

Projet scientifique issu des travaux des GT sur la fusion LIMSI+LRI-VALS – Juin 2019

Préambule

L'objectif ambitieux du projet scientifique présenté ici est d'aboutir à un centre de recherche pluridisciplinaire d'envergure, couvrant un large spectre de recherches fondamentales et appliquées dans les domaines des Sciences du Numérique et de l'Informatique, et des Sciences de l'Ingénieur, autour de la mécanique des fluides et de la mécanique énergétique.

Ce centre couvre à la fois des thématiques au cœur de la discipline informatique (algorithmes, données, connaissances, systèmes, réseaux, calculs ...) et des thématiques aux dimensions interdisciplinaires comme l'intelligence Artificielle, la science des données, l'interaction humain-machine, le traitement automatique des langues et de la parole, la bio-informatique et la mécanique des fluides et de l'énergétique. Du fait de son rayonnement, il pourra aussi agréger, à plus ou moins long terme, d'autres équipes ou unités dans son périmètre thématique.

Il rassemble des chercheurs et enseignants-chercheurs de cultures et d'horizons divers, relevant des Sciences du Numérique et des Sciences de l'Ingénieur, mais aussi des Sciences Humaines et Sociales, des Sciences Cognitives et des Sciences du mouvement, et entretient des liens étroits avec les autres disciplines de son environnement proche (mathématiques, physique, biologie, santé).

Ce centre de recherche relève principalement des sections 06, 07 et 10 du CNRS et des sections 27, 60 61 et 62 du CNU. Il comporte aussi des chercheurs en sciences du langage (section 34 du CNRS), en sociologie et sciences du droit (section 36 du CNRS) et des enseignants-chercheurs en psychologie (section 16 du CNU). Ce centre est rattaché principalement à l'INS2I et secondairement à l'INSIS, à l'INSHS et à l'INSB au CNRS et fera partie des Graduate Schools « Informatique et Sciences du Numérique » et « Ingénierie et Systèmes » de l'Université Paris-Saclay.

Dans le projet tel qu'il est envisagé actuellement, les équipes ou groupes des deux laboratoires actuels, LRI (sans VALS) et LIMSI, sont conservées. Leurs projets scientifiques, décrits dans leurs rapports respectifs à l'HCERES, restent d'actualité. Néanmoins, un travail important de regroupement de ces équipes en départements a été initié au cours de cette dernière année et a permis de faire émerger cinq départements décrits ci-après qui positionnent le centre de recherche de façon unique. Les départements Sciences et technologies des langues et Interaction avec l'humain font partie des plus importants pôles d'enseignement et de recherche dans leurs domaines respectifs aux niveaux national, européen et international. Les départements Sciences des Données d'une part et Algorithmes, Calculs et Apprentissage d'autre part montrent de fortes complémentarités avec une largeur de spectre thématique originale dans le paysage scientifique offrant ainsi des opportunités rares de s'attaquer aux nouveaux défis scientifiques de ces domaines. Enfin, la spécificité du département Mécanique est de développer une large gamme d'approches méthodologiques innovantes pour la simulation, l'analyse, l'estimation et le contrôle des écoulements. Ces méthodes sont mises en œuvre en conjonction avec théorie et expérimentation dans le contexte des Sciences de l'Ingénieur, avec une diversité de thématiques autour de la dynamique des fluides, des transferts de masse et de chaleur, des couplages multiphysiques et de l'énergétique.

Les travaux de recherche du centre visent non seulement à l'avancement des connaissances, mais aussi à

répondre à des défis sociétaux multiples de la Stratégie Nationale de Recherche (SNR). Les domaines applicatifs principaux du département Mécanique concernent l'efficacité énergétique (Défi 2, « Un énergie propre, sûre et efficace ») et la conception de moyens de transports plus sûrs et moins gourmands en ressources (Défi 6, « Transports et systèmes urbains durables »). Les autres départements s'inscrivent principalement dans le Défi 7 (« Société de l'information et de la Communication »), dont les quatre Orientations sont couvertes : les départements Sciences des Données, Sciences et technologie des langues, et Algorithmes, Calculs et Apprentissage abordent tous le traitement automatisé de grandes masses de données ; le département Sciences des Données traite également des réseaux de cinquième génération et des objets connectés ; le département Interaction avec l'humain traite notamment de la conception d'interfaces collaboratives entre humains et machines. Les questions de traitement de masses de données sont également présents dans le Défi 4 (« Santé et bien-être ») pour la bio informatique (Département Sciences des Données) et le TAL bio-médical (Département Sciences et Technologies des Langues), et dans le Défi 8 (« Sociétés innovantes, intégratives et adaptatives ») pour l'extraction de connaissances, ainsi que dans le programme d'actions « Big Data » (Département Sciences des Données). Au-delà de la Stratégie Nationale de Recherche, le centre aborde d'autres thématiques répondant à des enjeux de société importants, tels que l'aide à l'apprentissage et à l'acquisition de connaissances, l'assistance aux personnes âgées ou diminuées, la protection de la vie privée, ou la souveraineté des infrastructures de calcul et de données.

La pluridisciplinarité du centre met de fait en relations des objets et des pratiques de recherche très différents, impliquant des besoins matériels et humains différents (expérimentations, moyens de calcul, etc.). Le centre veillera donc à maintenir un bon équilibre de moyens et de personnels entre les différentes communautés, en interaction avec les tutelles. Il aura le souci de publier des postes sur l'ensemble des (sous-)disciplines présentes, de partager équitablement les moyens en fonction des besoins mais aussi, lorsque c'est possible, de mutualiser des équipements de recherche. Enfin il aura à cœur de construire un collectif scientifique, porteur d'un projet à long terme, alimenté par une vie de laboratoire riche et avec l'implication de tous dans sa gouvernance.

