Université Paris-Saclay Doctoral Programme in Artificial Intelligence (UDOPIA)
Call for applications 2020

1   General description of the programme

The **UDOPIA** doctoral program in Artificial Intelligence will build on the extensive forces of Université Paris-Saclay in AI and related areas to create a unique cohort of PhD students trained at the forefront of core AI topics, specialized AI topics, and applications of AI. The PhD student will benefit from a rich ecosystem with strong links to industry, and from the existing instruments such as the DataIa convergence institute in AI or the SaclayIA computing platform. Specific accompanying measures will foster interdisciplinarity, mobility, entrepreneurship, and the wide diffusion of research results to academia, industry and the general public.

2   Nature of the support

The doctoral program should bring together **25 to 50** doctoral students over **3 years**.

The selected PhD Students will be recruited on a fixed-term contract of **36 months** starting in October 2020, with a gross monthly salary of at least 1758 €.

Support from the National Research Agency has been requested for this programme. If successful, the programme could recruit up to 50 PhD students and in addition to the salary, for each thesis, direct support of 15,000 € will be provided to the research team to support the doctoral project.

3   Selection/recruitment process

3.1   General recruitment policy

Following the UPSaclay Doctoral Charter, the admission to doctoral schools is based on explicit and public criteria and is run in accordance with transparent, open and fair procedures, defined by UPSaclay in application of internationally-recognised principles (e.g. as per the *European Code of Conduct for the Recruitment of Researchers)*. The implementation of this policy is monitored; reports on recruitment campaigns are made public on an annual basis. This policy applies to each and every applicant, regardless of their disciplinary fields, their career path, or the funding of the related doctoral project.

The recruitment procedure will follow the UPSaclay process for selecting and matching PhD candidates with doctoral projects in AI. The following committees will proceed to the selection of the PhD candidates:

- The Admissions Committees of each UPSaclay Doctoral School, in charge of evaluating PhD candidates on a doctoral project relevant to the doctoral school,

- The UDOPIA Programme Selection Committee, in charge of arbitrating among doctoral projects and allocating resources.

3.2 Dissemination of offers

Prospective supervisors are invited to propose preliminary subjects or supervision offers, **in English** and in French, before **April 30** on the Adum platform, with their contacts so that the interested students can exchange with them and prepare together the application.

Two categories of funding will have to be selected by the prospective supervisors, as indicated below. If a thesis topic had been submitted without specifying both types of funding, the doctoral school will be able to correct the submission on request, provided that this request is done **before April 30**.
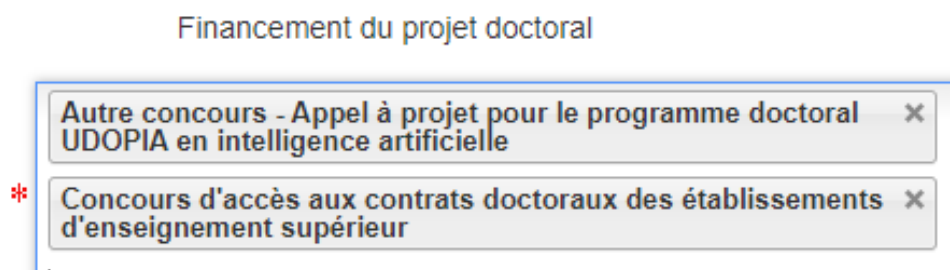
The subjects will be published on the Université Paris-Saclay website and other distribution channels so that interested students could easily find them.

→ http://www.adum.fr/psaclay/ptUDOPIA

Candidates are invited to contact the thesis directors, discuss the thesis topic and its progress and then can submit a complete application to the doctoral school.

3.3 Doctoral School Admission Committee

Each Doctoral School Admission Committee, appointed by the council of the doctoral school, includes as observer an elected representative of PhD students. The nominative composition of these committees, the dates of their meetings and of the Face-to-face interviews of the candidates, the eligibility and evaluation criteria are published no later than February. Face-to-face interviews are conducted by the Doctoral Schools Admissions Committees in the spring and results are made public in June.

→ https://www.universite-paris-saclay.fr/en/research/avant/fundings#doctoral-contracts

The Admission Committee of each Doctoral School is in charge of evaluating each applicant's research and communication abilities, as well as the positioning of her/his PhD topic and potential integration team with respect to the state of the art. In the case where the applicant does not hold a Master degree (or equivalent at the French or European level), the committees examine whether the candidate can still be enrolled in the doctoral program, on the basis of his/her skills and expertise.

3.4   The UDOPIA Programme Selection Committee

The UDOPIA Programme Selection Committee includes the programme supervision team and up to four external experts appointed by the programme supervision team. The selection committee strives to enforce thematic and gender balance, with a good disciplinary diversity and an external perspective through the presence of members from other higher education and research institutions. The program committee will have to arbitrate among the doctoral projects whose candidates have already been favorably evaluated by the doctoral schools according to the following criteria.

