

# Artificial Intelligence

Current challenges and Inria's engagement





## **O. Researchers in Inria project-teams and centres who contributed to this document (were interviewed, provided text, or both)<sup>1</sup>**

Abiteboul Serge\*, former DAHU project-team, Saclay  
Alexandre Frédéric\*\*, head of MNEMOSYNE project-team, Bordeaux  
Altman Eitan\*\*, NEO project-team, Sophia-Antipolis  
Amsaleg Laurent\*\*, head of LINKMEDIA project-team, Rennes  
Antoniou Gabriel\*\*, head of KERDATA project-team, Rennes  
Arlot Sylvain\*\*, head of CELESTE project-team, Saclay  
Ayache Nicholas\*\*\*, head of EPIONE project-team, Sophia-Antipolis  
Bach Francis\*\*\*, head of SIERRA project-team, Paris  
Beaudouin-Lafon Michel\*\*, EX-SITU project-team, Saclay  
Beldiceanu Nicolas\*, head of former TASC project-team, Nantes  
Bellet Aurélien\*\*, head of FLAMED exploratory action, Lille  
Bezerianos Anastasia \*\*, ILDA project-team, Saclay  
Bouchez Florent\*\*, head of AI4HI exploratory action, Grenoble  
Boujemaa Nozha\*, former advisor on bigdata for the Inria President  
Bouveyron Charles\*\*, head of MAASAI project-team, Sophia-Antipolis  
Braunschweig Bertrand\*\*\*, director, coordination of national AI research programme  
Brémond François\*\*\*, head of STARS project-team, Sophia-Antipolis  
Brodu Nicolas\*\*, head of TRACME exploratory action, Bordeaux  
Cazals Frédéric\*\*, head of ABS project-team, Sophia-Antipolis  
Casiez Géry\*\*, LOKI project-team, Lille  
Charpillet François\*\*\*, head of LARSEN project-team, Nancy  
Chazal Frédéric\*\*, head of DATASHAPE project-team, Saclay and Sophia-Antipolis  
Colliot Olivier\*\*\*, head of ARAMIS project-team, Paris  
Cont Arshia\*, head of former MUTANT project-team, Paris

---

<sup>1</sup> (\*): first edition, 2016; (\*\*): second edition, 2020; (\*\*\*) both editions

Cordier Marie-Odile\*, LACODAM project-team, Rennes

Cotin Stephane\*\*, head of MIMESIS project-team, Strasbourg

Crowley James\*\*\*, former head of PERVASIVE project-team, Grenoble

Dameron Olivier\*\*, head of DYLISS project-team, Rennes

De Charette, Raoul\*\*, RITS project-team, Paris

De La Clergerie Eric\*, ALMANACH project-team, Paris

De Vico Fallani Fabrizio\*, ARAMIS project-team, Paris

Deleforge Antoine\*\*, head of ACOUST.IA2 exploratory action, Nancy

Derbel Bilel\*\*, BONUS project-team, Lille

Deriche Rachid\*\*, head of ATHENA project-team, Sophia-Antipolis

Dupoux Emmanuel\*\*, head of COML project-team, Paris

Euzenat Jérôme\*\*\*, head of MOEX project-team, Grenoble

Fekete Jean-Daniel\*\*, head of AVIZ project-team, Saclay

Forbes Florence\*\*, head of STATIFY project-team, Grenoble

Franck Emmanuel\*\*, head of MALESI exploratory action, Nancy

Fromont Elisa, \*\*, head of HYAIAI Inria challenge, Rennes

Gandon Fabien\*\*\*, head of WIMMICS project-team, Sophia-Antipolis

Giavitto Jean-Louis\*, former MUTANT project-team, Paris

Gilleron Rémi\*, MAGNET project-team, Lille

Giraudon Gérard\*, former director of Sophia-Antipolis Méditerranée research centre

Girault Alain\*\*, deputy scientific director

Gravier Guillaume\*, former head of LINKMEDIA project-team, Rennes

Gribonval Rémi\*\*, DANTE project-team, Lyon

Gros Patrick\*, director of Grenoble-Rhône Alpes research centre

Guillemot Christine\*\*, head of SCIROCCO project-team, Rennes

Guitton Pascal\*, POTIOC project-team, Bordeaux

Horaud Radu\*\*\*, head of PERCEPTION project-team, Grenoble

Jean-Marie Alain\*\*, head of NEO project-team, Sophia-Antipolis

Laptev Ivan\*\*, WILLOW project-team, Paris

Legrand Arnaud\*\*, head of POLARIS project-team, Grenoble

Lelarge Marc\*\*, head of DYOGENE project-team, Paris

Mackay Wendy\*\*, head of EX-SITU project-team, Saclay

Malacria Sylvain\*\*, LOKI project-team, Lille

Manolescu Ioana\*, head of CEDAR project-team, Saclay

Mé Ludovic\*\*, deputy scientific director

Merlet Jean-Pierre\*\*, head of HEPHAISTOS project-team, Sophia-Antipolis

Maillard Odalric-Ambrym\*\*, head of SR4SG exploratory action, Lille

Mairal Julien\*\*, head of THOTH project-team, Grenoble

Moisan Sabine\*, STARS project-team, Sophia-Antipolis

Moulin-Frier Clément\*\*, head of ORIGINS exploratory action, FLOWERS project-team, Bordeaux

Mugnier Marie-Laure\*\*\*, head of GRAPHIK project-team, Montpellier

Nancel Mathieu\*\*, LOKI project-team, Lille

Nashashibi Fawzi\*\*\*, head of RITS project-team, Paris

Neglia Giovanni\*\*, head of MAMMALS exploratory action, Sophia-Antipolis

Niehren Joachim\*, head of LINKS project-team, Lille

Norcy Laura\*\*, European partnerships

Oudeyer Pierre-Yves\*\*\*, head of FLOWERS project-team, Bordeaux

Pautrat Marie-Hélène\*\*, director of European partnerships

Pesquet Jean-Christophe\*\*, head of OPIS project-team, Saclay

Pietquin Olivier\*, former member of SEQUEL project-team, Lille

Pietriga Emmanuel\*\*, head of ILDA project-team, Saclay

Ponce Jean\*, head of WILLOW project-team, Paris

Potop Dumitru\*\*, KAIROS project-team, Sophia-Antipolis

Preux Philippe\*\*\*, head of SEQUEL (SCOOOL) project-team, Lille

Roussel Nicolas\*\*\*, director of Bordeaux Sud Ouest research centre

Sagot Benoit\*\*\*, head of ALMANACH project-team, Paris

Saut Olivier\*\*, head of MONC project-team, Bordeaux

Schmid Cordelia\*, former head of THOTH project-team, Grenoble, now in WILLOW project-team, Paris

Schoenauer Marc\*\*\*, co- head of TAU project-team, Saclay

Sebag Michèle\*\*\*, co- head of TAU project-team, Saclay

Seddah Djamé\*, ALMANACH project-team, Paris

Siegel Anne\*\*\*, former head of DYLISS project-team, Rennes

Simonin Olivier\*\*\*, head of CHROMA project-team, Grenoble

Sturm Peter\*, deputy scientific director

Termier Alexandre\*\*\*, head of LACODAM project-team, Rennes

Thiebaut Rodolphe\*\*, head of SISTM project-team, Bordeaux

Thirion Bertrand\*\*, head of PARIETAL project-team, Saclay

Thonnat Monique\*, STARS project-team, Sophia-Antipolis

Tommasi Marc\*\*\*, head of MAGNET project-team, Lille

Toussaint Yannick\*, ORPAILLEUR project-team, Nancy

Valcarcel Orti Ana\*\*, coordination of national AI research programme

Vercouter Laurent\*\*, coordination of national AI research programme

Vincent Emmanuel\*\*\*, MULTISPEECH project-team, Nancy

# Index

0. Researchers in Inria project-teams and centres who contributed to this document (were interviewed, provided text, or both) .....	3
1. Samuel and his butler .....	8
2. A recent history of AI .....	11
3. Debates about AI .....	19
4. Inria in the national AI strategy .....	24
5. The Challenges of AI and Inria contributions .....	26
5.1 Generic challenges in artificial intelligence .....	30
5.2 Machine learning .....	33
5.3. Signal analysis, vision, speech .....	62
5.4. Natural language processing .....	77
5.5 Knowledge-based systems and semantic web .....	81
5.6 Robotics and autonomous vehicles .....	91
5.7 Neurosciences and cognition .....	104
5.8 Optimisation .....	116
5.9 AI and Human-Computer Interaction (HCI) .....	125
6. European and international collaboration on AI at Inria .....	139
7. INRIA REFERENCES: NUMBERS .....	145
8. Other references for further reading .....	146

## 1. Samuel and his butler<sup>2</sup>



7:15 a.m., Sam wakes up and prepares for a normal working day. After a quick shower, he goes and sits at the kitchen table for breakfast. *Toi.Net*<sup>3</sup>, his robot companion, brings warm coffee and a plate of fresh fruits. “Toi.Net, Pass me the sugar please”, Sam says. The robot brings the sugar shaker from the other end of the breakfast table – there is a sugar box in the kitchen cupboard but *Toi.Net* knows that it is much more convenient to use the shaker.

“Any interesting news?”, Sam asks. The robot guesses s/he must find news that correspond to Sam’s topics of interest. S/he starts with football.

*Toi.Net*: “Monaco beat Marseille 3-1 at home, it is the first time they score three goals against Marseille since the last twelve years. A hat trick by Diego Suarez.”

*Toi.Net*: “The Eurovision song contest took place in Ljubljana; Poland won with a song about friendship in social networks.”

---

<sup>2</sup> The title of this section is a reference to Samuel Butler, a 19<sup>th</sup>-century English novelist, author of *Erewhon*, one of the first books to speculate about the possibility of an artificial intelligence grown by Darwinian selection and reproduction among machines.

<sup>3</sup> Pronounce ‘*tʌnət*’, after the name of the maid-servant in Molière’s «*The imaginary invalid*»



Sam: "Please don't bother me again with this kind of news, I don't care about the Eurovision contest."

Toi.Net: "Alright. I won't."

Toi.Net: "The weather forecast for Paris is sunny in the morning, but there will be some heavy rain around 1:00p.m. and in the afternoon"

Toi.Net: "Mr. Lamaison, a candidate for the presidency of the South-west region, declared that the unemployment level reached 3.2 million, its highest value since 2004."

Sam: "Can you check this? I sort of remember that the level was higher in the mid 2010s."

Toi.Net (after two seconds): "You're right, it went up to 3.4 million in 2015. Got that from INSEE semantic statistics."

By the end of the breakfast, Sam does not feel very well. His connected bracelet indicates abnormal blood pressure and Toi.Net gets the notification. "Where did you leave your pills?" S/he asks Sam. "I left them on the nightstand, or maybe in the bathroom". Toi.Net brings the box of pills, and Sam quickly recovers.

Toi.Net: "It's time for you to go to work. Since it will probably be raining when you go for a walk in the park after lunch, I brought your half boots."

An autonomous car is waiting in front of the house. Sam enters the car, which announces "I will take a detour through A-4 this morning, since there was an accident on your usual route and a waiting time of 45 minutes because of the traffic jam".

Toi.Net is a well-educated robot. S/he knows a lot about Sam, understands his requests, remembers his preferences, can find objects and act on them, connects to the internet and extracts relevant information, learns from new situations. This has only been possible thanks to the huge progresses made in artificial intelligence: speech processing and understanding (to understand Sam's requests); vision and object recognition (to locate the sugar shaker on the table); automated planning (to define the correct sequences of action for reaching a certain situation such as delivering a box of pills located in another room); knowledge representation (to identify a hat trick as a series of three goals made by the same football player); reasoning (to decide to pick the sugar shaker rather than the sugar box in the cupboard, or to use weather forecast data to decide which pair of shoes Sam should wear); data mining (to extract relevant news from the internet, including fact checking in the case of the political declaration); Her/his incremental machine learning algorithm will make her/him remember not to mention Eurovision contests in the future; s/he continuously adapts her/his interactions with Sam by building her/him owner's profile and by detecting his emotions.

By being a little provocative, we can say that Artificial intelligence does not exist... but obviously, the combined power of available data, algorithms and computing resources

opens up tremendous opportunities in many areas. Inria, with its 200+ project-teams, mostly joint teams with the key French Universities, in eight research centres, is active in all these scientific areas. This white paper presents our views on the main trends and challenges in Artificial Intelligence (AI) and how our teams are actively conducting scientific research, software development and technology transfer around these key challenges for our digital sovereignty.

## 2. A recent history of AI

It's on everyone's lips. It's on television, radio, newspapers, social networks. We see AI in movies, we read about AI in science fiction novels. We meet AI when we buy our train tickets online or surf on our favourite social network. When we type its name on a search engine, the algorithm finds up to 16 million references ... Whether it fascinates us often or worries us sometimes, what is certain is that it pushes us to question ourselves because we are still far from knowing everything about it. For all that, and this is a certainty, artificial intelligence is well and truly among us. The last years were a period in which the companies and specialists from different fields (e.g. Medicine, Biology, Astronomy, Digital Humanities) have developed a specific and marked interest for AI methods. This interest is often coupled with a clear view on how AI can improve their workflows. The amount of investment of both private companies and governments is also a big change for research in AI. Major Tech companies but also an increasing number of industrial companies are now active in AI research and plan to invest even more in the future, and many AI scientists are now leading the research laboratories of these and other companies.

AI research produced major progress in the last decade, in several areas. The most publicised are those obtained in machine learning, thanks in particular to the development of deep learning architectures, multi-layered convolutional neural networks

learning from massive volumes of data and trained on high performance computing systems.

Be it in game resolution, image recognition, voice recognition and automatic translation, robotics..., artificial intelligence has been infiltrating a large number of consumer and industrial applications over the last ten years that are gradually

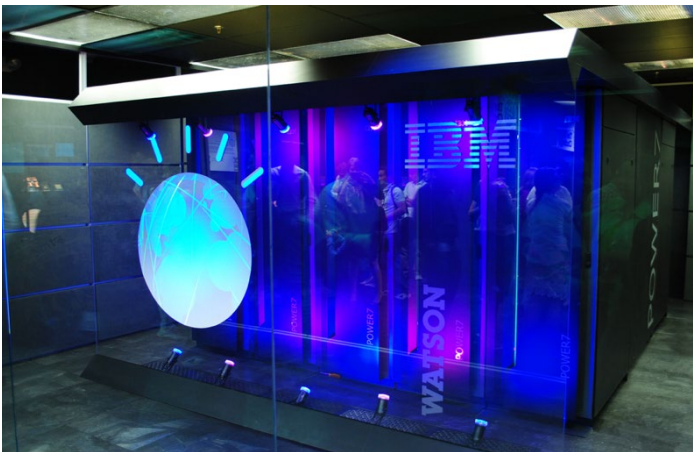


Figure 1: IBM Watson Computer

revolutionizing our relationship with technology.

In 2011, scientists succeeded in developing an artificial intelligence capable of processing and understanding language. The proof was made public when IBM Watson software won the famous game Jeopardy. The principle of the game is to provide the question to a given answer as quickly as possible. On average, players take

three seconds before answering. The program had to be able to do as well or even better in order to hope to beat the best of them: language processing, high-speed data mining, ranking proposed solutions by probability level, all with a high dose of intensive computing. In the line of Watson, Project Debater can now make structured argumentation discussing with human experts – using a mix of technologies (<https://www.research.ibm.com/artificial-intelligence/project-debater/>).

In another register, artificial intelligence shone again in 2013 thanks to its ability to master seven Atari video games (personal computer dating from the 1980-90s). Reinforcement learning developed in Google DeepMind's software allowed its program to learn how to play seven video games, and above all how to win by having as sole information the pixels displayed on the screen and the score. The program learned by itself, through its own experience, to continuously improve and finally win in a systematic way. Since then, the program has won about thirty different Atari games. The exploits are even more numerous on strategic board games, notably with Google Deepmind's AlphaGo which beat the world go champion in 2016 thanks to a combination of deep learning and reinforcement learning, combined with multiple trainings with humans, other computers, and itself. The algorithm was further improved in the following versions: in 2017, AlphaZero reached a new level by training only against itself, i.e. by self-learning. On a go, chess or checkers board, both players know the exact situation of the game at all times. The strategies are calculable in a way: according to the possible moves, there are optimal solutions and a well-designed program is able to identify them. But what about a game made of bluff and hidden information? In 2017, Tuomas Sandholm of Carnegie-Mellon University presented the Libratus program that crushed four of the best players in a poker competition using learning, see <https://www.cs.cmu.edu/~noamb/papers/17-IJCAI-Libratus.pdf>. By extension, AI's resolution of problems involving unknowns could benefit many areas, such as finance, health, cybersecurity, defence. However, it should be noted that even the board games with incomplete information that AI recently "solved" (poker, as described above, StarCraft, by DeepMind, Dota2 by Open AI) take place in a known universe: the actions of the opponent are unknown, but their probability distribution is known, and the set of possible actions is finite, even if huge. On the opposite, real world generally involves an infinite number of possible situations, making generalisation much more difficult.

Recent highlights also include the progress made in developing autonomous and connected vehicles, which are the subject of colossal investments by car manufacturers gradually giving concrete form to the myth of the fully autonomous vehicle with a totally passive driver who would thus become a passenger. Beyond the manufacturers' commercial marketing, the progress is quite real and also heralds a strong development of these technologies, but on a significantly different time scale. Autonomous cars have driven millions of kilometres with only a few major incidents happening. In a few years, AI has established itself in all areas of Connected Autonomous Vehicles (CAV), from perception to control, and through decision, interaction and supervision. This opened the way to previously ineffective solutions and opened new research challenges (e.g. end-to-end driving) as well. Deep Learning in particular became a common and versatile tool, easy to implement and to deploy.

This has motivated the accelerated development of dedicated hardware and architectures such as dedicated processing cards that are integrated by the automotive industry on board real autonomous vehicles and prototype platforms.

In its white paper, *Autonomous and Connected Vehicles: Current Challenges and Research Paths*, published in May 2018, Inria nevertheless warns about the limits of large-scale deployment: *"The first automated transport systems, on private or controlled access sites, should appear from 2025 onwards. At that time, autonomous vehicles should also begin to drive on motorways, provided that the infrastructure has been adapted (for example, on dedicated lanes). It is only from 2040 onwards that we should see completely autonomous cars, in peri-urban areas, and on test in cities,"* says Fawzi Nashashibi, head of the RITS project team at Inria and main author of the white paper. *"But the maturity of the technologies is not the only obstacle to the deployment of these vehicles, which will largely depend on political decisions (investments, regulations, etc.) and land-use planning strategies,"* he continues.

In the domain of health and medicine, see for example Eric Topol's book "Deep Medicine" which shows dozens of applications of deep learning in about all aspects of health, from radiography to diet design and mental remediation. A key achievement over the past three years is the performance of Deepmind in CASP (Critical Assessment of Structure Prediction) with AlphaFold, a method which significantly outperformed all contenders for the sequence to protein structure prediction. These results open a new era: it might be possible to obtain high-resolution structures for the vast majority of protein sequences for which only the sequence is known. Another key achievement is the standardization of knowledge in particular on biological regulations which are very complex to unify (BioPAX format) and the numerous knowledge bases available (Reactome, Rhea, pathwaysCommons...). Let us also mention the interest and energy shown by certain doctors, particularly radiologists, in the tools related to diagnosis and automatic prognosis, particularly in the field of oncology. In 2018, FDA permitted marketing of IDx-DR (<https://www.eyediagnosis.co/>), the first medical device to use AI to detect greater than a mild level of diabetic retinopathy in the eye of adults who have diabetes (<https://doi.org/10.1038/s41433-019-0566-0>).

In the aviation sector, the US Air Force has developed, in collaboration with the company Psibernetix, an AI system capable of beating the best human pilots in aerial combat<sup>4</sup>. To achieve this, Psibernetix combines fuzzy logic algorithms and a genetic algorithm, i.e. an algorithm based on the mechanisms of natural evolution. This allows AI to focus on the essentials and break down its decisions into the steps that need to be resolved to achieve its goal.

At the same time, robotics is also benefiting from many new technological advances, notably thanks to the Darpa Robotics Challenge, organized from 2012 to 2015 by the US Department of Defense's Advanced Research Agency (<https://www.darpa.mil/program/darpa-robotics-challenge> ). This competition

---

<sup>4</sup> [https://magazine.uc.edu/editors\\_picks/recent\\_features/alpha.html](https://magazine.uc.edu/editors_picks/recent_features/alpha.html)

proved that it was possible to develop semi-autonomous ground robots capable of performing complex tasks in dangerous and degraded environments: driving vehicles, operating valves, progressing in risky environments. These advances point to a multitude of applications be they military, industrial, medical, domestic or recreational.

Other remarkable examples are:

- Automatic description of the content of an image (“a picture is worth a thousand words”), also by Google (<http://googleresearch.blogspot.fr/2014/11/a-picture-is-worth-thousand-coherent.html>)
- The results of Imagenet’s 2012 Large Scale Visualisation Challenge, won by a very large convolutional neural network developed by University of Toronto (<http://image-net.org/challenges/LSVRC/2012/results.html>)
- The quality of face recognition systems such as Facebook’s, <https://www.newscientist.com/article/dn27761-facebook-can-recognise-you-in-photos-even-if-youre-not-looking#.VYkVxFzjZ5g>
- Flash Fill, an automatic feature of Excel, guesses a repetitive operation and completes it (programming by example). [Sumit Gulwani: Automating string processing in spreadsheets using input-output examples. POPL 2011: 317-330.](#)
- PWC-Net by Nvidia won the 2017 optical flow labelling competition on MPI Sintel and KITTI 2015 benchmarks, using deep learning and knowledge models. <https://arxiv.org/abs/1709.02371>
- Speech processing is now a standard feature of smartphones and tablets with artificial companions including Apple’s Siri, Amazon’s Alexa, Microsoft’s Cortana and others. Google Meet transcripts speech of meeting participants in real time. Waverly Labs’ Ambassador earbuds translate conversations in different languages, simultaneous translation has been present in Microsoft’s Skype since many years.

The screenshot displays a Google search result for 'Fermé', a restaurant in Paris. The main heading is 'Fermé' with the subtitle 'Mondol Kiri, Horaires du lundi'. Below this, there are links to TripAdvisor and Yelp, both providing a 4.5 rating and 158 reviews. A snippet from Yelp describes the restaurant as 'Cambodian' and lists its hours. To the right, there is a map showing the location at Avenue Edison and Parc de Choisy, along with a photo of the restaurant's exterior. Further down, the restaurant's name 'Mondol Kiri' is repeated, showing a 4.5 rating from 57 Google reviews and a description: 'Restaurant cambodgien'. The address is listed as 159 Avenue de Choisy, 75013 Paris, and the phone number is 01 53 79 75 96. The hours are listed as 'Fermé aujourd'hui'.

Figure 2 : Semantic Information added to Google search engine results

It is also worth mentioning the results obtained in knowledge representation and reasoning, ontologies and other technologies for the semantic web and for linked data:

- Google Knowledge Graph improves the search results by displaying structured data on the requested search terms or sentences. In the field of the semantic web, we observe the increased capacity to respond to articulated requests such as "Marie Curie daughters' husbands" and to interpret RDF data that can be found on the web.

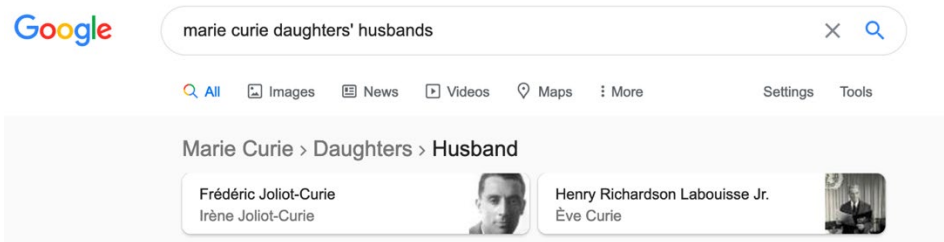


Figure 3: Semantic processing on the web

- Schema.org<sup>5</sup> contains millions of RDF (Resource Description Framework) triplets describing known facts: search engines can use this data to provide structured information upon request.
- The OpenGraph protocol – which uses RDFa – is used by Facebook to enable any web page to become a rich object in a social graph.

Another important trend is the recent opening of several technologies that were previously proprietary, in order for the AI research community to benefit from them but also to contribute with additional features. Needless to say that this opening is also a strategy of Big Tech for building and organizing communities of skills and of users focused on their technologies. Examples are:

- IBM's cognitive computing services for Watson, available through their Application Programming Interfaces, offers up to 20 different technologies such as speech-to-text and text-to-speech, concepts identification and linking, visual recognition and many others: <https://www.ibm.com/watson>
- Google's TensorFlow is the most popular open source software library for machine learning; <https://www.tensorflow.org/>. A good overview of the major machine learning open source platforms can be found on <http://aiindex.org>

---

<sup>5</sup> <https://schema.org/>

- Facebook opensourced its Big Sur hardware design for running large deep learning neural networks on GPUs: <https://ai.facebook.com/blog/the-next-step-in-facebooks-ai-hardware-infrastructure/>

In addition to these formerly proprietary tools, some libraries were natively developed as open source software. This is the case for example of the Scikit-learn library (see Section 5.2.5), one strategic asset in the Inria's engagement in the field.

Finally, let us look at a few scientific achievements of AI to conclude this chapter:

- Machine learning:
  - o Empirical questioning of theoretical statistical concepts that seemed firmly established. Theory had clearly suggested that the over-parameterized regime should be avoided to avoid the pitfall of over-learning. Numerous experiments with neural networks have shown that behaviour in the over-parameterized regime is much more stable than expected, and have generated a new effervescence to understand theoretically the phenomena involved.
  - o Statistical physics approaches have been used to determine fundamental limits to feasibility of several learning problems, as well as associated efficient algorithms.
  - o Embeddings (low-dimensional representations of data) were developed and used as input of deep learning architectures for almost all representations e.g. word2vec for natural language, graph2vec for graphs, math2vec for mathematics, bio2vec for biological data etc.
  - o Alignment of graphs or of clouds of points has made big progress both in theory and in practice, yielding e.g. surprising results on the ability to construct bilingual dictionaries in an unstructured manner.
  - o Transformers using very large deep neural networks and attention mechanisms have moved the state of the art of natural language processing to new horizons. Transformer-based systems are able to entertain conversations about any subject with human users.
  - o Hybrid systems which mix logic expressivity, uncertainty and neural network performance are beginning to produce interesting results, see for example <https://arxiv.org/pdf/1805.10872.pdf> by de Raedt et al.; This is also the case of works which mix symbolic and numerical methods to solve problems differently than what has been done for years e.g. "Anytime discovery of a diverse set of patterns with Monte Carlo tree search". <https://arxiv.org/abs/1609.08827>. See also the work of Serafini and d'Avila Garcez on "Logic tensor networks" that connect deep neural networks to constraints expressed in logic. <https://arxiv.org/abs/1606.04422>
- Image and video processing
  - o Since the revelation of deep learning performances in the 2012 Imagenet campaign, the quality and accuracy of detection and



tracking of objects (e.g. people with their posture) made significant progresses. Applications are now possible, even if there remain many challenges.

- Natural Language Processing (NLP)
  - NLP neural models (machine translation, text generation, data mining) have made spectacular progress with, on the one hand, new architectures (transformer networks using attentional mechanisms) and, on the other hand, the idea of pre-training word or sentence representations using unsupervised learning algorithms that can then be used profitably in specific tasks with extremely little supervised data.
  - Spectacular results have been obtained in unsupervised translation, and in the field of multilingual representations and in automatic speech recognition, with a 100-fold reduction in labelled data (10h instead of 1000h!), using unsupervised pretraining on unlabelled raw audio<sup>6</sup>.
  
- Generative adversarial networks (GAN)
  - The results obtained by generative adversarial neural networks (GANs) are particularly impressive. These are capable of generating plausible natural images from random noise. Although the understanding of these models is still limited, they have significantly improved our ability to draw samples from particularly complex data distributions. From random distributions, GANs can produce new music, generate realistic deepfakes, write understandable text sentences, and the like.
  
- Optimisation
  - Optimisation problems that seemed impossible a few years ago can now be solved with almost generic methods. The combination of machine learning and optimization opens avenues for complex problems solving in design, operation, and monitoring of industrial systems. To support this, there is a proliferation of tools and libraries for AI than can be easily coupled with optimisation methods and solvers.
  
- Knowledge representation
  - The growing interest in combining knowledge graphs and graph embeddings to perform (semantic) graph-based machine learning.
  - New directions such as Web-based edge AI.  
[https://www.w3.org/wiki/Networks/Edge\\_computing](https://www.w3.org/wiki/Networks/Edge_computing)

---

<sup>6</sup> <https://arxiv.org/abs/2006.11477>.

Of course, there are scientific and technological limitations to all these results, the corresponding challenges are presented later in Chapter 5.

On the other hand, these positive achievements have been balanced by some concerns about the dangers of AI expressed by highly recognised scientists, more globally by many stakeholders of AI, which is the subject of the next section.

### 3. Debates about AI

Debates about AI really started in the 20<sup>th</sup> century - for example, think of Isaac Asimov's Laws of Robotics – but increased to a much higher level because of the recent progresses achieved by AI systems as shown above. The Technological Singularity Theory claims that a new era of machines dominating humankind will start when AI systems become super-intelligent: *“The technological singularity is a hypothetical event related to the advent of genuine artificial general intelligence. Such a computer, computer network, or robot would theoretically be capable of recursive self-improvement (redesigning itself), or of designing and building computers or robots better than itself on its own. Repetitions of this cycle would likely result in a runaway effect – an intelligence explosion – where smart machines design successive generations of increasingly powerful machines, creating intelligence far exceeding human intellectual capacity and control. Because the capabilities of such a super intelligence may be impossible for a human to comprehend, the technological singularity is the point beyond which events may become unpredictable or even unfathomable to human intelligence »* (Wikipedia).

Advocates of the technological singularity are close to the transhumanist movement, which aims at improving physical and intellectual capacities of humans with new technologies. The singularity would be a time when the nature of human beings would fundamentally change, this being perceived either as a desirable event, or as a danger for mankind.

An important outcome of the debate about the dangers of AI has been the discussion on autonomous weapons and killer robots, supported by an open letter published at the opening of the IJCAI conference in 2015<sup>7</sup>. The letter, which asks for a ban of such weapons able to operate beyond human control, has been signed by thousands of individuals including Stephen Hawking, Elon Musk, Steve Wozniak and a number of leading AI researchers including some from Inria, contributors to this document. See also Stuart Russell's “Slaughterbots” video<sup>8</sup>.

Other dangers and threats that have been discussed in the community include: the financial consequences on the stock markets of high frequency trading, which now represents the vast majority of orders placed, where supposedly intelligent software (which is fact is based on statistical decision making that cannot really be qualified as AI) operate at a high rate leading to possible market crashes, as for the *Flash Crash* of 2010; the consequences of big data mining on privacy, with mining systems able to divulgate private properties of individuals by establishing links between their online operations or their recordings in data banks; and of course the potential unemployment caused by the progressive replacement of workforce by machines.

---

<sup>7</sup> see <http://futureoflife.org/open-letter-autonomous-weapons/>

<sup>8</sup> [https://www.youtube.com/watch?v=HipTO\\_7mUOW](https://www.youtube.com/watch?v=HipTO_7mUOW)



Figure 4: In the movie "Her" by Spike Jonze, a man falls in love with his intelligent operating system

The more we develop artificial intelligence the greater the risk of developing only certain intelligent capabilities (e.g. optimisation and mining by learning) to the detriment of others for which the return on investment may not be immediate or may not even be a concern for the creator of the agent (e.g. moral, respect, ethics, etc.). There are many risks and challenges in the large-scale coupling of artificial intelligence and people. In particular, if the artificial intelligences are not designed and regulated to respect and preserve humans, if, for instance, optimisation and performances are the only goal of their intelligence then this may be the recipe for large scale disasters where users are used, abused, manipulated, etc. by tireless and shameless artificial agents. We need to research AI at large including everything that makes behaviours intelligent and not only the most "reasonable aspects". This is beyond purely scientific and technological matters, it leads to questions of governance and regulation.

Dietterich and Horvitz published an interesting answer to some of these questions<sup>9</sup>. In their short paper, the authors recognise that the AI research community should pay moderate attention to the risk of loss of control by humans, because this is not critical in a foreseeable future, but should instead pay more attention to five near-term risks for AI-based systems, namely: bugs in software; cyberattacks; "The Sorcerer's Apprentice", that is, making AI systems understand what people intend rather than literally interpreting their commands; "shared autonomy", that is, the fluid

---

<sup>9</sup> Dietterich, Thomas G. and Horvitz, Eric J., Rise of Concerns about AI: Reflections and Directions, Communications of the ACM, October 2015 Vol. 58 no. 10, pp. 38-40

cooperation of AI systems with users, so that users can always take control when needed; and the socioeconomic impacts of AI, meaning that AI should be beneficial for the whole society and not just for a group of happy few.

In the recent years, the debates focused on a number of issues around the notion of responsible and trustworthy AI, we can summarize them as follows:

- **Trust:** Our interactions with the world and with each other are increasingly channeled through AI tools. How to ensure security requirements for critical applications, safety and confidentiality of communication and processing media? What techniques and regulations for the validation, certification and audit of AI tools need to be developed to build confidence in AI?
- **Data governance:** The loop from data, information, knowledge, and actions is increasingly automated and efficient. What data governance rules of all kinds, personal, metadata and aggregated data at various levels, are needed? What instruments would make it possible to enforce them? How can we ensure traceability of data from producers to consumers?
- **Employment:** The accelerated automation of physical and cognitive tasks has strong economic and social repercussions. What are its effects on the transformation and social division of labour? What are the impacts on economic exchanges? What proactive and accommodation measures would be required? Is this different from the previous industrial revolutions?
- **Human oversight:** We increasingly delegate more and more personal and professional decisions to PDAs. How to benefit from it without the risk of alienation and manipulation? How can we make algorithms intelligible, make them produce clear explanations and ensure that their evaluation functions reflect our values and criteria? How can we anticipate and restore human control when the context is outside the scope of delegation?
- **Biases:** Our algorithms are not neutral; they are based on the implicit assumptions and biases, often unintended, of their designers or present in the data used for learning. How to identify and overcome these biases? How to design AI systems that respect essential human values, that do not increase inequalities?
- **Privacy and security:** AI applications can pose privacy challenges, for example in the case of face recognition, a useful technology for an easier access to digital services, but a questionable technology when put into general use. How can we design AI systems that do not unnecessarily break privacy constraints? How can we ensure the security and reliability of AI applications which can be subject to adversarial attacks?
- **Sustainability:** machine learning systems use an exponentially increasing amount of computer power and energy, because of the amount of input data and of the number of parameters to optimise. How can we build increasingly sophisticated AI systems using limited resources?

Avoiding the risks is necessary but not sufficient to effectively mobilize AI at the service of humanity. How can we devote a substantial part of our research and

development resources to the major challenges of our time (climate, environment, health, education) and more broadly to the UN's sustainable development objectives?

