ici lister les divers comités éthiques qui font un réel

travail de dossier et de terrain, mais nous souhaitons évoquer la nécessité de l'ouverture de chacun à la réflexion transdisciplinaire (vers la philosophie, les sciences humaine, l'épistémologie...), socle d'une prise de recul responsable sur les avancées scientifiques qui émanent de nos disciplines.

Les sujets plus précis sur lesquels engager ce travail de réflexion sont nombreux. Plusieurs d'entre eux sont en particulier listés dans un document de synthèse produit par le Conseil Scientifique de l'INS2I en Septembre 2018 autour des algorithmes. Il est également important de soulever des questions légitimes sur les bonnes pratiques de recherche, la reproductibilité des travaux et le partage des données, l'impact écologique de nos productions et des comportements sociaux (comme chercheur, comme utilisateur futur). Et par la suite, comment mettre en place des mécanismes pour réfléchir et gérer au mieux la boucle à évolution rapide liant science et société ?

Ce sont ici simplement quelques éléments de réflexion éthique que le fonctionnement actuel de la recherche Française permet malheureusement assez peu. Il est ainsi souhaitable que les laboratoires de recherche, comme le CNRS, permettent et encouragent tout travail en ce sens.

Remerciements

Les membres de la section remercient chaleureusement les acteurs de la recherche française qui ont contribué à l'élaboration de ce rapport, en particulier les groupements de Recherche IGRV, IM, ISIS, MACS, MAGIS, MEGA, Robotique, Sécurité, SOC2, TAL, ainsi que l'AFIHM.

ANNEXE 1 : Sigles

S07: Section 7
IHM: Interfaces Homme-Machine
SOC: Systèmes sur puce (Systems-on-chips)
TALP: Traitement Automatique des Langues et de la Parole
IoT: Internet des Objets (Internet of Thing)

ANNEXE 2 : la Section 07 en chiffres

En septembre 2019, la section 07 compte 285 chercheurs actifs (60 femmes (F), 225 hommes (H)). Le corps des chargés de recherche compte 156 membres (37 F, 119 H), dont 10 Chargés de Recherche Hors-Classe. 71 personnes sont DR2 (13 F, 58 H), 46 sont DR1 (8 F, 38 H) et 12 sont DRCE (2F, 10 H). La proportion de femmes dans la section est de 21%. Un léger déséquilibre s'observe entre les CR (près de 24% de femmes) et les DR (19% de femmes) et la S07 est particulièrement attentive aux questions de genre dans l'ensemble des travaux qu'elle mène.

La grande majorité des chercheurs sont affectés dans des unités (principalement mixtes) de recherche de l'INS2I ou de l'INSIS. 26 sont affectés à des laboratoires rattachés à d'autres instituts (INSMI, INSB, INSHS, INP, INEE). Thématiquement, 53 chercheurs travaillent sur l'Automatique, 43 en Traitement du Signal, 41 en Traitement des Images, 24 en Informatique Graphique/IHM/RV, 55 en Traitement de la Parole, des Langues, de l'Audio, 39 en Robotique, 18 en Architecture Matérielle et SOC, en enfin 23 en sciences de l'information pour les sciences du vivant.

Notons que les thématiques de la S07 couvrent des thématiques des sections 27 et 61 du CNU. Les UMR de l'INS2I relevant de la S07 sont ainsi essentiellement peuplées d'enseignants chercheurs de ces sections, et dont les travaux de recherches sont couverts par ce rapport.