The arbitration meeting will be held on the **morning of June 23rd.**

The provisional composition of UDOPIA Programme Selection Committee in 2020 is as follows:

→ The director of the doctoral college, Sylvie Pommier and the deputy vice-rector for STEM research Michel Guidal.

→ Five directors of involved doctoral schools, Alain Denise (STIC), Frédéric Paulin (EDMH), Eric Cassan (EOBE), Eric Deutsch (CBMS), Alexandre Pery (ABIES),

→ An elected representative of doctoral students on the boards of the ICST, EDMH, EOBE or EDSP doctoral schools, acting as an observer, and whose research topics fall within the scope of the UDOPIA programme.

→ Six experts in artificial intelligence from Paris-Saclay University: Gilles Blanchard (LMO, UPSaclay), Claire Nédellec (MaIAge, INRA), Daniela Piana (ISP, ENS Paris-Saclay), Michèle Sebag (LRI, CNRS), Charles Soussen (L2S, CentraleSupélec), Frederic Pascal (L2S, CentraleSupelec), Bertrand Thirion (Inria-Saclay).

→ Four experts in artificial intelligence from outside Paris-Saclay University: Florence d'Alché-Buc (Telecom Paris, IP Paris), Sihem Amer-Yahia (LIG, Grenoble), Liva Ralaivola (LIF, Marseille), Pierre Sénellart (DI, ENS Paris).

Other experts may be invited to participate in the Jury depending on the doctoral projects submitted.

3.5   Eligibility criteria

→ The doctoral student's host research team must be part of the current IDEX Paris-Saclay perimeter. (list of the teams : https://bibliolabs.universite-paris-saclay.fr/Labs/Definition)

→ By the time the selection committee is held, the candidate must have successfully completed the admission tests of the doctoral school. Only candidates who have already been interviewed by the doctoral school and admitted to enroll subject to obtaining funding for their thesis are eligible.

3.6 Evaluation criteria

→ Integration of the doctoral project into the strategic axes of the UDOPIA doctoral programme;

→ Excellence of the application and scientific potential of the doctoral project for the field: the programme aims for excellence and encourages originality and scientific risk-taking, openness to the broader international and inter-disciplinary context, and the development of new areas;

→ Conditions for a successful completion of the doctoral project;

→ Potential outcomes of the doctoral project and socio-economic impact.

4 How to apply?

Explore the offer of research subjects and find a potential supervisor here :

→ French : https://www.adum.fr/psaclay/ptUDOPIA

→ English : https://www.adum.fr/psaclay/ptenUDOPIA

Get in touch with this potential supervisor and prepare your application with him/her.

Submit your application to the doctoral school according to the doctoral school's process before the doctoral school's deadlines in order to be auditioned at the doctoral school's annual Jury, if this takes place before June 16th or with a tailor-made jury otherwise.

It should be explicitly mentioned in your application in which of UDOPIA's thematic axes your doctoral project could fit in. You must also inform the doctoral school that you are applying to UDOPIA.

5 Strategic axes of the UDOPIA doctoral programme

Research in AI at UPSaclay focuses on fundamental issues related to AI, machine learning and data science, as well as on their operational applications in engineering, health and the environment, to name a few.

5.1 First circle: Basic AI and Data Science

UPSaclay covers a wide range of fundamental topics in AI, especially around Machine Learning, and more generally in Data Science.

i) Theoretical and algorithmic study of machine learning models, including functional approximations (deep learning & recursive neural networks, random forests, kernel methods), statistical learning theory (high-dimensional data, spectral methods), topological data analysis, uncertainty estimation (predictive uncertainty, systematic uncertainties), and causal modeling (identifiability, effects of interventions). A related thread of research concerns model selection and aggregation, which are crucial steps to allow efficient learning and high performance. Advances in these areas rest on the solid tradition of collaboration between statistics and probability theory at LMO, CMLA, L2S; some other relevant aspects include model verification (CEA-LIST, LRI, LSV).

ii) Development of algorithms robust to distribution drift, outliers or modelling errors, including transfer learning, zero-shot learning, concept drift, data streams, lifelong learning and reconciliation of simulated and experimental data (EdF, LAL). Another line of research addresses data augmentation and/or generative models (LRI). A hot research topic is the combination of data-integration methods in databases with statistical modeling and regularization from machine learning (Inria-Saclay), in particular to address missing data.

iii) Reinforcement learning is concerned with optimal sequential decision making, handling optimization under uncertainty or in an adversarial context (e.g. games). The challenges range from the scalability with respect to the state and action space, to improving the convergence speed, to acquiring multiple skills simultaneously or sequentially (LMO, LRI).

iv) Optimization for AI, aimed to algorithmic scalability with respect to high-dimensional data and big data: gradient-based methods, convex, sparse methods, dynamic programming, stochastic methods, meta-heuristics (LMO, LRI, Inria-Saclay, CMLA, L2S, CVN, MIA).

v) Meta-learning and AutoML: the automatic selection of the algorithm and hyper-parameters best suited to a given application is a key issue for transferring research results to industry, especially as computationally heavy algorithms are considered (e.g. in Deep Learning). AutoML and AutoDeep raise profound questions about the typology of learning problems (LRI, Inria-Saclay, IRT SystemX).