These and other issues must be the subject of citizen and political deliberations, controlled experiments, observatories of uses, and social choices. They have been documented in several reports providing recommendations, guidelines, principles for AI such as the Montreal Declaration for Responsible AI<sup>10</sup>, The OECD Recommendations on Artificial Intelligence<sup>11</sup>, the Ethics Guidelines for Trustworthy Artificial Intelligence by the European Commission's High-Level Expert Group<sup>12</sup> and many others including UNESCO, the Council of Europe, government, private companies, NGOs etc. Altogether there are more than a hundred such documents at the time of writing this white paper.

Inria is aware of these debates and acts as a national institute for research in digital science and technology, conscious of its responsibilities in front of the society. Informing the society and our governing bodies about the potentialities and risks of digital science and technologies is one of our missions.

Inria launched a reflexion about ethics long before the threats of AI were subject of debates in the scientific society. In the recent years, Inria:

- Contributed to the creation of Allistene's CERNA<sup>13</sup>, a think tank looking at ethics problems arising from research on digital science and technologies; the first two recommendations report published by CERNA concerned the research on robotics and best practices for machine learning;
- Set up a body responsible for assessing the legal or ethical issues of research on a case by case basis: the Operational Committee for the Evaluation of Legal and Ethical Risks (COERLE) with scientists from Inria and external contributors; COERLE's mission is to help identify risks and determine whether the supervision of a given research project is required;
- Was deeply involved in the creation of our national committee on the ethics of digital technologies<sup>14</sup>;
- Was put in charge of the coordination of the research component of our nation's AI strategy (see chapter 4);
- Was asked by the French government to organise the Global Forum on Artificial Intelligence for Humanity, a colloquium which gathered

---

<sup>10</sup> <https://www.montrealdeclaration-responsibleai.com/>

<sup>11</sup> <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

<sup>12</sup> European Commission High-Level Expert Group (2018). Ethics Guidelines for Trustworthy AI; Available at

<sup>13</sup> *Commission de réflexion sur l'Ethique de la Recherche en sciences et technologies du Numérique de l'Alliance des Sciences et Technologies du Numérique* : <https://www.allistene.fr/cerna/>

<sup>14</sup> <https://www.allistene.fr/tag/cerna/>

- leading world experts in AI and its societal consequences, in late 2019<sup>15</sup>, as a precursor for the GPAI (see below);
- Was given responsibility of the Paris Centre of Expertise of the Global Partnership on Artificial Intelligence, an international and multi-stakeholder initiative to guide the responsible development and use of artificial intelligence consistent with human rights, fundamental freedoms, and shared democratic values, launched by fourteen countries and the European Union in June 2020.

Moreover, Inria encourages its researchers to take part in the societal debates when solicited by press and media about ethical questions such as the ones raised on robotics, deep learning, data mining and autonomous systems. Inria also contributes to educating the public by investing in the development of MOOCs on AI and on some of its subdomains (“L’intelligence artificielle avec intelligence”<sup>16</sup>, “Web sémantique et web de données”<sup>17</sup>, “Binaural hearing for robots”<sup>18</sup>) and more generally by playing an active role in educational initiatives for digital sciences.

This being said let us now look at the scientific and technological challenges for AI research, and at how Inria contributes to addressing these challenges: this will be the subject of the next section.

---

<sup>15</sup> [https://www.youtube.com/playlist?list=PLJ1qHZpFsMsTXDBLLWikaUXQG\\_d5Ru3CT](https://www.youtube.com/playlist?list=PLJ1qHZpFsMsTXDBLLWikaUXQG_d5Ru3CT)

<sup>16</sup> <https://www.fun-mooc.fr/courses/course-v1:inria+41021+session01/about>

<sup>17</sup> <https://www.fun-mooc.fr/courses/course-v1:inria+41002+self-paced/about>

<sup>18</sup> <https://www.fun-mooc.fr/courses/course-v1:inria+41004+archiveouvert/about>

#### 4. Inria in the national AI strategy

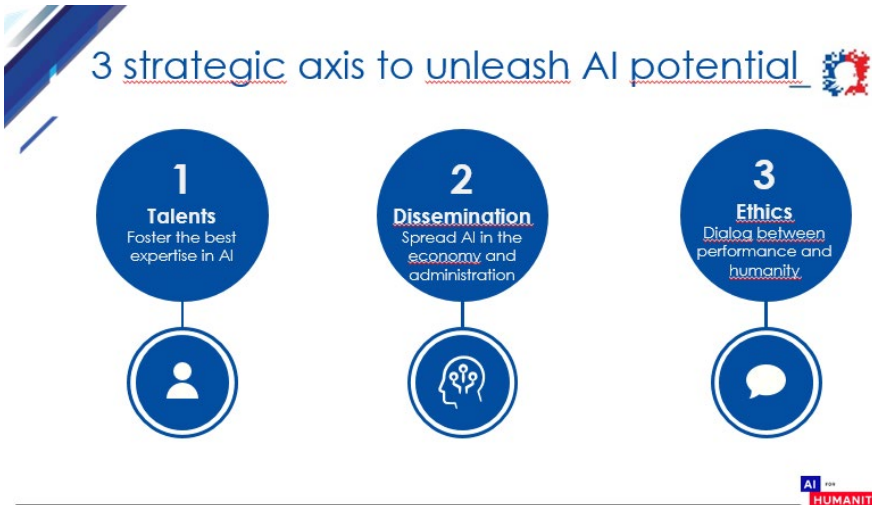
##### AI FOR HUMANITY: THE NATIONAL AI RESEARCH PROGRAMME

In the closing day of "AI for Humanity" debate held in Paris on March 29, 2018, the President of the French Republic presented an ambitious strategy for Artificial Intelligence (AI) and launched the National AI Strategy (<https://www.aiforhumanity.fr/en/>).

The National AI Strategy aims to make France a leader in AI, a sector currently dominated by the United States and China, and by emerging countries of the discipline like Israel, Canada and the United Kingdom.

The priorities that the President of the Republic set out are research, open data and ethical or societal issues. These measures come from the report written by the mathematician and Member of Parliament Cédric Villani, who conducted hearings with more than 300 experts from around the world. To conclude this project, Cedric Villani worked with Marc Schoenauer, research director and head of the TAU project-team at the Inria Saclay – Île-de-France research centre.

This National AI Strategy, with a budget of 1.5 billion € of public money for five years, gathers three axes: (i) achieving best-in-class level of research for AI, through training and attracting best global talent in the field; (ii) disseminating AI to the economy and society through spin-offs and public-private partnerships and data sharing; (iii) establishing an ethical framework for AI. Many measures have already been taken in these three areas.



*As part of the AI for Humanity Plan, Inria was entrusted with the coordination of National AI Research Programme. The research plan interacts with each of the three above-mentioned axes.*



The kick-off meeting for the research axis took place in Toulouse on 28 November 2018. The objective of the **National AI Research Programme** (<https://www.inria.fr/en/ai-mission-national-artificial-intelligence-research-program>) is twofold: to sustainably establish France as one of the top 5 countries in AI and to make France a European leader in research in AI.

To this aim, several actions will be carried out in a first stage lasting from the end of 2018 to 2022:

- Set up a national research network in AI coordinated by Inria;
- Initiate 4 Interdisciplinary Institutes for Artificial Intelligence;
- Promote programs of attractiveness and talent support throughout the country;
- Contribute to the development of a specific program on AI training;
- Increase the computing resources dedicated to AI and facilitate access to infrastructures;
- Boost public-private partnerships;
- Boost research in AI through the ANR calls;
- Strengthen bilateral, European and international cooperation;

The research axis also liaises with innovation initiatives in AI, in particular with the Innovation Council's Great Challenges (<https://www.gouvernement.fr/decouvrir-les-grands-defis>).

## 5. The Challenges of AI and Inria contributions

Inria's approach is to combine simultaneously two endeavours: understanding the systems at play in the world (from social to technological), and the issues they arise from their interactions; and acting on them to find solutions by providing numerical models, algorithms, software, technologies. This involves developing a precise description, for instance formal or learned from data, adequate tools to reason about it or manipulate it, as well as proposing innovative and effective solutions. This vision has developed over the 50 years of existence of the institute, favored by an organization that does not separate theory from practice, or mathematics from computer science, but rather brings together the required expertise in established research teams, on the basis of focused research projects.

The notion of "digital sciences" is not uniquely defined, but we can approach it through the dual goal outlined above, to understand the world and then act on it. The development of "computational thinking" requires the ability to define, organize and manipulate the elements at the core of digital sciences: Models, Data, and Languages. The development of techniques and solutions for the digital world calls for research in a variety of domains, typically mixing mathematical models, algorithmic advances and systems. Therefore, we identify the following branches in the research relevant for Inria:

- Algorithms and programming,
- Data science and knowledge engineering,
- Modeling and simulation,
- Optimisation and control.
- Architectures, systems and networks,
- Security and confidentiality,
- Interaction and multimedia,
- Artificial intelligence and autonomous systems.

As any classification, this presentation is partly arbitrary, and does not expose the many interactions between topics. For instance, network studies also involve novel algorithm developments, and artificial intelligence is very transverse in nature, with strong links to data science. Clearly, each of these branches is a very active area of research today. Inria has invested in these topics by creating dedicated project-teams and building strong expertise in many of these domains. Each of these directions is considered important for the institute.

AI is a vast domain; any attempt to structure it in subdomains can be debated. We will use the keywords hierarchy proposed by the community of Inria team leaders in order to best identify their contributions to digital sciences in general. In this hierarchy, Artificial Intelligence is a top-level keyword with eight subdomains, some of them specific, some of them referring to other sections of the hierarchy: see the following table.

**Knowledge**

- Knowledge bases
- Knowledge extraction & cleaning
- Inference
- Semantic web
- Ontologies

**Machine Learning**

- Supervised Learning
- Unsupervised learning
- Sequential and reinforcement learning
- Optimisation for learning
- Bayesian methods
- Neural networks
- Kernel methods
- Deep learning
- Data mining
- Massive data analysis

**Natural Language processing****Signal processing (speech, vision)**

- Speech
- Vision
  - Object recognition
  - Activity recognition
  - Search in image and video banks
  - 3D and spatiotemporal reconstruction
  - Objects tracking and movement analysis
  - Objects localisation
  - Visual servoing

**Robotics (including autonomous vehicles)**

- Design
- Perception
- Decision
- Action
- Robot interaction (environment/humans/robots)
- Robot fleets
- Robot learning
- Cognition for robotics and systems

**Neurosciences, cognitive sciences**

- Understanding and simulation of the brain and of the nervous system
- Cognitive sciences

**Algorithmic of AI**

- Logic programming and ASP
- Deduction, proof
- SAT theories
- Causal, temporal, uncertain reasoning
- Constraint programming
- Heuristic search
- Planning and scheduling

**Decision support**

We do *not* provide definitions of AI and of subdomains: there is abundant literature about them. Good definitions can also be found on Wikipedia, e.g.

[https://en.wikipedia.org/wiki/Artificial\\_intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence)

[https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)

<https://en.wikipedia.org/wiki/Robotics>

[https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing)

[https://en.wikipedia.org/wiki/Semantic\\_Web](https://en.wikipedia.org/wiki/Semantic_Web)

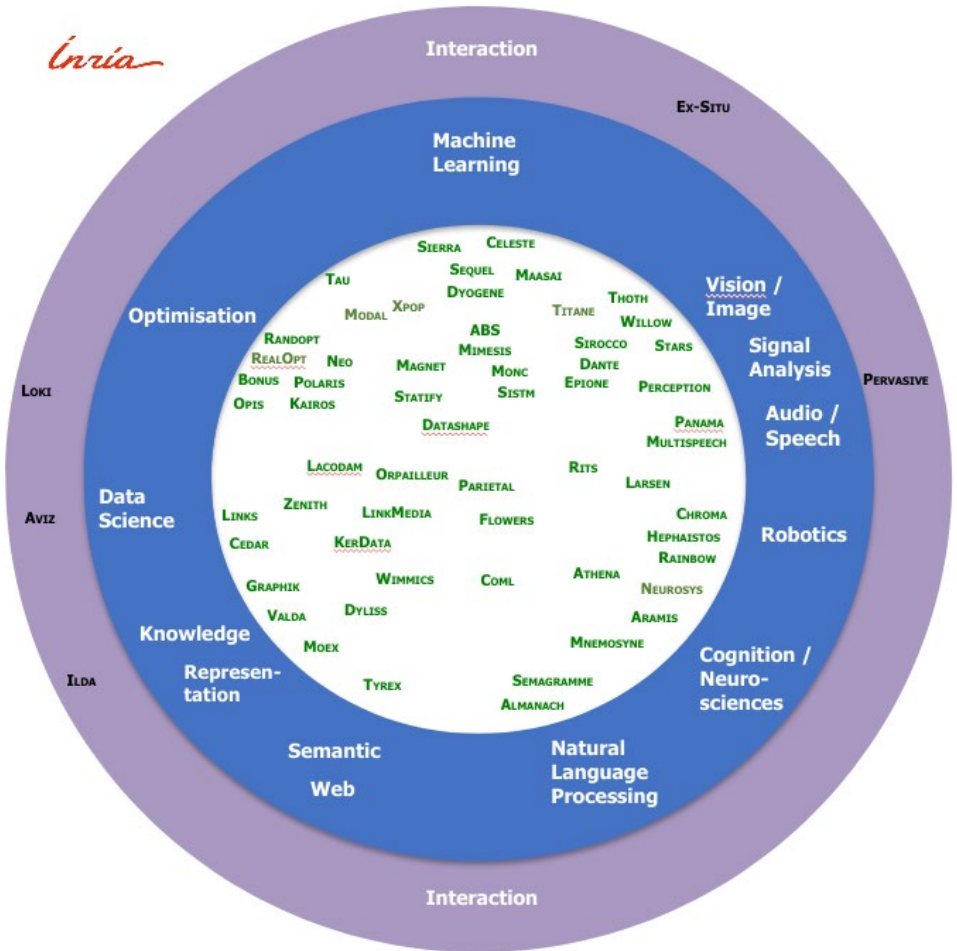
[https://en.wikipedia.org/wiki/Knowledge\\_representation\\_and\\_reasoning](https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning)

etc.

In the following, Inria contributions will be identified by *project-teams*.

Inria project-teams are autonomous, interdisciplinary and partnership-based, and consist of an average of 15 to 20 members. Project-teams are created based on a roadmap for research and innovation and are assessed after four years, as part of a national assessment of all scientifically-similar project teams. Each team is an agile unit for carrying out high-risk research and a breeding ground for entrepreneurial ventures. Because new ideas and breakthrough innovations often arise at the crossroads of several disciplines, the project team model promotes dialogue between a variety of methods, skills and subject areas. Because collective momentum is a strength, 80% of Inria's research teams are joint teams with major research universities and other organizations (CNRS, Inserm, INRAE, etc.) The maximum duration of a project-team is twelve years.

The project-teams' names will be written in SMALL CAPS, so as to distinguish them from other nouns.



After an initial subsection dealing with generic challenges, more specific challenges are presented, starting with machine learning and followed by the categories in the wheel above. The wheel has three parts: inside, the project-teams; in the innermost ring, subcategories of AI; in the outermost ring, teams in human-computer interaction with AI. Each section is devoted to a category, and starts with a copy of the wheel where teams identified to be fully in that category are underlined in dark blue and teams that have a weaker relation with that category are underlined in light blue.

## 5.1 Generic challenges in artificial intelligence

Some examples of the main generic challenges in AI identified by Inria are as follows:

**Trusted co-adaptation of humans and AI-based systems.** Data is everywhere in personal and professional environments. Algorithmic-based treatments and decisions about these data are diffusing in all areas of activity, with huge impacts on our economy and social organization. Transparency and ethics of such algorithmic systems, in particular AI-based system able to make critical decisions, become increasingly important properties for trust and appropriation of digital services. Hence, the development of transparent and accountable-by-design data management and analytics methods, geared towards humans, represents a very challenging priority.

- i) **Data science for everyone.** As the volume and variety of available data keep growing, the need to make sense of these data becomes ever more acute. Data Science, which encompasses diverse tasks including prediction and knowledge discovery, aims to address this need and gathers considerable interest. However, performing these tasks typically still requires great efforts from human experts. Hence, designing Data Science methods that greatly reduce both the amount and the difficulty of the required human expert work constitutes a grand challenge for the coming years.
- ii) **Lifelong adaptive interaction with humans.** Interactive digital and robotic systems have a great potential to assist people in everyday tasks and environments, with many important societal applications: cobots collaborating with humans in factories; vehicles acquiring large degrees of autonomy; robots and virtual reality systems helping in education or elderly people... In all these applications, interactive digital and robotic systems are tools that interface the real world (where humans experience physical and social interactions) with the digital space (algorithms, information repositories and virtual worlds). These systems are also sometimes an interface among humans, for example, when they constitute mediation tools between learners and teachers in schools, or between groups of people collaborating and interacting on a task. Their physical and tangible dimension is often essential both for the targeted function (which implies physical action) and for their adequate perception and understanding by users.
- iii) **Connected autonomous vehicles.** The connected autonomous vehicle (CAV) is quickly emerging as a partial response to the societal challenge of sustainable mobility. The CAV should not be considered alone but as an essential link in the intelligent transport systems (ITS) whose benefits are manifold: improving road transport safety and efficiency, enhancing access to mobility and preserving the environment by reducing greenhouse gas emissions. Inria aims at contributing to the design of

advanced control architectures that ensure safe and secure navigation of CAVs by integrating perception, planning, control, supervision and reliable hardware and software components. The validation and verification of the CAVs through advanced prototyping and in-situ implementation will be carried out in cooperation with relevant industrial partners.

In addition to the previous challenges, the following desired properties for AI systems should trigger new research activities beyond the current ones: some are extremely demanding and cannot be addressed in the near term but are worth considering.

### Openness to other disciplines

An AI will often be integrated in a larger system composed of many parts. Openness therefore means that AI scientists and developers will have to collaborate with specialists of other disciplines in computer science (e.g. modelling, verification & validation, networks, visualisation, human-computer interaction etc.) to compose the wider system, and with non-computer scientists that contribute to AI e.g. psychologists, biologists (e.g. biomimetics), mathematicians, etc. A second aspect is the impact of AI systems on several facets of our life, our economy, and our society: collaboration with specialists from other domains (it would be too long to mention them, e.g. economists, environmentalists, biologists, lawyers etc.) becomes mandatory.

### Scaling up ... and down!

AI systems must be able to handle vast quantities of data and of situations. We have seen deep learning algorithms absorbing millions of data points (signal, images, video etc.) and large-scale reasoning systems such as IBM's Watson making use of encyclopaedic knowledge; however, the general question of scaling up for the many V's (variety, volume, velocity, vocabularies, ...) still remains.

Working with small data is a challenge for several applications that do not benefit from vast amounts of existing cases. Embedded systems, with their specific constraints (limited resources, real time, etc.), also raise new challenges. This is particularly relevant for several industries and demands to develop new machine learning mechanisms, either extending (deep) learning techniques (e.g., transfer learning, or few-shot learning), or considering completely different approaches.

### Multitasking

Many AI systems are good at one thing but show little competence outside their focus domain; but real-life systems, such as robots must be able to undertake several actions in parallel, such as memorising facts, learning new concepts, acting on the real world and interacting with humans. But this is not so simple. The diversity of channels through which we sense our environment, the reasoning we conduct, the tasks we perform, is several orders of magnitude greater. Even if we inject all the data in the world into the biggest computer imaginable, we will be far from the capabilities of our brain. To do better, we will have to make specialized skills cooperate in sub-problems:

it is the set of these sub-systems that will be able to solve complex problems. There should be a bright future for distributed AI and multi-agent systems.

### Validation and certification

A mandatory component in mission-critical systems, certification of AI systems, or their validation by appropriate means, is a real challenge especially if these systems fulfil the previous expectations (adaptation, multitasking, user-in-the-loop): verification, validation and certification of classical (i.e. non-AI) systems is already a difficult task – even if there are already exploitable technologies, some being developed by Inria project-teams – but applying these tools to complex AI systems is an overwhelming task which must be approached if we want to put these systems in use in environments such as aircrafts, nuclear power plants, hospitals etc.

In addition, while validation requires comparing an AI system to its specifications, certification requires the presence of norms and standards that the system will face. Several organizations, including ISO, are already working on standards for artificial intelligence, but this is a long-term quest that has only just begun.

### Trust, Fairness, Transparency and accountability

As seen in chapter 3, ethical questions are now central to the debates on AI and even stronger for ML. Trust can be reached through a combination of many factors among which the proven robustness of models, their explanation capacity or their interpretability/auditability by human users, the provision of confidence intervals for outputs. These points are key towards the wide acceptance of the use of AI in critical applications such as medicine, transportation, finance or defence. Another major issue is fairness, that is, building algorithms and models that treat different categories of the population fairly. There are dozens of analysis and reports on this question, but almost no solution to it for the moment.

### Norms and human values

Giving norms and values to AIs goes far beyond current science and technologies: for example, should a robot going to buy milk for his owner stop on his way to help a person whose life is in danger? Could a powerful AI technology be used for artificial terrorists? As for other technologies, there are numerous fundamental questions without answers.

### Privacy

The need for privacy is particularly relevant for AIs that are confronted with personal data, such as intelligent assistants/companions or data mining systems. This need is valid for non-AI systems too, but the specificity of AI is that new knowledge will be derived from private data and possibly made public if not restricted by technical means. Some AI systems know us better than we know ourselves!



## 5.2 Machine learning

Even though machine learning (ML) is the technology by which Artificial Intelligence reached new levels of performance and found applications in almost all sectors of human activity, there remains several challenges from fundamental research to societal issues, including hardware efficiency, hybridisation with other paradigms, etc.

This section starts with some generic challenges in ML: ethical issues and trust – including resisting adversarial attacks –; performance and energy consumption; hybrid models; moving to causality instead of correlations; common sense understanding; continuous learning; learning under constraints. Next are subsections on more specific aspects i.e. fundamentals and theory of ML, ML and heterogeneous data, ML for life sciences, with presentation of Inria project-teams.

### Resisting adversarial attacks

It has been shown in the last years that ML models are very weak with respect to adversarial attacks, i.e. it is quite easy to fool a deep learning model by slightly modifying its input signal and thereby obtaining wrong classifications or predictions. Resisting such adversarial attacks is mandatory for systems that will be used in real life, but once more, generic solutions still have to be developed.

### Performance and energy consumption

As shown in the last AI Index<sup>19</sup> and in a number of recent papers, the computation demand of ML training has grown exponentially since 2010, doubling every 3.5 months – this means a factor of one thousand in three years, one million in six years. This is due to the size of data used, to the sophistication of deep learning models with billions of parameters or more, and to the application of automatic architecture search algorithms which basically consist in running thousands of variations of the models on the same data. The paper by Strubell et al.<sup>20</sup> shows that the energy used to train a big transformer model for natural language processing with architecture search is five times greater than the fuel used by an average passenger car over its lifetime. This is obviously not sustainable: voices are now heard that demand to revise the way in which machines learn so as to save computational resources and energy. One idea is of neural networks with parsimonious connections under robust and mathematically well understood algorithms, leading to a compromise between

---

<sup>19</sup> Raymond Perrault et al., *The AI Index 2019 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019

<sup>20</sup> Energy and Policy Considerations for Deep Learning in NLP; Strubell, Ganesh, McCallum; College of Information and Computer Sciences, University of Massachusetts Amherst, June 2019, arXiv:1906.02243v1

performance and frugality. It is also a question of ensuring the robustness of the approaches as well as the interpretability and explainability of the networks learned.

### Hybrid models, symbolic vs. continuous representations

Hybridisation consists in joining different modelling approaches in synergy: the most common approaches being the continuous representations used for Deep Learning, the symbolic approaches of the former AI community (expert and knowledge-based systems), and the numerical models developed for simulation and optimisation of complex systems. Supporters of this hybridisation state that such a combination, although not easy to implement, is mutually beneficial. For example, continuous representations are differentiable and allow machine-learning algorithm to approximate complex functions, while symbolic representations are used to learn rules and symbolic models. A desired feature is to embed reasoning into continuous representation, that is, find ways to make inferences on numeric data; on the other hand, in order to benefit from the power of deep learning, defining continuous representations of symbolic data can be quite useful, as has been done e.g. for text with *word2vec* and *text2vec* representations.

### Moving to causality

Most commonly used learning algorithms correlate input and output data - for example, between pixels in an image and an indicator for a category such as "cat", "dog", etc. This works very well in many cases, but ignores the notion of causality, which is essential for building prescriptive systems. Causality is a formidable tool for making such tools, which are indispensable for supervising and controlling critical systems such as a nuclear power plant, the state of health of a living being or an aircraft. Inserting the notion of causality into machine learning algorithms is a fundamental challenge; this can be done by integrating a priori knowledge (numerical, logical, symbolic models, etc.) or by discovering causality in data.

### Common sense understanding

Even if the performance of ML systems in terms of error rates on several problems are quite impressive, it is said that these models do not develop a deep understanding of the world, as opposed to humans. The quest for common sense understanding is a long and tedious one, which started with symbolic approaches in the 1980s and continued with mixed approaches such as IBM Watson, the TODAI robot project<sup>21</sup> (making a robot pass an examination to enter University of Tokyo), AllenAI's Aristo project<sup>22</sup> (build systems that demonstrate a deep understanding of the world, integrating technologies for reading, learning, reasoning, and explanation), and more recently IBM Project Debater<sup>23</sup>, a system able to exchange arguments on any subject with top human debaters. A system like Google's Meena<sup>24</sup> (a conversational agent that

---

<sup>21</sup> <https://21robot.org/index-e.html>

<sup>22</sup> <https://allenai.org/aristo>

<sup>23</sup> <https://www.research.ibm.com/artificial-intelligence/project-debater/>

<sup>24</sup> <https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html>

can chat about anything) can create an illusion when we see it conversing, but the deep understanding of its conversations is another matter.

### Continuous and never-ending (life-long) learning

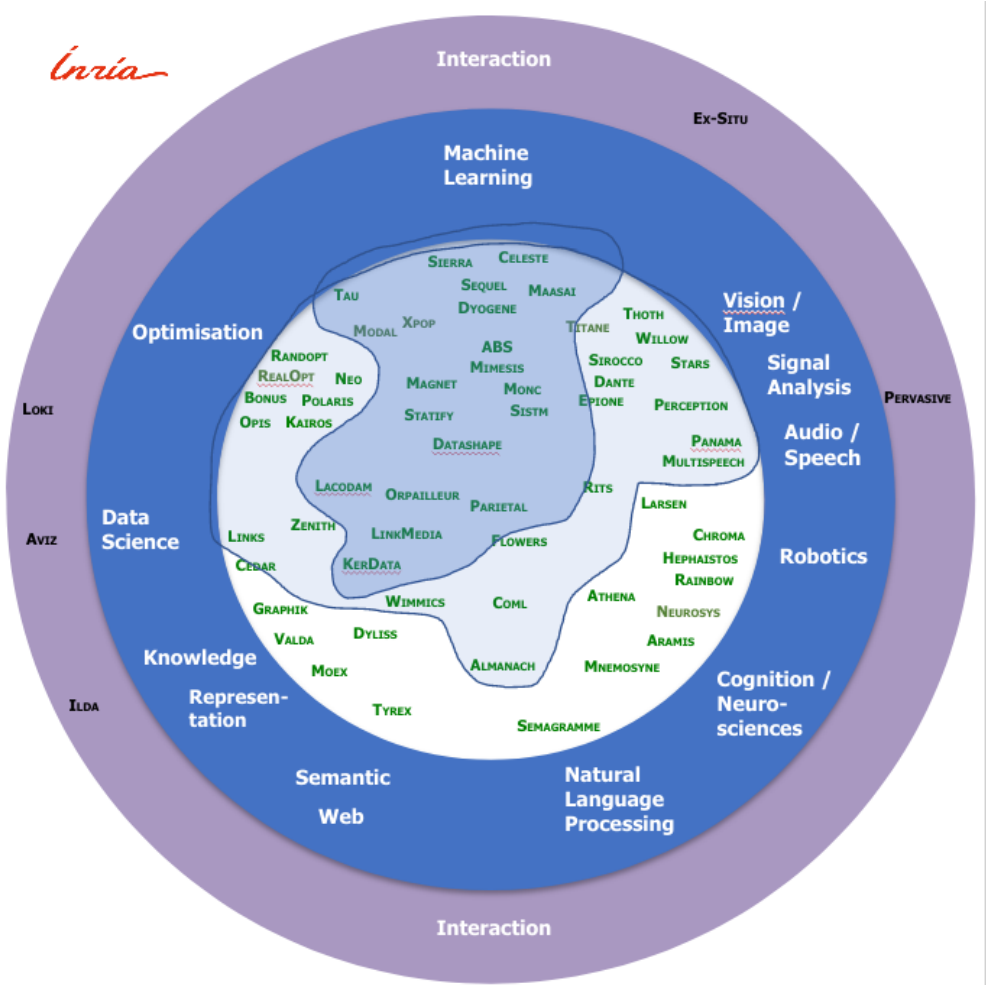
Some AI systems are expected to be resilient, that is, be able to operate on a 24/7 basis without interruptions. Interesting developments have been made for lifelong learning systems that will continuously learn new knowledge while they operate. The challenges are to operate online in real time and to be able to revise existing beliefs learned from previous cases, in a self-supervised way. These systems use some bootstrapping: elementary knowledge learned in the first stages of operation will be used to direct future learning tasks, such as in the NELL/Read the Web (never-ending language learning) system developed by Tom Mitchell at Carnegie-Mellon University<sup>25</sup>.

### Learning under constraints

Privacy is certainly the most important constraint that must be considered. The field of machine learning recently recognised the need to maintain privacy while learning from records about individuals; a theory of machine learning respectful of privacy is being developed by researchers. At Inria, several teams work on privacy: especially ORPAILLEUR in machine learning, but also teams from other domains such as PRIVATICS (algorithmics of privacy) and SMIS (privacy in databases). More generally speaking, machine learning might have to cope with other external constraints such as decentralised data or energy limitations – as mentioned above. Research on the wider problem of machine learning with external constraints is needed.

---

<sup>25</sup> <http://rtw.ml.cmu.edu/rtw/>



### 5.2.1 Fundamental machine learning and mathematical models

Machine learning raises numerous fundamental issues, such as linking theory to experimentation, generalisation, capability to explain the outcome of the algorithm, moving to unsupervised or weakly supervised learning, etc. There are also issues regarding the computing infrastructures and, as seen in the previous section, questions of usage of computing resources. A number of Inria teams are active in

fundamental machine learning, developing new mathematical knowledge and applying it to real world use cases.

### Mathematical theory

Learning algorithms are based on sophisticated mathematics, which makes them difficult to understand, use and explain. A challenge is to improve the theoretical underpinnings of our models, which are often seen externally as algorithmic black boxes that are difficult to interpret. Getting theory and practice to stick together as much as possible is a constant challenge, and one that is becoming more and more important given the number of applied researchers and engineers working in AI/machine learning: "state of the art" methods in practice are constantly moving away from what theory can justify or explain.

### Generalisation

A central challenge of machine learning is the one of generalisation: how a machine can predict/control a system beyond the data it has seen during training, especially beyond the distribution of the data seen during training. Moreover, generalisation will help need moving from systems that can solve a task to multi-purpose systems that can implement their capabilities in different contexts. This can also be by transfer (from one task to another) or adaptation.

### Explainability

One of the factors of trust in artificial systems, explainability is required for systems that makes critical predictions and decisions, when there are no other guarantees such as formal verification, certification or adhesion to norms and standards<sup>26</sup>. The quest for explainability of AI systems is a long one; it was triggered by DARPA's XAI (eXplainable AI)<sup>27</sup> programme, launched in 2017. There are many attempts to produce explanations (for example highlighting certain areas in images, doing sensitivity analysis on input data, transforming numerical parameters into symbols or if-then rules) but no one is fully satisfactory.

### Consistency of the algorithms' outputs.

These are the prerequisite for any development of legal frameworks necessary to the large testing and the deployments of AV's in real road networks and cities. the problem of statistical reproducibility: being able to assign a level of significance (for example a p-value) to the conclusions drawn from a machine learning algorithm. Such information seems indispensable to inform the decision-making process based on these conclusions.

### Differentiable programming

Beyond the availability of data and powerful computers that explain most recent advances in deep learning, there is a third reason which is both scientific and

---

<sup>26</sup> Some DL specialists claim that people trust their doctors without explanations, which is true. But doctors follow a long training period materialized by a diploma that certifies their abilities.

<sup>27</sup> <https://www.darpa.mil/program/explainable-artificial-intelligence>

technological: until 2010, researchers in machine learning derived the analytical formulas for calculating the gradients in the backpropagation mode. They then rediscovered automatic differentiation, which existed in other communities but had not yet entered the AI field. This opened up the possibility of experimenting with complex architectures such as the Transformers/BERTs that revolutionized natural language processing. Today we could replace the term "deep learning" with "differentiable programming", which is both more scientific and more general.

## **CELESTE**

### **Mathematical statistics and learning**

The statistical community has long-term experience in how to infer knowledge from data, based on solid mathematical foundations. The more recent field of machine learning has also made important progress by combining statistics and optimisation, with a fresh point of view that originates in applications where prediction is more important than building models.

The Celeste project-team is positioned at the interface between statistics and machine learning. They are statisticians in a mathematics department, with strong mathematical backgrounds behind us, interested in interactions between theory, algorithms and applications. Indeed, applications are the source of many of our interesting theoretical problems, while the theory we develop plays a key role in (i) understanding how and why successful statistical learning algorithms work -- hence improving them -- and (ii) building new algorithms upon mathematical statistics-based foundations.

Celeste aims to analyse statistical learning algorithms -- especially those that are most used in practice -- with our mathematical statistics point of view, and develop new learning algorithms based upon our mathematical statistics skills.

Celeste's theoretical and methodological objectives correspond to four major challenges of machine learning where mathematical statistics have a key role:

- First, any machine learning procedure depends on hyperparameters that must be chosen, and many procedures are available for any given learning problem: both are an estimator selection problem.
- Second, with high-dimensional and/or large data, the computational complexity of algorithms must be taken into account differently, leading to possible trade-offs between statistical accuracy and complexity, for machine learning procedures themselves as well as for estimator selection procedures.
- Third, real data are usually corrupted partially, making it necessary to provide learning (and estimator selection) procedures that are robust to outliers and heavy tails, while being able to handle large datasets.
- Fourth, science currently faces a reproducibility crisis, making it necessary to provide statistical inference tools (p-values, confidence regions) for assessing the significance of the output of any learning algorithm

(including the tuning of its hyperparameters), in a computationally efficient way.

## TAU

### Tackling the Underspecified

Building upon the expertise in machine learning (ML) and optimisation of the TaO team, the TaU project tackles some **under-specified challenges behind the New Artificial Intelligence wave**.

#### 1. A trusted AI

There are three reasons for the fear of the undesirable effects of AI and machine learning: (i) the smarter the system, the more complex it is and the more difficult it is to correct bugs (certification problem); (ii) if the system learns from data reflecting world biases (prejudices, inequities), the models learnt will tend to perpetuate these biases (equity problems); (iii) AI and learning tend to learn from predictive models (if conditions then effects); and decision-makers tend to use these models in a prescriptive manner (to produce such effects, seek to satisfy these conditions), which can be ineffective or even catastrophic (causal problems). Model certification. One possible approach to certifying neural networks is based on formal proofs. The main obstacle here is the perception stage, for which there is no formal specification or manipulable description of the set of possible scenarios. One possibility is to consider that the set of scenarios/perceptions is captured by a simulator, which makes it possible to restrict oneself to a very simplified, but well-founded problem.