1. Département Mécanique Énergétique

1.1. Thèmes

Les activités du département se regroupent autour de la Mécanique des Fluides, des Transferts de masse et de chaleur et de l'Énergétique. Nous menons des recherches à caractère plutôt fondamental, dont les applications se trouvent dans les secteurs de l'énergie, des transports, de la santé et de l'environnement. Nous approfondissons les théories existantes et en développons de nouvelles avec l'objectif d'améliorer la compréhension des phénomènes physiques fondamentaux de ces domaines, d'améliorer leur modélisation et leur simulation avec notamment une visée d'optimisation et de contrôle des processus. Les deux axes principaux pour atteindre ces objectifs sont le développement de méthodes numériques avancées permettant d'accroître les performances des outils de simulation, ainsi que la mise au point de techniques d'analyse efficaces pour la caractérisation de la dynamique des écoulements et des transferts et leur manipulation. Des réalisations expérimentales viennent en support de cette double approche méthodologique et phénoménologique.

1.2 Composition

Équipe ITAC

L'équipe ITAC (Instabilité, Transition et Turbulence) s'intéresse à la dynamique des fluides et en particulier à la maîtrise des instabilités et de la turbulence, qui caractérisent un grand nombre d'applications dans le domaine des sciences de l'ingénieur. Les travaux de l'équipe s'articulent autour de deux thèmes : d'une part, l'étude fondamentale des instabilités et de la turbulence développée, d'autre part l'analyse, la manipulation et

le contrôle des écoulements. Pour le premier thème, on s'intéresse à des écoulements cisailés, en présence ou non de rotation et d'entraînement par des parois. Les objectifs d'étude sont la caractérisation de la dynamique linéaire et non-linéaire, pour des régimes allant de l'instabilité primaire à la turbulence pleinement développée. Ces travaux s'appuient sur le développement de méthodes numériques, adaptées aux nouvelles générations de calculateurs, et capitalisées dans des logiciels numériques déposés et ouverts à la communauté.

Le second thème peut être décliné autour de trois axes : un premier axe concerne le traitement des données en Mécanique des Fluides dans un contexte de données massives, le deuxième axe s'intéresse aux méthodes d'estimation et en particulier à la quantification des incertitudes, tandis que le troisième axe concerne le contrôle par apprentissage des écoulements, des fondements théoriques à la mise en œuvre expérimentale.

Équipe DTCM

L'équipe DTCM (Dynamique des Transferts et Couplages Multiphysiques) s'intéresse à l'efficacité énergétique et à la gestion raisonnée de l'énergie. Les travaux cherchent à améliorer la connaissance de phénomènes physiques isolés, à comprendre le couplage entre différents phénomènes, et à augmenter l'efficacité énergétique d'un système ou d'un ensemble de systèmes. Ces travaux peuvent être organisés en quatre thèmes :

- La Convection regroupe des études fondamentales (convection naturelle, rayonnement, mélange de fluides miscibles) menées à l'aide de la simulation numérique, ainsi que plusieurs études à contexte applicatif fort sur l'efficacité énergétique (photovoltaïque, récupération de chaleur).
- La MagnétoHydroDynamique (MHD) s'intéresse au mouvement d'un fluide conducteur de l'électricité dans lequel les champs de vitesse et d'induction magnétique sont couplés et en particulier à la compréhension de l'apparition d'un champ magnétique par effet dynamo.
- Les Transferts aux Interfaces sont l'objet d'une double action : d'une part, le développement d'outils numériques pour les écoulements diphasiques et d'autre part l'étude des phénomènes physiques responsables du transfert thermique aux interfaces aux échelles micro-nanométriques à très basse température.
- La Dynamique en Fluide Oscillant et Thermoacoustique a un double intérêt, la maîtrise des écoulements sanguins dans le réseau vasculaire et l'énergétique dans les machines à conversion d'énergie de type thermoacoustique.

1.3. Positionnement

Le département de Mécanique des Fluides-Energétique comprend une trentaine de chercheurs, enseignants-chercheurs, et ingénieurs de recherche.

Les activités de recherche du département recouvrent un spectre large, allant du numérique à l'expérimental, des échelles nanométriques aux échelles géophysiques, des analyses fondamentales aux études applicatives, et comprennent de larges interfaces avec la physique, les mathématiques appliquées et l'informatique. Elles s'inscrivent dans de nombreuses collaborations nationales et internationales.

La simulation numérique représente une part importante des activités de recherche du département, et a conduit au développement d'outils numériques partagés et de supports informatiques mutualisés. La mise au point de solveurs parallèles performants a donné lieu à des collaborations avec l'équipe PARSYS.

La mise en œuvre de méthodes d'apprentissage pour le contrôle a conduit à une ANR avec l'équipe TLP. Les approches statistiques pour le traitement des données de Mécanique des Fluides bénéficient du soutien du Center for Data Science.

La dynamique de recherche globale du département s'inscrit résolument dans le contexte des Sciences de l'Ingénieur, en particulier pour l'aérodynamique et l'énergétique. Elle est soutenue sur le plateau de Saclay par diverses actions, telles que les projets LaSIPs, ainsi que par des collaborations avec le CEA. Le département entretient des liens forts avec le FAST, à travers notamment l'organisation d'un séminaire commun, ainsi que plusieurs collaborations exploitant la complémentarité des recherches menées dans les deux laboratoires. Des collaborations, concrétisées dans des ANR, existent également avec le GEEPS.