5.2 Second circle: Specialized ML and AI

This set of research topics addresses applications of the core AI technologies to bridge the gap with the industrial world. The main topics developed at Paris-Saclay are:

i) Management, analysis and protection of data: The performance of cloud-based applications is critical to AI but remains difficult to master. Optimizing such applications by integrating different types of parallelism and guaranteeing the desired level of performance and energy efficiency is still an open question. Architectural principles that grant isolation and integrity of remote processing and protection of data confidentiality are also critical. A recent development comes from the observation that Sensitive Training Data can leak into Deep Neural Networks. Ensuring privacy in the context of large, expressive models is thus particularly hard (L2S, Inria-Saclay).

ii) Natural Language Processing (NLP): In order to achieve accurate language understanding for all languages and modalities, research in NLP needs to progress in several directions. Covering more languages and modalities depends on the availability of annotated textual and spoken data resources. Domain adaptation, and more generally, weakly supervised, multi-task, multilingual and transfer learning have to be leveraged to generalize to more domains or genres. Multimodal language processing is another way to improve robustness by combining the input signal with complementary cues in multimedia documents. Deeper analyses of utterances, conversations, texts and documents requires to tackle ever more complex linguistic phenomena, and to connect underspecified semantic representations with large-scale symbolic knowledge bases or with physical actions and perceptions in the real-world. In particular, conversational agents need to analyze long-range contexts in question-answering and debating, and to handle social non-verbal signals. Such systems also have to be endowed with values and ethical norms. This research is particularly active at LIMSI.

iii) Human-Computer Interaction (HCI): AI systems will require new forms of interaction to create true human-computer partnerships where users are empowered by technology, not replaced by it. From autonomous cars to medical diagnostic, putting the "human in the loop" is not sufficient to avert problems and ensure acceptability. Instead a "computer-in-the-loop" approach can be designed, where humans needs and skills drive the design of interactive AI systems. Interactive techniques that use rich sensor data (3D camera, motion tracking) need AI for HCI to support real-time learning and recognition. AI visualization techniques, e.g. for large datasets and deep learning, will help understand AI algorithms and provide transparency for the wider public. This research direction is pursued at LRI and LIMSI.

iv) Signal, image, vision: As autonomous systems will embed different types of sensors to perceive the environment, deep models offer a generic and powerful framework to learn for heterogeneous visual data and fuse them, and research on this topic will play an important role in the design of future autonomous platforms. To address the data

scarcity challenge, smart automatic annotation methods, active learning and generative models open numerous perspectives to generate more informative training data. Data streams bring their own challenges and are better handled with dedicated architectures. An important and very active subdomain is that of medical image and signals, where modern AI is involved to improve acquisition efficiency (compressed sensing); to better reconstruct images and signals; to extract meaningful patterns for image interpretation and segmentation; to better interpret them, in the framework of big data analysis, in order to draw population-level inference or to perform individual diagnosis. This stream of research is active within CEA, CVN, MAS, Inria-Saclay, CMLA, in close partnership with some main players in the field (GE, Servier, Sanofi) and start-ups created by Paris-Saclay researchers (e.g. Therapixel, Terpanacea).

v) Architecture and hardware for AI: Self-driving cars, safety systems and devices with extensive and seamless interaction with users might require the processing to be done on-board, due to energy or privacy concerns. Such practical applications of AI require low-power, low-cost and high-performance AI accelerators. Innovative architectures inspired by neuroscience also provide new avenues for future AI hardware. AI can be very computationally expensive. For example, DeepMind's Alpha Zero requires up to 5000 TPUs, i.e. a processing power of about 500 POps (5×10^{17} operations per second). Currently, AI accelerators are from non-European countries and might be subject to export restrictions. This research is active at CEA, CNRS/Thales.

vi) Machine learning with a human focus. Sequential data, sequential learning, recommendation and reinforcement learning all require to integrate hypothesis confirmation within hypothesis learning; this integration is at the core of responsible AI and AI for Humanities (bias detection; self-fulfilling prophecies; causal learning). Such questions are addressed at LRI, CMLA and LIMSI.

vii) Trusted AI. The objective of "Trusted AI" is to develop responsible AI methods that should be compliant with regulations (GDPR) and ethical requirements, while being robust w.r.t. biased and corrupted data. This requires developing auditability frameworks, responsible-by-design approaches, and symbolic methods to enhance robustness and explainability. Moreover, equity-by-design approaches in supervised machine learning and progressive user-centric analytics, can contribute to ensure non-discrimination, explainability and transparency to avoid harm for people and society but also to align AI technical progress with human values. This line of research is developed at LIMSI, LRI, Inria-Saclay, CMLA.

5.3 Third circle: Application domains.

Based on the first two circles, AI research at Paris-Saclay address applications in the four national priorities, in both academic and industrial collaboration frameworks, with a particular focus on health on the one hand, and environment on the other.