Bias and fairness. In social sciences and humanities (e.g., links between a company's health and the well-being of its employees, recommendation of job offers, links between food and health) offer data that are biased. For example, behavioural data is often collected for marketing purposes, which may tend to over-represent one category or another. These biases need to be identified and adjusted to obtain accurate models.

Causality. Predictive models can be based on correlations (the presence of books at home is correlated with the good grades of children at school). However, these models do not allow for action to achieve desired effects (e.g. it is useless to send books to improve children's grades): only causal models allow for founded interventions. The search for causal models opens up major perspectives (e.g., the power of the causal model to influence the outcome). The search for causal models opens up major prospects (being able to model what would have happened if one had done otherwise, i.e. counterfactual modelling) for 'AI for Good'.

#### 2. Machine Learning and Numerical Engineering

A key challenge is to combine ML and AI with domain knowledge. In the field of mathematical modelling and numerical analysis in particular, there is extensive knowledge of description, simulation and design in the form of partial differential equations. The coupling between neural networks and numerical models is a strategic research direction, with first results in terms of i) complexity of underlying phenomena (multi-phase 3D fluid mechanics, heterogeneous hyperelastic materials, ...); ii) scaling-up (real-time simulation); iii) fine/adaptive

control of models and processes, e.g. control of numerical instabilities or identification of physical invariants.

### **3. A sustainable AI: learning to learn**

The Achilles' heel of machine learning, apart from a few areas such as image processing, remains the difficulty of fine-tuning models (typically for neural networks, but generally speaking). The quality of the models depends on the automatic adjustment of the whole learning chain, the pre-processing of the data to the structural parameters of the learning itself, the choice of the architecture for deep networks, the algorithms for classical statistical learning, and the hyperparameters of all the components of the processing chain.

The proposed approaches range from methods derived from information theory and statistical physics to the learning methods themselves. In the first case, given the very large size of the networks considered, statistical physics methods (e.g. mean field, scale invariance) can be used to adjust the hyperparameters of the models and to characterize the problem regions in which solutions can be found. In the second case, we try to model from empirical behaviour, which algorithms behave well on which data.

A related difficulty concerns the astronomical amount of data needed to learn the most efficient models of the day, i.e. deep neural networks. The cost of computation thus becomes a major obstacle for the reproducibility of scientific results.

### **Weakly supervised and unsupervised learning**

Most remarkable results obtained with ML are based on supervised learning, that is, learning from examples where the expected output is given together with the input data. This implies prior labelling of data with the corresponding expected outputs and can be quite demanding for large-scale data. Amazon's Mechanical Turk is an example of how corporations mobilise human resources for annotating data (which raises many social issues). While supervised learning undoubtedly brings excellent performance, the labelling cost will eventually become unbearable since both the dataset sizes constantly increase. Not to mention that encompassing all operating conditions in a single dataset is impractical. Leveraging semi or unsupervised learning is necessary to ensure scalability of the algorithms to the real world, where they ultimately face situations unseen in the training set. The holy Grail of artificial general intelligence is far from our current knowledge but promising techniques in transfer learning allow expanding training done in supervised fashion to new unlabelled datasets, for example with domain adaptation.

### **Computing Architectures**

Modern machine learning systems need high performance computing and data storage in order to scale up with the size of data and with problem dimensions; algorithms will run on Graphical Processing Units (GPUs) and other powerful architectures such as Tensor Processing Units – TPUs, Neural Processing Units – NPU, Intelligence Processing Units – IPU etc.; data and processes must be distributed over many processors. New research must address how ML algorithms and problem



formulations can be improved to make best usage of these computing architectures, also meeting sustainability questions (see above).

## **MAASAI**

### **Models and Algorithms for Artificial Intelligence**

Maasai is a research project-team at Inria Sophia-Antipolis, working on the models and algorithms of Artificial Intelligence. This is a joint research team with the laboratories LJAD (Mathematics, UMR 7351) and I3S (Computer Science, UMR 7271) of Université Côte d'Azur. The team is made of both mathematicians and computer scientists in order to propose innovative learning methodologies, addressing real-world problems, that are both theoretically sound, scalable and affordable.

Artificial intelligence has become a key element in most scientific fields and is now part of everyone life thanks to the digital revolution. Statistical, machine and deep learning methods are involved in most scientific applications where a decision has to be made, such as medical diagnosis, autonomous vehicles or text analysis. The recent and highly publicized results of artificial intelligence should not hide the remaining and new problems posed by modern data. Indeed, despite the recent improvements due to deep learning, the nature of modern data have brought specific issues. For instance, learning with high-dimensional, atypical (networks, functions, ...), dynamic, or heterogeneous data remains difficult for theoretical and algorithmic reasons. The recent establishment of deep learning has also open new questions such as: How to learn in an unsupervised or weakly-supervised context with deep architectures? How to design a deep architecture for a given situation? How to learn with evolving and corrupted data?

To address these questions, the Maasai team focuses on topics such as unsupervised learning, theory of deep learning, adaptive and robust learning, and learning with high-dimensional or heterogeneous data. The Maasai team conducts a research that links practical problems that may come from industry or other scientific fields, with the theoretical aspects of Mathematics and Computer Science. In this spirit, the Maasai project-team is totally aligned with the "Core elements of AI" axis of the Institut 3IA Côte d'Azur. It is worth noticing that the team hosts two 3IA chairs of the Institut 3IA Côte d'Azur.

## **SIERRA**

### **Statistical Machine Learning and Parsimony**

SIERRA addresses primarily machine learning problems, with the main goal of making the link between theory and algorithms, and between algorithms and high-impact applications in various engineering and scientific fields, in particular computer vision, bioinformatics, audio processing, text processing and neuro-imaging.

Recent achievements include theoretical and algorithmic work for large-scale convex optimisation, leading to algorithms that make few passes on the data while

still achieving optimal predictive performance in a wide variety of supervised learning situations. Challenges for the future include the development of new methods for unsupervised learning, the design of learning algorithms for parallel and distributed computing architectures, and the theoretical understanding of deep learning.

### Challenges in reinforcement learning

Making reinforcement learning more effective would allow to attack really meaningful tasks, especially stochastic and non-stationary ones. For this purpose, the current trends are to use transfer learning between tasks, and the possibility to integrate prior knowledge.

### Transfer learning

Transfer learning is useful when there is little data available for learning a task. It means using for a new task what has been learned from another task for which more data is available. It is a rather old idea (1993) but the results are modest because its implementation is difficult: it implies to abstract what the system has learned in the first place, but there is no general solution to this problem (what to abstract, how, how to re-use? ...). Another approach of transfer learning is the procedure known as "shaping": learning a simple task, then gradually complicate the task, up to the target task. There are examples of such process in the literature, but no general theory.

### **SCOOL**

The SCOOL project-team (formerly known as SEQUEL) works in the field of digital machine learning. SCOOL aims to study of sequential decision-making problems in uncertainty, in particular bandit problems and the reinforcement-learning problem.

SCOOL's activities span the spectrum from basic research to applications and technology transfer. Concerning basic and formal research, SCOOL focuses on modelling of concrete problems, design of new algorithms and the study of the formal properties of these algorithms (convergence, speed, efficiency ...). On a more algorithmic level, they participate in the efforts concerning the improvement of reinforcement learning algorithms for the resolution of larger and stochastic tasks. This type of tasks naturally includes the problem of managing limited resources in order to best accomplish a given task. SCOOL has been very active in the area of online recommendation systems. In recent years, their work has led to applications in natural language dialog learning tasks and computer vision. Currently, they are placing particular emphasis on solving these problems in non-stationary environments, i.e. environments whose dynamics change over time.

SCOOL now focuses its efforts and thinking on applications in the fields of health, education and sustainable development (energy management on the one hand, agriculture on the other).

## DYOGENE

### Dynamics of Geometric Networks

The scientific focus of DYOGENE is on geometric network dynamics arising in communications. Geometric networks encompass networks with a geometric definition of the existence of links between the nodes, such as **random graphs** and **stochastic geometric networks**.

- **Unsupervised learning for graph-structured data**

In many scenarios, data is naturally represented as a graph either directly (e.g. interactions between agents in an online social network), or after some processing (e.g. nearest neighbour graph between words embedded in some Euclidean space). Fundamental unsupervised learning tasks for such graphical data include graph clustering and graph alignment.

DYOGENE develops efficient algorithms for performing such tasks, with an emphasis on challenging scenarios where the amount of noise in the data is high, so that classical methods fail. In particular, they investigate: spectral methods, message passing algorithms, and graph neural networks.

- **Distributed machine learning**

Modern machine learning requires to process data sets that are distributed over several machines, either because they do not fit on a single machine, or because of privacy constraints. DYOGENE develops novel algorithms for such distributed learning scenarios that efficiently exploit communication resources between data locations, and storage and compute resources at data locations.

- **Energy networks**

DYOGENE develops control schemes for efficient operation of energy networks, involving in particular reinforcement learning methods and online matching algorithms.

## 5.2.2 Heterogeneous/complex data and hybrid models

In addition to the overall challenges in ML seen previously, the challenges for the teams putting the emphasis on data are to learn from heterogeneous data, available through multiple channels; to consider human intervention in the learning loop; to work with data distributed over the network; to work with knowledge sources as well as data sources, integrating models and ontologies in the learning process (see in section 5.4); and finally to obtain good learning performance with little data, in cases where big data sources are not common.

### Heterogeneous data

Data can be obtained from many sources: from distributed databases over the internet or over corporate information systems; from sensors in the Internet of Things; from connected vehicles; from large experimental equipment e.g. in materials science or astrophysics. Working with heterogeneous data is mandatory whatever the means are i.e. directly exploiting the heterogeneity, or defining pre-processing steps to homogenise.

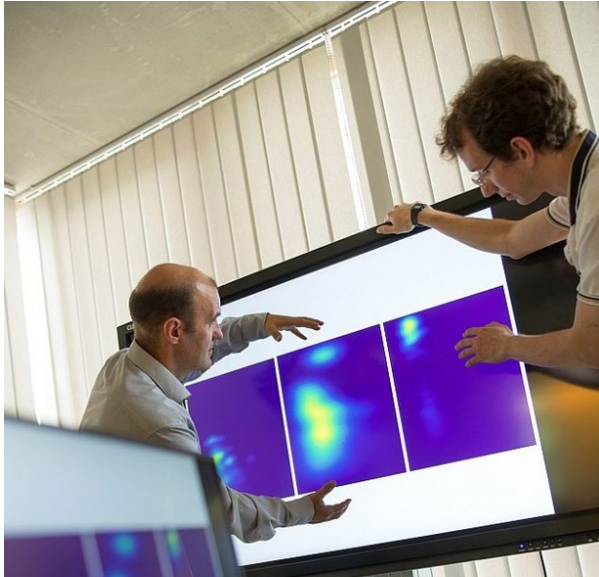
#### **DATASHAPE**

##### **Understanding the shape of data**

Modern complex data, such as time-dependent data, 3D images or graphs, reveals that they often carry an interesting topological or geometric structure. Identifying, extracting and exploiting the topological and geometric features or invariants underlying data has become a problem of major importance to better understand relevant properties of the systems from which they have been generated. Building on solid theoretical and algorithmic bases, geometric inference and computational topology have experienced important developments towards data analysis and machine learning. New mathematically well-founded theories gave birth to the field of Topological Data Analysis (TDA), which is now arousing interest from both academia and industry. During the last few years, TDA, combined with other ML and AI approaches, has witnessed many successful theoretical contributions, with the emergence of persistent homology theory and distance-based approaches, important algorithmic and software developments and real-world successful applications. These developments have opened new theoretical, applied and industrial research directions at the crossing of TDA, ML and AI.

The Inria DataShape team is conducting research activities on topological and geometric approaches in ML and AI with a double academic and industrial/societal objective. First, building on its strong expertise in Topological Data Analysis, DataShape designs new mathematically well-founded topological and geometric methods and algorithms for Data Analysis and ML and make them available to the

data science and AI community through the state-of-the-art software platform GUDHI. Second, thanks to strong and long-standing collaborations with French and international industrial partners, DataShape aims at exploiting its expertise and tools to address challenging problems with high societal and economic impact in particular in personalized medicine, AI-assisted medical diagnosis, or industry.



Topological data analysis

## **MAGNET**

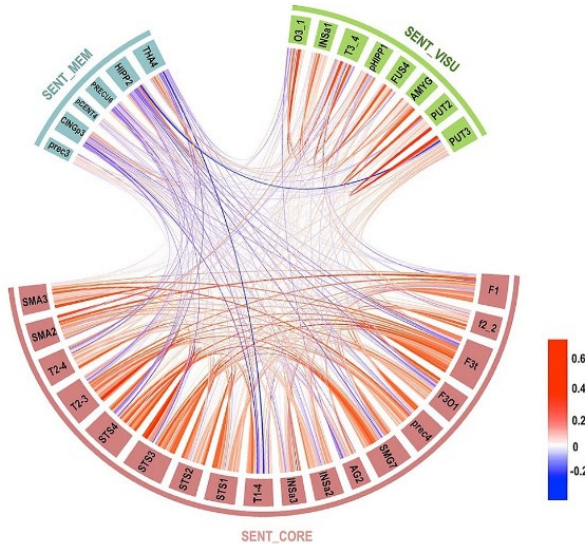
### **Machine Learning in Information Networks**

The Magnet project aims to design new machine learning based methods geared towards mining information networks. Information networks are large collections of interconnected data and documents like citation networks and blog networks among others. For this, they will define new structured prediction methods for (networks of) texts based on machine learning algorithms in graphs. Such algorithms include node classification, link prediction, clustering and probabilistic modelling of graphs. Envisioned applications include browsing, monitoring and recommender systems, and more broadly information extraction in information

networks. Application domains cover social networks for cultural data and e-commerce, and biomedical informatics. Specifically, MAGNET main objectives are:

- Learning graphs, that is graph construction, completion and representation from data and from networks (of texts)
- Learning with graphs, that is the development of innovative techniques for link and structure prediction at various levels of (text) representation.

Each item will also be studied in contexts where little (if any) supervision is available. Therefore, semi-supervised and unsupervised learning will be considered throughout the project.



Graph of extrinsic connectivity links

**STATIFY**

**Bayesian and extreme value statistical models for structured and high dimensional data**

The STATIFY team specializes in the statistical modelling of systems involving data with a complex structure. Faced with the new problems posed by data science and deep learning methods, the objective is to develop mathematically well-founded statistical methods to propose models that capture the variability of the systems under consideration, models that are scalable to process large dimensional data and with guaranteed good levels of accuracy and precision. The targeted applications are mainly brain imaging (or neuroimaging), personalized medicine,

environmental risk analysis and geosciences. STATIFY is therefore a scientific project centred on statistics and wishing to have a strong methodological and application impact in data science.

STATIFY is the natural follow-up of the MISTIS team. This new STATIFY project is naturally based on all the skills developed in MISTIS, but it consolidates or introduces new research directions concerning Bayesian modelling, probabilistic graphical models, models for high dimensional data and finally models for brain imaging, these developments being linked to the arrival of two new permanent members, Julyan Arbel (in September 2016) and Sophie Achard (in September 2019).

This new team is positioned in the theme "Optimisation, learning and statistical methods" of the "Applied mathematics, calculation and simulation" domain. It is a joint project-team between Inria, Grenoble INP, Université Grenoble Alpes and CNRS, through the team's affiliation to the Jean Kuntzmann Laboratory, UMR 5224.

### Human-in-the-learning-loop, explanations

The challenges are on the seamless cooperation of ML algorithms and users for improving the learning process; in order to do so, machine-learning systems must be able to show their progress in a form understandable by humans. Moreover, it should be possible for the human user to obtain explanations from the system on any result obtained. These explanations would be produced during the system's progression and could be linked to input data or to intermediate representations; they could also indicate confidence levels as appropriate.

### **LACODAM**

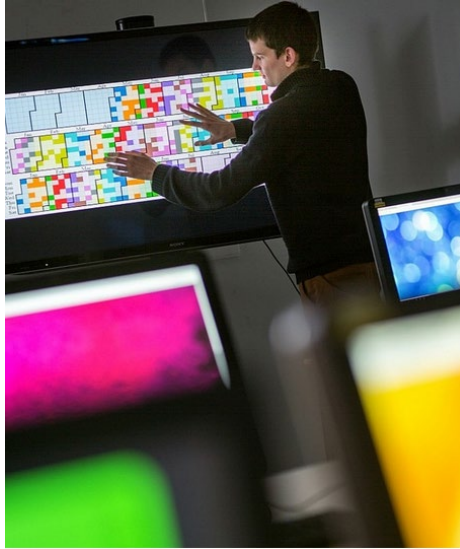
#### **Large scale Collaborative Data Mining**

The objective of the Lacodam team is to facilitate the process of making sense out of (large) amounts of data. This can serve the purpose of deriving knowledge and insights for better decision-making. The team mostly studies approaches that will provide novel tools to data scientists, that can either performs tasks not addressed by any other tools, or that improve the performance in some area for existing tasks (for instance reducing execution time, improving accuracy or better handling imbalanced data).

One of the main research areas of the team are novel methods to discover patterns inside the data. These methods can fall within the fields of data mining (for exploratory analysis of data) or machine learning (for supervised tasks such as classification).

Another key research interest of the team is about *interpretable* machine learning methods. Nowadays, there are many machine learning approaches that have excellent performances, but which are very complex: their decisions cannot be

explained to human users. An exciting recent line of work is to combine performance in the machine learning task while being able to justify the decisions in an understandable way. It can for example be done with *post-hoc interpretability* methods, which for a given decision of the complex machine learning model will approximate its (complex) decision surface *around that point*. This can be done with a much simpler model (ex: linear model), that is understandable by humans.



Detection and characterization of user behaviour in the context of Big data

## LINKMEDIA

### Creating and exploiting explicit links between multimedia fragments

LINKMEDIA focuses on machine interpretation of professional and social multimedia content across all modalities. In this framework, artificial intelligence relies on the design of content models and associated learning algorithms to retrieve, describe and interpret messages edited for humans. Aiming at multimedia analytics, LINKMEDIA develops machine-learning algorithms primarily based on statistical and neural models to extract structure, knowledge, entities or facts from multimedia documents and collections. Multimodality and cross-modality to reconcile symbolic representations (e.g., words in a text or concepts) with continuous observations (e.g., continuous image or signal descriptors) is one of the key challenges for LINKMEDIA, where neural networks embedding appear as a



promising research direction. Hoax detection in social networks combining image processing and natural language processing, hyperlinking in video collections simultaneously leveraging spoken and visual content, interactive news analytics based on content-based proximity graphs are among key subjects that the team addresses.

"User-in-the-loop" analytics, where artificial intelligence is at the service of a user, is also central to the team and raises challenges for humanly supervised machine-based multimedia content interpretation: humans need to understand machine-based decisions and to assess their reliability, two difficult issues with today's data-driven approaches; knowledge and machine learning are strongly entangled in this scenario, requiring mechanisms for human experts to inject knowledge into data interpretation algorithms; malicious users will inevitably temper with data to bias machine-based interpretation in their favour, a situation that current adversarial machine learning can poorly handle; last but not least, evaluation shifts from objective measures on annotated data to user-centric design paradigms that are difficult to cast into objective functions to optimize.

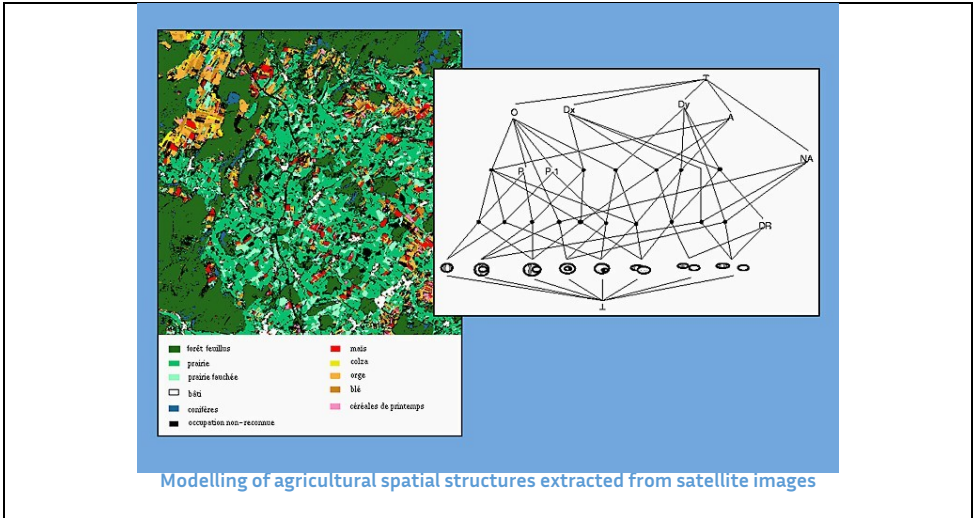
## **ORPAILLEUR**

### **Knowledge discovery, knowledge engineering**

ORPAILLEUR is a project-team at INRIA Nancy-Grand Est and LORIA since the beginning of 2008. It is a rather large and special team as it includes computer scientists, but also a biologist, chemists, and a physician. Life sciences, chemistry, and medicine, are application domains of first importance and the team develops working systems for these domains.

Knowledge discovery in databases –hereafter KDD– consists in processing a large volume of data in order to discover knowledge units that are significant and reusable. Assimilating knowledge units to gold nuggets, and databases to lands or rivers to be explored, the KDD process can be likened to the process of searching for gold. This explains the name of the research team: in French "orpailleur" denotes a person who is searching for gold in rivers or mountains. Moreover, the KDD process is iterative, interactive, and generally controlled by an expert of the data domain, called the analyst. The analyst selects and interprets a subset of the extracted units for obtaining knowledge units having a certain plausibility. As a person searching for gold and having a certain knowledge of the task and of the location, the analyst may use its own knowledge but also knowledge on the domain of data for improving the KDD process.

A way for the KDD process to take advantage of domain knowledge is to be in connection with ontologies relative to the domain of data, for making a step towards the notion of knowledge discovery guided by domain knowledge or KDDK. In the KDDK process, the extracted knowledge units have still "a life" after the interpretation step: they are represented using a knowledge representation formalism to be integrated within an ontology and reused for problem-solving needs. In this way, knowledge discovery is used for extending and updating existing ontologies, showing that knowledge discovery and knowledge representation are complementary tasks and reifying the notion of KDDK.



### Data distributed over the network

There are issues of performance with distributed data, as shown in the KERDATA presentation below. But there is a more fundamental issue linked to privacy. Federated learning has been developed so as to meet privacy requirements when learning with sensible data: the need to ensure "by design" GDPR-compatible processing (e.g. respecting confidentiality with regard to persons whose image is captured by cameras).

#### **KERDATA**

##### **Scalable Storage for Clouds and Beyond**

##### **The HPC-Big Data-AI convergence and the digital continuum**

The tools and cultures of High Performance Computing and Big Data Analytics have evolved in divergent ways. This is to the detriment of both. However, big computations generate Big Data and powerful computational resources are needed to analyse Big Data. More recently, machine learning strongly emerged as a powerful means to enable relevant data analytics at scale. As scientific research increasingly depends on both high-speed computing and data analytics, the potential interoperability and scaling convergence of the corresponding ecosystems (HPC, Big Data, AI) is crucial to the future. In particular, a key milestone will be to achieve convergence through common abstractions and techniques for data storage and processing in support of complex workflows combining simulations, analytics and learning. Such application workflows will need such a convergence to run on hybrid infrastructures combining HPC systems, clouds and edge devices, in a complete digital continuum.

##### **Support AI across the digital continuum**

Integrating and processing high-frequency data streams from multiple sensors scattered over a large territory in a timely manner requires high-performance computing techniques and equipment. For instance, a machine learning earthquake detection solution has to be designed jointly with experts in distributed computing and cyber-infrastructure to enable real-time alerts. Because of the large number of sensors and their high sampling rate, a traditional centralized approach that transfers all data to a single point (e.g., an HPC system or a traditional cloud datacentre) may be impractical. The KerData project-team investigates innovative solutions for the design of efficient data processing architecture across hybrid infrastructures combining supercomputers, clouds and edge systems, in support of distributed machine learning (and, more generally, of scalable distributed data analytics).

In particular, building on the team's previous results in the area of efficient stream processing systems, the goal now is to explore approaches for unified data storage, processing and machine-learning based analytics across the whole digital continuum (i.e., for highly distributed applications deployed on hybrid edge/cloud/HPC infrastructures). Typical target applications include complex workflows combining simulations and analytics, for instance data-enhanced digital twins

### **Machine Learning in the context of Edge stream processing**

This recent Kerdata research axis is worked out in close collaboration with the group of Manish Rutgers University, and with the LACODAM team. It aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. EEW systems are designed to detect and characterize medium and large earthquakes before their damaging effects reach a certain location.

Traditional EEW methods based on seismometers fail to accurately identify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robustness of EEW systems.

In practice, EEW can be seen as a typical classification problem in the machine learning field: multi-sensor data are given in input, and earthquake severity is the classification result. We introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel machine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes.

DMSEEW is based on a new stacking ensemble method that has been evaluated on a real-world dataset validated with geoscientists. The system builds on a geographically distributed infrastructure (deployable on clouds and edge systems), ensuring an efficient computation in terms of response time and robustness to partial infrastructure failures. Our experiments show that DMSEEW is more accurate than the traditional seismometer-only approach and the combined-sensors (GPS and seismometers) approach that adopts the rule of relative strength. These results have been acknowledged by the international AI community through an "Outstanding Paper Award - Special Track on AI for Social Impact" at AAAI-20, an "A\*" conference in the area of Artificial Intelligence:

- Kévin Fauvel, Daniel Balouek-Thomert, Diego Melgar, Pedro Silva, Anthony Simonet, et al.. A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, Feb 2020, New York, United States. pp.1-9.

Other project-teams in this domain: MODAL (Lille), XPOP (Saclay)

### 5.2.3 Machine Learning for Biology and Health

This section lists four project teams using and developing some aspects of machine learning to problems in Biology and Health. Other teams can be found in the section on neurosciences and cognition.

Many applications of Deep Learning have been highlighted in the literature (e.g. in Eric Topol's book "Deep Medicine") or in practical use of technological devices including some machine learning. Life sciences is one of the most complicated fields but an ideal field of application: there are strong (and positive) societal and economic stakes, there are already large amounts of data and knowledge available and formalised. Talking about life-critical applications, the demands are even stronger than for other domains in terms of verification & validation, transparency and traceability, explainability, in order to establish trust.

#### **ABS**

#### **Algorithms, Biology, Structure**

Computational structural biology (CSB) is concerned with the elucidation of the relationship between the structure, dynamics and functions of biomolecules. CSB is fuelled by experimental data of several kinds. On the one hand, genome sequencing projects give access to protein sequences, and ~ 120 millions of sequences have been archived in UNiProtKB/TrEMBL. On the other hand, structure determination experiments (notably X ray crystallography and cryo-electron microscopy) give access to geometric models of molecules – atomic coordinates. Alas, only ~ 150,000 structures have been solved. With one structure for ~ 1000 sequences, we hardly know anything about biological functions at the atomic/molecular level. This state of affairs owes to the high dimensionality of molecular systems. More specifically, recall the following three ingredients.

First, the conformation of a molecule with  $n$  atoms is characterized by  $3n$  Cartesian coordinates and  $3n - 6$  degrees of freedom – one needs to quotient out by rigid motions. In practice,  $n \in [103,105]$ .

Second, to each conformation is associated a potential energy landscape (PEL). The PEL is defined by a function from  $R^{3n} \rightarrow R$ , which is extremely complex – the number of critical points is exponential in the dimension.

Third, molecules deform continuously, and their macroscopic properties depend on ensemble - average values computed over regions of the PEL, as statistical physics tells us. Therefore, estimating structural, thermodynamic, and dynamic properties are very hard problems

Summarizing, there are three main challenges in CSB:

- Predict the 3-dimensional structure of a protein from its amino-acid sequence. This challenge is investigated in the context of the biennial community wide experiment Critical Assessment of Protein Structure Prediction (CASP) –see below.

- Estimate thermodynamic and kinetic properties of a protein or protein complex from its structure.

- Reconstruct the structure of molecular machines involving up to hundreds of subunits – a prerequisite to study their function.

The ABS project team develops original methods to shed new light on these problems. These methods borrow and contribute to several disciplines in computer science and applied mathematics:

- Geometry and topology, since structural models are graphs embedded in 3D.

- Combinatorial optimisation, since graphs are ubiquitous representations both for molecules and molecular networks.

- Machine learning, both supervised (regression, classification) and non-supervised M-(clustering, dimensionality reduction, numerical mathematics).



Modelisation of the influenza virus polymerase

## MIMESIS

### Computational Anatomy and Simulation for Medicine

MIMESIS develops new solutions in the field of surgical training and computer-aided interventions to reduce risk and improve image- and signal-guided therapies.

**Real-time patient-specific computational models** – We are developing computationally efficient, stable, and accurate simulations of (i) soft tissue deformation and other biophysical phenomena to provide instant feedback and visual augmentation during surgery; (ii) electric brain activity and mammalian behaviour to improve medical neuromodulation therapies in patients. Our research also addresses model parametrization to describe patient-specific characteristics of (i) soft tissue (shape, material, conductivity, etc.); (ii) electromagnetic observations of brain activity (electro-/magnetoencephalography, local field potentials, single neuron activity). By extension, we also develop numerical models of tissue-tool interactions, a key component of surgical training systems.

**Data-driven simulation** – This research direction aims at bridging the gap between medical imaging and clinical routine by adapting pre-operative data to the time of the procedure. We address this challenge by combining Bayesian methods with advanced physics-based techniques to handle uncertainties in signal- and image-driven simulations. We are also developing neural networks that can predict the

complex physics of soft tissues and combine them with classical methods to ensure the prediction's explainability and accuracy.



Computer-aided intervention

## MONC

### Mathematical modelling for Oncology

The Monc project team is working in the field of data-driven medicine against cancer. We couple coupling mathematical models and AI with data to address relevant challenges for biologists and clinicians.

It has the following objectives:

- Improve our understanding in cancer biology and pharmacology,
- Assist the development of novel therapeutic approaches,
- Develop personalized decision-helping tools for monitoring the disease and evaluating therapies.

More precisely, we are developing mathematical models – involving partial differential equations (PDE) and built from a precise biological and medical knowledge – combined with novel data assimilation techniques, image processing, statistical methods and artificial intelligence (machine learning, deep learning) – in order to build numerical tools based on available quantitative data about cancer follow-up.

Each type of cancer is different and the models are specifically targeting a limited number of pathologies (\*e.g.\* brain and lung metastases, meningioma, gliomas, soft-tissue sarcoma, lung tumours).





Mathematical modelling for Oncology - Predicting tumour growth and estimating response to treatment

## SISTM

### Statistics In System biology and Translational Medicine

SISTM stands for Statistics in Systems Biology and Translational Medicine. The Research performed in this team is applied to the field of medical sciences and more precisely in infectious diseases and immunology. Specific methods are required to deal with the high dimensional data generated in this field. Specifically, biotechnological improvements allow to measure the various types of cells and their activity in a much more precise way. Hence, in a single sample of blood of a given patient, millions of types of cells ( $2^{40}$ ) can potentially be determined by mass cytometry, expression of 20 000 genes by RNA-sequencing and production of hundreds to thousands of proteins by Multiplex or spectrometry. Hence, the analysis of these data requires dimension reduction approaches (1,2), unsupervised (3), or supervised (e.g. based on Random forests (4) classification in multidimensional space, adapted statistical tests for high dimensional setting (5). The results obtained from these high dimensional spaces provides much more knowledge from single clinical studies which is very useful for the development of vaccines for instance (6). The adaptation of the interventions based on the data collected over time during the trials is the next step (7).

1. Sutton M, Thiébaud R, Liqueur B. Sparse partial least squares with group and subgroup structure. *Stat Med* (2018) 37:3338–3356. doi:10.1002/sim.7821
2. Lorenzo H, Misbah R, Odeber J, Morange PE, Saracco J, Tregouet DA, Thiébaud R. High-dimensional multi-block analysis of factors associated with thrombin generation potential. in *Proceedings - IEEE Symposium on Computer-Based Medical Systems (Institute of Electrical and Electronics Engineers Inc.)*, 453–458. doi:10.1109/CBMS.2019.00094
3. Hejblum BP, Alkassim C, Gottardo R, Caron F, Thiébaud R. Sequential dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. *Ann Appl Stat* (2019) 13:638–660. doi:10.1214/18-AOAS1209

4. Capitaine L, Genuer R, Thiébaud R. Fréchet random forests. (2019) Available at: <http://arxiv.org/abs/1906.01741> [Accessed June 4, 2020]
5. Agniel, Denis, Hejblum B. Variance component score test for time-course gene set analysis of longitudinal RNA-seq data | Biostatistics | Oxford Academic. Available at: <https://academic.oup.com/biostatistics/article/18/4/589/3065599> [Accessed June 5, 2020]
6. Rechten A, Richert L, Lorenzo H, Martrus G, Hejblum B, Dahlke C, Kasonta R, Zinser M, Stubbe H, Matschil U, et al. Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV. Cell Rep (2017) 20:2251–2261. doi:10.1016/j.celrep.2017.08.023
7. Pasin C, Dufour F, Villain L, Zhang H, Thiébaud R. Controlling IL-7 Injections in HIV-Infected Patients. Bull Math Biol (2018) 80:2349–2377. doi:10.1007/s11538-018-0465-8

## 5.2.4 Exploratory Actions (AEx) and Inria Challenges

### Inria Challenge- “Hybrid Approaches for Interpretable Artificial Intelligence” (HyAIAI)

Project teams: LACODAM, TAU, SCOOOL, MAGNET, ORPAILLEUR, MULTISPEECH

There is an emerging research trend aiming to provide interpretations for the decision of “black box” ML algorithms such as Deep Learning (DL) ones.

In the HyAIAI Inria Challenge, we claim that there is a need for two-way communication between a DL model and a user: of course, the user must understand the DL decisions, but when the user participates in the training of the DL model, s/he must also be able to provide expressive feedback to the model. We believe that this two-way communication requires a hybrid approach: complex numerical models must play the role of the learning engine due to their performance, but they must be combined with symbolic models in order to ensure an effective communication with the user.

### Inria Challenge "HighPerformance Computing and Big Data" (HPC-BigData)

See <https://project.inria.fr/hpcbigdata/> for the full list of project-teams.

Big Data analytics is becoming more compute-intensive thanks to deep learning, while data handling is becoming a major concern for scientific computing. The Challenge HPC-BigData gathers teams from the HPC, Big Data and Machine Learning areas to work at the intersection between these domains.

### AEx-AI4HI – Artificial Intelligence for human intelligence

Project team: CORSE

The objective of AI4HI is to bring together advances in Artificial Intelligence (classification, statistical approaches, deep learning) and compilation and teaching skills in order to improve teaching by automatically generating exercises and recommending them to students. The project focusses on the teaching of programming and debugging to beginners.