2. Département Sciences et Technologies des Langues

2.1. Thèmes

Le département « Sciences et Technologies des Langues » étudie des questions fondamentales relatives aux systèmes linguistiques, en exploitant de larges corpus collectés, annotés et enrichis de manière automatique et semi-automatique. L'exploitation de ces corpus par les outils d'apprentissage statistique demande de nouveaux développements méthodologiques pour les adapter au matériau linguistique ; elle permet en retour d'étudier le fonctionnement des langues, leur variation synchronique et diachronique ainsi que de soulever des questions relatives à leur acquisition en tant que langues maternelles ou secondes ainsi que l'apprentissage des langues par les humains. Enfin, un dernier axe de recherche concerne les grandes applications du traitement des langues : reconnaissance vocale, recherche d'information, agents conversationnels, qui portent des enjeux sociétaux (sauvegarde des langues en danger, outillage de personnes en situation de handicap) et éthiques de plus en plus importants. Le thème langue couvre ainsi un large spectre, des recherches les plus fondamentales aux plus appliquées, dans toutes les modalités (écrit, parole et signe), et sur une grande variété de supports, certains mêlant d'ailleurs ces différentes modalités. Ces recherches ont un caractère fortement pluridisciplinaire et elles rassemblent des communautés diverses relevant des sciences de l'ingénieur et sciences humaines et sociales, qui ont en commun la pratique de l'exploitation de corpus et la culture de l'évaluation quantitative, deux marqueurs forts de l'activité du thème. Elles impliquent également des collaborations multiples, aussi bien au sein du périmètre Paris-Saclay que plus largement en France, en Europe et dans le monde.

2.2 Composition

Équipe TLP

Le groupe TLP mène des recherches visant à comprendre les processus de communication humaine par la parole et à développer des modèles pour le traitement automatique de la parole. Nos travaux de recherche sont interdisciplinaires par nature, allant du traitement du signal à la sémantique. L'un des principaux objectifs est l'analyse automatique de la parole visant une description complète du signal audio, qu'il s'agisse d'analyser la scène audio, de détecter la présence de parole, d'identifier la langue parlée, de caractériser le ou les locuteurs (leur identité et détecter leur état émotionnel), de transcrire la parole en texte, d'identifier des entités spécifiques dans la transcription, ou encore de traduire la parole. Les domaines applicatifs ciblés sont par exemple l'audiovisuel (sous-titrage, indexation, et traduction), l'interaction dialogique (robotique) et la sécurité (identification et analyse). Nos travaux s'intéressent également aux corrélats linguistiques des mesures observées sur des grands corpus, qu'il s'agisse de la variation phonétique synchronique, des marqueurs dialectaux, des changements diachroniques ou encore des facteurs extra-linguistiques (par exemple, les émotions) actualisés par des traits spécifiques.

Les recherches du groupe sont actuellement structurées en six thèmes interdépendants : Caractérisation du

locuteur dans un contexte multimédia (Thème 1); Dimensions affectives et sociales des interactions parlées avec des (ro)bots et enjeux éthiques (Thème 2); Perception et traitement automatique de la variation de la parole (Thème 3); Traduction et apprentissage automatiques (Thème 4); Reconnaissance de la parole (Thème 5); Ressources langagières (Thème 6).

Équipe ILES

Le groupe ILES se consacre au traitement automatique des langues écrites et signées. Nous étudions la sémantique et les informations véhiculées par des énoncés langagiers, comment les représenter, les détecter, ou encore les générer. Les méthodes employées vont de la modélisation à l'apprentissage automatique. Nous travaillons sur la modélisation des représentations associées aux traitements, qui s'instancie dans la construction de corpus annotés. En langue des signes, cela repose sur une modélisation en linguistique et en sciences du mouvement.

Nos axes clés sont la compréhension et la transformation des productions langagières. Ils s'instancient plus spécifiquement dans l'analyse automatique d'énoncés pour extraire, représenter ou accéder aux informations qu'ils contiennent ou dialoguer avec un utilisateur, ainsi que dans le traitement de la similarité textuelle (synonymes, paraphrases, simplification, implication). En langue des signes, ces axes sont complétés par une problématique de génération d'énoncés signés.

Ces travaux s'appuient sur l'apprentissage de représentations continues, notamment en exploitant conjointement des textes et des bases de connaissances. Ils se placent souvent dans un contexte où les données sont en quantité insuffisante, ce qui motive l'étude et la conception de méthodes pour palier ce manque.

Le domaine biomédical est un champ applicatif important pour nos travaux.

2.3. Positionnement

Le département « Sciences et Technologies des Langues » a pour ambition de contribuer au progrès des sciences et technologies de la langue. Fort de 80 personnes dont 30 CEC permanents, il constitue l'un des pôles majeurs de ce domaine en France et plus largement en Europe et est un acteur reconnu à l'international : projets américains, projets européens (T4Me, QT21, MiRoR, ...), fellows ISCA et AMIA, rôle dans les associations internationales, etc.

Ce département se caractérise par la considération de toutes les modalités : langue parlée, écrite et signée. Les thématiques sont abordées de manière pluridisciplinaire, du traitement du signal à la sémantique : chercheurs en sciences de l'information (CNRS 07) et en sciences du langage (CNRS 34), collaborations avec les sciences du mouvement (CNRS 26), etc.