Health and well-being: One of the main societal challenges is better health through personalized medicine. This new paradigm based on the digital representation of a patient creates a shift from optimized treatment based on an "average patient" to personalized diagnostic and therapy. The integration and analysis of multi-omic data, linked with imaging and other data are essential to address medical problems in a comprehensive and integrated way. This can be applied to cancer, neural or rare diseases. Medical image analysis and medical data mining will also be applied to neuroscience, epidemiology and nutrition. A particular spot of interest is that of medical images, for which deep learning has proven to be particularly efficient. Another thread of activity concerns the development of AI systems for agronomy, including the representation of knowledge on biological systems and its use to better analyse data, the analysis of causal factors affecting plant growth, the analysis of microbiote, all of which now rely on data-intensive analysis. This application area builds upon the Paris-Sud hospital group (Kremlin-Bicêtre, P. Brousse, Béclère), INSERM, INRA-Jouy en Josas, AgroParistech, Institut Gustave Roussy and AP-HP, Neurospin/NeuroPSI facilities, and industry partners such as Servier and GE Healthcare, and the Therapixel start-up.

Energy, climate and environment: Research on intelligent and autonomous decision-making systems will enable efficient smart grid and sustainable production systems and circular material flows towards carbon-neutral societies. Key partners in this area include EDF, RTE, Total and ENEDIS, as well as SME Artelys: the PEPER project aims to learn predictive models regarding the use of renewable energy; RTE aims to learn control policies enforcing the safety of the energy network under perturbations; it has participated in the organization of the European See4C Challenge, and further developments aim to cast energy management as a reinforcement learning problem (LRI). Other topics well represented at UPSaclay (LMO, MIA, MaIAGE) range from electricity consumption forecast (in partnership with EDF), toecology (analysis of omics data, paleoclimate identification, epidemiology and diffusion phenomena).

Transportation and mobility: Two major goals are the safety and certification of AI algorithms for autonomous vehicles, and the study of cooperative intelligent transport systems using traffic modelling, distributed and interactive traffic management and optimal decision. Understanding urban mobility and its impact on the quality of life calls for combining AI techniques with domain knowledge and human expertise in social sciences and economics. Such multi-disciplinary approaches will enable the development of simulation algorithms for new mobility services and infrastructures, and the prediction of new usages. This application area builds upon the VEDECOM ITE, the System-X IRT and industry partners including Renault, PSA and Valeo. Most importantly, deploying AI-based algorithms in autonomous vehicles requires some guarantees to be established about their robustness under attacks; the hybridization of deep learning and formal methods is an emerging and most promising domain (CEA-LIST, LRI, LSV).

Security and defense: This area involves a large panel of AI-based technologies such as

dynamic scene recognition, operations planning monitoring and decision support for applications to autonomous robots, fleets of drones and space oversight, and information processing systems for collecting, processing and distilling multilingual, multimedia and multimodal information sources for intelligence collecting purposes. AI robustness and reliability as well as portable AI devices are key topics investigated by the Saclay ecosystem (CEA, L2S, LRI, LSV, CMLA) in partnership with IRT SystemX and Thales, among others. Furthermore, AI-based defense applications such as Radar detection, SAR image processing and smart sensors networks developments are important topics investigated by UPSaclay in connection with ONERA, the French Aerospace lab, and the DGA.

6 Presentation of the university and the teams and departments involved.

6.1 Presentation of Université Paris-Saclay

Université Paris-Saclay UPSaclay), founded by 14 partner institutions, unifies the expertise, culture and best practices of 3 Universities, 4 "Grandes Ecoles" and 7 National Research Organisations. Since 2015, it has been offering graduate cursus across 20 doctoral schools and 350 Master tracks.

Innovation and Excellence. UPSaclay represents about 13 % of France's academic forces in R&D. It includes Université Paris-Sud (fully integrated into UPSaclay as of Jan. 2020), which, with 2 Nobel Prizes and 10 Field Medals, is the top French University (9th European and 37th worldwide) in the 2019 Shanghai ranking. In terms of research, it includes 300 laboratories and 9,000 professors and researchers, one third of whom with the *Habilitation à Diriger des Recherches*. With about 1,000 PhDs per year, UPSaclay represents around 8% of all PhDs defended in France. The number of industrial CIFRE PhDs represents around 10% of all CIFREs in France. UPSaclay also features outstanding scientific facilities, such as Neurospin MRI, Apollo Laser, and SOLEIL Synchrotron.

Doctoral training and research-oriented Masters is a top priority of UPSaclay, with 55% of the students at the graduate level (43,000 Master's, grandes écoles and doctoral students). UPSaclay offers excellent higher education programmes in Computer Science, Mathematics, Engineering, as well as in Medicine and Pharmacy, Life Sciences, Biochemistry, Chemistry, Food, Environment, Climate, Law, Public Policy, Bioethics, Sociology and Demographics. Each academic track is designed to foster exchanges with industrial partners and start-ups and facilitate entry to the job market.