### AEx-MALESI - MACHine LEarning for Simulation

Project team: TONUS

Physical simulations require the ultra-precise resolution of partial differential equations (PDE). Current numerical schemes can generate significant numerical

pollution. The project aims to develop image-based learning methods to correct these numerical shortcomings while demonstrating the important properties of convergence and universality.

### AEx-SR4SG : Sequential collaborative learning of recommendations for sustainable gardening

Project team: SCOOl

The objective of the SR4SG is twofold: federate an ambitious mixed community around the theme "Reinforcement Learning for Sustainable Gardening" and provide a common application platform to integrate progressively the research expertise of all stakeholders (sequential learning, ontology, hci, distributed computing, data certification, botany, functional ecology, epidemiology, agronomy, agro ecology, etc).

### AEx-TRACME – Multi-scale causal pathways

Project team: GEOTSTAT

This project focuses on modelling a physical system from measurements on that system. How, starting from observations, to build a reliable model of the system dynamics? When multiple processes interact at different scales, how to obtain a significant model at each of these scales? How to relate these models to physical quantities, such as the amount of energy, or that of information, which are processed at each scale? This project proposes to identify causally equivalent classes of system states, then model their evolution with a stochastic process. Renormalizing these equations is necessary in order to relate the scale of the continuum to that, arbitrary, at which data are acquired. Applications primarily concern natural sciences.

### AEx-FLAMED – Federated learning and analytics on Medical Data

Project team: MAGNET

FLAMED aims to explore a decentralised approach to Artificial Intelligence applied to health. In close collaboration with the university-affiliated hospital of Lille, FLAMED objective is to carry out data analysis and machine learning (decentralised federated learning) tasks involving several hospitals while allowing each site to keep its data internally and guaranteeing confidentiality.

### AEx-MAMMALS - Memory-augmented Models for low-latency Machine-learning Serving

Project team: NEO

MAMMALS aims to provide low-latency inferences by running—close to the end user—simple machine learning models that can also take advantage of a (small) local data store of examples. The focus is on algorithms to learn online what to store locally to improve inference quality and achieve domain adaptation. MAMALS will lead to deepen the understanding of the relation between memorization and generalization that is still wanting even in the static setting.



## 5.2.5 Software: SCIKIT-LEARN

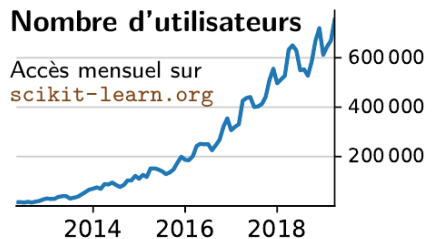
### The Python reference library for Machine Learning

Worldwide, scikit-learn is the first open source machine learning software led by a research community. It rivals in popularity the tools developed by the GAFA.

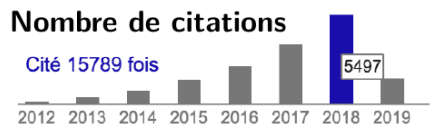
**The scikit-learn vision:** scikit-learn has been developed by the Inria Parietal team since 2010 in order to provide access to statistical learning to as many people as possible, particularly neuroscientists. By providing an effective tool, simple to use and very well documented with hundreds of examples, the developers of scikit-learn have contributed to the democratization of statistical learning that fuelled the current artificial intelligence revolution. With an impact much wider than neurosciences, the Inria researchers and engineers behind scikit-learn's success have allowed the use of statistical learning in all experimental sciences from chemistry, biology and physics, as well as in many industrial applications.

**Scikit-learn: a reference in statistical learning.** Scikit-learn brings together more than 180 different statistical learning models. It encompasses many aspects of this discipline of the applied mathematics and provides a set of algorithmic reference tools, as found in books on the subject. Its documentation -<http://scikit-learn.org>- is itself an introduction to statistical learning. It is considered a pedagogical tool and would be over a thousand pages on paper format. Scikit-learn does not directly include deep learning architectures but can be connected to DL libraries as needed.

**Usage Metrics.** As scikit-learn is a free software, it is difficult to have exact figures of its number of users. However, the website statistics have shown more than 42 million visits in 2018 and 700,000 monthly users (figure on the right).



GitHub, which hosts the project's source code, reports close to 17,000 forks and 35,000 stars. Scikit-learn represents 39 years\*person of work. It is the third most popular open source machine learning software, behind two software tools developed by Google (source). A survey conducted a few years ago identified 63% of users in industry, and 34% in academia. The academic paper of reference has been cited 25,000 times on Google scholar since 2012 with 8200 citations in 2019 (figure on the right).

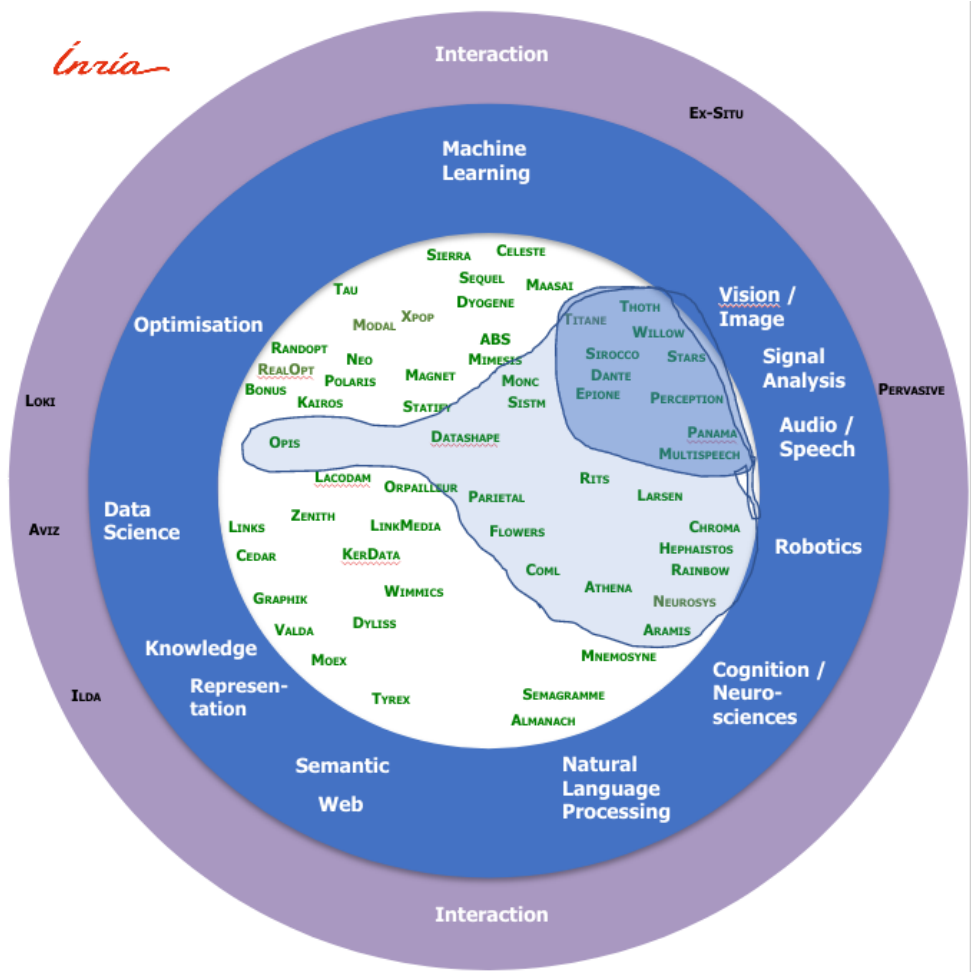


The scikit-learn consortium hosted by the Inria foundation was born in September 2018 with the support of 7 companies: Microsoft, BCG, AXA, BNP Paribas-Cardif, Intel, NVIDIA, and Dataiku, joined by Fujitsu. This partnership/sponsorship demonstrates the industrial impact of scikit-learn and will enable the long-term financing of the software.

### 5.3. Signal analysis, vision, speech

Signal analysis, in particular vision and pattern recognition, is the starting point of the current hype on Deep Learning: since 2012, Deep learning systems won *\*all\** the challenges in vision and pattern recognition, something that convinced almost all researchers and practitioners in the field to convert to Deep Learning. These successes also reached speech recognition, and gradually became utmost popular in most fields of Computer Science, while being quickly transferred to the corresponding industry: the MobilEyevision system empowers cars' self-driving abilities, while voice-guided assistants such as Siri, Cortana, or Amazon Echo are put in use every day by millions of users.

Object recognition —or, in a broader sense, scene understanding— is the ultimate scientific challenge of computer vision: After 40 years of research, even though huge progresses have been made in identifying the familiar objects (chair, person, pet), scene categories (beach, forest, office), and activity patterns (conversation, dance, picnic) depicted in family pictures, news segments, or feature films, human-like understanding of complete scenes is still beyond the capabilities of today's vision systems in part because of the lack of common sense (i.e., general a priori knowledge) of all current learning systems. However, the impact of current and future object recognition and scene understanding technology will continue to grow in application domains as varied as defence, entertainment, health care, human-computer interaction, image retrieval and data mining, industrial and personal robotics, manufacturing, scientific image analysis, surveillance and security, and transportation.



The challenges in **signal analysis for vision** are: (i) scaling up; (ii) from still images to video; (iii) multi-modality; (iv) introduction of a priori knowledge.

**Scaling up**

Modern vision systems must be able to deal with high volume and high frequency data at inference time: for example, surveillance systems in public places, robots moving in unknown environments, web search engines in images have to process huge quantities of data. Vision systems must not only process these data at high speed, but need to reach high levels of precision in order to free operators from

checking the results and post-processing. Even precision rates of 99.9% for image classification on mission-critical operations are not enough when processing millions of images, as the remaining 0.1% will need hours of human processing.

### From images to video

Despite the limitations of today's scene understanding technology, tremendous progress has been accomplished in the past ten years, due in part to the formulation of object recognition as a statistical pattern matching problem. The emphasis is in general on the features defining the patterns and on the algorithms used to learn and recognize them, rather than on the representation of object, scene, and activity categories, or the integrated interpretation of the various scene elements.

### Multi-modality

Understanding vision data can be improved by different means: on the web, metadata provided with images and videos can be used to filter out several hypotheses, and to guide the system towards the recognition of specific objects, events, situations. Another option is to use multimodality, that is, signals coming from various channels e.g. infrared, laser, magnetic data etc. it is also desirable to use a combination of auditory signal with vision (images or video) if available.

### Introduction of a priori knowledge

Another option for improving vision applications is to introduce a priori knowledge in the recognition engine. One example consists in adding information about the anatomy and pathology of a patient for better analysis of biomedical images; in other domains, contextual information, information about a situation, about a task, localisation data, etc. can be used for disambiguating candidate interpretations. However, the question of how to provide this a priori knowledge is not solved in the general case: specific methods and specific knowledge representations must be established for dealing with a target application in vision understanding.

## WILLOW

### Models of visual object recognition and scene understanding

WILLOW addresses fundamental computer vision problems such as three-dimensional perception, computational photography, and image and video understanding. It investigates new models of image content (what makes a good visual vocabulary?) and of the interpretation process (what is a good recognition architecture?).

Despite the tremendous progress in visual recognition in the last 10 years, current visual recognition systems still require large amounts of carefully annotated training data, often use black-box architectures that do not model the 3D physical nature of the visual world, and do not capture real-world semantics. WILLOW addresses these limitations by developing models of the entire visual



understanding process that are learnable without the need for direct supervision, support complex reasoning about visual data, and are grounded in interactions with the physical world. More concretely, WILLOW addresses fundamental scientific challenges along four research axes: (i) visual recognition in images and videos with an emphasis on weakly supervised learning; (ii) learning embodied visual representations for robotic manipulation and locomotion; (iii) image restoration and enhancement; and (iv) 3D object and scene modelling, analysis and retrieval.

Recent achievements of the team include theoretical work on the geometric foundations of computer vision, new advances in image restoration tasks such as deblurring, denoising, or upsampling, and weakly supervised methods of learning powerful representation for text-video retrieval and temporal action localization. WILLOW members collaborate closely with the SIERRA and THOTH teams at Inria, and researchers at places such as Carnegie-Mellon University, UC Berkeley, or Facebook AI Research, in efforts that reflect the strong synergy between machine learning and computer vision, with new opportunities in domains ranging from archaeology to robotics. Challenges for the future include the development of minimally supervised models for visual recognition in large-scale image and video datasets, and vision-driven autonomous agents.



## STARS

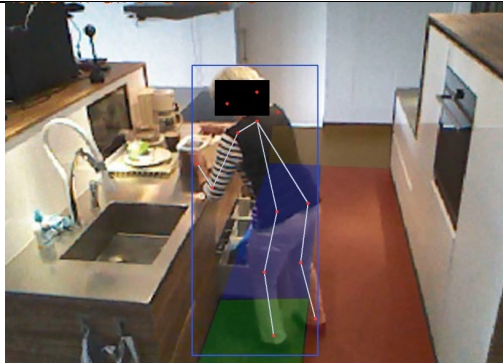
### Spatio-Temporal Activity Recognition Systems

Many advanced studies have been done in Computer Vision and in particular in Scene Understanding during these last few years. Scene Understanding is the process, often real time, of perceiving, analysing and elaborating an interpretation of a 3D dynamic scene observed through a network of sensors (e.g. video cameras). This process consists mainly in matching signal information coming from sensors observing the scene with models which humans are using to understand the scene. Based on that, scene understanding is both adding and extracting semantic from the sensor data characterizing a scene. This scene can contain a number of physical objects of various types (e.g. people, vehicle) interacting with each other or with their environment (e.g. equipment) more or less structured. The scene can last few instants (e.g. the fall of a person) or few months (e.g. the depression of a person), can be limited to a laboratory slide observed through a microscope or go beyond the size of a city. Sensors include usually cameras (e.g. omni-directional, infrared, Depth), but also may include microphones and other sensors (e.g. optical cells, contact sensors, physiological sensors, accelerometers, radars, smoke detectors, smart phones).

Scene understanding is influenced by cognitive vision and it requires at least the melding of three areas: **computer vision, machine learning and software engineering**. Scene understanding can achieve five levels of generic computer vision functionality of detection, localization, tracking, recognition and understanding. But scene understanding systems go beyond the detection of visual features such as corners, edges and moving regions to extract information related to the physical world which is meaningful for human operators. Its requirement is also to achieve more robust, resilient, adaptable computer vision functionalities by endowing them with a cognitive faculty: the ability to learn, adapt, weigh alternative solutions, and develop new strategies for analysis and interpretation.

Concerning scene understanding, STARS team has developed original automated systems to understand human behaviours in a large variety of environment for different applications:

- in metro stations, in streets and on-board trains: fighting, abandoned luggage, graffiti, fraud, crowd behaviour,
- on airport aprons: aircraft arrival, aircraft refuelling, luggage loading/unloading, marshalling,
- in bank agencies: bank attack, access control in buildings, using ATM machines,
- homecare applications for monitoring older people activities: cooking, sleeping, preparing coffee, watching TV, preparing pill box, falling,
- smart home, office behaviour monitoring for ambient intelligence: reading, drinking,
- supermarket monitoring for business intelligence: stopping, queuing, picking up object,
- biological applications: wasp monitoring.
- biometrics: facial expression
- dementia and cognitive disorder: early diagnostic based on behaviour and emotion monitoring



Preparing coffee

To build these systems, the STARS team has designed novel technologies for video generation [Wang 2020], people Re-Identification [Chen 2021] and for the recognition of human activities using in particular 2D or 3D video cameras. More specifically, they have combined 4 categories of algorithms to recognise human activities:

- Recognition engines using hand-crafted ontologies based on rules modelling expert knowledge. These activity recognition engines are easily extensible and allow later integration of additional sensor information when available [Crispim 2016].
- Supervised learning methods based on positive/negative samples representative of the targeted activities which have to be specified by users. These methods are usually based on Deep Learning computing robust spatio-temporal descriptors [Das 2019].
- Unsupervised (fully automated or weakly or partially supervised) learning methods based on clustering of frequent activity patterns on large datasets which can generate/discover new activity models [Negin 2019].
- Attention mechanisms (self-supervision or focus on the spatial or temporal dimension) to guide the learning methods to focus on the most salient information within a video [Das 2020].

C. Crispim-Junior, K. Avgerinakis, V. Buso, G. Meditskos, A. Briassouli, J. Benois-Pineau, Y. Kompatsiaris and F. Bremond. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition, Transactions on Pattern Analysis and Machine Intelligence - PAMI 2016.

S. Das, R. Dai, M. Koperski, L. Minciullo, L. Garattoni, F. Bremond and G. Francesca. Toyota Smarthome: Real-World Activities of Daily Living with supplementary. In Proceedings of the 17th International Conference on Computer Vision, ICCV 2019, in Seoul, Korea, October 27 to November 2, 2019.

F. Negin and F. Bremond. An Unsupervised Framework for Online Spatiotemporal Detection of Activities of Daily Living by Hierarchical Activity Models, in Sensors 2019, 19, 1-27, doi:10.3390/s19194237; 29 September 2019.

Y. Wang, P. Bilinski, F. Bremond and A. Dantcheva. G<sup>3</sup>AN: Disentangling appearance and motion for video generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle-online, US, June 14-19, 2020.

S. Das, S. Sharma, R. Dai, F. Bremond and M. Thonnat. VPNet: Learning Video-Pose Embedding for Activities of Daily Living. In Proceedings of the 16th European Conference on Computer Vision, ECCV 2020, arXiv:2007.03056, online, UK, 23-28 August 2020.

H. Chen, B. Lagadec and F. Bremond. Enhancing Diversity in Teacher-Student Networks via Asymmetric branches for Unsupervised Person Re-identification. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Virtual, January 5-9, 2021.

## THOTH

### **Learning visual models from large-scale data**

The quantity of digital images and videos available on-line continues to grow at a phenomenal speed: home users put their movies on YouTube and their images on Flickr; journalists and scientists set up web pages to disseminate news and research results; and audio-visual archives from TV broadcasts are opening to the public. In 2021, it is expected that nearly 82% of the Internet traffic will be due to videos, and that it would take an individual over 5 million years to watch the amount of video that will cross global IP networks each month by then. Thus, there is a pressing and in fact increasing demand to annotate and index this visual content for home and professional users alike. The available text and audio metadata is typically not sufficient by itself for answering most queries, and visual data must come into play. On the other hand, it is not imaginable to learn the models of visual content required to answer these queries by manually and precisely annotating every relevant concept, object, scene, or action category in a representative sample of everyday conditions—if only because it may be difficult, or even impossible to decide a priori what are the relevant categories and the proper granularity level.

The main goal of THOTH is to automatically explore large collections of data, select the relevant information, and learn the structure and parameters of visual models. There are three main challenges: (1) designing and learning structured models capable of representing complex visual information; (2) on-line joint learning of visual models from textual annotation, sound, image and video; and (3) large-scale learning and optimisation. Another important focus is (4) data collection and evaluation.

Today's object recognition and scene understanding technology operates in a very different setting; it mostly relies on fully supervised classification engines, and visual models are essentially (piecewise) rigid templates learned from hand labeled images. The sheer scale of on-line data and the nature of the embedded annotation call for a departure from this fully supervised scenario. The main idea of the Thoth project-team is to develop a new framework for learning the structure and parameters of visual models by actively exploring large digital image and video sources (off-line archives as well as growing on-line content, with millions of images and thousands of hours of video), and exploiting the weak supervisory signal provided by the accompanying metadata. This huge volume of visual training data will allow us to learn complex non-linear models with a large number of parameters, such as deep convolutional networks and higher-order graphical models. This is an ambitious goal, given the sheer volume and intrinsic variability of the visual data available on-line, and the lack of a universally accepted formalism for modeling it. Yet, the potential payoff is a breakthrough in visual object recognition and scene understanding capabilities. Further, recent advances at a smaller scale suggest that this is realistic. For example, it is already possible to determine the identity of multiple people from news images and their captions, or to learn human action models from video scripts. There has also been recent progress in adapting supervised machine learning technology to large-scale

settings, where the training data is very large and potentially infinite, and some of it may not be labeled. Methods that adapt the structure of visual models to the data are also emerging, and the growing computational power and storage capacity of modern computers are enabling factors that should of course not be neglected.



Learning Motion Pattern in Videos

## SIROCCO

### Analysis representation, compression and communication of visual data

The research agenda of the Sirocco team is the design of mathematical models and algorithms for computational imaging, leveraging signal processing and machine learning methods, with a recent focus on emerging modalities such as high dynamic range imaging, light fields and omni-directional imaging. The research problems addressed by the team are at the intersection between signal processing, computer vision, machine learning and information theory. More precise research topics are:

- Visual data analysis with computer vision problems such as scene depth and scene flow estimation

- Signal processing and learning methods for visual data representation and compression. This includes sparse, low rank and graph-based models for different imaging modalities,
- Algorithms for inverse problems in visual data processing such as compressive acquisition, restoration, super-resolution.
- Information theoretic tools and coding for interactive communication



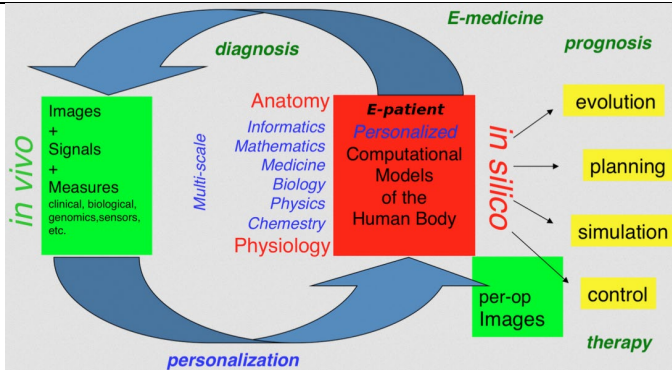
Learning Scene Depth from a Flexible Subset of Dense and Sparse Light Field Views

## EPIONE

### E-Patient: Images, Data & MOdels for e-MedicINE

The EPIONE long-term goal is to contribute to the development of what it is call the e-patient (digital patient) for e-medicine (digital medicine).

- the **e-patient** (or digital patient) is a set of computational models of the human body able to describe and simulate the anatomy and the physiology of the patient's organs and tissues, at various scales, for an individual or a population. The e-patient can be seen as a framework to integrate and analyze in a coherent manner the heterogeneous information measured on the patient from disparate sources: imaging, biological, clinical, sensors...
- **e-medicine** (or digital medicine) is defined as the computational tools applied to the e-patient to assist the physician and the surgeon in their medical practice, to assess the diagnosis/prognosis, and to plan, control and evaluate the therapy.



The models that govern the algorithms designed for e-patients and e-medicine come from various disciplines: informatics, mathematics, medicine, statistics, physics, biology, chemistry, etc. The parameters of those models must be adjusted to an individual or a population based on the available images, signals and data. This adjustment is called personalization and usually requires the resolution of difficult inverse problems.

EPIONE's research objectives are organized along 5 scientific axes:

1. Biomedical Image Analysis & Machine Learning
2. Imaging & Phenomics, Biostatistics
3. Computational Anatomy, Geometric Statistics
4. Computational Physiology & Image-Guided Therapy
5. Computational Cardiology & Image-Based Cardiac Interventions

## DANTE

### Dynamic Networks: Temporal and Structural Capture Approach

The DANTE team develops machine learning techniques and signal processing algorithms with the main objective of endowing them with solid theoretical foundations, physical interpretability and resource-efficiency.

With a culture rooted at the interface of signal processing and machine learning, the team's expertise leverages the notion of parsimony and its structured variants – and noticeably that of graphs – which play a fundamental role to warrant the identifiability of decompositions in latent spaces, such as inverse problems in high dimensional signal processing.

Recent achievements of the team include distributed algorithms to learn from highly compressed data representations with privacy guarantees, and techniques to exploit random walks on graphs for semi-supervised learning in difficult settings. A major challenge is to leverage these ideas to ensure not only resource-efficient methods, but also explainable decisions and interpretable learnt parameters, all being major societal challenges to make "algorithmic decisions" reliable and acceptable.

The challenges in **signal analysis for speech and sound** have a lot in common with the previous list: scaling up, multimodality, introduction of prior knowledge, are relevant for audio applications too. The target applications are speaker identification, speech understanding, dialogue – including for robots, source separation (in the case of multiple conversations), emotion recognition and synthesis, and automatic translation in real time. In the case of audio signals, it is also mandatory to develop or to have access to high volume data for machine learning. Online incremental learning might be needed for real time speech processing.

## PERCEPTION

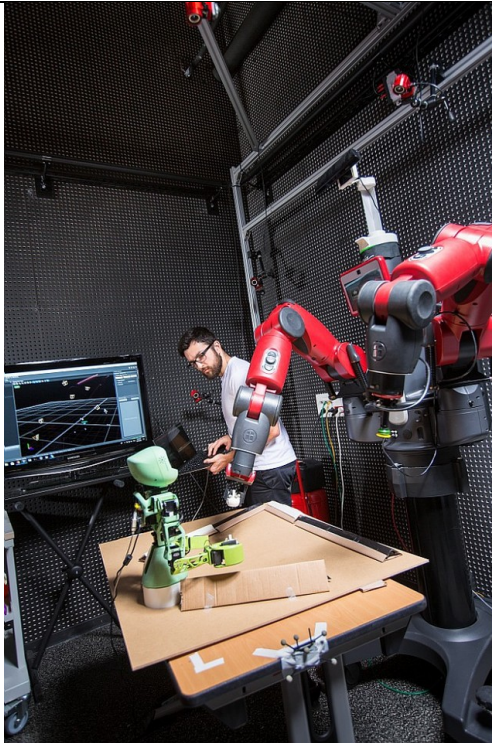
### Interpretation and Modelling of Images and Sounds

The research agenda of the PERCEPTION group is the investigation and implementation of computational models for mapping images and sounds onto meaning and onto actions. PERCEPTION team members address this challenging problem with an interdisciplinary approach that spans the following topics: computer vision, auditory signal processing, audio scene analysis, machine learning, and robotics. In particular, we develop methods for the representation and recognition of visual and auditory objects and events, audio-visual fusion, recognition of human actions, gestures and speech, spatial hearing, and human-robot interaction.

Research topics:

- **Computer vision:** spatio-temporal representation of 2D and 3D visual information, action and gesture recognition, analysis of human faces, 3D sensors, binocular vision, multiple-camera systems, person and object tracking in video sequences
- **Auditory scene analysis:** binocular hearing, multiple sound source localization, tracking and separation, speech communication, sound-event classification, speaker diarization, acoustic signal enhancement.
- **Machine learning:** probabilistic mixture models, linear and non-linear dimension reduction, manifold learning, graphical models, Bayesian inference, neural networks and deep learning.
- **Robotics:** robot vision, robot hearing, human robot interaction, data fusion, software architectures.





Poppy torso learning to speak with Baxter mommy

Specific challenges on the field of **speech** are:

#### Use of pre-trained self-supervised models for speech recognition,

The application of self-supervised pre-training methods to speech could give in the coming years results as spectacular as for text with many applications in the field of automatic speech processing in low-resource languages (some of which have no text resources). In general, the application of machine learning to economically non-dominant languages or cultures is very important to avoid widening the digital divide.

#### Process “real-world” audio signals

Automatic processing of the actual audio signal is an unresolved problem (contrary to what one seems to think). Source separation does not work well 'in the wild'. As a result, the drop in performance of automatic language processing with ecological data does not allow for a whole range of medical or educational applications. Generally speaking, the learning machine must learn to go outside the boxed data

framework, and face the difficult problem of real data head-on if it is to be used in concrete applications.

## MULTISPEECH

### Speech Modeling for Facilitating Oral-Based Communication

Beyond supervised black box learning – MULTISPEECH studies fundamental challenges relating to deep learning. For instance, they explore hybrid methods combining deep learning with statistical modeling, signal processing, or symbolic reasoning to increase performance and explainability, they design weakly supervised learning or transfer learning methods to exploit noisy labels or out-of-domain data, and they explore speech anonymization methods to preserve the data subjects' privacy.

**Speech production** - MULTISPEECH develops an articulatory speech synthesis system based on modeling the dynamics of the vocal tract, and a highly realistic talking head based on dynamic animation of the mouth and facial expressions. Applications include computer animation, and language learning for children with difficulties or the hearing impaired.

**Speech in its environment** - MULTISPEECH designs algorithms to enhance speech in the presence of acoustic echo, reverberation, noise, and competing speakers, and to achieve robust speech and speaker recognition in such conditions. They model semantics in order to further improve recognition and to classify the spoken contents. Finally, they develop methods to estimate the room's acoustic properties and to detect ambient sound events. Beyond spoken communication, these methods have many applications in sound monitoring, robot audition, building acoustics, augmented reality, or social media monitoring.



A highly realistic talking head based on dynamic animation of the mouth and facial expressions

## PANAMA

### Parsimony and New Algorithms for Signal and Audio Modeling

At the interface between audio modeling and mathematical signal processing, the global objective of PANAMA is to develop mathematically founded and

algorithmically efficient techniques to model, acquire and process high-dimensional signals, with a strong emphasis on acoustic data.

Applications fuel the proposed mathematical and statistical frameworks with practical scenarios, and the developed algorithms are extensively tested on targeted applications. PANAMA's methodology relies on a closed loop between theoretical investigations, algorithmic development and empirical studies.

The scientific foundations of PANAMA are focused on sparse representations and probabilistic modeling, and its scientific scope is extended in three major directions:

- The extension of the sparse representation paradigm towards that of "sparse modeling", with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning.
- A focus on sophisticated probabilistic models and advanced statistical methods to account for complex dependencies between multi-layered variables (such as in audiovisual streams, musical contents, biomedical data, remote sensing ...).
- The investigation of graph-based representations, processing and transforms, with the goal to describe, model and infer underlying structures within content streams or data sets.

## Exploratory actions (AExs)

### AEx- Ayana - AI and Remote Sensing on board for the New Space

The AYANA AEx is an interdisciplinary project using knowledge in stochastic modeling, image processing, artificial intelligence, remote sensing and embedded electronics/computing. The aerospace sector is expanding and changing ("New Space"). It is currently undergoing a great many changes both from the point of view of the sensors at the spectral level (uncooled IRT, far ultraviolet, etc.) and at the material level (the arrival of nano-technologies or the new generation of "Systems on chips" (SoCs) for example), that from the point of view of the carriers of these sensors: high resolution geostationary satellites; Leo-type low-orbiting satellites; or mini-satellites and industrial cube-sats in constellation. AYANA will work on a large number of data, consisting of very large images, having very varied resolutions and spectral components, and forming time series at frequencies of 1 to 60 Hz. For the embedded electronics/computing part, AYANA will work in close collaboration with specialists in the field located in Europe, working at space agencies and/or for industrial contractors.

### AEx- ACOUST.IA - Artificial Intelligence to support Building Acoustics

Project team: MULTISPEECH

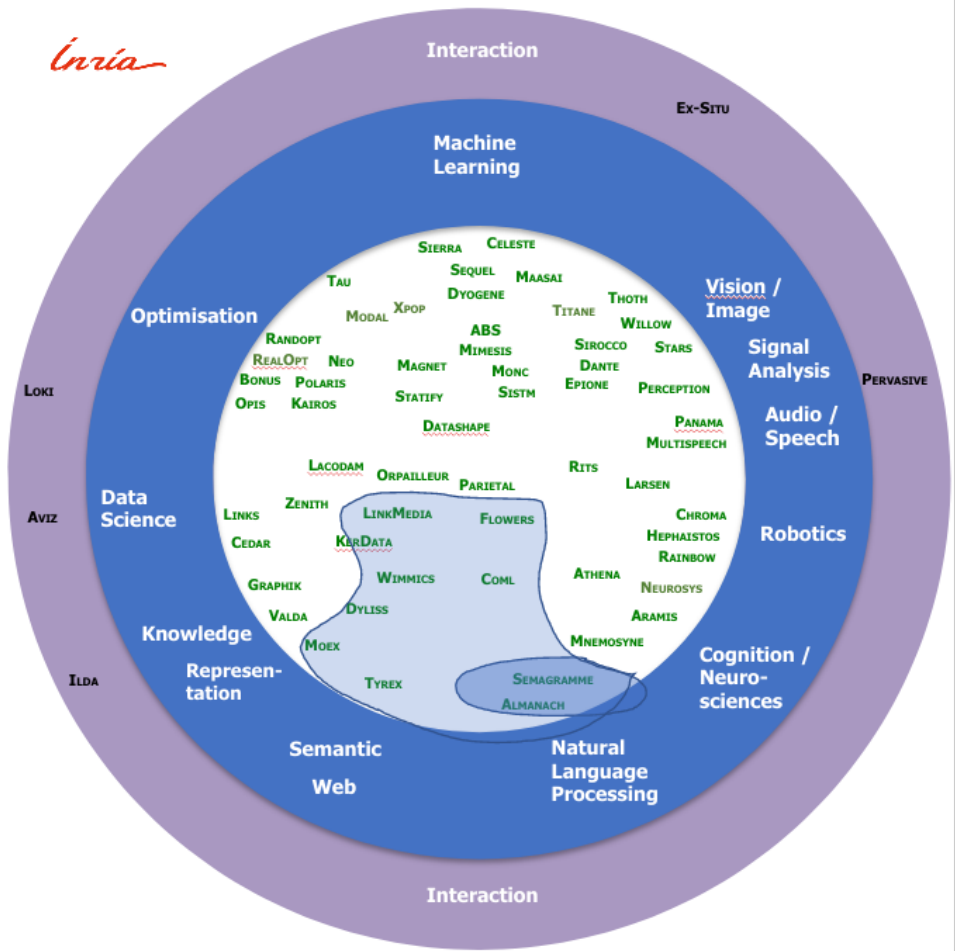
Is it possible to establish the acoustic profile of a room by simply recording a clap? This is the objective of ACOUST.IA, which aims to radically simplify and improve the accuracy of acoustic diagnosis of buildings, an important public health issue, thanks to artificial intelligence and signal processing. Innovative approaches combining

supervised learning, statistical and physical modelling, and multi-channel audio processing will be developed to overcome the limitations of the manual, costly and iterative approaches currently used.

Other project-teams in this domain: TITANE (Sophia Antipolis), MORPHEO (Grenoble)(

## 5.4. Natural language processing

The field of Natural Language Processing (NLP) goes back to the 1950s. Yet it is still of crucial importance today for the new information society. Its goal is to process natural language texts, either for analysing existing texts/generating new texts or for achieving human-like language processing for a range of tasks or applications. These applications, regrouped under the term 'language engineering', include machine translation, question answering, information retrieval, information extraction, text mining, reading and writing aid, and many others. From a more research-oriented point of view, empirical linguistics and digital humanities can be also viewed as application domains of NLP.



NLP is a transdisciplinary domain; it requires an expertise in formal and descriptive linguistics (to develop linguistic models of human languages), in computer science and algorithmic (to design and develop efficient programs that can deal with such models) and in applied mathematics (to automatically acquire linguistic or general knowledge). Processing natural language texts is a difficult task, in particular because of the large amount of ambiguity in natural language, the specificities of individual languages and dialects and because many users do not necessarily conform to grammatical and spelling conventions, when such conventions exist.

The first decades of NLP mostly focused on symbolic approaches, also contributing major notions to Computer Science, especially in formal grammar theory and parsing techniques. Linguistic knowledge was mostly encoded in the form of manually developed grammars and lexical databases. Over the last two decades statistical and machine learning based approaches (word embedding, RNN, Transformers) have greatly renewed the field, bringing annotated corpora to centre stage, and significantly improving the state of the art.

