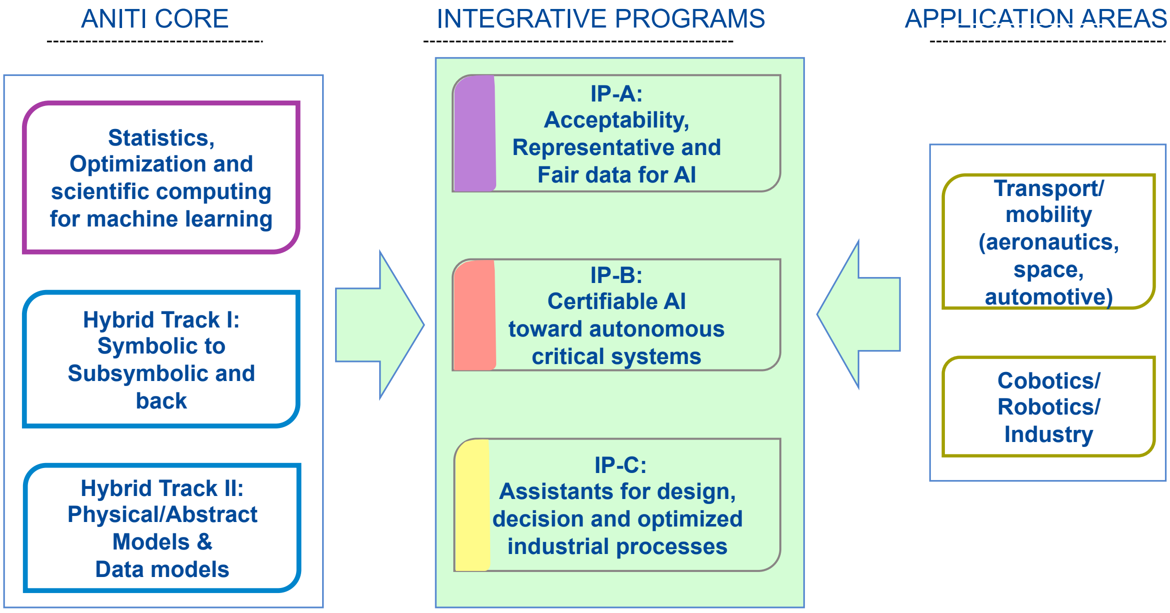
Institut 3iA - Artificial and Natural Intelligence Toulouse Institute

IPs and Proposed Chairs

1) ANITI Integrated Programs

As illustrated in the following figure, ANITI's scientific project is structured around three integrative programs (IP), based on advances in three core areas of artificial intelligence and in relation to the application sectors targeted by the project (Transport and mobility, Cobotics, robotics, industry 4.0).



The core subjects will explore solutions potentially allowing to obtain different properties that one seeks to satisfy, in particular related to the challenges derived from the application areas covered by the project: the robustness, the reliability, the explicability of the modeled phenomena or of the behavior of the system, the adaptability of the algorithms to changes of the environment, compliance with normative constraints, social or economic acceptability, or certifiability. Other criteria are also important and include the need to reduce the learning effort or optimize the performance of proposed solutions. In particular, two hybrid tracks are distinguished. Hybrid Track 1 addresses relations between subsymbolic representations, for example representations in a continuous, high dimensional space, and symbolic representations, for example a set of logical constraints. Hybrid Track 2 examines challenges related to how to complement and even replace analytical models of complex physical systems with data-driven ones exploiting machine learning.

In the following, we describe the three proposed IP with the proposed Chairs

1) A short synopsys of the IP taken from the project proposal
2) The list of proposed chairs with the names of the Principal investigator (PI), the topic and proposed starting date. The appendix provides more details for each chair. While the Chairs are all part of an IP and thus are linked both to the pull of application problems and questions and to the push of theoretical research, we distinguish broadly

three levels in how closely they are tied to applications or theory. The Chairs are grouped into three categories identified by three different colors:

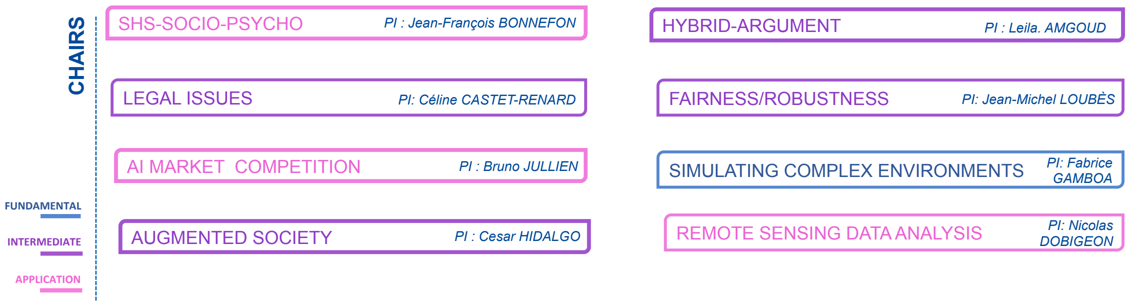
   a. The Blue chairs are closely tied to theoretical concerns whether they have to do with the mathematical foundations of machine learning, to hybrid systems or to symbolic based representations and reasoning
   b. The Purple chairs are somewhere in the middle between theoretical concerns and applications: they are driven by a mixture of each.
   c. The Pink chairs are more oriented application of AI. Note that an application of AI need not be an industrial concern, however. It might be an application of AI to legal thinking, for instance.

3) A more detailed description of the IP outlining the objectives and the connection with the proposed chairs

Note that some chairs can contribute to two IPs.

IPA: Acceptability, Fair, representative data for AI

Synopsis: This program addresses two main set of challenges.
The first set of challenges deals with different facets of the problem of social, economic, legal and ethical acceptability of systems integrating artificial intelligence algorithms. Acceptability is intimately related to the ability to explain and interpret the results and behaviors of these algorithms and the confidence that can be placed in them. It also involves trade-offs in the performance of these algorithms and their economic impact.
The second set focuses on data-related challenges. It aims to develop techniques to tackle problems related to 1) the existence of bias in the selection and interpretation of data used by models of learning 2) the prevention of the risk of not taking into account rare events in the learning process 3) automatic annotation and semantic representation of heterogeneous, multi-source, multi-scale, and time-varying data masses. These locks are common to several areas of application.



Chairs: see Appendix for details for each chair

Description:

The first part of this integrative program deals with several facets of AI acceptability and argues that the combination of outputs from ANITI core areas with research-based insights from social and human sciences can make AI systems more likely to be accepted. This has repercussions for all application areas.

AI acceptability depends to a large extent on its explainability, although all of the desirable properties for AI systems are in some sense relevant. We will study how to extract explanations and arguments from AI systems and how this relates to persuasion and bias (L. Amgoud chair). AI systems will typically be better accepted if they mimic human behavior/decisions. In particular we need to compare algorithmic decisions with human decisions in controlled environments(J.F. Bonnefon chair). Generally, elements of human cognition or "natural intelligence" like those discussed in Hybrid track 2 need to be integrated into AI systems to improve their acceptability. But explainability and "human-like" behavior of systems may involve trade-offs with other desirable properties (speed, quality). This IP will also investigate such tradeoffs

This IP will also study whether AI's potentially negative economic effects can be prevented (e.g. through regulation). In particular, AI's effects on the labor market are still not well understood and there are some potential dangers regarding the way AI affects competition. Moreover, the use of AI-based data analytics by firms will generate "network effects" that will affect their strategies and the benefits that consumers reap from using their products (B. Jullien chair).