In 2019, UPSaclay joined forces with 4 European universities (in Lund, Munich, Porto and Szeged) to create a pilot European University, EUGLOH, committed to higher education and training on challenges in global health and well-being, in particular based on data science, AI and robotics.

Integration in the ecosystem, socio-economic impact and science-society relations.
UPSaclay, in close interaction with its regional ecosystem, aims for the top 10 of international universities for economic impact. It has close links with companies, public services and citizens from the nearby cities, including tight links with large companies (Air Liquide, EDF, IBM, PSA, Renault, Safran, Thalès, …) and SMEs; 100 start-ups are created every year in this rich ecosystem. UPSaclay has developed a series of instruments dedicated to student entrepreneurship and innovation with high socio-economic impact: networks of incubators, of FabLabs; Seed Fund; TTO ("SATT Paris-Saclay").

UPSaclay is strongly supported by the Greater Paris Region, in particular via the research networks (DIM) *Réseau Francilien en Sciences Informatiques* and *MathInnov*, and through the new AI-oriented programme Paris Region PhD2. *La Diagonale,* UPSaclay's transdisciplinary service, fosters the link between the University and citizens, NGOs and high schools by promoting and popularizing science and the arts / science connection, and by relating them to societal challenges.

The Research Ethics and Scientific Integrity Council (POLÉTHIS), created in 2017 at UPSaclay, encourages awareness, in particular for doctoral candidates and supervisors, on emerging ethical issues, such as those typically associated with AI goals and impacts on society and people. It acts as IRB (Institutional Review Board) for approving research that involves personal data.

6.2 Research in Artificial Intelligence at UPSaclay and presentation of the teams involved in UDOPIA

According to Elsevier ("Artificial Intelligence: How knowledge is created, transferred, and used: Trends in China, Europe, and the United States", Jan 2019), UPSaclay is the European leader in AI.

Artificial intelligence (AI) is one of the 7 interdisciplinary research priorities of UPSaclay. Specifically, out of the 4,170 PhDs defended at UPSaclay during the 2015-2019 period (8% of all PhDs in France), more than 500 (over 12%) are associated with one of the 5 most frequent AI keywords[3]. The scientific expertise of UPSaclay in AI covers a broad spectrum ranging from fundamental, algorithmic and methodological research in AI to AI applications in various domains, as demonstrated by the profiles of potential supervisors (See CVs in Appendix). On the application side, UPSaclay is particularly strong in the sectors of health and medicine, including bioinformatics, and energy and environment, where interaction with AI methods has become widespread.

The investment and strength of UPSaclay in AI has led to a number of landmarks (see Section 2.3): UPSaclay created the internationally renowned open-source scikit-learn library; UPSaclay has invested in high-performance computing facilities dedicated to AI

(Saclay-IA) as well as in the transdisciplinary DataIa institute; UPSaclay has a strong network of AI-related industrial partners.

The research units involved in UDOPIA include Computer Science, Mathematics and Engineering research units, as well as research units in other disciplines, notably Health and Life Sciences. Some of the main research units quoted below are: CEA-LIST (Laboratoire d'Intégration de Systèmes et de Technologies du Centre d'Etudes Atomiques), CESP (Centre de recherche en Epidémiologie et Santé des Populations, INSERM), CMLA (Centre de Mathématique et leurs applications, CNRS-ENS Paris-Saclay), CVN (Centre de Vision Numérique CentraleSupélec-Inria), Inria-Saclay (Centre de Saclay, Institut National de recherche en Informatique et Automatique), L2S (Laboratoire des Signaux et Systèmes, CNRS-CentraleSupélec), LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, CNRS), LMO (Laboratoire de Mathématique d'Orsay UPSud-CNRS), LRI (Laboratoire de recherche en Informatique CNRS-UPSud), LSV (Laboratoire Spécification et Vérification, CNRS-ENS Paris-Saclay), MaIAGE (Mathématique et Informatique Appliqué au Génome et à l'Environnement, Institut National de la Recherche Agronomique (INRA), Centre de Jouy-en-Josas), MAS (Laboratoire de Mathématiques Appliquées aux Systèmes, CentraleSupélec), MIA (Laboratoire de Mathématique, Informatique et Applications AgroParisTech-INRA), MSH (Maison des Sciences de l'Homme, CNRS-UPSaclay).

6.3 Graduate college presentation

As part of the IDEX Paris-Saclay excellence initiative, founded in 2011, the member institutions of UPSaclay have created a doctoral college, supported by research and training teams at the highest international level. The doctoral college defines, coordinates and monitors doctoral training across the University. It consists of twenty doctoral schools, which were very favourably evaluated by the High Council for Evaluation of Research and Higher Education (Hceres).

The main doctoral schools involved in UDOPIA are STIC (Information and Communications Sciences and Technologies), EDMH (Hadamard Mathematics Doctoral School), EOBE (Electrical, Optical and Bio Engineering) and Interfaces, as well as SHS (Human and Social Sciences), EDSP (Public Health), ABIES (Agriculture Food Biology Environment and Health), SDSV (Structure and Dynamic of Living Systems), SEIF (Ile-de-France Environmental Sciences), AAIF (Astronomy and Astrophysics, Ile-de-France).