Capitalisant sur les résultats de ses recherches et sur les compétences de ses chercheurs, il est également engagé dans une démarche d'innovation et valorisation (projets partenariaux, maturation, participation à la création d'entreprise, tremplin Carnot). Ses activités sont concentrées dans les groupes ILES et TLP.

Les travaux menés au sein du département s'attaquent à des données complexes, naturelles, hétérogènes et multidimensionnelles, de grande dimension et d'une grande variété. Ils sont par essence pluridisciplinaires et donnent lieu à des collaborations en interne et en externe, avec des linguistes, pour la langue des signes des spécialistes du mouvement, pour le Traitement des langues appliqué au domaine biomédical des professionnels de santé.

L'essentiel des activités de recherche porte des enjeux sociétaux (sauvegarde des langues en danger, outillage de personnes en situation de handicap, aide au traitement de l'information et de la connaissance médicale) et

éthiques de plus en plus importants. Ce sont des thématiques extrêmement importantes pour le tissu industriel, avec lequel nous entretenons des collaborations suivies. Nos activités couvrent ainsi un large spectre, des recherches les plus fondamentales aux plus appliquées, dans toutes les modalités (écrit, parole et signe), et sur une grande variété de supports, certains mêlant d'ailleurs ces différentes modalités. Nos recherches ont un caractère pluridisciplinaire. Elles impliquent également des collaborations multiples, aussi bien au sein du périmètre Paris-Saclay que plus largement en France, en Europe et dans le monde.

3. Département Interaction avec l'humain (*nom provisoire*)

3.1. Thèmes

Le département *Interaction avec l'humain* rassemble des chercheurs et enseignants-chercheurs actuellement répartis au LIMSI (groupes AMI, CPU et VENISE), au LRI (équipes-projets communes UPSUD-CNRS-Inria EX-SITU et ILDA) et à Inria (équipe-projet AVIZ qui souhaite devenir équipe-projet commune). Les équipes du département s'intéressent à l'interaction entre l'humain et la machine avec un regard pluridisciplinaire alliant informatique, traitement du signal, psychologie, psychologie sociale et ergonomie cognitive. Cette thématique englobe non seulement les multiples moyens matériels et logiciels qui peuvent être mobilisés pour établir un lien entre l'humain et l'ordinateur, mais aussi la compréhension de l'humain pour la conception de nouvelles interfaces et la compréhension des interactions sociales entre humains pour assister leur collaboration grâce à l'ordinateur. Le département constitue l'un des principaux pôles européens sur cette thématique.

Les équipes de ce département explorent de multiples formes d'interaction : interaction visuelle, tangible, haptique, auditive, gestuelle... ; interaction multimodale ; visualisation interactive d'information ; réalité virtuelle ou augmentée ; interactions affectives et sociales ; interactions collaboratives en situation co-localisée ou distante. Elles développent des méthodes de conception (comme la conception participative), des méthodes d'analyse (comme l'analyse d'image pour l'interaction) et elles abordent des domaines d'application variés (comme la prise en compte du handicap). L'éventail de ces recherches permet au département de développer des travaux qui sont à la fois très fondamentaux, en cherchant à modéliser les phénomènes en cours lors de ces interactions, et tournés vers des applications pour évaluer la pertinence de ces modèles et de certains dispositifs dans leur usage réel.

3.2 Composition

Équipe AMI

AMI regroupe des chercheurs et enseignants-chercheurs spécialisés dans différents modes d'interaction tels que la vision, le mouvement, l'interaction tactile ainsi que dans les interfaces multimodales qui combinent différentes modalités pour parvenir à une interaction plus « naturelle ». Les recherches de l'équipe s'articulent autour de trois axes: le traitement d'image et la visualisation d'information appliqués à différents domaines tels que la réalité augmentée vidéo-projetée, l'analyse d'activité, l'aide au diagnostic ; les environnements ambiants et connectés avec des applications dans le domaine des services à la personne et la santé; l'étude de la multimodalité dans le contexte du handicap sensoriel, moteur et cognitif.

Équipe CPU

À la frontière entre informatique, psychologie et ergonomie, CPU (Cognition, Perception et Usages) est composé d'enseignants-chercheurs de ces trois disciplines. L'équipe conduit des recherches pluridisciplinaires sur la cognition humaine et artificielle et la conception, l'évaluation et les usages des technologies. Ces recherches ont un double objectif : 1) étudier l'humain pour concevoir des systèmes adaptés et adaptatifs aux caractéristiques humaines dans leurs situations d'usage; 2) concevoir des

interactions homme-machine pour étudier et mieux comprendre l'humain. Ces recherches sont appliquées à l'analyse et à l'apprentissage de compétences sociales et professionnelles, à la simulation de comportements et aux technologies motivationnelles/persuasives.

Équipe VENISE

VENISE (Virtual & augmented ENvIronments for Simulation and Experiments) a pour objectif scientifique de comprendre et faire progresser le domaine de la Réalité Virtuelle et Augmentée (RV&A). Les recherches de l'équipe visent, d'une part des travaux génériques concernant la conception et l'évaluation de nouveaux modèles d'interaction en RV&A (thème 1), et d'autre part, l'élaboration des modèles de données ou de simulation pour accroître l'utilisabilité de la RV&A dans différentes applications phares (thème 2). Transversalement à ces deux thèmes, l'équipe étudie les problématiques perceptives et cognitives qui sont propres à ce domaine, telles que les incohérences sensorimotrices, la "cybersickness", les sensations d'immersion ou de présence, et elle travaille sur les évolutions de leurs formalisations théoriques.