### Hybridisation between ML and symbolic models

Despite important developments made in recent years, natural dialogue tasks continue to yield unimpressive results. They suffer from many problems (e.g. poorly posed problem, lack of evaluation metrics and difficulty in generalizing outside the training set). But one of the central problems is also to consider dialogue as a pure machine learning problem, whereas putting the human being in the loop is essential, which implies dialogue with other disciplines (social sciences, cognitive sciences, etc.). Symbolic approaches retain specific advantages, and best results could be obtained when leveraging all types of resources within hybrid systems coupling symbolic and statistical techniques.

## ALMANACH

### Automatic Language Modelling and Analysis & Computational Humanities

The ALMAnaCH project-team (ALMAnaCH was created as an Inria team (“équipe”) on the 1st January, 2017 and as a project-team on the 1st July 2019.) brings together specialists of a pluri-disciplinary research domain at the interface between computer science, linguistics, statistics, and the humanities, namely that of **natural language processing, computational linguistics and digital and computational humanities and social sciences**.

**Computational linguistics** is an interdisciplinary field dealing with the computational modelling of natural language. Research in this field is driven both by the theoretical goal of understanding human language and by practical applications in **Natural Language Processing (NLP)** such as linguistic analysis (syntactic and semantic parsing, for instance), machine translation, information extraction and retrieval and human-computer dialogue. Computational linguistics and NLP, which date back at least to the early 1950s, are among the key sub-fields of **Artificial Intelligence**.

**Digital Humanities and social sciences (DH)** is an interdisciplinary field that uses computer science as a source of techniques and technologies, in particular NLP, for exploring research questions in social sciences and humanities. **Computational Humanities** and computational social sciences aim at improving the state of the art in both computer sciences (e.g. NLP) and social sciences and humanities, by involving computer science as a research field.

One of the main challenges in computational linguistics is **to model and to cope with language** variation. Language varies with respect to domain and genre (news wires, scientific literature, poetry, oral transcripts...), sociolinguistic factors (age, background, education; variation attested for instance on social media), geographical factors (dialects) and other dimensions (disabilities, for instance). But language also constantly evolves at all-time scales. Addressing this variability is still an open issue for NLP. Commonly used approaches, which often rely on supervised and semi-supervised machine learning methods, require very large amounts of annotated data. They still suffer from the high level of variability found for instance in **user-generated content, non-contemporary texts**, as well as in **domain-specific documents** (e.g. financial, legal).

## SEMAGRAMME

### Semantic Analysis of Natural Language

Computational linguistics is a discipline at the intersection of computer science and linguistics. On the theoretical side, it aims to provide computational models of the human language faculty. On the applied side, it is concerned with natural language processing and its practical applications.

The research program of Sémagramme aims to develop models based on well-established mathematics. We seek two main advantages from this approach. On the one hand, by relying on mature theories, we have at our disposal sets of mathematical tools that we can use to study our models. On the other hand, developing various models on a common mathematical background will make them easier to integrate, and will ease the search for unifying principles.

The main mathematical domains on which we rely are formal language theory, symbolic logic, and type theory.

Formal language theory studies the purely syntactic and combinatorial aspects of languages, seen as sets of strings (or possibly trees or graphs). Formal language theory has been especially fruitful for the development of parsing algorithms for context-free languages. We use it, in a similar way, to develop parsing algorithms for formalisms that go beyond context-freeness. Language theory also appears to be very useful in formally studying the expressive power and the complexity of the models we develop.

Symbolic logic (and, more particularly, proof-theory) is concerned with the study of the expressive and deductive power of formal systems. In a rule-based approach to computational linguistics, the use of symbolic logic is ubiquitous. As we previously said, at the level of syntax, several kinds of grammars (generative, categorial...) may be seen as basic deductive systems. At the level of semantics,



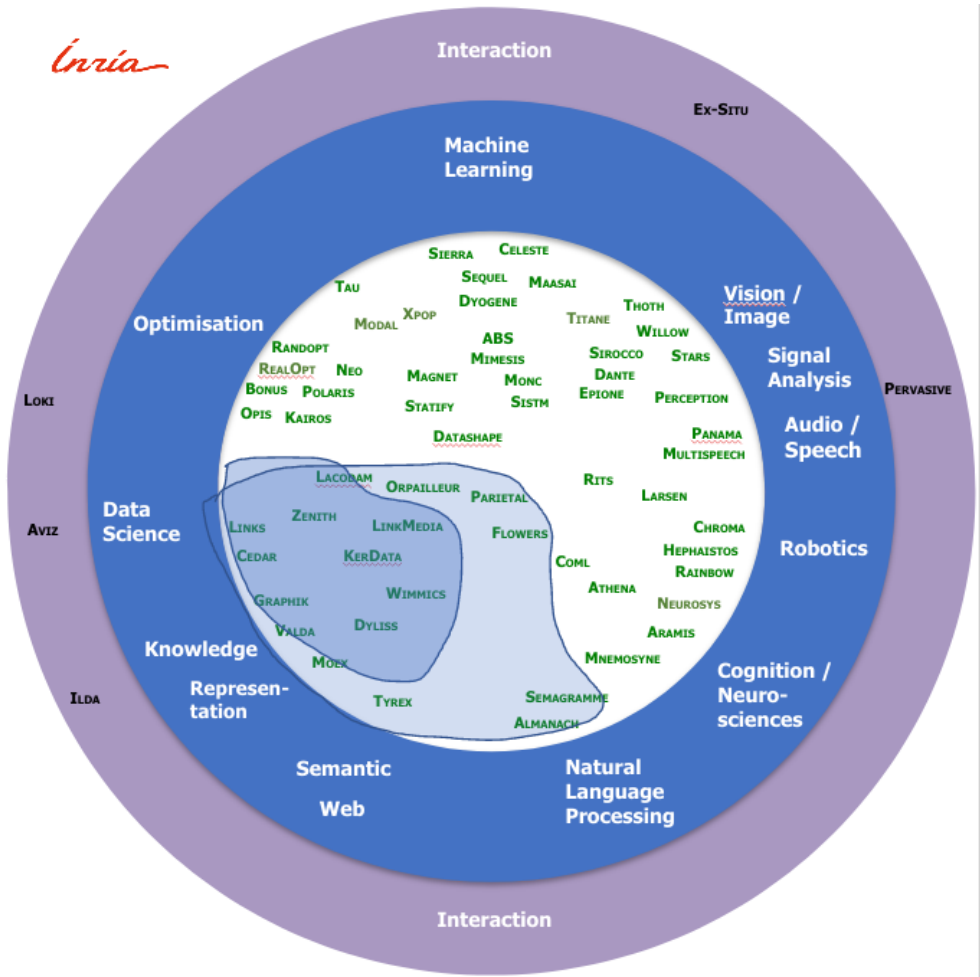
the meaning of an utterance is captured by computing (intermediate) semantic representations that are expressed as logical forms. Finally, using symbolic logics allows one to formalize notions of inference and entailment that are needed at the level of pragmatics.

Among the various possible logics that may be used, Church's simply typed  $\lambda$ -calculus and simple theory of types (a.k.a. higher-order logic) play a central part. On the one hand, Montague semantics is based on the simply typed  $\lambda$ -calculus, and so is our syntax-semantics interface model. On the other hand, as shown by Gallin, the target logic used by Montague for expressing meanings (i.e., his intensional logic) is essentially a variant of higher-order logic featuring three atomic types (the third atomic type standing for the set of possible worlds).

## 5.5 Knowledge-based systems and semantic web

From Tim Berners-Lee's initial definition, "*the Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation*". The semantic tower builds upon URIs and XML, through RDF schemas representing data triplets, up to ontologies allowing reasoning and logical processing.

Inria teams involved in Knowledge representation, reasoning and processing address the following challenges in different manners: (i) dealing with large volumes of information from heterogeneous distributed sources; (ii) building bridges between massive data stored in data bases using semantic technologies; (iii) developing semantically based applications on top of these technologies.



Dealing with large volumes of information from heterogeneous distributed sources

With the ubiquity of the Internet we are now faced with the opportunity and challenge of moving from local artificial intelligent systems to massively distributed artificial intelligences and societies. Designing and running reliable and efficient systems combining linked data from distant sources through workflows of distributed services remains an open problem. The data quality and their processes traceability, the precision of their extraction and capture, the correctness of their alignment and integration, the availability and quality of shared models (ontologies,

vocabularies) to represent, exchange and reason on them, etc. all these aspects need to be addressed on a large scale and continuously.

A second aspect is underlined by the Web which does not provide only a universal application framework for Internet but also a hybrid space where humans and software agents can interact on large scales and form mixed communities. Millions of users and artificial agents now interact daily in online applications resulting in very complex systems to be studied and designed. We need models and algorithms that generate justifications and explanations and accept feedbacks to support interactions with very different users. We need to consider complex systems including the users as an intelligent component that will interact with other components (e.g. artificial intelligence in interfaces, natural language interaction), participate to the process (e.g. human computing, crowdsourcing, social machines) and may be augmented by the system (intelligence amplification, cognitive augmentation, augmented intelligence, extended mind and distributed cognition)

### **WIMMICS**

#### **Web-Instrumented Man-Machine Interactions, Communities and Semantics**

The Web provide virtual spaces (e.g. Wikipedia) where persons and software interact in mixed communities exchanging and using formal knowledge (e.g. ontologies, knowledge bases) and informal content (e.g. texts, posts, tags).

The WIMMICS team studies models and methods to bridge formal semantics and social semantics on the web. It follows a multidisciplinary approach to analyse and model these spaces, their communities of users and their interactions. It also provides algorithms to compute these models from traces on the web including, knowledge extraction from text, semantic social network analysis, argumentation theory.

In order to formalise and reason on these models, the WIMMICS team then proposes languages and algorithms relying on and extending graph-based knowledge approaches for the semantic web and linked data on the web - e.g. graph models of the Resource Description Framework (RDF). Together, these contributions provide analysis tools and indicators, and support new functionalities and management tasks in epistemic communities.

The research objectives of Wimmics can be grouped according to four topics that we identify in reconciling social and formal semantics on the Web:

**Topic 1 - users modelling and designing interaction on the Web:** The general research question addressed by this objective is *"How do we improve our interactions with a semantic and social Web more and more complex and dense?"* Wimmics focuses on specific sub-questions: *"How can we capture and model the users' characteristics?"* *"How can we represent and reason with the users' profiles?"*

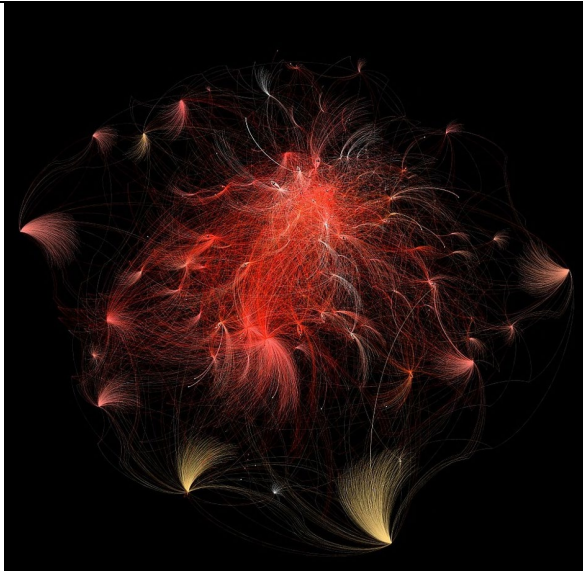
"How can we adapt the system behaviours as a result?" "How can we design new interaction means?" "How can we evaluate the quality of the interaction designed?"

**Topic 2 - communities and social interactions analysis on the Web:** The general question addressed in this second objective is "*How can we manage the collective activity on social media?*" Wimmics focuses on the following sub-questions: "How do we analyse the social interaction practices and the structures in which these practices take place?" "How do we capture the social interactions and structures?" "How can we formalize the models of these social constructs?" "How can we analyse and reason on these models of the social activity?"

**Topic 3 - vocabularies, semantic Web and linked data based knowledge representation and Artificial Intelligence formalisms on the Web:** The general question addressed in this third objective is "*What are the needed schemas and extensions of the semantic Web formalisms for our models?*" Wimmics focuses on several sub-questions: "What kinds of formalism are the best suited for the models of the previous section?" "What are the limitations and possible extensions of existing formalisms?" "What are the missing schemas, ontologies, vocabularies?" "What are the links and possible combinations between existing formalisms?" In a nutshell, an important part of this objective is to formalize as typed graphs the models identified in the previous objectives in order for software to exploit them in their processing (in the next objective).

**Topic 4 - artificial intelligence processing: learning, analysing and reasoning on heterogeneous semantic graphs on the Web:** The general research question addressed in this last objective is "*What are the algorithms required to analyse and reason on the heterogeneous graphs we obtained?*" Wimmics focuses on several sub-questions: "How do we analyse graphs of different types and their interactions?" "How do we support different graph life-cycles, calculations and characteristics in a coherent and understandable way?" "What kind of algorithms can support the different tasks of our users?"

These research results are integrated, evaluated and transferred through generic software (e.g. semantic web factory CORESE) and dedicated applications (e.g. CREEP for detecting cyberbullying). The ultimate goal of the team is to make the Web a place where to seamlessly link natural and artificial intelligence.



Data graph of the Discovery hub exploratory search engine

Indeed, the produced data and extracted knowledge is constantly changing, hence agents and processes consuming it must be able to adapt their own knowledge.

## **MOEX**

### **Evolving Knowledge**

MOEX studies the principles by which the knowledge of social agents evolves. These agents may be programs observing the (semantic) web, selecting and exchanging interesting information or social robots communicating with humans and other robots. Toi.Net seems to cover both cases. Agents are faced with changing environments (Sam not interested in Miss ceremonies any more, new knowledge about coronaviruses) and may have to interact with other agents (Sam, new friends of Sam or other robots).

The behaviour of such agents is governed by knowledge that may be represented in a variety of ways. In a changing situation, agents should not wait for a programmer to update their knowledge or many examples to be generated, and as many mistakes to be made. They must adapt their knowledge to behave adequately. Mechanisms for adapting knowledge respond to the external pressure, exerted by the environment and society in which agents evolve, and internal pressure to warrant knowledge coherence.

The ambition is to answer, in particular, the following questions:

- How do agent populations adapt their knowledge representation to their environment and to other populations?

- How must this knowledge evolve when the environment changes and new populations are encountered?
- How can agents preserve knowledge diversity and is this diversity beneficial?

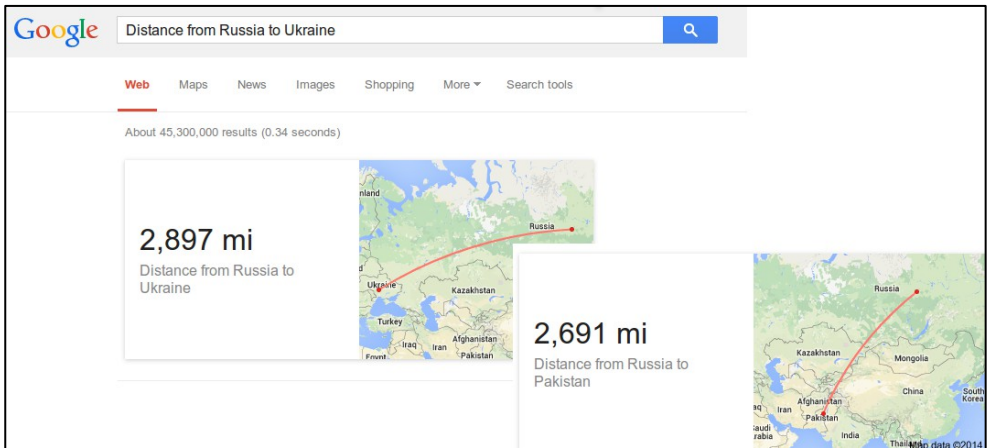
For that purpose, we combine knowledge representation and cultural evolution methods. The former provides formal models of knowledge; the latter provides a well-defined framework for studying situated evolution. We consider knowledge as a culture and study the global properties of local adaptation operators applied by populations of agents by jointly:

- experimentally testing the properties of adaptation operators in various situations using experimental cultural evolution, and
- theoretically determining such properties by modelling how operators shape knowledge representation.

We aim at acquiring a precise understanding of knowledge evolution through the consideration of a wide range of situations, representations and adaptation operators.

### Building bridges between massive data stored in data bases using semantic technologies

The semantic Web addresses the massive integration of very different data sources (e.g. sensors of smart cities, biological knowledge extracted from scientific articles, event descriptions on social networks) and using very different vocabularies (e.g. relational schemas, lightweight thesauri, formal ontologies) in very different reasoning (e.g. decision making by logical derivation, enrichment by induction, analysis through mining, etc.). On the Web, the initial graph of linked pages has been joined by a growing number of other graphs and is now mixed with sociograms capturing the social network structure, workflows specifying the decision paths to be followed, browsing logs capturing the trails of our navigation, service compositions specifying distributed processing, open data linking distant datasets, etc. Moreover, these graphs are not available in a single central repository but distributed over many different sources and some sub-graphs are public (e.g. dbpedia <http://dbpedia.org>) while others are private (e.g. corporate data). Some sub-graphs are small and local (e.g. a users' profile on a device), some are huge and hosted on clusters (e.g. Wikipedia), some are largely stable (e.g. thesaurus of Latin), some change several times per second (e.g. social network statuses), etc. Each type of network of the Web is not an isolated island, they interact with each other: the social networks influence the message flows, their subjects and types, the semantic links between terms interact with the links between sites and vice-versa, etc. There is a huge challenge not only in finding means to represent and analyse each kind of graphs, but also means to combine them and combine their processing.



From the paper "Why the Data Train Needs Semantic Rails" by Janowicz et al., AI Magazine, 2015.  
Without semantics, Russia appears closer to Pakistan than the Ukraine

## **CEDAR**

### **Rich Data Exploration at Cloud Scale**

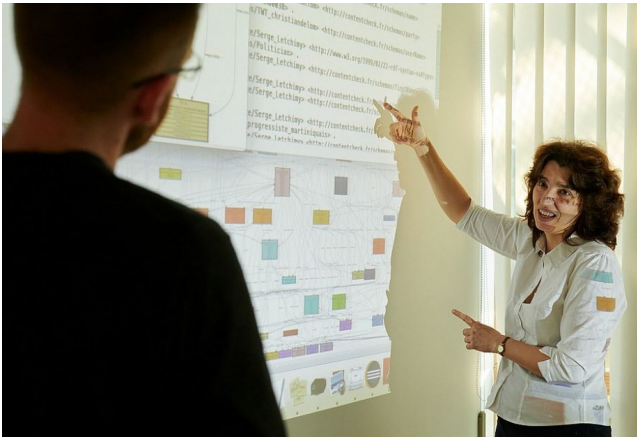
Making sense of "Big Data" requires interpreting it through the prism of knowledge about the data content, organization, and meaning. Moreover, domain knowledge is often the language closest to the users, be they specialized domain experts or novice end users of a data-intensive application. Expressive and scalable tools for OBDA (Ontology-Based Data Access) are thus a key factor in the success of Big Data applications.

Cedar works at the interface between knowledge representation formalisms (such as some description logics or classes of existential rules) and database engines. The team builds highly efficient OBDA tools with a particular focus on scaling up to very large databases; this can be seen as augmenting database engines with reasoning capabilities, and deploying them in a cloud setting for scale. Cedar also investigates novel ways of interacting with large, complex data and knowledge bases such as those referenced in the Linked Open Data cloud (<http://lod-cloud.net>). Semantics is also investigated as a means to integrate and make sense of heterogeneous, complex content, in repositories of rich, heterogeneous Web data, in particular applied to journalistic fact checking.

Optimisation and performance at scale: this topic is at the heart of Y. Diao's ERC project "Big and Fast Data", which aims at optimisation with performance guarantees for real-time data processing in the cloud. Machine learning techniques and multi-objectives optimisation are leveraged to build performance models for data analytics the cloud. The same goal is shared by our work on efficient evaluation of queries in dynamic knowledge bases.

Data discovery and exploration: today's Big Data is complex; understanding and exploiting it is difficult. To help users, we explore: compact summaries of knowledge bases to abstract their structure and help users formulate queries;

interactive exploration of large relational databases; techniques for automatically discovering interesting information in knowledge bases; and keyword search techniques over Big Data sources.



Data graph mining

## Graphik

### GRAPHS for Inferences and Knowledge representation

The main research domain of GraphIK is *Knowledge Representation and Reasoning* (KR), which studies paradigms and formalisms for representing knowledge and reasoning on these representations. A large part of our work is strongly related to *data management* and *database theory*.

We develop logical languages, which mainly correspond to fragments of first-order logic. However, we also use graphs and hypergraphs (in the graph-theoretic sense) as basic objects. Indeed, we view labelled graphs as an *abstract representation* of knowledge that can be expressed in many KR languages: different kinds of conceptual graphs —historically our main focus—, the Semantic Web language RDFS, expressive rules equivalent to so-called tuple-generating-dependencies in databases, some description logics dedicated to query answering, etc. For these languages, reasoning can be based on the structure of objects (thus on graph-theoretic notions) while being sound and complete with respect to entailment in the associated logical fragments. An important issue is to study *trade-offs* between



the expressivity and computational tractability of (sound and complete) reasoning in these languages.

GraphIK focuses on some of the main challenges in KR:

- ontological query answering: querying large, complex or heterogeneous datasets, provided with an ontological layer;
- reasoning with rule-based languages;
- reasoning in presence of inconsistency and
- decision making.

An important feature of knowledge-based techniques is their explanatory power, i.e., their potential ability to explain drawn conclusions. Being able to explain, justify or argue is a mandatory requirement in many AI applications in which the users need to understand the results of the system, in order to trust and control it. Moreover, it becomes a crucial concern with respect to ethical issues as soon as the automated decisions may impact human beings.

## **LINKS**

### **Linking Dynamic Data**

The appearance of linked data on the web calls for novel database management technologies for linked data collections. The classical challenges from database research need to be now raised for linked data: how to define exact logical queries, how to manage dynamic updates, and how to automatize the search for appropriate queries. In contrast to mainstream linked open data, the LINKS project focuses on linked data collections in various formats, under the assumption that the data is correct in most dimensions. The challenges remain difficult due to incomplete data, uninformative or heterogeneous schemas, and the remaining data errors and ambiguities. We develop algorithms for evaluating and optimizing logical queries on linked data collections, incremental algorithms that can monitor streams of linked data and manage dynamical updates of linked data collections, and symbolic learning algorithms that can infer appropriate queries for linked data collections from examples.

Research themes

We develop algorithms for answering logical querying on heterogeneous linked data collections in hybrid formats, distributed programming languages for managing dynamic linked data collections and workflows based on queries and mappings, and symbolic machine learning algorithms that can link datasets by inferring appropriate queries and mappings. Our main objectives are structured as follows:

- **Querying heterogeneous linked data.** We develop new kinds of schema mappings for semi-structured datasets in hybrid formats including graph databases, RDF collections, and relational databases. These induce recursive queries on linked data collections for which we investigate evaluation algorithms, static analysis problems, and concrete applications.

- **Managing dynamic linked data.** In order to manage dynamic linked data collections and workflows, we develop distributed data-centric programming languages with streams and parallelism, based on novel algorithms for incremental query answering, we study the propagation of updates of dynamic data through schema mappings, and investigate static analysis methods for linked data workflows.
- **Linking graphs.** Finally, we develop symbolic machine learning algorithms, for inferring queries and mappings between linked data collections in various graphs formats from annotated examples.

### Developing applications on top of these technologies

All teams mentioned in this section develop knowledge-based applications. The last team presented, DYLISS, is fully dedicated to bioinformatics. Increasingly powerful technologies (e.g. sequence analysis) have accelerated the progress towards a complete map of biological process at molecular and cellular levels. The knowledge represented in these biological models must be shared (between software tools and between software and users) in ways that preserve the semantics of the knowledge. Standardization of knowledge, particularly on biological regulations that are very complex to unify (format BioPAX), and using the numerous knowledge bases available (Reactome, Rhea, pathwaysCommons...) will ensure reliable semantic interoperability.

#### DYLISS

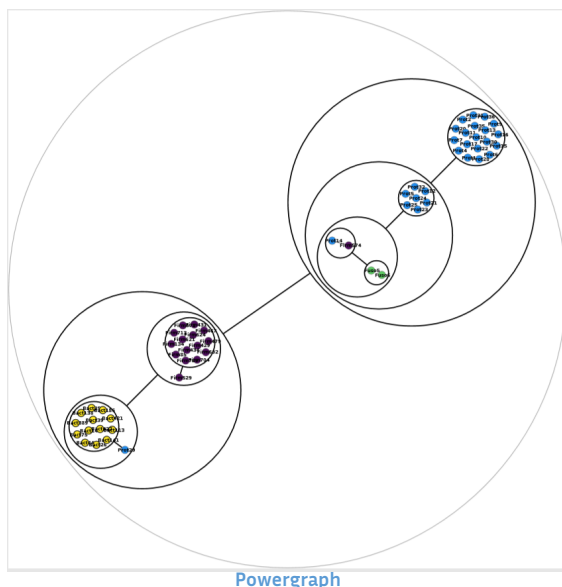
##### **Dynamics, Logics and Inference for biological Systems and Sequences**

Experimental sciences are undergoing a data revolution due to the multiplication of sensors that allow for measuring the evolution of thousands of interdependent physical or biological components over time. When measurements are precise and various enough, they can be integrated in a machine learning framework to highlight the top-ranking entities within the considered datasets. However, **the biological interest lies in the explanation of the ranking, more precisely in identifying the biological processes leading the specificity of the selected entities** with respect to the considered phenotype. This requires to take into account the existing domain knowledge about the chains of biological compounds involved in the data sources, together with their regulators.

This raises several issues: first, we need to **integrate the various project-specific data sources**, both together as well as with the reference domain data and knowledge bases. Second, we need to **extract explanation-supporting models** for the role of the entities of interest, which have to be consistent with domain knowledge.

Importantly, even if we can acquire unprecedented amounts of data, they are still no match for the biological complexity. This results in large numbers of models (even only considering the minimal ones) all equally compatible with the observations and the domain knowledge. Avoiding the bias of greedy approaches

and streetlight effect raises a third issue: **consider the exhaustive family of consistent models** and assists domain experts for exploring and analysing them. To address these issues, Dyliss develops knowledge-based data-analysis and reasoning methods. A first axis is to develop data-structuration and integration methods to unify data sources and knowledge corpora into knowledge-graphs. This is supported by the Semantic Web technologies and the resources from the Linked Open Data initiative (more than 1,600 knowledge repositories for life sciences). A second axis is to take advantage of structured data to extract families of models that explicitly explain the role of the molecules: this is achieved with a combination of learning methods from examples, query-based approaches and logical programming methods involving dynamical systems constraints viewed as optimisation rules. In the third axis, these methods also assist domain experts for exploring and analysing exhaustively the family of models.



Other project-teams in this domain: TYREX, Grenoble; VALDA, Paris; ZENITH, Montpellier.

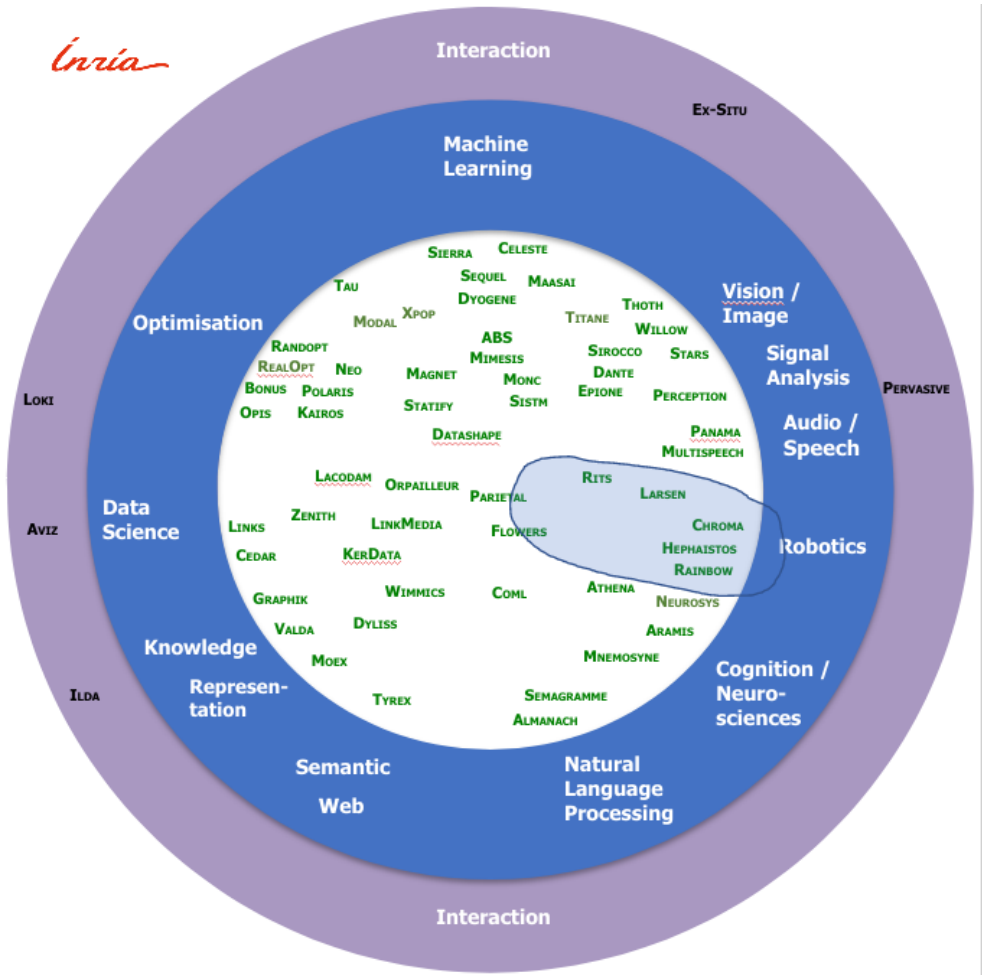
## 5.6 Robotics and autonomous vehicles

Robotics combine many sciences and technologies, from the “lower level” mechanics, mechatronics, electronics, control, to the “upper level” of perception, cognition, collaboration and reasoning; in this section, even though artificial intelligence in

robotics might imply to dig into the lower level functions for some processing features, we only deal with the upper levels, those who directly relate to the field of AI.

Recent progress made by robotics is impressive. Humanoid robots can walk, run, move in known and unknown environments, perform simple tasks like grasping objects or manipulating devices; bio-inspired robots are able to mimic behaviours of a wealth of quite diverse living creatures (insects, birds, reptiles, rodents ...) and use these behaviours for efficiently solving complex problems. Boston Dynamics' Atlas (<http://www.bostondynamics.com/robot-Atlas.html>) biped robot, using simple perception and efficient control mechanisms, can move efficiently in outdoor rough terrain and carry heavy objects, following the same company's four-legged robot BigDog.

On the cognitive side, thanks to the progresses in speech processing, vision and scene understanding from many sensors, and thanks to the reasoning capacities implemented, robots can play music, welcome visitors in shopping malls, converse with children. With coordination features among a fleet of robots, they are able to play football together – but no robot team is yet able to beat a team of low-skilled humans. Autonomous vehicles are able to behave safely over long periods of time, and some countries and US states might allow them to drive on public roads in the near future, even though a lot of open questions – including ethical ones – remain.



The challenges addressed by Inria teams developing research on robots and self-driving vehicles are: (i) situation understanding from multisensory input; (ii) reasoning under uncertainty, resilience; (iii) combining several approaches for decision-making. For a deeper analysis of autonomous and connected vehicles, refer to Inria's white paper<sup>28</sup> (in French), which states that fully autonomous cars will not be of general use before 2040.

<sup>28</sup> <https://www.inria.fr/sites/default/files/2019-10/inrialivrebblancvac-180529073843.pdf>

## Situation understanding from multisensory input

For a robot to move in unknown areas, for a self-driving car in traffic, for a personal assistance robot such as Toi.Net (see section 1), it is essential to perceive the environment and to characterise the situation. This is done using input from multiple sensors (vision, laser, sound, internet, ... , road2car data in the case of vehicles). Situations can be simple symbols, ontologies, or more sophisticated representations of actors and objects present in an environment. A good characterisation of the situation can help the robot to make decisions - even in some case to infringe the law or a regulation for saving the car's passengers lives.

## Reasoning under uncertainty, resilience

Robots are active in the physical world and have to cope with defaults of many sorts: network shutdowns, defective sensors, electronic hazards, etc. Some sensors provide incomplete information or have error margins generating uncertainty on the data. However, an autonomous mobile robot must perform its operation continuously without any human intervention, and for long periods of time. A challenge for robot architectures and software is to deal with uncertain or missing information, and with information only available at separate acquisition times. Anytime algorithms that provide an output on demand can be a solution in the case of fast decision-making needed even though the decision is not perfect.

## Combining several approaches for decision-making

A variety of data and information can be available for a robot to make a decision. Data from different sensors, information about the environment in the form of a situation assessment, memories of past decisions made, rules and regulations implemented in the robot's memory: there is a need to combine these facts and data and to conduct hybrid reasoning from numeric data, continuous or discrete, and from semantic representations. Moreover, as seen above, this reasoning must also consider uncertainty: the research on decision-making for robots has to address this challenge. One possible solution is unsupervised machine learning and reinforcement learning of situations and semantic interpretations.

## Human-Robot collaboration

In most real-life situations, such as assistance to the elderly, autonomous driving, operation in factories, robot must properly interact with human users and operators. This interaction is needed both ways: obviously, for robots to understand the goals and actions of humans (see for example Stuart Russell's book on the subject<sup>29</sup>), but also for humans to understand the goals and actions undertaken by robots in their presence. A good example of the latter is given in a report on safety for automated driving published by a consortium of stakeholders including major German manufacturers<sup>30</sup>, which states that: "HMI should be carefully designed to consider the

---

<sup>29</sup> Stuart Russell. *Human compatible, AI and the problem of control*. Penguin books, 2019.

<sup>30</sup> Safety first for Automated Driving, Aptiv, BMW, Baidu Continental, Daimler et al, July 2019

psychological and cognitive traits and states of human beings with the goal of optimizing the human's understanding of the task and situation and of reducing accidental misuse or incorrect operations".

## HEPHAISTOS

### HEXapode, PHysiology, AssISTance and RobOTics

The goal of the project HEPHAISTOS is to set up a generic methodology for the design and evaluation of an adaptable and interactive assistive ecosystem for the elderly and the vulnerable persons that provides furthermore assistance to the helpers, on-demand medical data and may manage emergency situations. More precisely our goals are to develop devices with the following properties:

- they can be adapted to the end-user and to its everyday environment
- they should be affordable and minimally intrusive
- they may be controlled through a large variety of simple interfaces
- they may eventually be used to monitor the health status of the end-user in order to detect emerging pathology

Assistance will be provided through a network of communicating devices that may be either specifically designed for this task or be just adaptation/instrumentation of daily life objects.