As of Jan 2020, Master degrees, training programs, doctoral schools and research teams will be organized into Graduate Schools. UDOPIA will rely primarily on UPSaclay's internationally attractive Masters, with 4 tracks in Mathematics (including "Mathematiques, Vision, Apprentissage", created 20 years ago, the most visible and sought-after french Master in mathematical AI), 5 tracks in Computer Science. Masters in Life Sciences and Health, Social and Human Sciences, Electrical Engineering and in Law also include modules relevant to data science and AI. These Masters will

ensure that many high profile PhD candidates can apply to the present doctoral program.

7 Description of the Programme supervision team

7.1 Programme supervision team:

The role of the programme supervision team is:

- To promote the doctoral programme, both internally (laboratories, researchers) and externally (for potential applicants);

- To ensure the quality of the recruitment: the members of the supervision team will participate in the selection committee, appoint other members where necessary and organize their training and information to carry out this mission;

- To animate the AI research community composed of the members of the doctoral programme, the doctoral candidates and their supervisors; and

- To monitor the quality of the doctoral programme together with the doctoral schools, report to the members of the doctoral programme, to the doctoral college, to the academic council of UPSaclay, and to ANR.

The programme supervision team is composed of seventeen members :

- The director of the doctoral college, Sylvie Pommier and the deputy vice-rector for STEM research Michel Guidal.

- Four directors of involved Doctoral Schools: Alain Denise (STIC: Information and Communication Sciences and Technologies), Frédéric Paulin (EDMH: Mathematics), Eric Cassan (EOBE: Electrical, Optical, Bio-physics and Engineering), Florence Menegaux (SP: Public Health).

- One doctoral candidate, member of the doctoral programme, representative of the doctoral candidates in the programme.

- Six experts in AI from Paris-Saclay : Gilles Blanchard (LMO, Univ. Paris-Sud), Claire Nédellec (MaIAge, INRA), Daniela Piana (ISP, ENS Paris-Saclay), Michèle Sebag (LRI, CNRS), Charles Soussen (L2S, CentraleSupélec), Frederic Pascal (L2S, CentraleSupelec), Bertrand Thirion (Inria-Saclay).

- Four experts in AI from outside UPSaclay: Florence d'Alché-Buc (Telecom Paris, IP Paris), Sihem Amer-Yahia (LIG, Grenoble), Liva Ralaivola (LIF, Marseille), Pierre Sénellart (DI, ENS Paris).

7.2 PhD Students' Supervision Team

The students' supervision team includes several dozens of researchers, members of the

research labs and graduate colleges involved in UDOPIA. Altogether, their expertise covers a wide thematic spectrum ranging from core AI issues to AI applications (Section 2), as evidenced by the list of CVs, which represents a subset of the team (see Appendix).

UPSaclay has spelled out good practices in terms of supervision through flexible guidelines and accompanying actions, including :

- The doctoral charter emphasises the active part that each one (PhD student, supervisor, admission committee, monitoring committee) has to play and the shared responsibilities that the doctoral pathway entails, sets out a reference framework and highlights good practices;

- In order to ensure adequate availability, the number of PhD students placed under the guidance and responsibility of any given thesis supervisor shall not exceed five (three in some Doctoral Schools);

- An annual statistical review is carried out at the UPSaclay level and widely disseminated. It includes a follow-up on the number of doctoral students per supervisor, co-supervision, dropouts, etc.

- Annual satisfaction surveys are conducted among doctoral students and among supervisors and give rise to awareness notes focused on issues that need improvement;

- 2-day training courses to learn how to supervise doctoral students are offered to future supervisors 3-5 times a year, in small groups, with role-playing activities;

- "Supervisors' cafés" are organized once a month to share experiences and good practices in terms of supervision. Practical information sheets are prepared using the material of these cafés-debates on topics as varied as "conflict management", "preventing loss of trust", "the challenges of internationalization", "hiring", "Individual Thesis Monitoring Committee"...

- etc.

8 Accompanying measures

Beyond the various research activities described above, the Saclay ecosystem is developing many initiatives (software development projects, challenges) and hosts major computational resources that support the implementation of AI systems by academic partners.

8.1 Local environment for AI PhD students

Saclay-IA. UPSaclay hosts two national high-performance computing (HPC) centers

(CNRS-IDRIS, CEA-TGCC), as well as smaller-size centers (ENS-/CS-Fusion, Inria-Saclay, EP), and laboratory equipment, providing the scientific community with a wide range of computing facilities. Regarding AI specifically, the site hosts the Saclay-IA computing platform: this equipment, initially funded in the context of the Saclay Center for Data Science, has grown from a dozen to about a hundreds of recent GPU cards, associated with the appropriate data-storage facilities. Bi-located on CNRS and CEA premises, this equipment is opened to all the academic teams of the site and can also handle requests from industry[5]. By the end of 2019, the site will also host the French national computing platform for AI, the largest one in Europe (over 1,000 GPUs), providing accessible computational resources for at-scale experiments with computationally hungry machine learning techniques (chiefly Deep Learning).