Équipe ILDA

Dans un nombre croissant de domaines, les utilisateurs de systèmes informatiques ont à faire face à de grands jeux de données. Ces jeux de données sont de plus en plus interconnectés et organisés sous forme de structures complexes, grâce notamment aux nouveaux modèles de données que nous voyons apparaître, par exemple, avec le Web des données. Les travaux de recherche de l'équipe ILDA ont pour but de tirer parti de la richesse de ces modèles de données pour concevoir de nouveaux systèmes interactifs qui assistent au mieux les utilisateurs dans leurs tâches de compréhension et de traitement de ces grandes quantités de données.

Équipe EX-SITU

ExSitu explore les limites de l'interaction en étudiant des utilisateurs "extrêmes", tels que les professionnels créatifs, artistes et designers qui repoussent les limites de la technologie pour atteindre leurs objectifs créatifs, ou les scientifiques qui essaient de comprendre les phénomènes complexes par l'exploration créative de données complexes. L'étude de ces utilisateurs permet non seulement d'anticiper les besoins des usages routiniers de demain, mais aussi de mieux comprendre le phénomène de l'interaction lui-même et de créer de nouvelles formes d'interaction. Le but de l'équipe est de créer des partenariats humains-machines permettant aux utilisateurs de maîtriser la technologie, fondés sur des principes issus de notre approche théorique de l'interaction.

Équipe AVIZ

Aviz a pour objectif d'améliorer les méthodes, techniques et outils basés sur la visualisation interactive. Traditionnellement, la visualisation a été segmentée en questions spécialisées : visualisation scientifique, techniques d'interaction, ou nouvelles représentations. En contraste, Aviz veut étudier le continuum d'activités entre la modélisation, l'exploration, la présentation, la prise de décision et l'éducation. De plus, lorsque la quantité de données augmente, elle dépasse les capacités de transfert des seuls canaux visuels. Aviz travaille aussi sur le "Visual Analytics": un ensemble de méthodes pour combiner l'analyse de données avec la visualisation afin de permettre l'exploration de données massives. Un des axes d'Aviz est le passage à l'échelle dans l'exploration des données à l'aide d'un nouveau paradigme de calcul : l'analyse progressive de données. Explorer des données nécessite une latence inférieure à dix secondes, que les systèmes actuels ne peuvent garantir. L'analyse progressive produit des estimations de résultats dont le temps de calcul est borné et améliore ces estimations itérativement. Ce nouveau paradigme a des implications à toutes les étapes du traitement des données, de leur stockage, analyse et visualisation. Aviz s'intéresse à ce nouveau paradigme qui est un lien fort avec plusieurs autres départements du laboratoire, comme Sciences des Données et Algorithmes, Calcul et Apprentissage.

3.3. Positionnement

Avec près de 40 chercheurs et enseignant-chercheurs, le département *Interaction avec l'humain* constitue l'un des plus importants pôles d'enseignement et de recherche dans ce domaine au niveau national, européen et international. Sur le plan thématique, les plus grosses équipes sont à Grenoble, avec un spectre assez large mais centré sur le génie logiciel des systèmes interactifs, et à Toulouse, avec une spécialisation sur l'interaction pour les systèmes critiques et le handicap. Au niveau international, les plus gros centres sont surtout aux États-Unis (CMU, Georgia Tech, U. Washington, MIT ...). En dehors de ces centres, peu d'Universités ont un spectre thématique aussi large et un nombre de chercheurs aussi importants.

Le spectre large du département *Interaction avec l'humain* lui confère un rôle structurant dans le domaine, en s'appuyant des chercheurs très reconnus dans leurs thématiques au niveau national (projets ANR, collaborations industrielles) et international (projets européens, ERC, collaborations internationales majeures). Deux masters d'audience internationale sont également portés par des enseignants et chercheurs très impliqués, ce qui contribue à l'aura des équipes du département.

Le département s'appuie sur un réseau d'équipements exceptionnels, développé notamment au travers de l'Equipex Digiscope, dont il assure la direction scientifique et dont il héberge trois des 10 plateformes (EVE, WILD et WILDER), ainsi que le FabLab.

Le caractère pluridisciplinaire des recherches du département *Interaction avec l'humain* est visible à travers différents niveaux de collaboration : à l'intérieur d'une équipe entre plusieurs disciplines (ex. informaticiens et psychologues dans CPU), entre différentes équipes du département (ex. VENISE et ExSitu avec les dispositifs Digiscope) ou encore entre des équipes appartenant à plusieurs départements du laboratoire (ex. des spécialistes du Traitement de la Langue, des Sciences de Données, des Mécaniciens ou des Algorithmiciens).

De fortes collaborations ont également été construites au sein de l'Université Paris-Saclay, par exemple avec la Fédev pour les travaux sur le mouvement ou à travers les actions « Arts et Sciences » dans lesquelles les équipes du département ont toujours été fortement impliquées. Au niveau national, le département est fortement impliqué dans l'animation de groupes de travail thématiques (la journée commune de l'AFIHM et de l'AFIA en 2019 a ainsi été organisée par des chercheurs et enseignants-chercheurs du département) et de conférences scientifiques (ex. la conférence IHM). Nous avons aussi construit de nombreuses collaborations avec des acteurs industriels qui financent régulièrement des thèses CIFRE et des travaux de recherche fondamentale ou appliquée (ex. deux chaires industrielles, participation à l'Institut Carnot « Cognition », collaborations avec EDF R&D). Enfin, au niveau international, nos équipes collaborent avec les plus grandes universités dans le domaine de l'Interaction Humain-Machine dans le monde, animent la communauté de la RV&A à l'échelon européen (EuroVR) et participent à l'organisation des conférences phares du domaine (CHI 2013 à Paris par exemple avec 3500 participants, et IEEE VIS 2014 avec 1200 participants).