The targeted population is limited to people with mobility impairments (for the sake of simplicity this population will be denoted by elderly in the remaining of this document although our work deal also with a variety of people (e.g. handicapped or injured people, ...)) and the assistive devices will have to support the individual autonomy (at home and outdoor) by providing complementary resources in relation with the existing capacities of the person. Personalization and adaptability are key factor of success and acceptance. Our long-term goal will be to provide robotized devices for assistance, including smart objects, which may help disabled, elderly and handicapped people in their personal life.

Assistance is a very large field and a single project-team cannot address all the related issues. Hence HEPHAISTOS will focus on the following main societal challenges:

- **mobility:** previous interviews and observations in the HEPHAISTOS team have shown that this was a major concern for all the players in the ecosystem. Mobility is a key factor to improve personal autonomy and reinforce privacy, perceived autonomy and self-esteem

- **managing** emergency situations: emergency situations (e.g. fall) may have dramatic consequences for elderly. Assistive devices should ideally be able to prevent such situation and at least should detect them with the purposes of sending an alarm and to minimize the effects on the health of the elderly
- **medical monitoring:** elderly may have a fast changing trajectory of life and the medical community is lacking timely synthetic information on this evolution, while available technologies enable to get raw information in a non intrusive and low cost manner. We intend to provide synthetic health indicators that take measurement uncertainties into account, obtained through a network of assistive devices. However respect of the privacy of life, protection of the elderly and ethical considerations impose to ensure the confidentiality of the data and a strict control of such a service by the medical community.
- **rehabilitation and biomechanics:** our goals in rehabilitation are 1) to provide more objective and robust indicators, that take measurement uncertainties into account to assess the progress of a rehabilitation process 2) to provide processes and devices (including the use of virtual reality) that facilitate a rehabilitation process and are more flexible and easier to use both for users and doctors. Biomechanics is an essential tool to evaluate the pertinence of these indicators, to gain access to physiological parameters that are difficult to measure directly and to prepare efficiently real-life experiments



MARIONET-ASSIST, cable parallel robot for the assistance of persons with reduced mobility

LARSEN

Lifelong Autonomy and interaction skills for Robots in a Sensing ENvironment



The Larsen team aims to combine recent advances in artificial intelligence, machine learning and decision making with those of robotics to design robots that are smarter, more flexible and capable of cooperating with humans. The goal is to move beyond traditional robotics, which is limited to repetitive tasks in highly controlled environments where humans have little place in.

To achieve this goal, the team is developing methods to endow robots with long-term autonomy skills, allowing them to operate 24/7, and with skills that allow them to interact naturally with humans while taking into account the embedded and external sensors in the environment.

The team benefits from a rich testing infrastructure: an apartment equipped with sensors, a robotic arena with motion capture, a flight arena for drones with motion capture, and many robots: iCub and Talos humanoid robots, a quadruped, two hexapods, two mobile manipulators, two industrial manipulators, etc.

Larsen aims at designing robots having the ability to:

- handle dynamic environment and unforeseen situations;
- cope with physical damage;
- interact physically and socially with humans;
- collaborate with each other;
- exploit the multitude of sensors measurements from their surrounding;
- enhance their acceptability and usability by end-users without robotics background.

All these abilities can be summarized by the following two objectives:

- life-long autonomy: continuously perform tasks while adapting to sudden or gradual changes in both the environment and the morphology of the robot;
- natural interaction with robotics systems: interact with both other robots and humans for long periods of time, considering that people and robots learn from each other when they live together.



Creativ'Lab robotic arm

## RAINBOW

### Sensor-based Robotics and Human Interaction

The long-term vision of the Rainbow team is to develop the next generation of sensor-based robots able to navigate and/or interact in complex unstructured environments together with human users. Clearly, the word “together” can have very different meanings depending on the particular context: for example, it can refer to mere co-existence (robots and humans share some space while performing independent tasks), human-awareness (the robots need to be aware of the human state and intentions for properly adjusting their actions), or actual cooperation (robots and humans perform some shared task and need to coordinate their actions).

One could perhaps argue that these two goals are somehow in conflict since higher robot autonomy should imply lower (or absence of) human intervention. However, we believe that our general research direction is well motivated since:

- despite the many advancements in robot autonomy, complex and high-level cognitive-based decisions are still out of reach. In most applications involving tasks in unstructured environments, uncertainty, and interaction with the physical world, human assistance is still necessary, and will most probably be for the next decades. On the other hand, robots are extremely capable at autonomously executing specific and repetitive tasks, with great speed and precision, and at operating in dangerous/remote environments, while humans possess unmatched cognitive capabilities and world awareness which allow them

to take complex and quick decisions;

- the cooperation between humans and robots is often an implicit constraint of the robotic task itself. Consider for instance the case of assistive robots supporting injured patients during their physical recovery, or human augmentation devices. It is then important to study proper ways of implementing this cooperation;
- finally, safety regulations can require the presence at all times of a person in charge of supervising and, if necessary, take direct control of the robotic workers. For example, this is a common requirement in all applications involving tasks in public spaces, like autonomous vehicles in crowded spaces, or even UAVs when flying in civil airspace such as over urban or populated areas.

Within this general picture, the Rainbow activities will be particularly focused on the case of **(shared) cooperation between robots and humans** by pursuing the following vision: on the one hand, empower robots with a large degree of autonomy for allowing them to effectively operate in non-trivial environments (e.g., outside completely defined factory settings). On the other hand, include human users in the loop for having them in (partial and bilateral) control of some aspects of the overall robot behaviour. We plan to address these challenges from the **methodological, algorithmic** and **application-oriented** perspectives. The main research axes along which the Rainbow activities will be articulated are: three supporting axes (**Optimal and Uncertainty-Aware Sensing; Advanced Sensor-based Control; Haptics for Robotics Applications**) that are meant to develop methods, algorithms and technologies for realizing the central theme of **Shared Control of Complex Robotic Systems**.



Moving an Intelligent Wheelchair in Virtual Reality

### Autonomous vehicles

The first fundamental problems in the use of AI in the Autonomous Vehicles (AV) field are those of explainability and consistency of the algorithm's outputs. These are the prerequisite for any development of legal frameworks necessary to the large testing and the deployments of AV's in real road networks and cities. On the technical level, the first challenges are computational costs as well as energy consumption if dedicated AI architectures (cards and others) are widely deployed.

Other algorithmic challenges are related to the need of large annotated multi-sensors and multi-scenario datasets. In the last years, the global effort to publish reproducible research led to an increasing number of open source codes and public datasets -- paving the way to exciting results. KITTI in 2012 was the first large-scale dataset for autonomous driving with vision and since then public datasets such as

ScanNet (2018) for 3D processing, nuScenes (2019) for multi-sensors driving, SemanticKITTI (2019) for 3D driving scenes, and many others all together allowed a great performance leap. In fact, many researches displayed the benefit from pre-training deep networks on these large public datasets for a large variety of tasks, demonstrating that high-level features can be shared even for tasks of different nature.

Still, the current research line suffers from following this supervised paradigm that requires large datasets (in order of thousands/millions of data) which annotation is both tedious and menial. While supervised learning undoubtedly brings the best performance, the labelling cost will eventually become unbearable since both the dataset size and the number of sensors constantly increase. Not to mention that encompassing all conditions (lightings, traffic scenarios, weathers, etc.) in a single dataset is impractical. For example, not a single available dataset encompasses dangerous driving scenarios. Leveraging semi or unsupervised learning is necessary to ensure scalability of the algorithms to the real outside world, where they ultimately face situations unseen in the training set. The holy Grail of artificial general intelligence is far from our current knowledge but promising techniques in transfer learning allow expanding training done in supervised fashion to new unlabelled datasets, for example with domain adaptation. Exciting experiments in RITS team and other research labs demonstrated the ability to apply such strategy for example to transfer learning to changing lighting conditions (training on day data and testing on night data), weathers (clear to rain driving), or even nature of data (simulator to real driving).

Today, ML is extensively used in the AV field for the perception systems. However, other AI techniques seem as promising as ML, in addition of being easier to interpret. AI certainly paves the way to new research areas and demonstrates great ability to solve long-standing problems crucial for autonomous driving (e.g. semantic labelling of complex outdoor environments).

## **RITS**

### **Robotics & Intelligent Transportation Systems**

The project-team RITS is a multidisciplinary project at Inria, working on Robotics for Intelligent Transportation Systems. It seeks in particular to combine artificial intelligence and mathematical modelling to design advanced intelligent robotics systems for autonomous and sustainable mobility.

Among the scientific topics covered:

- Cross-modal techniques for scene understanding from camera, laser data, GPS, etc...,

- Unsupervised or weakly supervised training (domain adaptation, distillation),
- Low and high level vehicle control,
- Decision making for autonomous driving,
- Large-scale traffic modelling and simulation,
- Control and optimisation of road transport systems,
- Development and deployment of automated vehicles (cyber cars, private vehicles,...).

The goal of these studies is to **improve road transportation** in terms of safety, efficiency, and comfort and also to minimize nuisances. The technical approach is based on driver's assistance, going all the way to full driving automation. The project-team provides to the different partner teams some important means such as a fleet of a dozen computer driven vehicles, various sensors and advanced computing facilities including a simulation tool. An experimental system based on fully automated vehicles has been installed on the Inria grounds at Rocquencourt for demonstration purposes.



One of the autonomous driving platforms of RITS

## CHROMA

### Cooperative and Human-aware Robot Navigation in Dynamic Environments

The overall objective of Chroma is to address fundamental and open issues that lie at the intersection of the emerging research fields called “Human Centred Robotics” [1]. More precisely, the goal is to design algorithms and develop models allowing mobile robots to navigate and cooperate in dynamic and human-populated environments. Chroma is involved in all decision aspects pertaining to single and multi-robot navigation tasks, including perception and motion-planning.

The general objective is to build robotic behaviours that allow one or several robots to operate safely among humans in partially known environments, where time, dynamics and interactions play a significant role. Recent advances in embedded computational power, sensor and communication technologies, and miniaturized mechatronic systems, make the required technological breakthroughs possible (including from the scalability point of view).

Chroma is clearly positioned in the “Artificial Intelligence and Autonomous systems” research theme of the Inria 2018-2022 Strategic Plan. More specifically we refer to the “Augmented Intelligence” challenge (connected autonomous vehicles) and to the “Human centred digital world” challenge (interactive adaptation).

[1] Montreuil, V.; Clodic, A.; Ransan, M.; Alami, R., "Planning human centred robot activities," in Systems, Man and Cybernetics, 2007



Mini-UAV Crazyflies 2.0, controlled by ultra wide band (UWB)

## 5.7 Neurosciences and cognition

AI and cognition have a long collaboration history. AI paradigms most often rely on concepts taken from research in cognition, and can in turn contribute to progresses in cognition science e.g. experiencing with large neural networks can be a tool for neuroscientists to check new models of the brain. The intersection between AI, neurosciences and cognition motivated some of the largest research projects undertaken by mankind, such as the Human Brain Project Flagship funded by the European Commission, or the BRAIN Initiative of the NIH in the USA.

An emerging trend in AI is to follow Nobel laureate Daniel Kahneman's proposal to model human thinking as the continuous interaction of two systems, namely System 1 and System 2.

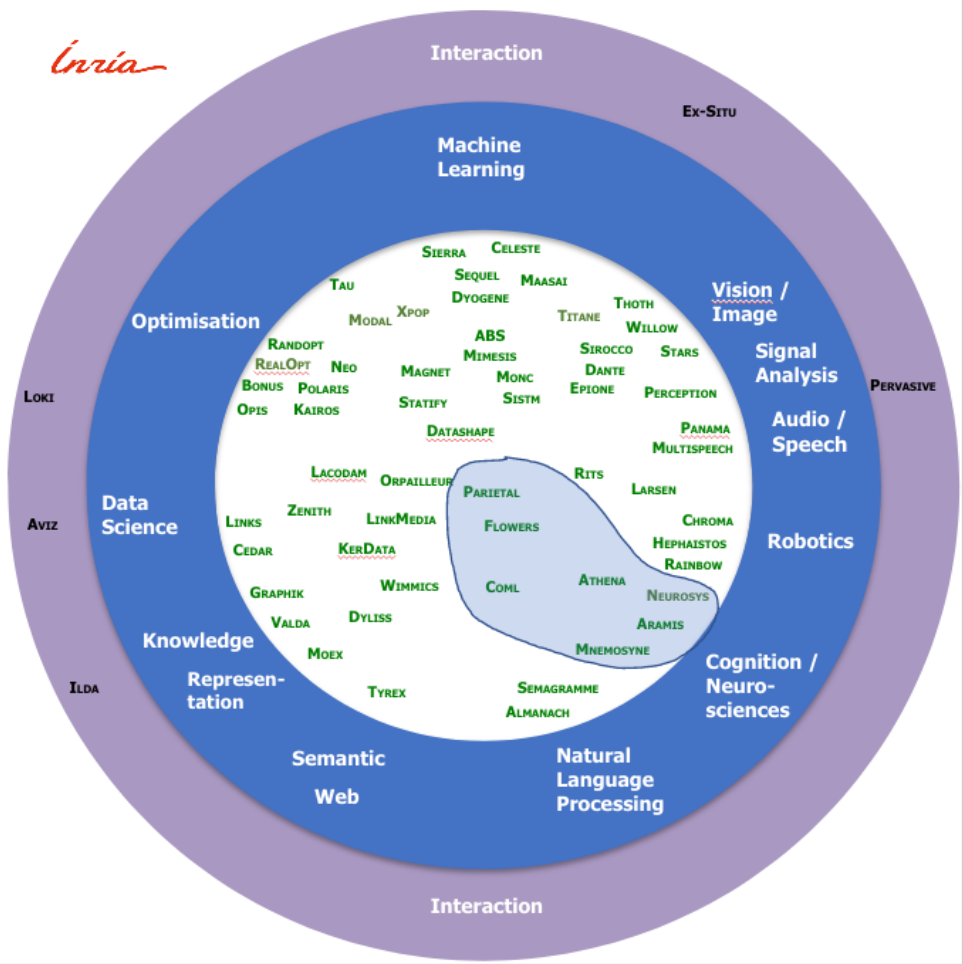
*From Kahneman's book, Thinking, Fast and Slow:*

*System 1 thinking is FAST, AUTOMATIC, happens UNCONSCIOUSLY and requires MINIMAL EFFORT*

*System 2 thinking is SLOWER, requires EFFORT, and happens CONSCIOUSLY and DELIBERATELY*

Most ML systems using neural networks can be allocated to System 1 e.g. in the case of vision, speech recognition, autonomous driving etc. The question on how to develop System 2 capacities is subject of debate: some authors believe that these capacities can be obtained using more sophisticated models of the brain i.e. more complex neural networks; others are convinced that complementary AI approaches such as semantic and knowledge-based reasoning will be useful for this purpose. Mid-2020, this debate is in its infancy, more research and experimentation is needed and this will take years if not decades.





Within Inria, a few research teams are at the intersection of AI and neurosciences. Their work can be qualified as contributions to both System 1 and System 2 thinking, even if some of them might be more closely related to one of them.

Their main scientific challenges are the following:

**Build better models of the brain**

This challenge is shared by all teams in this domain, as it is the most fundamental problem in neurosciences and cognition. It can concern the healthy brain as well as brain diseases. For this purpose, various modelling paradigms are exploited and

matched with diverse data including MRI, EEG and MEG. Models are developed for individual cells, clusters of cells, connectivity structures as well as activity patterns stored in dictionaries.

### Towards Common sense

Common sense reasoning is an overarching motivation for AI. It remains a distant goal for all approaches even after major investments and years of research such as Doug Lenat's CYC<sup>31</sup> project in the 1990s. Research in neurosciences and cognition can ultimately contribute new understandings of common sense human reasoning, but our not-so-recent history invites some modesty on the matter.

### Access to higher order executive functions/autonomy

Higher executive functions (temporal organization of behaviour, ability to generalize, manipulation of implicit and explicit knowledge, etc.) as well as real autonomy (continuous learning, flexibility, learning with one or few examples) remain major challenges that we are only beginning to address.

## ARAMIS

### Algorithms, models and methods for images and signals of the human brain

Multiple characteristics of brain diseases can now be measured in living patients thanks to the tremendous progress of neuroimaging, genomic and biomarker technologies. Collection of multimodal data in large patient databases provide a comprehensive view of brain alterations, biological processes, genetic risk factors and symptoms. A major challenge is now **to build numerical models of brain diseases from multimodal patient data** based on the development of specific data-driven approaches. Such models shall help to deepen our understanding of neurological diseases and to design effective systems to assist in clinical decisions.

The aim of the Inria ARAMIS project team is to design new machine learning and data analysis approaches for modelling brain diseases and decision support systems to assist clinicians. To this end, we develop approaches that can integrate multiple types of data acquired in the living patient including neuroimaging, peripheral biomarkers, clinical and omics data. A first line of research is devoted to the detection of alterations in brain imaging data and the design of AI systems to assist radiologists [2]. A second thread concerns the analysis of temporal phenomena from longitudinal data. This involves the development of sophisticated mixed effects models using tools from Riemannian geometry [3]. Such models can reconstruct scenarios of disease progression at the individual and population levels. They are implemented in the freely available software tools

---

<sup>31</sup> <https://en.wikipedia.org/wiki/Cyc>

Leaspy<sup>1</sup> and Deformetrica<sup>2</sup>. A third axis aims to model the functional interactions between distant brain areas that underlie cognitive processes. This is based on approaches that can model the organization of complex brain networks [1]. They are applied to the design of new devices, brain-computer interfaces and neurofeedback, for the rehabilitation of neurological patients. The team devotes many efforts to the transfer of these tools to clinical studies, through the development of the Clinica software platform<sup>3</sup>. Finally, we also provide guidelines and frameworks for reproducible research in the field. Three team members (N. Burgos, O. Colliot, S. Durrleman) are chairs in the PRAIRIE 3IA Institute.

[1] De Vico Fallani F, Richiardi J, Chavez M, and Achard S, Graph analysis of functional brain networks: practical issues in translational neuroscience., *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 369:1653, 2014.

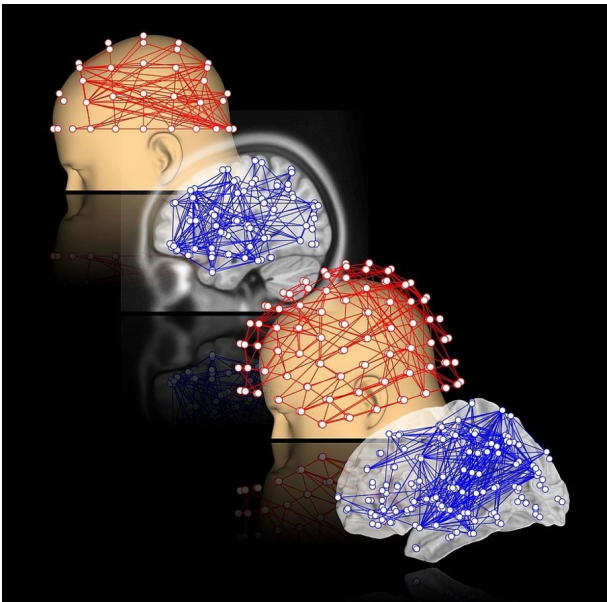
[2] Samper-González J, Burgos N, Bottani S, Fontanella S, Lu P, Marcoux A, Routier A, Guillon J, Bacci M, Wen J, Bertrand A, Bertin H, Habert M-O, Durrleman S, Evgeniou T, and Colliot O, Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data., *NeuroImage*, 183, 504–521, 2018

[3] Schiratti J-B, Allasonnière S, Colliot O, and Durrleman S, A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations., *Journal of Machine Learning Research*, 18:133, 1–33, 2017

<sup>1</sup> <https://gitlab.com/icm-institute/aramislab/leaspy>

<sup>2</sup> <https://www.deformetrica.org/>

<sup>3</sup> <http://www.clinica.run>



## **ATHENA**

### **Computational Imaging of the Central Nervous System**

Although exceptional progress has been obtained for exploring the human brain during the past decades, it is still terra-incognita and calls for specific research efforts to better understand its architecture and functioning.

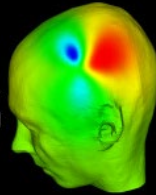
The ATHENA project-team has the overall objective to better understand the human brain structure and function by developing a new generation of computational models and methodological breakthroughs for brain connectivity mapping. To solve the limited view of the brain provided just by one imaging modality, and recover the brain structural and functional connectivity, the models built by the team are solidly grounded on advanced and complementary integrated non invasive and in-vivo imaging modalities: diffusion Magnetic Resonance Imaging (dMRI) and Electro & Magneto-Encephalography (EEG & MEG).

The main research directions of the team are :

1. Develop rigorous mathematical and computational tools for the acquisition, processing and combined analysis of Diffusion MRI and MEG & EEG data.
2. Push forward the state-of-the-art in Computational Brain Connectivity Mapping and Brain Computer Interfaces (BCI).
3. Develop and address, with our collaborators, clinical and BCI applications.

This will greatly help to better understand and reconstruct the structural and functional brain connectivity and to provide a clinical added value to better identify and characterize abnormalities in brain connectivity. While BCI is advocated as a means to communicate and help restore mobility or autonomy for very severe cases of disabled patients, it is also a new tool for interactively probing and training the human brain.

One third of the burden of all the diseases in Europe is due to problems caused by diseases affecting the brain. The objectives of ATHENA represent a fantastic scientific challenge as well as a pressing clinical need that, when solved, will positively impact the unacceptable burden of brain diseases and open new perspectives in neuroscience.



Brain mapping

## MNEMOSYNE

### Mnemonic Synergy

At the frontier between Artificial Intelligence and Computational Neuroscience, the MNEMOSYNE team proposes to model the main forms of memory and learning in the brain and to study how they are organized and implement complex cognitive functions. In neuroscience, a major dichotomy is reported between explicit (e.g. semantic, episodic) and implicit (e.g. procedural, habitual) memories and learning. Key mechanisms to understand such cognitive functions as reasoning, decision-making, attentional processes and language rely on competition, cooperation and transfer between these different ways to learn and memorize information: they are presently the topic of major progresses in different fields of neuroscience.

The MNEMOSYNE team designs models of the underlying neuronal structures and circuits under this functional view of brain organization and dynamics. Models are based on different kinds of neural architectures (feedforward, recurrent, convolutional, generative) with the challenge of mimicking the loops between the prefrontal cortex and the basal ganglia, and their interactions with the sensory cortex, hippocampus, amygdala and other cerebral structures, reported to be the substratum for the targeted cognitive functions. These models are the bases for collaborations of the team with the neuroscience and medical communities; they are also the ground for its original positioning in Machine Learning, towards Artificial General Intelligence. The team considers it a major challenge to propose computational models, embodied into virtual or real agents interacting on-line with the environment and able to autonomously extract structures to build a distributed model of the world, flexibly select the best strategy to reach internal and external goals and learn from their errors.

Recent topics of investigation concern language acquisition and the extraction of

syntax, goal encoding in motivated behaviour, transfer from goal-directed to habitual behaviour, planning and reasoning with a working memory and retrospective and prospective deliberation. These models are built in tight interaction with neuroscientists, in association with experimental protocols; they are exploited to consider pathological cases in the medical domain. They are also transferred to the socio-economic world with industrial applications and their impact in social science and humanities is also actively investigated, particularly in joint projects with educational science, linguistics, economics and philosophy.

## PARIETAL

### **Modelling brain structure, function and variability based on high-field MRI data.**

Artificial intelligence is a multi-faceted field, and the study of the brain through brain imaging offers an almost unique opportunity to explore these different facets. The Parietal team, member of the largest French brain-imaging platform, Neurospin, explores the links between brain, imaging, and cognition.

First, data acquired on the brain is provided as **signals** (electrophysiology recordings) or **images**, such as those acquired in Magnetic Resonance Imaging. Correctly exploiting these data involves large-scale estimation and statistical problems, which are nowadays solved by optimisation and statistical learning methods (machine learning), one of the areas of AI. For example, reconstructing brain electrical activity from measurements of electromagnetic fields taken at the scalp surface requires the solution of an **ill-posed inverse problem**, for which large-scale regression tools offer optimal solutions. The Parietal team has developed particularly efficient models and algorithms for parsimonious regression. Similarly, reconstructing a MRI image of the brain from a limited number of measurements to reduce the acquisition time amounts to solving a formally similar inverse problem. For these two problems, Parietal's researchers develop methods based on deep learning, leading to faster solvers for large-scale analysis.

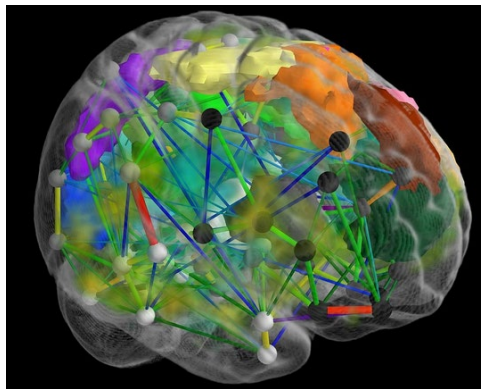
On the other hand, it is sometimes necessary to extract patterns present in the brain activity data to build much simpler models of the data based on these patterns. The Parietal team has developed **dictionary learning techniques** and, by working on the structure of the estimators, they have developed very efficient algorithms that can analyse millions of images of the brain in a reasonable amount of time. The same method also allows extraction of patterns from time series.

On other methodological aspects, work on the analysis of **statistical guarantees** is ongoing: when one asserts that the activity of a brain region predicts a person's behaviour, how to guarantee that this is the case, and that this is not an erroneous interpretation? It is difficult to prove that a given region plays a role in the

prediction when many other areas could have the same effect. Parietal researchers develop techniques to find confidence intervals to establish that the statistical relationships highlighted in the images are indeed credible.

Functional images of the brain represent activation when the subject performs particular tasks, such as watching a movie. But while describing in detail the mental operations that follow one another when watching a movie or listening to a story is complicated, we now have artificial neural networks that do it as well as or even better than humans. It is therefore exciting to study **whether certain regions of the brain could react like artificial neurons**. Parietal's researchers have shown that certain areas of the visual cortex behave like successive layers of a deep neural network! We are now studying whether modern language processing systems can explain the response observed in the brain when listening to a story.

Knowledge of the brain does not stop with image and signal processing: experiments produce results that need to be integrated into **knowledge bases**, so that they can be incorporated in unifying theories or can be reused to better analyse new data. Until now, this work has been done by reading publications in the field. Parietal's recent research contributed to automate the acquisition and use of knowledge from publications ([neuroquery.org](http://neuroquery.org)), but also to test the results of several dozens of cognitive neuroscience experiments in order to integrate them into a model. In this way, we can synthesize the experimental information collected into a model of the brain's organization, which becomes more precise as more data is added. In addition, to make it possible to question the role, the structure and relationships between different parts of the brain, Parietal's researchers have created a domain-specific language Neurolang that allows data sets to be queried to automatically identify brain structures in a new brain image. This language has formal guarantees, and allows probabilistic information to be produced with a limited degree of certainty.



Functional connectivity between brain regions

## New models of human learning

Teams in this domain study how machines can acquire knowledge models by interacting with their environment, pushed by artificial curiosity mechanisms (otherwise called *developmental robotics*). This is an important challenge connected to the question of sustainability of AI, by learning with a small set of examples as opposed to the huge datasets currently used by deep learning systems with the now well-known consequences in terms of computing resources and energy consumption.

### FLOWERS

#### Flowing Epigenetic Robots and Systems

FLOWERS studies models of open-ended development and learning. These models are used as tools to help us understand better how children learn, as well as to build developmental machines that learn like children, with applications in robotics, human-computer interaction and educational technologies.

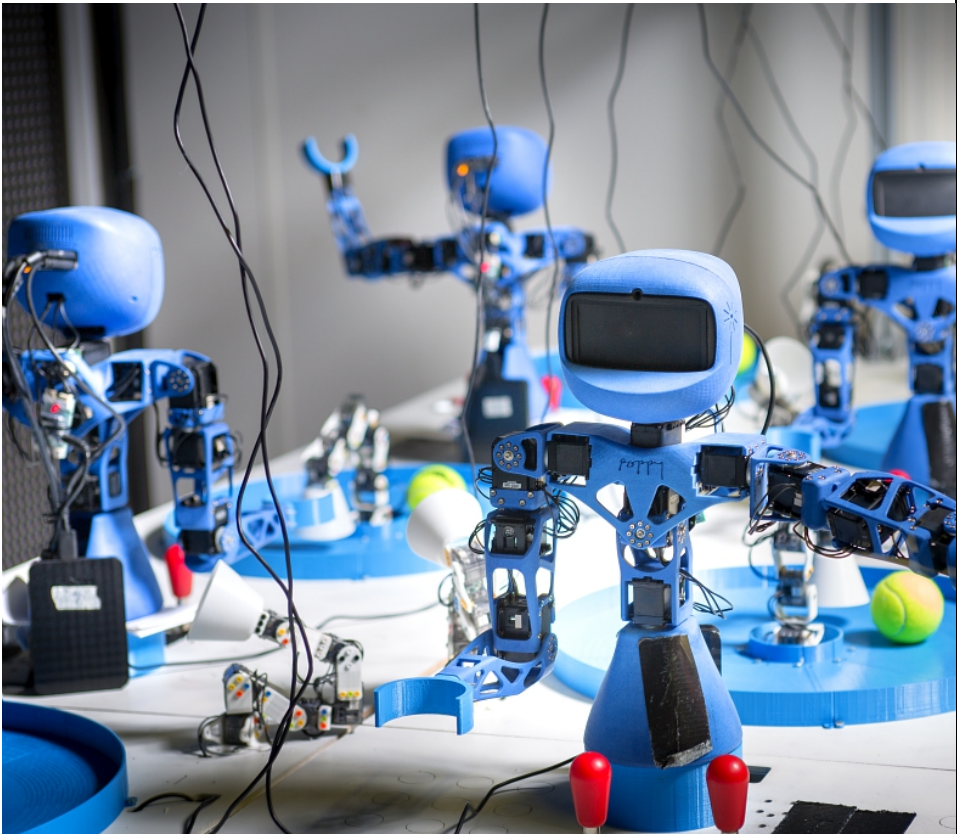
A major scientific challenge in artificial intelligence and cognitive sciences is to understand **how humans and machines can efficiently acquire world models, as well as open and cumulative repertoires of skills over an extended time span.** Processes of sensorimotor, cognitive and social development are organised along ordered phases of increasing complexity, and result from the complex interaction between the brain/body with its physical and social environment.

To advance the fundamental understanding of mechanisms of development, the FLOWERS team has developed computational models that leverage advanced machine learning techniques such as **intrinsically motivated deep reinforcement learning**, in strong collaboration with developmental psychology and neuroscience. In particular, the team has focused on models of intrinsically motivated learning and exploration (also called curiosity-driven learning), with mechanisms enabling agents to learn to represent and generate their own goals, self-organizing a learning curriculum for efficient learning of world models and skill repertoire under limited resources of time, energy and compute. The team also studies how autonomous learning mechanisms can enable humans and machines to acquire **grounded language skills, using neuro-symbolic architectures for learning structured representations and handling systematic compositionality and generalization.**

Beyond leading to new theories and new experimental paradigms to understand human development in cognitive science, as well as new fundamental approaches to **developmental machine learning**, the team has also explored how such models can find applications in robotics, human-computer interaction and educational technologies. In robotics, the team has shown how artificial curiosity combined with imitation learning can provide essential building blocks allowing



robots to acquire multiple tasks through natural interaction with naive human users, for example in the context of assistive robotics. The team also showed that **models of curiosity-driven learning can be transposed in algorithms for intelligent tutoring systems**, allowing educational software to incrementally and dynamically adapt to the particularities of each human learner, and proposing personalised sequences of teaching activities. In human-computer interaction, the team has shown how incremental learning algorithms can be used to remove the calibration phase in certain brain-computer Interfaces.



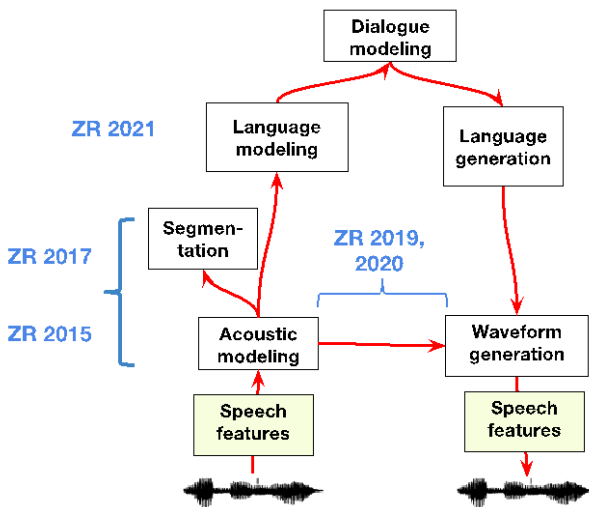
Poppy torso : curiosity driven learning

## CoML

### Cognitive Machine Learning

The general aim of CoML is to bridge the gap in cognitive flexibility between humans and machines learning in language processing and commonsense reasoning by reverse engineering how young children between 1 and 4 years of age learn from their environment. CoML conducts work along two axes: the first one, **Developmental AI** is focused on building infant-inspired machine learning algorithms. The second axis, **Quantitative studies of human learning**, uses these algorithms to conduct large scale quantitative analyses of human infants learning in the wild across diverse environments.

Developmental AI rests on the idea that it might be simpler to build a machine that learns as an infant than to build an adult one (A. Turing, 1950). Developmental research shows that infants spontaneously and autonomously learn language, social cognition, and common sense from limited uncurated and unlabelled multimodal data, and in most cultures, with only sparse direct adult supervision. We study how **self-supervised** or **weakly supervised** algorithms can discover representations or discrete units like phonemes or words from the raw acoustic signal, without any expert label (*zero resource speech learning*). We explore the **inductive biases** of neural systems by studying the conditions of language emergence (*zero data language learning*). We establish metrics and datasets for unsupervised/self-supervised systems and put together **benchmarks and challenges** in order to help building an international community in this general area.



The Zero Resource Challenge Series: learning speech and language representations by self-supervision from raw audio ([www.zerospeech.com](http://www.zerospeech.com)).

In quantitative studies of human learning, we analyse naturalistic longform recordings of infants-parents interactions to provide **upper and lower bounds** on the data that can yield successful language learning through self- or weak supervision (for instance, a 4 year old requires between only 2k and 5k hours of directed speech to learn a functioning spoken language dialogue system). We construct **causal models** of language growth that predict infant vocabulary given their input. We also model second language acquisition in adults. The team develops a hardware and software **platform** to help with data collection, annotation, and analysis on a large scale while preserving privacy and security (BeHive project).

### ***Exploratory actions (AEx)***

[AEx- ORIGINS - Grounding Artificial Intelligence in the origins of human behaviour](#)

Project team: FLOWERS

One of the most ambitious goal in Artificial Intelligence (AI) is the realisation of a so-called Artificial General Intelligence (AGI), i.e. an AI that is not limited to the realisation of a predefined set of tasks but is able to generalise its capabilities to any cognitive task that can be solved by human intelligence. However, although AGI is fundamentally related to the characteristics of human intelligence, research in this field rarely considers the processes that may have guided the emergence of complex cognitive capacities during the evolution of the species. The AEx ORIGINS will address this gap by extracting computational principles from the literature in Human Behavioural Ecology and applying them in AI to improve the acquisition of complex behaviour in artificial agents.