Scikit-learn. Scikit-learn is a machine-learning library in Python, developed within UPSaclay research laboratories, which has over the years matured from a research project into an industrial open-source software, with a world-wide base of users, thanks to sustained support from the INRIA foundation, the University and its founding members. Scikit-learn nowadays powers many applications of AI and data science, with application domains ranging from medical imaging to product recommendation. The quality of scikit-learn, its algorithms, its interfaces, its documentation, are universally recognized and praised. Importantly, the scikit-learn ecosystem includes academic teams who contribute novel algorithms and use-cases either in the core library or in affiliated projects. This development presents an opportunity for students to get involved (participation to coding sprints, doctoral missions, links with industrial partners of the project).

IPOL. IPOL is an Open Science and Reproducible Research journal of image processing and image analysis. It contains many algorithms and their source codes useful for PhD Students, with an online demonstration facility and an archive of experiments. IPOL is maintained by CMLA at ENS Paris-Saclay, and contains many contributions in AI image processing by PhD students.

Digiscope. UPSaclay is the home of Digiscope, a unique network of ten interactive visualization platforms featuring ultra-high resolution wall-sized displays, immersive CAVE environments and large 3D displays, interconnected by a telepresence network that supports remote collaboration. The platforms can be used by researchers and Ph.D. students to visualize large and complex datasets and, using advanced interaction devices, manipulate and explore them. Digiscope supports research activities in Human-Computer Interaction, Big Data, Modelling and Simulation, and is well adapted to visualizing data sets and algorithms used in AI.

8.2 Accompanying measures

PhD AI days. Two annual days dedicated to the AI PhD program will be implemented to foster collaborations between PhD students and hosted laboratories

in UPSaclay as well as to make a real-time roadmap of the training program. These days will be decomposed in a half-day seminar to present research progress and a half-day workshop where groups of PhD students will develop future projects on a given topic.

Student training and mobility. UDOPIA will strongly encourage student training and mobility. For this purpose, we will take advantage of two major initiatives of UPSaclay: The **Digicosme** Excellence Laboratory Center in digital science, and The **DataIa** convergence institute. Digicosme develops three main themes that will define future information and communication systems: reliability and safety, future networks, data intelligence. It carries out actions in the fields of Education, Research and Innovation. DataIa is structured around several programs to foster activity in AI. Both foster activity in AI, in particular targeting students (summer schools, junior conference, master stipends, exchange programs, in- and out-going mobility) but also the ecosystem in general (invited professors, biannual conference, support for challenges etc).

Data Challenges. An important line of development for modern AI is based on Challenges. Since the famed Netflix challenge, challenges have been increasingly used to explore real-world problems, through the competition of international teams prototyping efficient solutions. In particular, Université Paris-Sud maintains and extends the Codalab platform ([https://codalab.lri.fr/,](https://codalab.lri.fr/) formerly developed at Microsoft), with a strong international visibility (40,000 users, 1000 competitions as of July 2019). Most interestingly, Codalab is used for teaching at the under-graduate and graduate levels, as designing one's own challenge is a fun and efficient way to understand *what is a well-posed machine learning problem*. Codalab is also used in industrial and scientific challenges (e.g. continuing the Higgs Boson Challenge in partnership with the LAL and CERN). Complementarily, the UPSaclay *Center for Data Science* has developed the Ramp framework to handle 1-day challenges in experimental sciences.

Industrial partnership. An important asset of UPSaclay is its strong and rapidly growing industrial network. Some partners have a long standing tradition of industrial chairs, and most build upon their own array of industrial collaborations. For example:

- ENS Paris-Saclay has launched the *AI for industry* chair with ATOS, BERTIN, CEA, MICHELIN, SNCF. This chair organizes the startup challenge on AI sponsored by Région Ile-de-France;

- Inria-Saclay has developed bilateral collaborations with Fujitsu and Nokia; it has built a consortium to develop the *scikit learn* library that includes AXA, BNP Paribas, Cardif, BCG, Nvidia, Dataiku and Intel.

- Centrale Supelec has launched 3 education chairs related to AI with Randstad,

Givaudan and RTE.

- IBM has installed an R&D center at Saclay and is developing collaborations with UPSaclay on the topic of AI (Aida project).

To further foster exchange with industrial partners, the Dataia institute is setting up an industrial affiliation plan: discussions are ongoing with Renault, Total, GE Healthcare, Servier, IBM, AirLiquide. This framework will allow UPSaclay to monitor changes in AI-related jobs and their impact on the organization of training in the field, so as to avoid the pitfalls of a short-term vision in a fast-changing discipline. We plan to set up an "AI-jobs monitoring board" to bring together members of the Industrial Affiliate Program of Dataia with training directors and to organize events such as job fairs in AI.