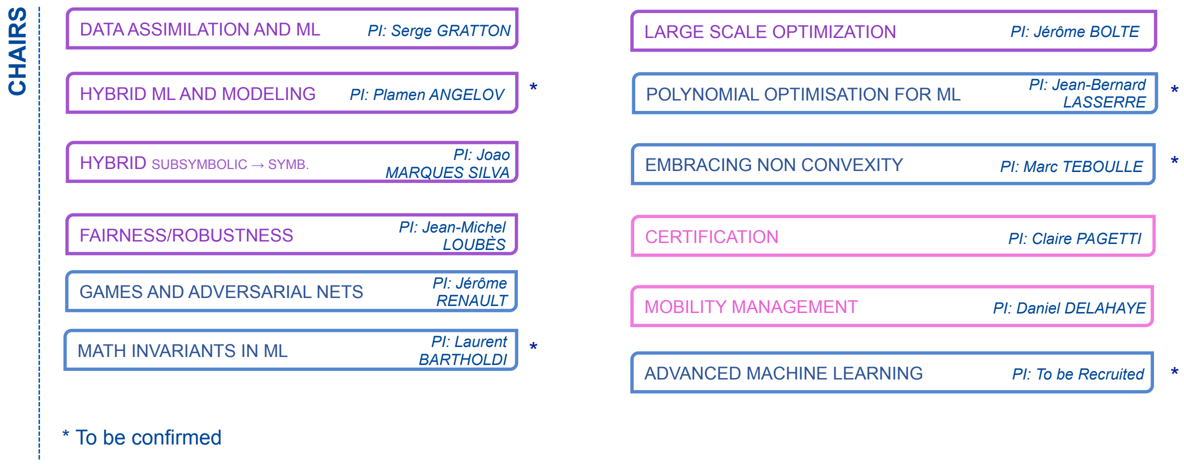
Another factor that could affect AI's acceptability are the risks for individuals and societies: explainability, fairness, non-discrimination, transparency, privacy, security, safety, and democracy concerns are the most frequently mentioned. AI needs to be human-centric, and respectful of fundamental rights. In a context of rapid technological change, it is essential that trust encourages the innovation and deployment of technologies. Trust in AI means to trust in a robust technology, as well as in the rules, laws, and norms that govern it. Most of them are outdated and we will develop a human-centric framework for trustworthy and competitive AI based on new legal rules, accountability governance of AI, and ethical purposes. These topics will be addressed by our local chair C.Castets-Renard and external chair C. Hidalgo.

The second part of this program will provide several methods for improving data sets for ML. First we want fair data and algorithms. Biases can affect fairness and we will develop methods for controlling the bias of a learning algorithm (optimization under fairness constraints) as well as removing the bias of a learning sample (via optimal transport). We also will model bias in the learning sample, via independency constraints (local chair J.M. Loubes). This problem also has legal implications (local chair C. Castet-Renard). We will extend our model with theoretical studies of interpretive bias, which is part of Amgoud's chair. This will give us a fuller picture of ways to find unwanted biases and avoid them.

We will also address the problem of how to prevent the erasing of improbable but extremely important events for the learning process, for which hybrid models may be useful. Another component will apply our expertise in image analysis to multi source, multi-scale and temporally dynamic data, which is especially relevant for climate analysis that is needed to support various application sectors such as transportation and mobility as well as many other fields (N. Dobigeon local chair). Fabrice Gamboa chair will explore computer simulations for physical, chemical or biological phenomena, and seek to improve their analysis with application to various data driven deep learning model. To help with this, we will also study hybrid methods for the semantic labeling of data.

IPB: Certifiable AI functions toward autonomous safety critical systems

Synopsis: This program aims to develop new methods, models and tools based on hybrid AI to design and validate adaptive autonomous critical systems for which strong guarantees are required in particular from certification authorities (e.g. in Aeronautics), e.g., dependability, explicability, etc. Different levels will be considered, ranging from a component or embedded basic function, to the autonomous vehicle, to a fleet of vehicles or to the global transport system. In addition to the autonomy of the operations, the work will concern the autonomy of the maintenance activities including the automation of the monitoring activities of the systems and the associated recovery and reconfiguration actions.



Chairs: see Appendix for details for each chair

Description:

This integrative program focuses on certifiable AI for safety assured autonomous systems that will become a ubiquitous feature of future transport and mobility. We will develop methods, models and tools for hybrid AI solutions that define the functions required for adaptable autonomous systems, from sensors to systems of systems (e.g. fleets). We will also study tools to define and to verify the properties required for their proper functioning.

Optimisation methods are crucial for all deep learning approaches from convolutional networks to the most sophisticated GANs. Their training boils down in the final analysis to some high dimensional optimisation problem with an often complex objective or loss function. While most of the currently fashionable optimisation methods have been around for a while, we still don't understand why simple methods like gradient descent or various first order methods work so well.