Le domaine de l'interaction est en constante mutation du fait des évolutions technologiques qui continuent à modifier le regard qu'on peut avoir sur une interaction avec une machine. L'introduction dans la vie de tous les jours de ce qui n'apparaît pas toujours comme étant une machine mais avec laquelle il faut malgré tout interagir nécessite de définir de nouveaux paradigmes. Le département souhaite être l'un des acteurs majeurs de cette transformation.

4. Département Sciences des Données

4.1. Thèmes

Le département Science des Données a pour objet la modélisation et l'analyse de données et de connaissances.

Des traces numériques de l'ensemble des activités humaines sont aujourd'hui disponibles, de la santé à l'éducation, de la finance et des assurances à la production industrielle, des sciences expérimentales aux sciences humaines. Les données disponibles sont souvent massives, hétérogènes, dynamiques et de qualité variable (les 4 V : Volume, Variété, Vitesse, Vérité). Leur exploitation conduit à la définition d'un quatrième paradigme scientifique où la conception de modèles et d'algorithmes repose sur la prise en compte des caractéristiques de ces données et la validation de ces modèles et algorithmes sur une interaction forte avec les experts du domaine.

Le département *Science des données* a donc pour objectif de prendre en compte les caractéristiques des données et de répondre aux défis de passage à l'échelle et de résistance aux biais, faisant naître de nouvelles problématiques informatiques dans le domaine du stockage, de la communication, de l'analyse et de l'optimisation des traitements, de l'interrogation et de l'enrichissement des données et de la découverte des connaissances, au confluent de quatre équipes d'expertises complémentaires.

4.2 Composition

Équipe LaHDAK

L'équipe LaHDAK (Large-scale Heterogeneous Data and Knowledge) s'attaque aux défis inhérents à la disponibilité croissante de données et de connaissances, impactant tous les domaines d'innovation, dans l'ère du BigData. Elle propose des solutions pour gérer des données massives, complexes ou interconnectées, sémantiquement hétérogènes, incertaines, incomplètes et évolutives, provenant de sources variées, telles que le Web social, le Linked Open Data ou les données scientifiques. L'équipe développe des solutions visant à résoudre les problèmes d'efficacité et de passage à l'échelle, de pertinence sémantique et de robustesse dans les divers scénarios d'accès à l'information. Elle rassemble des expertises complémentaires en bases de données, systèmes de recommandation, intelligence artificielle, représentation des connaissances, raisonnement et Web Sémantique. Ses travaux de recherche sont organisés autour de trois axes : gestion des données, gestion de connaissances, fouille de données et découverte de connaissances.

Équipe ROCS

L'équipe ROCS est composée de chercheurs en Réseaux et en Optimisation Combinatoire Stochastique. Elle se décline en trois activités: Réseaux, Optimisation Combinatoire et Stochastique, Optimisation dans les réseaux.

L'objectif commun à ces 3 activités est de traiter des problèmes issus du large domaine de l'Internet des Objets avec la programmation mathématique (pas seulement employée comme un simple outil) afin de proposer des méthodes de résolution exactes qui serviront de point de référence lors de l'élaboration d'heuristique au moment du passage à l'échelle ou du distribué. La méthodologie générale de l'équipe ROCS consiste dans un premier temps, à avoir une approche théorique, puis les performances sont étudiées par simulation, domaine dans lequel ROCS a développé des compétences fortes en mettant en œuvre des outils logiciels efficaces (en particulier sous NS et OPNET, CPLEX). Les solutions sont également implantées dans des systèmes réels pour conduire des expérimentations. Dans ce cadre, plusieurs plateformes ont été développées.

Équipe BioInformatique

L'objectif de l'équipe Bioinfo est de concevoir et de développer de nouvelles approches informatiques pour répondre à de véritables questions biologiques et bioinformatiques. Le principal atout de ce groupe est non seulement d'être capable de comprendre les problèmes posés par l'étude des systèmes biologiques depuis l'échelle moléculaire jusqu'à celle des cellules et des populations, mais aussi de concevoir des solutions impliquant des compétences issues de domaines très variés (algorithmique, combinatoire, bases de données,

représentation des connaissances, extraction de données, analyse statistique, apprentissage automatique, méthodes formelles et simulation). Notre programme scientifique est structuré en trois thèmes, fortement corrélés : Analyse de séquences et de structures biologiques ; Biologie des systèmes et biologie de synthèse ; Intégration et analyse de données biologiques.

Équipe Apprentissage et Optimisation (équipe transverse avec le Dpt ACA).

L'équipe A&O -- équipe commune Paris-Saclay, CNRS et INRIA-Saclay -- s'intéresse à l'apprentissage de modèles à partir des données, en mettant l'accent sur trois types de propriétés :

- La robustesse statistique, qui vise la stabilité et la qualité prédictive des modèles face au bruit et à la variabilité des données ;
- L'équité des modèles, qui s'intéresse aux impacts des décisions fondées sur les modèles et à l'éthique (recherche des modèles causaux ; analyse des biais des données et des modèles ; traçabilité des décisions) ;
- L'interprétation des représentations internes de l'apprentissage, en lien avec les structures et les invariances des données, au coeur de l'apprentissage profond.