Furthermore, the site can also take advantage of the existence of several Carnot institutes managed by key institutions of the site, which all have AI in their scope. These institutes can facilitate collaborations with industry partners. Carnot institutes active on the site are CentraleSupelec (Science des Systèmes), LIST-CEA, Inria-Saclay, CNRS (Tremplin Carnot Cognition). Inria-Saclay is currently developing a program for start-up development. Finally, UPSaclay collaborates on AI topics with IRT SystemX, Pole de compétitivité Systematic, and ITE Vedecom (sustainable mobility).

All these actions offer opportunities for the students to get acquainted with the types of activities currently developed in industry, as well as the skills to develop in order to meet applicative challenges. The critical asset of Université Paris-Saclay is the breadth of these collaborations, making it possible for students to find their place in a complex and stimulating ecosystem.

9 Impact and dissemination

The AI needs, as expressed by various research domains and by industrial partners, span a wide range of contexts. For instance, in the areas of personalized medicine and human resources, the goal is to deliver efficient and bias-free models and to handle uncertainty in the face of large, complex, heterogeneous and sensitive data. At another extreme, e.g. in autonomous vehicles, it is of primary importance to secure representative data from simulators, to achieve multi-scale learning for the sake of efficiency, and to assess the scope of validity of the learned models. The students, exposed to such a variety of fundamental and practical challenges, will have every opportunity to acquire and deploy a first rate expertise of data-scientist, addressing the much feared shortage of modern AI skills.

All participants in the PhD programme will be expected to reach out and disseminate their achievements along the following dimensions:

9.1   Dissemination to the international, national and local AI scientific community

Classical indicators of success include publications in the core AI international venues (non exhaustive list: AIJ, JMLR, MLJ, OR, EJOR; IJCAI, ECAI, NeurIPS, ICML, ICLR, KDD, ECMLPKDD, ICCV, ECCV, COLT, AISTATS, ICDM). Publication in both the above venues and in the relevant application domains will demonstrate the fundamental and applicative quality of the work. A HAL collection will be set up to group together all of UDOPIA's scientific production in an open archive.

The organization of symposia focusing on junior researchers (such as the Junior Conference on Data Science and Engineering organized in May 19) or workshops (such as those on Formal Methods for AI in June 19 and on Safety in AI in Sept. 19) will foster the dissemination of the advances and contribute to the creation of a (pluri-disciplinary) AI network at the UPSaclay scale. An invitation program, possibly in partnership with Institut Pascal or Maison des Sciences de l'Homme will support thematic weeks or months, focused on an emerging research question.

The UDOPIA programme will organise at least two symposia with proceedings:

- The UDOPIA symposium, bringing together the members of the program (doctoral students and supervisors) in a residential setting and preceded by an interdisciplinary thematic school in artificial intelligence. The course materials of the thematic school and the proceedings of the symposium will be widely disseminated.

- An international   conference of doctoral students organised within the framework of the European University EUGLOH around the theme of artificial intelligence for global health, well-being and environment issues, bringing together the doctoral students of the programme and the European University.

9.2   Dissemination to companies and to the general public

Doctoral students will be informed of the possibility of carrying out consultancy missions in AI with partner companies. An appropriate and transparent framework will be defined to

i) enable companies and in particular SMEs to identify the skills they need ;

ii) to implement a smooth contractualization process, establishing a bridge between the academic and entrepreneurial worlds.

In addition, the doctoral college offers a range of training and activities to doctoral students to encourage the dissemination and exploitation of their research work (MT180, Doct'Iliens festival, Docto'preneuriales...). From 2020 onwards, these activities should be devoted successively to each of the University's 7 major challenges, including AI.

Ma thèse en 180 secondes, inspired by *Three minute thesis* (3MT®, initially designed at the University of Queensland, Australia) lets doctoral students present their research topic, in French, to a general and diverse audience. For the past five years UPSaclay has organized one final round each year, widely attended by local citizens. Some doctoral schools organize a preliminary round to encourage and prepare candidates for the final round. Our plan is to generalize this widely popular initiative, focusing on each of the 7 University challenges including AI.

The *"Doct'Ilien : Ideas for... "* festival series is designed to invite PhD students to consider the possible exploitation of their research in a fun setting. Pairs of PhD students from different disciplines are formed during a short "creativity" training session. They are invited to imagine together a potential use of their research and trained to pitch their project. A hackathon, organized with societal and industrial partners, selects 4 to 6 such projects, presented during a public event. An edition of the "Doct'Ilien : Ideas for... " festivals series will be devoted to Artificial Intelligence.

Finally, the *Docto'preneurials* (two 1-week sessions respectively devoted to a Doctoral Camp and a Valorisation Camp) are intended for doctoral students who have a real business creation or transfer project. The purpose is to introduce themselves to the entrepreneurial ecosystem (SATT, Seed Fund, Incuballiance...), to meet entrepreneurs and to receive tailored training adapted to their project. Sessions focusing on AI and targeting the PhDs in the programme will be organized.