To determine exactly what are the limits of reliability and robustness for various learning systems, the local optimisation chairs (J. Bolte, J.B. Lasserre), reinforced with the external chair Teboulle, will examine problems of non-convexity, convergence rates of optimization methods, statistical risk bounds, adversarial regret bounds, and information-theoretic risk lower bounds. Using advanced mathematical ideas from geometry, game theory and topology the local chairs of Loubes/Renault and the Bartholdi external chair will also contribute to our understanding of learning architectures. To further strengthen research in this area, we also plan to hire one of our eminent visitors (Kwiatkowska, Anandkumar, Mausam) in machine learning to reinforce this area.

However, providing absolute guarantees of reliability and robustness is particularly difficult if not impossible to achieve for many modern machine learning systems when considered in isolation. Using translation techniques from subsymbolic to symbolic representations, the external hybrid track 1 chair Marques-Silva will verify or guarantee the robustness and explainability of learning models relevant to autonomous systems. This challenge can also be addressed when such systems are embedded in a larger context---for instance, when integrated in complementarity with other models as suggested in Hybrid track 2; this work will be carried out by chairs Angelov and Gratton. Integrating this work with practical applications will be the subject of the chairs Pagetti/Delahaye, as well as the engineers associated with this IP.

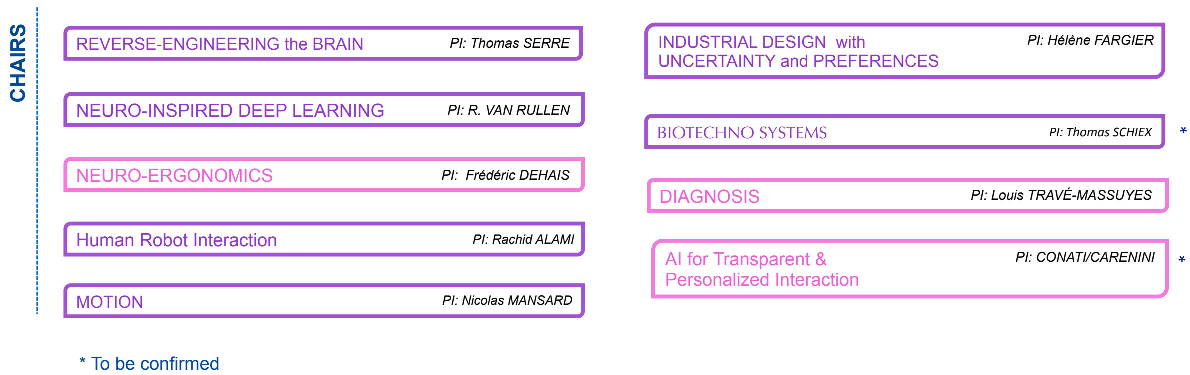
Autonomy concerns not only operations but also maintenance. As an example, we have in mind the automation of at least components of 100 hour inspections and their consequences. This will require techniques of similarity analysis that exploit both physical models and data-driven techniques, as well as symbolic reasoning and decision making. Another example is automated troubleshooting and reconfiguration of various systems like satellite constellations and aircraft.

IP-C: Assistants for design, decision and optimized industry processes

Synopsis: This program will examine the interaction between humans and AI systems and their complementarity in order to increase the performance of the activities implemented, thanks to hybrid AI. Two parts will be explored. In the first, the system integrating AI is a cognitive assistant that contributes to different human activities: a driver or a passenger in transport; a doctor, a multi-disciplinary medical team or a patient or user needing to be guided in his or her health care path, in an administration or a banking institution. These assistants will incorporate advanced dialogue and interaction capabilities, the design of which will be based on models characterizing human behavior from the cognitive sciences. The second part deals with the monitoring of complex systems in order to model their behavior and predict their evolution, to detect precursor symptoms that make it possible to anticipate incidents and to carry out corrective actions proactively. These issues are important for all ANITI application areas.

This program also focuses on the study and design of autonomous mobile robots with human, cognitive and physical interaction capabilities (manipulate / move objects, apply forces, ...), in order to perform complex tasks in collaborative way. This requires advances in several areas 1) the perception of the environment (location and detection of the presence of humans, fusion and integration of multi-modal and multi-sensory data, etc.) 2) the generation of motion for anthropomorphic robots with multiple degrees of freedom and which are subject to several constraints inherent in their mode of mobility and dynamics) the physical and cognitive interaction between humans and robots which involves the development of learning techniques by discussion and by observation and interpretation of actions performed by humans. These topics have applications in several areas including industry 4.0 and services.