[AEx – ODiM - Computerised tools to assist the diagnosis of mental illness](#)

Project team: SEMAGRAMME

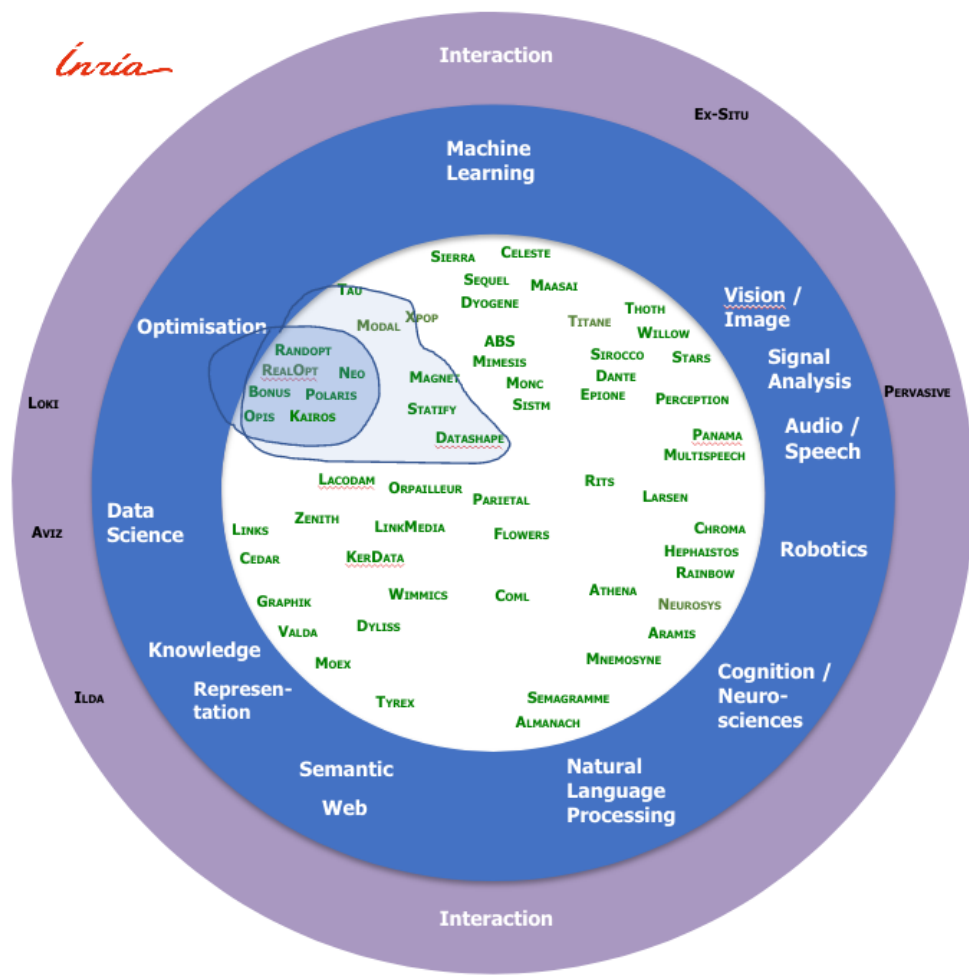
ODiM is an interdisciplinary project at the interface between psychiatry- psychopathology, linguistics, formal semantics and digital sciences. It aims to develop novel approaches to help diagnose and screening of psychotic disorders by broadening the long-term methods used in psychiatry. Production of tools is planned so that a maximum number of users from to the Mental Health Sector (doctors-psychiatrists, psychologists, speech therapists...) are able to use them.

Other project-team in this domain: NEUROSYS (Nancy)

## 5.8 Optimisation

The turn of the century has seen the development of optimisation technology in the industry and the corresponding scientific field, at the border of Constraint Programming, Mathematical Programming, Local Search and Numerical Analysis. Optimisation technology is now assisting public sector, companies and people to some extent for making decisions that use resources better and match specific requirements in an increasingly complex world. Indeed, computer aided decision and optimisation is becoming one of the cornerstones for aiding all kinds of human activities.

In the more or less near future, quantum computing is expected to revolutionise the field of optimisation, making it possible to solve problems that are intractable today.



## OPTIMISATION AND MACHINE LEARNING

Machine Learning relies on numerical optimisation for the adjustment of model parameters (billions of them in the case of deep learning), therefore close links have been established for decades between both paradigms. The use of ML as a component of optimisation is a more recent trend, where machine learning models – usually neural networks, thanks to their differentiability properties, allow an end-to-end optimization using simple gradient methods, provided enough data is available. Some challenges are at the intersection of both approaches.

## Scaling up

Models and data continue to grow exponentially as problem sizes increase. It is mandatory to design methods and algorithms able to cope with larger and larger problems without using exponentially increasing computer resources. This is true for all kinds of optimisation paradigms i.e. continuous, discrete or hybrid and for all machine learning approaches.

## Complex structures

ML and optimisation deal with complex objects i.e. not only 1-D to 3-D signal (sound, images, videos etc.) but also structures like graphs, trees, semantic networks etc. Even if in many cases these complex structures can be represented by vectors thanks to the development of specialised embeddings, this is not true for all structures, in particular working directly with graphs can be particularly useful, but this remains a challenging question.

## Proofs, confidence

When dealing with real-world applications, all elements supporting confidence in the AI/optimisation systems used are welcome. In the beginning of this chapter, we addressed the generic question of trust and confidence in AI – in particular in the case of ML. There is a need to produce proofs of convergence or confidence intervals for optimisation systems within a reasonable amount of resources used or computing time.

## Proper use of surrogate models

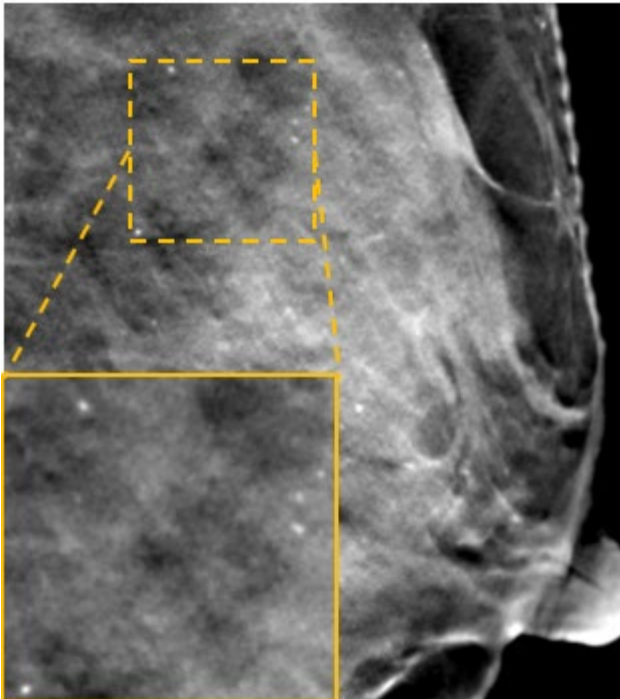
The first historical use of ML within an optimisation framework, still widely used and profoundly useful, has been to provide a surrogate model for the complex system at hand, which can be used efficiently and faithfully instead of running the real system – which in some cases is not even thinkable. The use of such surrogate models implies to develop tools and methods providing guarantees that the model is close enough to reality so that the results can be put into use.

### **OPIS**

#### **Optimisation for large Scale biomedical data**

OPIS is a new Inria-Saclay project that aims at addressing challenges raised by **advanced optimisation methods for processing large scale biomedical data**. Optimisation methods are at the core of many recent advances in artificial intelligence since one of the main brain functionalities is to provide optimal responses to problems we face. OPIS seeks optimisation methods able to tackle data with both a large sample-size ("big N" e.g.,  $N=109$ ) and/or many measurements ("big P" e.g.,  $P=104$ ). The methodologies to be explored will be grounded on nonsmooth functional analysis, fixed point theory, parallel/distributed strategies, and neural networks. The new optimisation tools that will be developed will be set in the general framework of graph signal processing, encompassing both regular graphs (e.g., images) and non-regular graphs (e.g., gene regulatory networks). More precisely, OPIS is working on three fronts:

1. New algorithms are designed for solving high-dimensional problems (sometimes involving up to billions of variables) that are encountered in inverse problems e.g., image reconstruction or restoration, for medical applications.
2. Novel strategies are proposed to address data mining problems that are formulated over graphs. Graph structures allow us to capture complex system interactions such as those existing in biological networks.
3. Deep learning methods are investigated by putting emphasis on robustness guarantees and the ability to account for prior information. Proposing better neural network models is of crucial importance in the context of the diagnosis or prognosis of diseases from medical images.



Digital Breast Tomosynthesis reconstruction based on machine learning techniques to increase the detectability of microcalcifications (collaboration with GE Healthcare)

**RANDOPT**

**Randomized Optimisation**

The RandOpt team at Inria's Saclay – Ile-de-France research centre, joint team with the CMAP at Ecole Polytechnique, deals with the analysis, development and implementation of randomized blackbox optimisation methods in the continuous domain. RandOpt is in particular focusing on CMA-ES type methods and are interested in benchmarking.

The specificity in black-box optimisation is that methods are intended to solve problems characterized by a non-property—non-convex, non-linear, non-smooth. This contrasts with gradient-based optimisation and poses on the one hand some challenges when developing theoretical frameworks but also makes it compulsory to complement theory with empirical investigations.

RandOpt ultimate goal is to provide software that is useful for practitioners. They see that theory is a means for this end (rather than an end in itself) and it is also RandOpt's firm belief that parameter tuning is part of the designer's task.

This shapes, on the one hand, four main scientific objectives:

1. **develop novel theoretical frameworks** for guiding (a) the design of novel black-box methods and (b) their analysis, allowing to
2. provide **proofs off-key features** of stochastic adaptive algorithms including the state-of-the-art method CMA-ES: linear convergence and learning of second order information.
3. develop **stochastic numerical black-box algorithms** following a **principled design** in domains with a strong practical need for much better methods namely **constrained, multiobjective, large-scale and expensive optimisation**. Implement the methods such that they are easy to use. And finally, to
4. **set new standards in scientific experimentation, performance assessment and benchmarking** both for optimisation on continuous or combinatorial search spaces. This should allow in particular to advance the state of **reproducibility of results of scientific papers** in optimisation.

## OPTIMISATION AND PERFORMANCE

In terms of the design of effective Artificial Intelligence techniques dealing with complex tasks and optimisation problems, the main challenges are:

- (1) gaining a more fundamental understanding of what makes a task/problem difficult to solve,



- (2) accommodating the broad range of complex tasks/problems with respect to the broad range of specialized solving techniques in an abstract, flexible and efficient manner,
- (3) cross-fertilizing the knowledge from other disciplines, such as HPC, operation research, etc, for an increased accuracy and efficiency,
- (4) Dealing with large scale and computationally expensive tasks/problems,
- (5) Incorporating the multi-objective nature of many practical tasks/problems, and scaling on (ultra-scale) modern supercomputers.

## **BONUS**

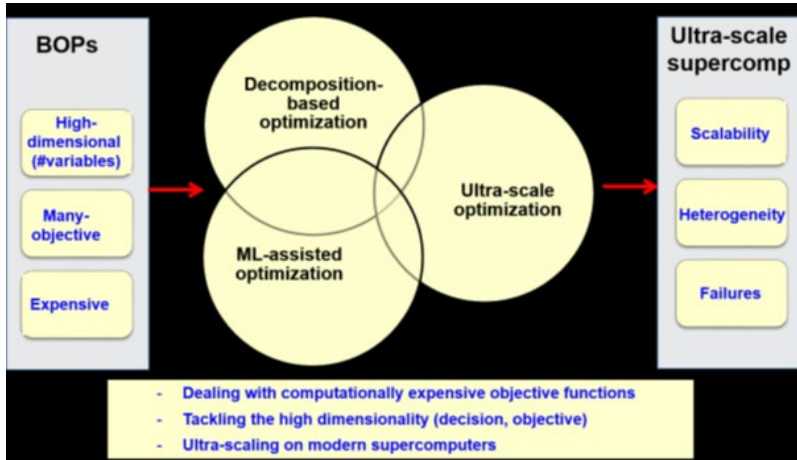
### **Big Optimisation and Ultra Scale computing**

BONUS is a joint research team between Inria Lille - Nord Europe, CRISTAL (UMR 9189, Univ Lille, CNRS, EC Lille) and the University of Lille. The team addresses big optimisation problems, defined by a large number of parameters, of decision variables, and/or many computationally expensive objective functions. The focus is on the design of effective solving techniques from computational intelligence (stochastic local search, evolutionary computation) and exact combinatorial search (branch-and-bound) following three research lines:

1. **Decomposition-based optimisation:** Given the particularly large scale of big optimisation problems in terms of variables and objectives, BONUS develops new decomposition techniques by breaking up the original target problem into smaller subproblems that are easier to solve, and loosely coupled or independent. Solving these subproblems simultaneously and cooperatively is essential to address the curse of dimensionality.
2. **Machine learning-assisted optimisation:** When dealing with high-dimensional problems and objective(s) coming from simulations or other black-box systems, BONUS is coupling computational intelligence techniques with surrogate meta-models and other machine learning algorithms in order to speed-up the convergence of the optimisation process and to cope with the computationally expensive nature of big optimisation problems.
3. **Ultra-scale optimisation:** In order to benefit from the massive parallelism offered by modern supercomputers, BONUS relies on ultra-scale computing for the effective resolution of big optimisation problems, such as handling the large amount of subproblems generated by decomposition, or the parallel evaluation of simulation-based objectives and meta-models.

From the software standpoint, BONUS objective is to integrate the approaches BONUS will develop in ParadisEO [3] (ParadisEO: <http://paradiseo.gforge.inria.fr/>) framework in order to allow their reuse inside and outside the Bonus team. The

major challenge will be to extend ParadisEO in order to make it more collaborative with other software including machine learning tools, other (exact) solvers and simulators.



BONUS closely collaborates with international researchers from the University of Mons (Belgium), the University of Coimbra (Portugal), Shinshu University (Japan), City University (Hong Kong), Monash University, and University of Luxembourg in an effort to reflect the strong synergy between optimisation, computational intelligence and parallel computing.

## NEO

### Network Engineering and Operations

**NEO is positioned at the intersection of Operations Research and Network Science.** NEO researchers model situations arising in several application domains, involving networking and distributed systems in one way or the other, with the goal to take (possibly) optimal decisions using the tools of Stochastic Operations Research. Modern AI is also involved with decision, taken (or suggested) by machines based upon some data (machine learning). Quite naturally then, distributed AI has become one of NEO research topics along the following axes:

1. Semi-supervised learning on graph structures and its distributed implementations.
2. Design of Internet-scale distributed machine learning systems, both for training and inference, with a focus on the trade-off between performance and economic and environmental costs.
3. Multi-agent learning models based on game theory. This includes evolutionary game theory whose equilibrium consists of restpoints of Darwinian-type dynamics, dynamic non-cooperative games in which

cooperation may be induced by threats and punishments, and matching games that have been applied for recommendation networks.

4. Analysis of the fundamental limits of the influence of information-provisioning policies (recommender systems, media, social networks, etc.) on decision takers involved in competitive interactions (markets, shared-resource systems).

The team collaborates on these topics with many industrial partners, including Qwant, Nokia, Accenture, MyDataModels, Azursoft.

Other related NEO research topics are: resource allocation in communication networks, social networks, green computing and communications, and sustainable development.

## **POLARIS**

### **Performance analysis and Optimisation of LARge Infrastructures and Systems**

The goal of the POLARIS project is to contribute to the understanding (from the observation, modeling and analysis to the actual optimisation through adapted algorithms) of the performance of very large-scale distributed systems such as supercomputers, cloud infrastructures, wireless networks, smart grids, transportation systems, or even recommendation systems.

A first line of research is devoted to the use statistical learning techniques (Bayesian inference) to model the expected performance of distributed systems to build aggregated performance views, to feed simulators of such systems, or to detect anomalous behaviours.

In a distributed context it is also essential to design systems that can seamlessly adapt to the workload and to the evolving behaviour of its components (users, resources, network). Obtaining faithful information on the dynamic of the system can be particularly difficult, which is why it is generally more efficient to design systems that dynamically learn the best actions to play through trial and errors. A key characteristic of the work in the POLARIS project is to leverage regularly game-theoretic modeling to handle situations where the resources or the decision is distributed among several agents or even situations where a centralised decision maker has to adapt to strategic users.

The POLARIS members are thus particularly interested in the design and analysis of adaptive learning algorithms for multi-agent systems, i.e. agents that seek to progressively improve their performance on a specific task (see Figure). The resulting algorithms should not only learn an efficient (Nash) equilibrium but they should also be able of doing so quickly (low regret), even when facing the difficulties associated to a distributed context (lack of

coordination, uncertain world, information delay, limited feedback, ...).

An important research direction in POLARIS is thus centered on reinforcement learning (Multi-armed bandits, Q-learning, online learning) and active learning in environments with one or several of the

following features:

- Feedback is limited (e.g., gradient or even stochastic gradients are not available, which requires for example to resort to stochastic approximations);
- Multi-agent setting where each agent learns, possibly not in a synchronised way (i.e., decisions may be taken asynchronously, which raises convergence issues);
- Delayed feedback (avoid oscillations and quantify convergence degradation);
- Non stochastic (e.g., adversarial) or non stationary workloads (e.g., in presence of shocks);
- Systems composed of a very large number of entities, that we study through mean field approximation (mean-field games and mean field control). As a side effect, many of the gained insights can often be used to dramatically improve the scalability and the performance of the implementation of more standard machine or deep learning techniques over supercomputers.

## **KAIROS**

### **Temps Logique Multiforme pour Conception de Systèmes Cyber-Physiques**

Machine Learning (ML) techniques (e.g. Deep Neural Networks) have benefited from efficient implementation platforms (GPUs and TPUs) and from compilation methods developed by the High Performance Computing (HPC) community to gain practical feasibility and recognition.

Meanwhile, Safety-Critical (often Real-Time) Embedded systems identified themselves as a place of choice for real-life ML applications (e.g. automated driving, digital twin models). Therefore it becomes tempting and profitable to combine both domains, and in particular to federate:

1. the optimized compilation methods for data parallel specifications, developed in the HPC/ML community, and
2. the methods developed in the embedded real-time community to provide worst-case resource consumption guarantees for task parallel specifications.

Based on the deep proximity between intermediate formalisms of HPC/ML compilers (MLIR/SSA) and formalisms used in real-time design (Lustre), the Kairos team explores methods for the specification and (safe and efficient) implementation of ML-friendly high-performance embedded applications.

Other project-team in this domain: REALOPT (Bordeaux)

## 5.9 AI and Human-Computer Interaction (HCI)

Humans can now delegate tasks such as driving a car or piloting a plane, and AI systems are regularly touted to be "better than humans" at various high-level tasks. AI systems are not perfect however, and humans have been kept or put "in the loop" of many AI-based safety-critical systems to protect against unexpected system behaviours. Unfortunately, this arrangement has led to some dire consequences, as exemplified by recent accidents such as the crashes of two Boeing 737 Max commercial planes, where the anti-stall system made the planes nose-drop twenty-six times in a row in less than ten minutes without giving the pilots the necessary information and control to save the plane. Such accidents are the consequence of an unfettered trust in technology over human skills, and a shift from situations where humans delegate tasks but remain in control to those where the computer treats the human as a source of input to an algorithm. The "human in the loop" is essentially a cog in the machine, who takes the blame when things go wrong. Such systems do not take optimal advantage of human talent and system abilities, but rather assume that the computer can always compute an optimal solution. *Thus, a major challenge for both AI and HCI is to create a better division of labor between humans and computers, harnessing their respective powers and capabilities while acknowledging their limitations and weaknesses.*

Another strand that interweaves AI and HCI relates to the massive quantities of personal data analysed by powerful machine learning algorithms. Our interaction with the digital world has been fundamentally redefined -- our decisions are monitored, nudged and often manipulated, which threatens not only our privacy but also democracy and basic human rights. Here too, human control over computer processes has been traded for computer control over human behaviour. *A second major challenge is how to bring true transparency and explainability to AI systems through appropriate user interfaces and visualizations.*

Current applications of AI techniques to fields such as medical diagnostic, justice sentencing or automated driving tend to deskill expert users: by automating tasks once performed by humans, it may be possible to improve productivity for "normal" situations. But computers are extremely bad at handling exceptional cases, and it is illusory to think that a "better" AI will significantly change this situation. Humans, on the other hand, are very good at handling exceptional cases, as long as they can stay trained, but are notoriously bad at monitoring activities. *A third major challenge is how to combine interactive and AI systems so that each takes advantage of the other's strengths at the appropriate time, while minimizing each other's limitations.*

Modern AI systems are becoming so complex that engineers require new tools simply to monitor and manage their development, evolution, debugging, and generally understand what is happening "under the hood". For example, large ML environments come with sophisticated tools to design and program them<sup>32</sup>. Most steps involved in AI systems require tools to assess quality of data, features, training, and decisions; to understand the behaviour of an AI system at any particular point; to monitor and improve its quality; to discover biases and uncertainty in the results; and to deliver the results to target users in a meaningful way. *A fourth major challenge is to create better, more user-centred tools for experts who create and evaluate AI systems.*

### HCI to Improve AI

In addition to tools to improve AI, HCI should also help create more transparent AI systems so they can be assessed by experts in their application domains. For example, bank loan management is more and more assisted by AI tools and has a direct impact on the life of citizens. Some automated decisions have been subject to structural biases difficult to foresee by AI engineers but certainly detectable by loan experts<sup>33</sup>. However, addressing these biases require communication tools between the two kinds of experts to find-out the causes and agree on remedies. For the loans, causes have been found in faulty proxy measures used to score people, and in unbalanced training data misrepresenting women or minorities. Discovering these biases requires human judgement, and can be very different in kind.

Transparency is also more than explaining decisions or showing the machinery, it also consists in explaining or taking into account the capabilities of a system and its limitations. Self-driving cars are good in some standard situations but unreliable in others. They should provide a warning to the driver to take back the control when needed, which requires AI systems to be aware of their own level or reliability (something they rarely do), and to gracefully hand the control over to humans, something that is notoriously difficult and will require more research.

---

<sup>32</sup> K. Wongsuphasawat et al., "Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow," in IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 1, pp. 1-12, Jan. 2018,

<sup>33</sup> C. O'Neil, Weapons of Math Destruction, Crown Publishing, 2016

Finally, novel machine learning systems try to learn continuously from humans through interaction with them to complete their knowledge. A system such as Google search improves its precision by monitoring the rank of the results that the user reads (clicks on) after a search query. This method is only effective at improving the "precision" of the search engine, but not its recall (if a result is not shown, it cannot be ranked). Finding methods to learn interactively and measure the increase in quality and usability remains a complex problem needing more research.

## **Aviz**

### **Analysis and Visualization**

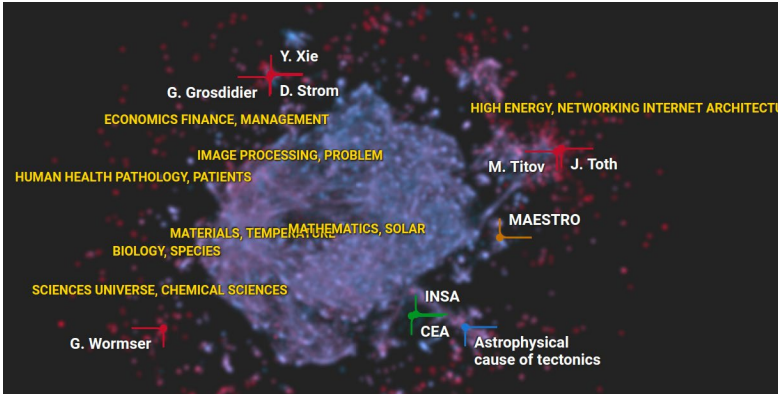
Aviz is a multidisciplinary project that seeks to improve visual exploration and analysis of large, complex datasets by tightly integrating analysis methods with interactive visualization.

Our work has the potential to affect practically all human activities for and during which data is collected and managed and subsequently needs to be understood. Often data-related activities are characterized by access to new data for which we have little or no prior knowledge of its inner structure and content. In these cases, we need to interactively explore the data first to gain insights and eventually be able to act upon the data contents. Interactive visual analysis is particularly useful in these cases where automatic analysis approaches fail and human capabilities need to be exploited and augmented.

Within this research scope Aviz focuses on five research themes:

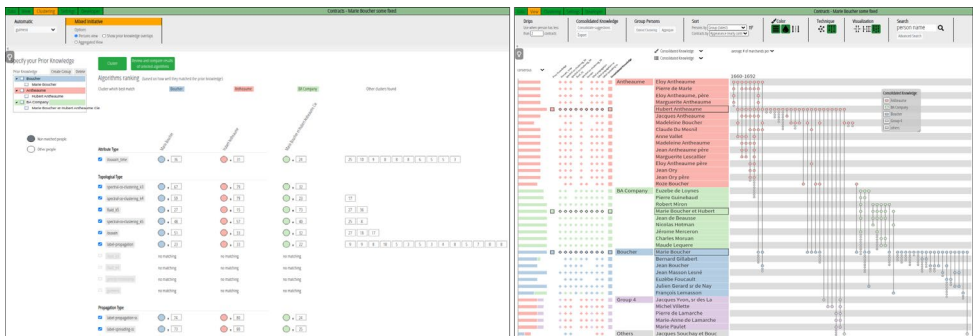
- Methods to visualize and smoothly navigate through large datasets;
- Efficient analysis methods to reduce huge datasets to visualisable size;
- Visualization interaction using novel capabilities and modalities;
- Evaluation methods to assess the effectiveness of visualization and analysis methods and their usability;
- Engineering tools for building visual analytics systems that can access, search, visualize and analyze large datasets with smooth, interactive response.

In collaboration with the TAU project-team, Aviz visualizes the HAL repository, containing all the publications of public French research institutions, using multidimensional projections to create a "map", resulting from Natural Language Processing analysis (topic modelling), clustering to collect thematic regions over the map and find meaningful labels. All these techniques related to AI are gathered using a web-based user interface to let researchers of any domain explore the publications around topics or authors, allowing complex AI techniques to be explored by a large audience of users. See [Philippe Caillou, Jonas Renault, Jean-Daniel Fekete, Anne-Catherine Letournel, Michèle Sebag. Cartolabe: A Web-Based Scalable Visualization of Large Document Collections. IEEE CG&A 2020, to appear] and <https://cartolabe.fr> .



Cartolabe visualizing HAL, with 208984 authors (red) and 827156 articles (blue)

Aviz is also working on network analysis and visualization, to let network researchers such as historians, sociologists, or brain researchers incorporate their prior knowledge into ensemble clustering methods [Alexis Pister, Paolo Bueno, Jean-Daniel Fekete, Catherine Plaisant, Paola Valdivia. Integrating Prior Knowledge in Mixed Initiative Social Network Clustering. IEEE TVCG 2021, to appear]. With PK-Clustering, users with little understanding of the clustering algorithms can still introduce some of their prior knowledge to better select or steer algorithms, instead of blindly believing the results of one particular algorithm.



PK-Clustering, showing the results of nine clustering algorithms shown as columns of dots on the left (each cluster has a colour), applied to the network on the right, and consolidated on the rightmost column against prior knowledge.

AI to improve HCI



## LOKI

### Technology & Knowledge for Interaction

LOKI envisions computers as tools that could ultimately empower people, focusing on how such tools can be designed and engineered. By better understanding phenomena that occur at each level of interaction and their relationships, we gather the necessary knowledge and technological bricks to reconcile the way interactive systems are engineered for, around, and with human abilities. Our scope of research encompasses a broad set of interactive environments (desktop computers, mobile devices, VR, BCI...) and borrows its methods from fields as varied as psychology and neuroscience, AI, or design and engineering.

In our goals to better understand users and to design systems that adequately respond to their abilities, we frequently make use of recent AI contributions, notably machine learning and optimization. We played an instrumental role in the design of the new French keyboard layout standard [NF Z 71-300. <http://norme-azerty.fr/>] commissioned by the French Ministry of Culture, using state-of-the-art combinatorial optimization methods [A. Feit et al, *Élaboration de la disposition AZERTY modernisée*. 2018. <https://hal.inria.fr/hal-01826476>]. In collaboration with Aalto University and the Max Planck Institute, we developed a workflow that allowed non-technical typography and linguistics experts to iterate and evaluate layout ideas with an optimizer. That optimizer was in turn able to express the consequences of these ideas in understandable terms of ergonomics and typing performance [A. Feit et al., *AZERTY amélioré: Computational Design on a National Scale*. In *Communications of the ACM* (In press)].

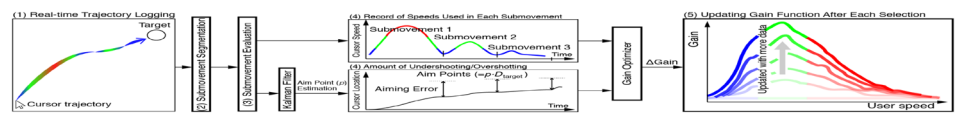
Using a different approach, AI methods can also be leveraged to dynamically adapt user interfaces depending on the user's profile, context of interaction, or needs. As an example, with colleagues from University College London, we used hierarchical clustering methods to adapt displayed content to user's profile, in the context of mobile news reading [Constantinides et al, *Exploring mobile news reading interactions for news app personalisation* <https://hal.inria.fr/hal-01252631>]. We also plan to explore the use of computational methods to dynamically anticipate users' needs in the context of rich-software interaction and help them discover novel features they are not yet aware of (ANR project DISCOVERY).

Many interactive contexts could benefit from a synergy between user input and system intelligence. One of our hypotheses is that users are more likely to accept a solution suggested by an AI when they have directly contributed to the development of that solution (e.g., through occasional explicit inputs), while the AI provides "honest" feedback that acknowledges its possible imprecision. We are exploring this question in the context of archival of old handwritten documents, which currently combines document scanning with manual or automatic transcription in a sequential manner. Following the same Human-AI partnership paradigm, we are currently

exploring with colleagues from the University of Waterloo how users rely on AI-suggested words when typing text. We investigate how users manage the trade-off between typing words with a virtual keyboard and using the suggestions proposed by the AI, depending on the accuracy of the suggestions and the efficiency of the interface. This will help inform the design of interactive systems by providing ways to automate the user's task [Roy et al. under review CHI 2021].



Interacting with a system in real time requires the ability to gather and interpret continuous data streams that can be noisy or that can lack semantics. AI allows us to better leverage these rich signals and to solve known interface issues in novel and efficient ways. Latency for instance, whether noticeable or not [R. Jota et al., How Fast is Fast Enough? A Study of the Effects of Latency in Direct-touch Pointing Tasks. In Proc. of ACM CHI '13], is a scourge of interaction performance. Up until recently, its only cure was to wait for hardware to improve — which is however inevitably followed by more demanding software, bringing latency back to where it started. We tried another, more hardware-independent approach: we applied state-of-the-art optimization and estimation techniques to tune an algorithm capable of accurately predicting cursor movements in the near future, which we used to visually compensate end-to-end latency for relative pointing [M. Nancel et al., Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation. In Proc. of ACM UIST '18. <https://hal.inria.fr/hal-01893310>]. Also using optimization algorithms, and in collaboration with Aalto University and KAIST, we designed a tool able to adapt in real time the acceleration profile of a cursor to the user's pointing skills and habits, be it controlled by a mouse, a trackpad, or even by hand gestures in mid-air [B. Lee et al. AutoGain: Gain Function Adaptation with Submovement Efficiency Optimization. In Proc. ACM CHI '20. <https://hal.inria.fr/hal-02918581>].



## Human-AI Partnerships

Early thinkers such as J.C.R. Licklider and D. Engelbart have put forward the concept of “human-machine symbiosis”<sup>34</sup> or the vision of “augmenting human intellect”<sup>35</sup> where computer systems use AI to serve, rather than replace, human intelligence and expertise. Creating such successful *human-AI partnerships* is key when combining AI and HCI.

Human-Computer Interaction focuses on the interaction between the user and a system, which we assume is a dynamic relationship that changes over time. When we deal with intelligent systems, both the user and the system can have agency. One of the key interaction design challenges is how to manage this shared agency, ideally leaving the user in control of the interaction, but at least giving them ‘informed consent’ as to what is happening. The standard ‘human-in-the-loop’ perspective treats the human user as input to the algorithm, and success is defined in terms of creating faster, higher performing algorithms. While creating better algorithms remains a desirable goal, it is critical that we also take a human-centered perspective that defines success in more qualitative, user-oriented terms, which includes increased human performance, but also increased human capabilities and satisfaction. This perspective also colors how we view mixed-initiative approaches. Instead of trying to replace the human user with an algorithm, they emphasize the on-going role of the user within the interaction. Most of today’s mixed-initiative research still focuses on the algorithm, rather than enhancing human skills. Human-AI partnerships seek to leverage the best characteristics of human users and intelligent systems, where the combination exceeds what can be accomplished by either alone.

### **ExSitu Extreme Situated Interaction**

ExSitu explores the limits of interaction — how extreme users interact with technology in extreme situations. We are particularly interested in creative professionals, artists and designers who rewrite the rules as they create new works, and scientists who seek to understand complex phenomena through creative exploration of large quantities of data. Studying these advanced users today will not only help us to anticipate the routine tasks of tomorrow, but to advance our understanding of interaction itself.

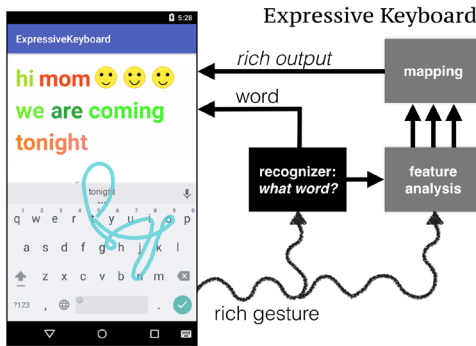
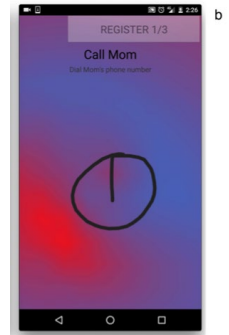
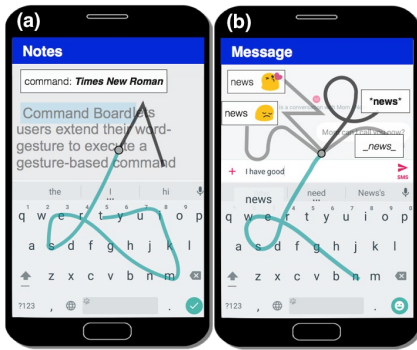
---

<sup>34</sup> <http://memex.org/licklider.pdf>

<sup>35</sup> [https://www.doungengelbart.org/pubs/papers/scanned/Doug\\_Engelbart-AugmentingHumanIntellect.pdf](https://www.doungengelbart.org/pubs/papers/scanned/Doug_Engelbart-AugmentingHumanIntellect.pdf)

In creative practices, *human-centred machine learning* facilitates the workflow for creatives to explore new ideas and possibilities. We have compiled recent research and development advances in human-centred machine learning and AI in creative industries [B. Caramiaux et al. *AI in the media and creative industries*, New European Media (NEM), April 2019, pp. 1-35. <https://hal.inria.fr/hal-02125504>]. We have also explored the use of Deep Reinforcement Learning in the context of sound design by comparing manual exploration versus exploration by reinforcement. We showed that an algorithmic sound explorer learning from human preferences enhances the creative process by allowing holistic and embodied exploration as opposed to the analytic exploration afforded by standard interfaces.