4.3. Positionnement

Le département Science des Données comprend des enseignants-chercheurs dont l'expertise couvre le spectre large du domaine : bases de données, fouille de données, Web sémantique, représentation des connaissances, algorithmique, combinatoire, optimisation stochastique et distribuée, apprentissage statistique (supervisé, non supervisé, structuré) réseaux, simulation, validation et transfert.

Les membres du département ont une expertise approfondie de la recherche interdisciplinaire et du dialogue avec les experts des domaines d'applications (particulièrement biologie, médecine, sciences humaines et sociales, physique expérimentale), permettant un accès privilégié aux données d'intérêt et à l'évaluation des modèles et des algorithmes.

Les membres de ce département sont très impliqués dans l'animation de l'Université Paris-Saclay, compte tenu de l'importance centrale de l'IA et de la science des données : dans le Labex DigiCosme (axes Data Science et Comex); dans les IRS B5GI (Beyond 5G Initiative, avec une forte composante IoT), Nutriperso (modélisation des relations entre santé et alimentation) et CDS 2.0, et dans l'Institut de Convergence DataIA.

Les membres du département sont aussi très présents dans l'IRT SystemX (membre du conseil scientifique, formation), et ont de nombreux contacts directs avec les industriels notamment des liens forts avec Airbus, IBM, Orange et Safran ; parmi les thèses Cifre citons EdF, Peugeot, Renault, Thalès, RTE, IFPEN.

Sur le plan national, les équipes du département sciences des données sont fortement impliquées :

- dans le GDR de la science des données MaDICS (pilotage d'une action de reproductibilité d'expériences bioinformatiques, portage du dossier de renouvellement 2020, et membres actifs de l'action RoD de découverte de connaissances) ainsi que dans le GDR RO (porteur du groupe de travail programmation mathématiques) et GDR BiM (Bioinformatique Moléculaire, membre du conseil scientifique) ;
- dans les sociétés savantes en IA (AFIA), en apprentissage automatique (SSFAM) et en informatique (SIF), ROADEF (vice présidence communication) ;
- dans les IPL (recherches transversales majeures) d'Inria.

Sur le plan international, les équipes du département science des données ont des liens forts avec les centres de sciences des données Alan Turing (Londres), Ivado (Montréal), JST (Tokyo), CDS de New York ainsi

qu'avec les départements de science des données de Berkeley, Manchester et plus généralement avec de nombreuses universités et partenaires étrangers.

Les équipes font partie de plusieurs réseaux internationaux :

- NEWCOM, réseau européen d'excellence dans la communication sans fil ;
- EIT Digital, institut Européen pour l'éducation la recherche et l'innovation ;
- ELIXIR, le réseau Européen de coordination des ressources bioinformatiques (bases de données, logiciels, clouds, etc) ;
- Steering Committee of European Machine Learning and Knowledge Discovery from Data.

5. Département : Algorithmes, Apprentissage, Calcul.

5.1. Thèmes

L'algorithmique et le calcul, au coeur de l'informatique, s'intéressent à tous les aspects théoriques et pratiques, logiciels et matériels de la discipline. Depuis une décennie, la conception des algorithmes est repensée par l'intelligence artificielle et l'apprentissage, visant la construction automatique de modèles et de processus de calcul en lien étroit avec les données, l'expert, l'utilisateur et/ou l'environnement.

Le département Algorithmes, Apprentissage, Calcul a pour objet la conception, la mise en oeuvre et l'analyse d'algorithmes en prêtant attention à quatre types de propriétés : i) leur complexité théorique, en particulier en lien avec les paramètres spécifiques aux des problèmes considérés ; ii) leur convergence, leur passage à l'échelle et leur robustesse par rapport aux divers critères de qualité ; iii) les caractéristiques des algorithmes, des données disponibles a priori et générées a posteriori, et des processus de calcul en termes de théorie de l'information ; iv) la performance des calculs et leur implémentation optimale sur les architectures parallèles les plus récentes.

Les principaux axes de recherche du département concernent les modèles de calcul (du calcul haute performance au calcul quantique en passant par les réseaux neuronaux et les algorithmes répartis), les architectures de traitement (graphes, traitement distribué, synchrone ou asynchrone), et les méthodes (e.g., optimisation continue, combinatoire, stochastique; théorie de l'information).

L'analyse et la conception des modèles et des processus fait une large part aux approches mathématiques (du discret et du continu, en passant par les probabilités et statistiques, et la combinatoire) et de la physique statistique (en particulier en lien avec les phénomènes de transition de phase des systèmes complexes).

Les applications comprennent le calcul scientifique (e.g., algèbre linéaire, calcul tensoriel, optimisation numérique, systèmes dynamiques, simulation d'algorithmes quantiques, physique computationnel, etc.), le calcul distribué et ses applications (e.g., cloud, ordonnancement, monnaie virtuelle, informatique ubiquitaire, robots autonomes, circuits micro-biologiques, etc.), l'analyse des données et les preuves de robustesse et de non-divulgaration des modèles.