The last topic covered by this program concerns the investigation of AI methods for automatically exploring, evaluating and optimizing designs of new systems and products (e.g., aircraft design based on market prediction and competitor assessment and other products whose design involves beliefs and preferences of consumers or users of the product)



Chairs: see Appendix for details for each chair

Description:

This IP will address three main topics, namely: 1) AI systems as assistants or monitors, 2) AI for advanced robot motion and interaction capabilities, and 3) AI for design of innovative systems and products

<u>1) AI systems as Assistants or Monitors</u>: We will examine the interaction between - and exploit the complementarity of - humans and AI systems as monitors or assistants. Monitoring will enhance predictive maintenance in mobility/aerospace that avoids operation disruption and

moves from fleet level systematic inspection to on-condition individual maintenance as well as predictive and personalized medicine that builds digital models of individuals and populations to reduce hospital readmissions and stays in an efficient, and patient-centered manner.

AI systems can assist with many human activities, and cognitive assistants are a showpiece for hybrid AI. While learning basic interactions involving speech and movements seems feasible with current deep learning methods, the ability to persuade a human counterpart or to take decisions in complex systems will probably rely on higher level logical representations advanced language and dialogue capabilities, but also natural intelligence architectures.
For this reason, we will exploit cognitive science based models (T. Serre external chair, R. van Rullen and F. Dehais local chairs). They will work on multimodal processing, vision and language to develop cognitive science inspired AI algorithms. Amgoud's and Van Rullen's chairs will be relevant for language and dialogue capabilities, as will international fellows Carenini and Conati. We are also proposing a local senior chair F. Dehais for enhancing human interactions with artificial agents through neuro-ergonomics.

Applications will include monitoring and predicting the evolution of complex systems, diagnosis of faults, detecting anomalies and anticipating incidents, by physical models, symbolic reasoning and decision making, and data-driven techniques, prognostics & health management of a vehicle of aircraft, vehicles, robots and machines in manufacturing (L. Travé chair) or mobility and traffic management of autonomous UAV systems (F. Delahaye chair).

<u>2) AI for advanced robot motion and interaction capabilities</u>: This research will endow mobile robots with the capability to move and interact autonomously—acting physically on the environment (applying forces, carrying tools, manipulating objects,…) but also interacting cognitively with humans. This requires advanced computational abilities and integrating motion plans and sensor-based control loops to deal with problems of perception, interpretation, motion generation and cognitive reasoning.

Mobile robots are complex systems that include many degrees of freedom and are subject to numerous constraints inherent in their dynamics and their locomotion mode. Though model-based trajectory planning methods guarantee awareness of the physics of systems, they become extremely tedious as soon accurate representations are required. Furthermore, they do not allow robots to exploit knowledge from past motion experience. We will rely on model-based trajectory planning to explore possible motion strategies and reinforce the motion generator, using what has already been learned in an incremental way, thus providing a particular kind of hybrid model based AI. Nicolas Mansard is our local candidate chair for this topic.

In the context of Industry 4.0, service robots must move in a robust, largely autonomous, adaptive and safe way amidst humans (local chair R. Alami). To enhance interaction, we will take natural language capabilities and theory of mind and integrate them with situated robotic knowledge. This will help capture referential meaning, a component of meaning needed for sophisticated AI systems, by injecting constraints about the physical world linking internal content with motion, action and situational intelligence. This will also extend Toulouse research on situated conversation.

<u>3) AI for design of innovative systems and products</u>: This topic is about the use of AI methods for automatically exploring, evaluating and optimizing designs of innovative systems and products — systems aircraft design based on market prediction and competitor assessment and other products whose design involves beliefs and preferences of consumers or users of the product (local chair H. Fargier and J. Renault), proteins (local chair Schiex).

These efforts will use work on mixed and multidisciplinary optimization in integrating physical and data-driven models for mission-driven product design. We will also deal with design issues involving beliefs and preferences of consumers or users of the designed system. Designing distributed systems with both cooperative and non-cooperative agents is important for many businesses, but especially the air transport industry (aircraft design based on market prediction and competitor assessment) and the automotive industry (autonomous vehicles mixing with piloted vehicles). We will combine symbolic game theoretic models with complex, high dimensional data sets describing actual decisions.