5.2 Composition

Équipe Apprentissage & Optimisation (A&O)

L'équipe A&O - équipe commune LRI, CNRS et INRIA-Saclay - s'intéresse à l'apprentissage de modèles à partir des données, en mettant l'accent sur quatre axes fondamentaux :

i) L'apprentissage adversarial repose sur l'interaction de deux ou plusieurs agents apprenants, remplaçant la

fonction objectif inconnue par une approche de type min-max (théorie des jeux, test de Turing) ;

ii) La sélection et la configuration d'algorithmes a priori, en fonction des données disponibles, aussi appelée AutoML, constitue non seulement un passage obligé pour la démocratisation de l'IA, mais aussi un défi ouvert depuis 40 ans, lié à la définition des paramètres d'ordre des données ;

iii) L'apprentissage de modèles complexes pose l'identification des régularités permettant l'augmentation bien fondée des données dans les nombreux domaines d'application où les données sont petites (small data) par rapport à la complexité des modèles recherchés ;

iv) Enfin, loin de remplacer la connaissance par des modèles purement issus des données, un objectif consiste à dialoguer avec les connaissances du domaine, par exemple exprimées par des équations aux dérivées partielles ; l'enjeu est ici de faire le pont entre Apprentissage et Numerical Engineering, en lien avec le département Mécanique.

Équipe Graphes, Algorithmes et Combinatoires (GALaC)

Motivés par les problèmes issus de la bio-informatique, des réseaux sociaux et des réseaux de capteurs, les enseignants et chercheurs de l'équipe GALaC s'intéressent aux modèles mathématiques discrets pertinents pour l'analyse et la conception d'algorithmes pour les systèmes en réseaux, les problèmes combinatoires et la théorie des graphes.

En théorie des graphes, les problèmes considérés (e.g. coloration, domination, Hamiltoniens) et la preuve de propriétés structurelles de graphes sont connus pour être (NP) difficiles. La question devient ainsi de proposer des algorithmes approchés avec garanties (en vitesse et/ou en qualité d'approximation) et d'identifier les paramètres d'ordre contrôlant la croissance exponentielle de la complexité d'algorithmes exacts (algorithmes FPT).

L'algorithmique en réseau s'intéresse à la conception, l'analyse et l'optimisation d'algorithmes distribués, répondant à des objectifs appliqués ou à des questions théoriques.

Le thème combinatoire s'intéresse d'une part à des questions algébriques et énumératives (algèbres de Hopf, théorie des représentations, partitions) et d'autre part aux systèmes symboliques (mots, pavages, automates cellulaires), en relation avec les systèmes dynamiques, les systèmes de numération et l'analyse de complexité calculatoire.

Équipe Systèmes Parallèles (ParSys)

L'équipe ParSys est spécialisée dans le calcul haute performance et les algorithmes distribués. La recherche sur les algorithmes parallèles et distribués porte à la fois sur la théorie et sur l'implémentation. Ceci inclut la conception et l'analyse de complexité d'algorithmes pour les systèmes distribués et le calcul quantique ainsi que le développement de solveurs de calcul haute-performance (HPC). L'activité de l'équipe implique également l'utilisation des architectures parallèles les plus récentes pour obtenir une performance optimale des codes développés en vue des futures plate-formes Exascale.

Les principaux défis concernant les algorithmes étudiés dans l'équipe ParSys sont : l'optimisation des performances, le passage à l'échelle et l'équilibrage de charge, la tolérance aux défaillances et l'économie d'énergie.

Les applications (allant du calcul scientifique et de l'analyse de données aux protocoles réseaux en passant par les circuits micro-biologiques) répondent à des besoins industriels ou scientifiques essentiels et font l'objet de coopérations approfondies.

5.3. Positionnement

Les chercheurs impliqués dans ce département porte une attention sur les algorithmes, tant du point théorique (complexité, comportement, analyse) que de leur implémentation est au cœur de la recherche de ce département. Les champs d'application principaux sont: l'apprentissage automatique et l'analyse de données, le calcul scientifique, combinatoire ou quantique, les graphes, les réseaux...

Le développement de bibliothèques logicielles du domaine public est une activité commune aux équipes du département dans les domaines de l'algèbre linéaire (numérique et exacte), tenseurs, apprentissage automatique, calcul combinatoire...Ces Bibliothèques utilisent le parallélisme massif permis par les processeurs récents et l'utilisation des accélérateurs de type GPU est indispensable pour bénéficier de leur performance de calcul en précision réduite (par exemple pour le deep learning ou les solveurs en prévision mixte).

Les méthodes d'apprentissage (GANs, l'apprentissage par renforcement, par exemple) exigent un effort computationnel considérable. L'utilisation des GPUs est une bonne solution (la pratique montre qu'une GPU réalise des calculs 10 fois plus vite qu'une CPU). L'évaluation de performance des méthodes d'apprentissage nécessite de nombreuses expériences afin d'obtenir des résultats statistiquement fiables. Pour ces raisons une conjugaison de l'apprentissage automatique avec des méthodes de calculs de haute performance est indispensable pour mener à bien des études d'une grande envergure.

Une piste sur le croisement de la théorie de la complexité paramétrée et des méthodes de l'apprentissage automatique a un grand potentiel. Elle consiste à classer des instances d'un problème de satisfaction de contraintes (CSP, Constraint Satisfaction Problem) selon le critère de leur difficulté, par exemple leur optimum peut être trouvé en temps FPT selon un certain paramètre.

Récemment, l'optimisation et l'apprentissage automatique sont de plus en plus liés: l'optimisation concerne l'efficacité des algorithmes dans le pire des cas tandis que l'apprentissage automatique se concentre sur l'efficacité via des formulations efficaces en termes de calcul et évolutives, adaptables pour une classe bien définie d'instances.

Les objectifs, les méthodes, et les nouveaux défis de ces deux champs de recherche semblent se rejoindre, s'intensifier, et se nourrir (comme l'attestent les publications des articles dans les conférences de référence des deux domaines).