We are also interested in designing effective *human-computer partnerships*, in which expert users control their interaction with technology. Rather than treating human users as the 'input' to a computer algorithm, we explore human-centered machine learning, where the goal is to use machine learning and other techniques to increase human capabilities. Our specific goal is to create *co-adaptive systems* that are discoverable, appropriable and expressive for the user. The CREATIV ERC Advanced project developed this approach and created a series of prototypes designed to increase the user's power of expression on mobile devices: CommandBoard [J. Alvina et al. CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboards. Proc. UIST 2017. <http://hal.inria.fr/hal-01679137>], FieldWard [J. Malloch et al. Fieldward and Pathward: Dynamic Guides for Defining Your Own. Proc. CHI 2017. <http://hal.inria.fr/hal-01614267>], Expressive Keyboard [J. Alvina et al. Expressive Keyboards: Enriching Gesture-Typing on Mobile Devices. Proc. UIST 2016. <http://hal.inria.fr/hal-01437054>] (figure below).



*CommandBoard (left) lets users enter complex commands with gestures; Fieldward (center) lets users define their own gestures while ensuring that they are recognizable by the system; and Expressive Keyboard (right) extracts expressive characteristics of the user's gesture to generate rich, expressive output, including dynamically modifying color, font characteristics and even emoji expressions.*

When we work with creative professionals, we focus not on trying to make them more creative — they are already creative — but rather on providing tools that support their own, personal creative process. Such tools include the use of interactive paper to support composers [Musink, Polyphony] and designers [StickyLines, Enact]. We have also explored how mood board designers and intelligent systems can effectively share agency according to their in-the-moment needs with Semantic Collage [J. Koch et al. (2020) Semantic Collage. In Proc. DIS'20. <https://dl.acm.org/doi/10.1145/3357236.3395494>] and ImageSense: [J. Koch et al. (2020) ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. In Proc. ACM on Human Computer Interaction, Issue CSCW. <https://hal.archives-ouvertes.fr/hal-02867303/>], joint with Aalto University.

In the *Bayesian Information Gain* (BIG) project, joint with Telecom Paris, we use a technique based on Bayesian Experimental Design where the criterion is to maximize the information-theoretic concept of mutual information: rather than simply interpret user commands, BIG uses user input to update its knowledge about the user's intended goal and provides an output that maximizes the expected information gain from the next input. In other words, the system challenges the user in order to make interaction more efficient. We have applied BIG to multiscale navigation [W. Liu et al. BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. Proc. CHI 2017. <http://hal.inria.fr/hal-01677122>] and to file retrieval [W. Liu et al. . BIGFile: Bayesian Information Gain for Fast File Retrieval. Proc. CHI 2018. <http://hal.inria.fr/hal-01791754>] and demonstrated performance gains of up to 40% compared to conventional navigation techniques.

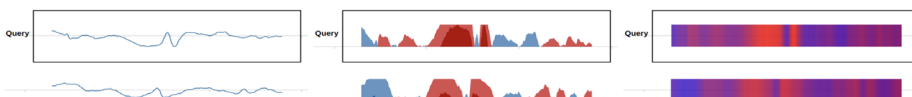
## ILDA

### Interacting with Large Data

ILDA designs data-centric interactive systems that provide users with the right data at the right time and enable them to effectively manipulate and share these data. Our work focuses on the design, development and evaluation of novel interaction and visualization techniques to empower users in both mobile and stationary contexts involving a variety of display devices, including: smartphones and tablets, augmented reality headsets, desktop workstations, table tops, ultra-high-resolution wall-sized displays. Our research themes include novel forms of input and display for both groups and individuals, as well as novel ways to interact with novel data models that enable diverse structuring and querying strategies, give machine-processable semantics to the data and ease their interlinking. We investigate ways to leverage this richness from the users' perspective, designing interactive systems adapted to the specific characteristics of data models and data semantics, with a focus on mission critical systems and the exploratory analysis of scientific data.

With colleagues from Paris-Descartes and the ExSitu team we investigated human-AI partnerships in the domain of neuroscience and time series analysis (EEG signals). We first explored how to aid expert neuroscientists evaluate epileptiform patterns found in EEG signals, by combining visualization and automated processing in the form of similarity search algorithms. We examined how using different visualizations can affect the similarity perception in EEG signals, and how different visualizations can better match similarity measures [A.Gogolou, et al. Comparing Similarity Perception in Time Series Visualizations. IEEE TVCG 2019 (Proc InfoVis 2018), <https://hal.inria.fr/hal-01845008>]. We thus showed that the notion of similarity is visualization-dependent, and the need to match automated processes with appropriate visual representations. Other work also helps experts query massive data series collections (such as EEG databases) within interaction times. We provided

progressive similarity search results on large time series collections (100 GB) and showed how these can cut waiting times for users, as we observed that high-quality approximate answers are found very early, e.g., in less than a second [A.Gogolou et al. Progressive Similarity Search on Time Series Data. Proc BigVis 2019, <https://hal.inria.fr/hal-02103998v1>]. Nevertheless, it is important for users to be able to determine the quality of these early answers and to decide if they need to wait further for better matches. To this end, we have worked on providing probabilistic distance and error bounds, to help analysts evaluate the quality of their progressive results [A.Gogolou et al. Data Series Progressive Similarity Search with Probabilistic Quality Guarantees. Proc ACM SIGMOD 2020, <https://hal.inria.fr/hal-02560760v1>].



Three time series visualizations compared in order to understand if we perceive similarity differently with each one (Line Chart left, Horizon Graph middle, Colorfield right).

We also have a long-lasting collaboration with colleagues from INRAe, where we combine visual exploration with evolutionary computation to help guide experts in exploring large multi-dimensional datasets. Our framework (Evolutionary Visual Exploration - EVE), uses an interactive evolutionary algorithm to steer the exploration of multidimensional datasets towards two dimensional projections that are of interest to the analyst [N.Boukhelifa et al. Evolutionary Visual Exploration: Evaluation of an IEC Framework for Guided Visual Search Evolutionary Computation, In Evolutionary Computation, MIT Press, 2018]. Our method smoothly combines automatically calculated metrics and user input in order to propose pertinent views to the user. This work has led to a prototype application that has been used by domain experts in different fields to formulate interesting hypotheses and reach new insights when exploring freely [N.Boukhelifa et al. Evolutionary Visual Exploration: Evaluation With Expert Users. In Computer Graphics Forum 2013, <https://hal.inria.fr/hal-02005699v1>]; has acted as a collaborative platform for teams of researchers to explore trade-offs [N.Boukhelifa et al. An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts. Proc ACM CHI 2019, <https://hal.inria.fr/hal-02005699v1>]; and has initiated investigations about how to best test and evaluate frameworks such as EVE, which incorporate human and artificial intelligence that work together to reach decisions.

## Signal+AI as input to HCI

Interactive systems increasingly take advantage of sensors that capture rich user input such as voice, gaze, gestures or brain activity. HCI uses AI techniques, particularly machine learning, to analyse, recognize and/or classify these signals. The context of interaction creates specific constraints that push the limits of current AI

techniques: processing must occur in real time, at the scale of the human perception-action loop (typically under 100ms and sometimes much less); models often need to be trained with very few examples, e.g. a user is only willing to show a gesture once or twice and expect the system to robustly recognize it from then on; the model must adapt to changes in user behaviour over time. In many cases, recognition must occur progressively, as the signal arrives, so that the system can provide real-time feedback and feed-forward, as exemplified by the Octopocus dynamic guide for gesture input<sup>36</sup>. In addition, continuous input, e.g. movement data from a Kinect sensor, must be segmented in real-time in addition to the segments being recognized. Interactive Machine Learning, Reinforcement Learning, Active Learning and Online Learning all provide potential approaches to address these problems.

## PERVASIVE

The Inria project PERSVASIVE INTERACTION develops theories and models for context aware, sociable interaction with systems and services that are composed from ordinary objects that have been augmented with abilities to sense, act, communicate and interact with humans and with the environment (smart objects). The ability to interconnect smart objects makes it possible to assemble new forms of systems and services in ordinary human environments.

Pervasive Interaction explores the use of situation models as a foundation for situated behaviour by smart objects. Research is driven by experiments with situated interaction with people, with environments, and with pervasive computing.

The research program addresses the question: can situation modelling provide a theory for situated behaviour by smart objects? The program is driven by the following four research questions:

Q1: What are the most appropriate computational techniques for acquiring and using situation models for situated behaviour by smart objects?

Q2: What perception and action techniques are most appropriate for situated smart objects?

Q3: Can we use situation modelling as a foundation for sociable interaction with smart objects?

Q4: Can we use situated smart objects as a form of immersive media?

It is organized as four interacting research areas responding to these research questions:

RA1. Acquiring and Using Situation Models (Q1)

RA2. Perception of People, Activities and Emotions (Q2)

---

<sup>36</sup> O.Bau & W. Mackay. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. UIST 2008. <http://dl.acm.org/citation.cfm?id=1449724>].



## Explainable AI

Explainable AI is usually characterized in terms of explaining to users how an algorithm works. However, a true human-computer interaction perspective shifts the focus, arguing that users rarely care about the details of how the *algorithm* works, and instead are more concerned with how such algorithms may *affect* them personally as well as on their ability to accomplish the task at hand. Thus, the key challenge of user-centred explainable AI is how to reveal information to the user in terms that users understand. Users must be able to visualize how the AI system is currently interpreting and reacting to their behaviour, as well as what decisions it is making and why. Users should be able to intervene in the process, not simply to discover how and why the AI performed a particular interaction, but also have easy ways to inform the AI when those decisions are incorrect and suggest better solutions. Systems such as Fieldward and Pathward<sup>37</sup> provide both visual feedback and progressive feedforward as the user draws a proposed new gesture command. The AI dynamically interprets the gesture as it is drawn and provides a continuous classification that is revealed via a changing coloured heatmap or gesture continuations. This shows the user both how the AI has interpreted the gesture as of that instant and suggests alternative strategies for successfully generating a new, unique command.

## Cognitive Biases, Ethics, and Legal Issues

Fairness, explainability and accountability are critical properties for the acceptability of AI systems in a wide range of domains. These properties, however, must be assessed from a human perspective, not just from a system perspective. For example, Tversky and Kahneman's seminal experiments in behavioral economics show that human perception of fairness is not always rational and depends heavily on contextual information such as how the question is asked. More generally, many cognitive biases are known to affect human decision making and reasoning, such as confirmation bias and anchoring. This implies that we need to adopt HCI-centric experimental methods that involve participants, rather than relying solely on the simulations and measurements common in AI research. However this also raises ethical questions about whether and how AI systems should account for human biases, either by reproducing them or, on the contrary, combating them.

Another type of bias involves the training sets for intelligent systems. Recent studies have shown that face detection algorithms are extremely accurate for white men (over 98%), less accurate for white women, and less than 30% accurate for black

---

<sup>37</sup> J. Malloch et al. Fieldward and Pathward: Dynamic Guides for Defining Your Own. Proc. CHI 2017. <http://hal.inria.fr/hal-01614267>

women. When young, white male engineers select training sets of people who look like them, the result will be biased when applied to the general population, such as using this data for identifying potential criminals or potential job candidates.

Delegating tasks and decisions to AI systems raises additional ethical and legal questions, in particular about accountability and responsibility. While there seems to be consensus that humans should ultimately be responsible for the decisions made by AI systems, the temptation is to blame the user rather than the system designer, as exemplified by the accident that killed the driver of an autonomous car. A key question here is whether the interface to the AI system provided the user with sufficient information to avoid the accident, and whether it accounted for human traits and behavior. Assuming that users will always remain in a high state of alert after hours of accident-free driving is a fundamentally poor design decision, not a fault of the human user. Ethical issues must be addressed within the larger socio-technical environment in which the system operates.

## 6. European and international collaboration on AI at Inria

### COLLABORATIONS IN AI: INRIA'S VISION

Inria's European and international cooperation actions aim to promote exchange between Inria and the most dynamic geographical areas, whilst upholding European values for a human-centric AI<sup>38</sup>. The context is well known: the race for investment in certain areas of the world, the role of China and the United States in AI, the race for talent by prestigious foreign academic institutions and by private actors in AI.

This context encourages the institute to reinforce collaborations that are likely to boost the quality of Inria's work, to guarantee the visibility and positioning of teams at the best European and international level, but also to enrich the institute's debate on the impact of AI on our societies.

In addition to the links that are naturally established between researchers through informal collaborations and exchanges, Inria, as a national public institute on digital technology, builds its international policy through targeted agreements with partners, taking into account the orientations of France's international strategy, the specific constraints it faces, and the European framework.

### CONTRIBUTION TO EUROPEAN R&I EFFORTS IN AI

Europe's strengths lie in the quality of its researchers and engineers, training and applications. Aware of the challenges of sovereignty, the EU has adopted a human-centred strategy, advocating ethical principles<sup>39</sup>. Inria's involvement in European AI efforts relies on three dimensions: integration into networks, participation in large-scale projects and a solid contribution to exploratory research, notably through ERC-funded projects.

#### (i) Integration into networks

Inria is a member of BDVA (Big Data Value Association) and EU Robotics, which are European associations bringing together industrial and academic partners active in the fields of data and robotics, respectively coordinating the corresponding Public-Private Partnerships (PPPs). Moreover, INRIA participates in the AI/Data/Robotics PPP proposal to be submitted to the European Commission in 2021.

In addition, a number of academic oriented networks emerged in Europe that include:

- networks at the initiative of scientific communities, such as, CLAIRE (Confederation of Artificial Intelligence Research Laboratories in Europe) and, ELLIS (European laboratory for learning and intelligent systems). Inria

---

<sup>38</sup> The Ethics Guidelines for Trustworthy Artificial Intelligence (AI), AI HLEG, April 2019

<sup>39</sup> White Paper on Artificial Intelligence: a European approach to excellence and trust, EC, February 2020

institutionally supports the CLAIRE initiative, but acknowledges support to the ELLIS initiative from part of its researchers;

- networks at the initiative of the European Commission to help structure the various AI communities and stimulate dialogue and convergence between them.

With respect to the network supported by the European Commission through the Horizon 2020 programme, Inria is involved in three projects that started on Sept. 1<sup>st</sup>, 2020: the TAILOR and HumanAI R&I projects and the VISION support and coordination action. These projects lay the foundations for a world-class European research and innovation ecosystem, to implement safe, reliable AI that respects the values advocated by the European Union. Some Inria researchers are also members of the ELISE project.

TAILOR aims to reinforce links between academic, public and industrial research actors to develop the scientific basis for trusted AI. It does so, by combining learning, optimisation and reasoning to produce AI systems that guarantee the requirements of reliability, safety, transparency and respect for human activities, and optimising the expected benefits while reducing possible harm.

HumanAINet is to develop an AI that is safe, reliable, and capable of adapting to real environments and interact appropriately in complex social contexts. The objective is to promote AI systems that enhance human capabilities and provide support to individuals and society as a whole, while respecting human autonomy and self-determination.

ELISE gathers the best European research in machine learning to create a network of artificial intelligence. Where ELISE starts from machine learning as current core technology of AI, the network is inviting all ways of reasoning, considering all types of data, applicable for almost all sectors of science and industry.

VISION intends to coordinate the activity of the four European networks of excellence in AI (TAILOR, HumanAINet, ELISE and AI4Media), to help position European research as a major player in AI. This requires overcoming the fragmentation of the AI community in Europe, and stimulate synergies for the emergence of the next generation of reliable AI tools and systems, based on methods covering a wider range of AI techniques.

## **(ii) Collaborations through large research projects**

Large-scale projects complement and extend the work carried out at Inria:

AI4EU is the project that aims to build the European on demand platform, which is to render AI technology accessible to all, and as such reduce barriers to innovation, stimulate technology transfer and facilitate the growth of start-ups and SMEs in all economic sectors.

TRUST-AI and ALMA are two fundamental research projects that seek to advance human-centric AI. More precisely, TRUST-AI aims to integrate the notion of explicability into the learning phase of "black box" models, without compromising their performance. ALMA relies on the Algebraic Machine Learning (AML) paradigm, which produces generalizing models from the semantic integration of data into discrete algebraic structures, which has a number of advantages over statistical learning models.

### (iii) Scientific excellence promoted by ERC

Since the launch of the ERC (European Research Council) in 2007, Inria has obtained 59 individual grants (Starting, Consolidator, Advanced), 2 Synergy grants and 9 Proof of Concept (PoC) grants. In the field of AI, Inria has 17 ERC laureates, one of whom obtained a PoC funding in addition to his individual grant (see table below and list in appendix).

Thematic distribution:

Machine Learning & its applications	Francis Bach, Julien Mairal, Alessandro Rudi, George Drettakis (application)
Computer Vision & Signal-Image Processing	Cordelia Schmid, Ivan Laptev, Josef Sivic, Jean Ponce, Rémi Gribonval, Radu Horaud, Alexandre Gramfort, Emilie Chouzenoux
Medical imagery	Nicolas Ayache, Stanley Durrleman, Rachid Deriche
Robotics	Pierre-Yves Oudeyer, Jean-Baptiste Mouret

### INRIA'S INTERNATIONAL PARTNERSHIPS IN AI

Since 2017, we observe an increase in public policies and national strategies on AI that are issued by national authorities and often include an international dimension. This gives rise to multiple demands. These contacts can thus generate agreements to explore the opportunities and challenges of collaboration, in a top-down approach.

For example, through Inria Chile<sup>40</sup>, the institute is participating in actions and projects in the field of AI or its applications. Inria Chile, in partnership with local institutions, contributes to the definition of Chilean AI policy conducted by the Ministry of Science, Technology, Knowledge and Innovation and the Senate.

In addition, Inria supports international collaborations, in a bottom-up approach, thanks to ad hoc incentives (Inria International Labs, Associated Teams, mobility programmes), which enable Inria to remain responsive to cooperation opportunities.

Finally, as AI advances come largely from the private sector, Inria sometimes chooses to establish collaborations with international industrial players with significant R&D

---

<sup>40</sup> <https://www.inria.fr/fr/centre-inria-chile>

capacities (cf. Inria - Fujitsu long-term research program on AI and big data processing).

In addition to this international watch policy, Inria is currently focusing its collaborative efforts in the field of AI on three geographical areas: bilateral Europe, Asia and North America.

## BILATERAL EUROPE

### **Inria-DFKI Partnership**

Following the Treaty of Aachen of 22 January 2019 signed between Germany and France promoting joint efforts in the field of AI, Inria and the DFKI concluded a memorandum of understanding in January 2020, in which they commit to implement a joint research and innovation programme. This programme covers the areas of AI for industry 4.0, AI for portable technologies, AI and cybersecurity, and human-robot cooperation. The Memorandum of Understanding is also part of a joint commitment within the CLAIRE network.

### **Inria-University College of London partnership**

Signed at the end of 2019, the agreement between Inria and University College of London (UCL) formalizes the collaboration between Inria and UCL. This collaboration is set to grow and expand to include other London partners.

## ASIA

Two countries are now considered to be a priority for the Institute in establishing cooperation in artificial intelligence in Asia: Japan and Singapore.

### **Japan**

Many similarities exist between the Japanese and French (and European) visions of AI: the Japanese "human-centric AI" approach echoes the French strategy's AI for humanity concept, and the secure sharing of data and resources between trusted partners is considered to gain competitiveness.

Furthermore, in both national strategies, the mobility and health sectors are identified as priority sectors for the application of AI. Finally, the two countries also converge on the use of AI to improve productivity, the consideration of environmental issues and the need to train more talent in the field.

In June 2019, Inria signed a four-year Memorandum of Understanding with the Department of Information Technology and Human Factors of the National Institute for Advanced Science and Industrial Science and Technology-AIST, which gathers eight research centres, including the Artificial Intelligence Research Centre (AIRC). This agreement aims to strengthen Inria-AIST cooperation, particularly in the field of

AI and robotics, through the development of scientific exchanges and joint research projects.

## **Singapore**

A cooperation agreement was signed in 2018 between the National University of Singapore (NUS), as operator of the AI Singapore plan, and Inria, the CNRS and INSERM. This agreement aims to promote the development of joint activities in AI and intelligent digital technologies, in the areas of cooperation in AI and Health; explainable AI; federated learning; automatic natural language processing; and confidentiality, security and responsibility in data sharing.

## **NORTH AMERICA**

Following long-term cooperation between the Inria project-teams and North American researchers in the field of AI, the Institute has for several years been formalizing partnerships with highly visible players on the international scene and renowned researchers in the field, mainly in the field of fundamental methods and tools for learning and data analysis.

### **United States**

The Centre for Data Science and Courant Institute of Mathematical Sciences is strongly involved in the New York University - Inria agreement signed in May 2017 for a period of 5 years. The joint programme has made it possible to fund collaborative projects and visits by researchers and doctoral students and the long-term stay of an Inria senior researcher (Jean Ponce).

### **Canada**

Inria and CIFAR (Canadian Institute for Advanced Research) signed an agreement in January 2015, which is currently being renewed. Inria is involved in the "Neural Computing and Adaptive Perception" program, now called "Machine Learning, Biological Learning". This program is co-coordinated by Yann Le Cun (NYU & Facebook) and Yoshua Bengio (Université de Montréal). The WILLOW and SIERRA project-teams participate in the activities of this group. Its main objective is to understand the principles underlying natural and artificial intelligence, and to elucidate the mechanisms by which learning can lead to the emergence of intelligence.

In addition to these two partnerships, five collaborations are supported within the framework of Inria's Associated Teams programme:

- Carnegie Mellon University (GAYA Associate Team on Semantic and Geometric Models for Video Interpretation);
- University of Southern California (LEGO Associate Team on Automatic Language Processing);

- Stanford University (Meta&Co Associate Team on Machine Learning and Automatic Language Processing for Meta-Analysis of Neuro-Cognitive Associations
- and Geomstat Associate Team on algorithmic anatomy - application of learning methods in neuroscience);
- the Argonne National Laboratory (UNIFY Associate Team on AI aspects as a complement to optimize hybrid workflows coupling computationally intensive simulation and massive data analysis).

## LATIN AMERICA

### **Brazil**

Inria and LNCC, the Brazilian National Scientific Computing Laboratory, have a long history of scientific cooperation. A partnership agreement was signed in 2020 on several research fields, including AI.



## 7. INRIA REFERENCES: NUMBERS

Over the 2013-2019 period, Inria researchers published more than **450 AI journal articles and more than 1800 AI conference papers** in the following list of journals and conferences. Indeed, Inria is among the top 20 entities in the 2019 AI Research Ranking. The 2019 edition of the AI Research Ranking analyzed publications at the Annual conference of Neural Information Processing Systems (NeurIPS) and the International Conference on Machine Learning (ICML). Using the 2019 conference proceedings, they went into each of the 2200 accepted papers, compiled the list of authors and their affiliated organizations and released the ranking of the top countries and organizations. Inria comes 16th in the overall ranking of the public research organizations. Only 3 other European public entities appear in the list (Oxford University, ETH and EPFL).

## 8. Other references for further reading

This section contains other references identified to be relevant for further reading, grouped in categories. It does not claim to be exhaustive but simply gives some reading additional to those mentioned in the previous chapters and to the publications of Inria project teams.

### *Generic AI*

One Hundred Year Study on Artificial Intelligence (AI100), Stanford University, August 2016, <https://ai100.stanford.edu>.

AI for humanity. French strategy for AI. <https://www.aiforhumanity.fr/en/>

Alan Turing. *Intelligent Machinery, a Heretical Theory*. *Philosophia Mathematica* (1996) 4 (3): 256-260. Original article from 1951.

Yves Caseau et al., *Renouveau de l'Intelligence artificielle et de l'apprentissage automatique*, Commission technologies de l'information et de la communication, Rapport de l'Académie des technologies, 2018

Ernest Davis and Gary Marcus. *Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence*. *Communications Of The ACM* Vol. 58 No. 9. 2015

Olivier Ezratty, *Les usages de l'intelligence artificielle*, 2020 edition, downloadable at <http://www.oezratty.net/>

Michael A. Goodrich and Alan C. Schultz. *Human–Robot Interaction: A Survey. Foundations and Trends® in Human–Computer Interaction* Vol. 1, No. 3 (2007) 203–275

Jonathan Grudin. *AI and HCI: Two Fields Divided by a Common Focus*. *AI magazine*, 30(4), 48-57. 2008

Kevin Kelly. *The Three Breakthroughs That Have Finally Unleashed AI On The World*. <http://www.wired.com/2014/10/future-of-artificial-intelligence>. 2014

Yang Li, Ranjitha Kumar, Walter S. Lasecki, Otmar Hilliges. *Artificial Intelligence for HCI: A Modern Approach*. *CHI*, 2020.

Pierre Marquis, Odile Papine, Henri Prade (eds). *Panorama de l'Intelligence Artificielle. ses bases méthodologiques, ses développements*. 3 vols. Cepaduc. 2014.

Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Nieves, *The AI Index 2019 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. <http://aima.cs.berkeley.edu/>

Terry Winograd. *Shifting viewpoints: Artificial intelligence and human-computer interaction*. *Artificial Intelligence* 170(18):1256-1258. 2006.

### *Debates about AI*

Dario Amodè, Chris Olah et al, Concrete Problems in AI Safety, arXiv:1606.06565v2, 2016

Ronald C. Arkin. *The Case for Ethical Autonomy in Unmanned Systems*. *Journal of Military Ethics* 12/2010; 9(4)

Anne Bouverot, Thierry Delaporte et al., *Algorithmes : contrôle des biais*, S.V.P., Institut Moutaigne, 2020

Bertrand Braunschweig and Malik Ghallab, editors, *Reflections on AI for Humanity*, book to be published, Springer, 2020

Erik Brynjolfsson, Daniel Rock and Chad Syverson, Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics Working Paper 24001 <http://www.nber.org/papers/w24001>

Samuel Butler. *Erewhon*. Free eBooks at Planet eBook.com, 1872.

Lettre du CICDE N°10. Emploi opérationnel de l'intelligence artificielle. April 2018. <https://www.irsem.fr/data/files/irsem/documents/document/file/2934/20180412-NP-CICDE-Lettre-CICDE-AVRIL-2018.pdf>

Kate Crawford, Roel Dobbe, Theodora Dryer et al. *AI Now 2019 Report*. AI Now Institute, 2019, [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.html](https://ainowinstitute.org/AI_Now_2019_Report.html)

Dominique Cardon. *A quoi rêvent les algorithmes*. Seuil, 2015.

Dominique Cardon, Jean-Philippe Cointet and Antoine Mazières, La revanche des neurones, L'invention des machines inductives et la controverse de l'intelligence artificielle, La Découverte «Réseaux» 2018/5 n° 211, pp 173-220, 2018

Thomas G. Dietterich and Eric J. Horvitz. *Rise of Concerns about AI: Reflections and Directions*. Communications of the ACM | October 2015 Vol. 58 No. 1

Virginia Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, Springer, 2019.

Jessica Fjeld, Nele Achten et al., Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI , <https://cyber.harvard.edu/publication/2020/principled-ai>, 2020

Carl Benedikt Frey and Michael A. Osborne, The future of employment: how susceptible are jobs to computerisation ? , 2013

Malik Ghallab, Responsible AI: Requirements and Challenges, by request to the author, LAAS-CNRS, University of Toulouse, [malik.ghallab@laas.fr](mailto:malik.ghallab@laas.fr), 2020

Thilo Hagendorff. The Ethics of AI Ethics -- An Evaluation of Guidelines. Minds & Machines, 2020.

High Level Expert Group on AI. Ethics guidelines for trustworthy AI. 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, Thomas Dandres. Quantifying the Carbon Emissions of Machine Learning. 2019 <https://arxiv.org/abs/1910.09700>

OECD (2019); Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO); available at <https://www.oecd-ilibrary.org/>  
Stuart Russell. *Human compatible, AI and the problem of control*. Penguin books, 2019.

Roy Schwartz, Jesse Dodge, Noah A. Smith, Oren Etzioni. Green AI. 2019 <https://arxiv.org/abs/1907.10597>

Ion Stoica, Dawn Song, Raluca Ada Popa, David A. Patterson, Michael W. Mahoney, Randy H. Katz, Anthony D. Joseph, Michael Jordan, Joseph M. Hellerstein, Joseph Gonzalez, Ken Goldberg, Ali Ghodsi, David E. Culler and Pieter Abbeel. *A Berkeley View of Systems Challenges for AI*. EECS Department, University of California, Berkeley, 2017.

UNESCO (2019); Preliminary Study on the Ethics of Artificial Intelligence. SHS/COMEST/EXTWG-ETHICS-AI/2019/1; Available on <https://unesdoc.unesco.org/>

Moshe Vardi. *On Lethal Autonomous Weapons*. Communications of the ACM, December 2015 vol. 58 no. 12.

### *Machine learning*

Martin Abadi et al. *Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Software available from tensorflow.org. 2015.

Nicholas Ayache. AI and Healthcare: towards a Digital Twin?. *MCA 2019 - 5th International Symposium on Multidisciplinary Computational Anatomy*, 2019 <https://issuu.com/univ-cotedazur/docs/ayache-ai-summit-2018-vl10-uca>

Alejandro Barredo Arrieta and Natalia Díaz-Rodríguez and Javier Del Ser and Adrien Bennetot and Siham Tabik and Alberto Barbado and Salvador García and Sergio Gil-López and Daniel Molina and Richard Benjamins and Raja Chatila and Francisco Herrera. *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*. Information fusion, 2020.

Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Cléménçon, Florence d'Alché-Buc, et al. , *Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach*, Hal-02506409, 2020

Tarek R. Besold et al., *Neural-Symbolic Learning and Reasoning: a Survey and Interpretation*, arXiv:1711.03902v1, 2017

Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

Léon Bottou: *From machine learning to machine reasoning: an essay*, Machine Learning, 94:133-149, January 2014.

Mathieu Causse, Cameron James, Mohamed Masmoudi and Houcine Turki, *Parsimonious Neural Networks*, Adagos company, 2019

Pedro Domingos. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Penguin books, 2015.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. *A Survey of Methods for Explaining Black Box Models*. ACM Comput. Surv. 2018.

Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, Lalana Kagal. *Explaining Explanations: An Overview of Interpretability of Machine Learning*. 2019 <https://arxiv.org/abs/1806.00069>

Demis Hassabis, Dhharshan Kumaran, Christopher Summerfield and Matthew Botvinick, Neuroscience-Inspired Artificial Intelligence, *Neuron* 95, pp. 245-258, 2017

Michael I. Jordan and Tom M. Mitchell. *Machine learning: Trends, perspectives, and prospects*. *Science*, Vol 349 Issue 6245. 2015.

Peter Kairouz, H. Brendan MacMahan et al., Advances and Open Problems in Federated Learning, arXiv:1912.04977v1, 2019

Nan Rosemary Ke et al., Learning neural causal models from unknown interventions, arXiv:1910.01075v1, 2019

Yann Le Cun. *The Unreasonable Effectiveness of Deep Learning*. Facebook AI Research & Center for Data Science, NYU. <http://yann.lecun.com> , 2015

Yann Le Cun. Quand la machine apprend, La révolution des neurones artificiels et de l'apprentissage profond (French). Odile Jacob, 2019.

Volodymyr Mnih et al. *Human-level control through deep reinforcement learning*. *Nature* 518, 529–533. 2015

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al.. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Microtome Publishing, 2011.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf, Elements of Causal Inference: Foundations and Learning Algorithms, MIT Press, 2017

David Rolnick et al., Tackling Climate Change with Machine Learning, arXiv:1906.05433v1, 2020

Ribana Roscher, Bastian Bohn, Marco F. Duarte, Jochen Garcke. *Explainable Machine Learning for Scientific Insights and Discoveries*. IEEE Access, 2020.

Bernhard Schölkopf, Causality for machine learning, arXiv:1911.10500v1, 2019

Michèle Sebag. *A tour of Machine Learning: an AI perspective*. AI Communications, IOS Press, 2014, 27 (1), pp.11-23.

Thomas Serre. *Deep Learning: The Good, the Bad, and the Ugly*. Annual Reviews, 2019

Emma Strubell Ananya Ganesh Andrew McCallum, Energy and Policy Considerations for Deep Learning in NLP, arXiv:1906.02243v1, 2019

Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz, PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume, arXiv:1709.02371v2, 2017

Neil C. Thompson et al, The Computational Limits of Deep Learning, arXiv:2007.05558v1, 2020

## Vision

Nicholas Ayache. *Des images médicales au patient numérique*, Leçons inaugurales du Collège de France. Collège de France / Fayard, March 2015.

Yasutaka Furukawa, Jean Ponce. *Accurate, Dense, and Robust Multiview Stereopsis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.

Sancho McCann, David G. Lowe. *Efficient Detection for Spatially Local Coding*. Lecture Notes in Computer Science Volume 9008 pp 615-629. 2015.

Farhood Negin, Serhan Cosar, Michal Koperski, François Bremond. *Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition*. Asian conference on pattern recognition (ACPR 2015), 2015.

A. Rosenfeld, R. Zemel, J.K. Tsotsos, *The Elephant in the Room*, 2018 <https://arxiv.org/abs/1808.03305>

Oriol Vinyals, Alexander Toshev, Samy Bengio & Dumitru Erhan. *Show and Tell: A Neural Image Caption Generator*, 2015. <https://arxiv.org/pdf/1502.03044.2015>

## Knowledge representation, semantic web, data

Bettina Berendt, Fabien Gandon, Susan Halford, Wendy Hall, Jim Hendler, Katharina Kinder-Kurlanda, Eirini Ntoutsi, and Steffen Staab. *Web Futures: Inclusive, Intelligent, Sustainable, The 2020 Manifesto for Web Science*, Dagstuhl Manifesto, pp. 1–44, issn:2193-2433 <https://www.webscience.org/wp-content/uploads/sites/117/2020/07/main.pdf>

Tim Berners-Lee, James Hendler and Ora Lassila. *The Semantic Web*. Scientific American, May 2001.

Fabien Gandon. *A Survey of the First 20 Years of Research on Semantic Web and Linked Data*. Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information, Lavoisier, 2018.

Fabien Gandon. *The three 'W' of the World Wide Web call for the three 'M' of a Massively Multidisciplinary Methodology*. Valérie Monfort; Karl-Heinz Krempels. 10th International Conference, WEBIST 2014, Barcelona, Spain. Springer International Publishing, 226, Web Information Systems and Technologies. 2014

Janowicz, K.; Hitzler, P.; Hendler, J.; and van Harmelen, F. *Why the Data Train Needs Semantic Rails*. AI Magazine, 36(5-14). 2015

Antonella Poggi et al. *Linking Data to Ontologies*. Journal on data semantics X Pages 133-173. Springer-Verlag Berlin, Heidelberg. 2008

### *Robotics and self-driving cars*

Safety First for Automated Driving – a new cross-industry white paper, 2019. <https://www.bmwgroup.com/en/company/bmw-group-news/artikel/Safety-First-for-Automated-Driving.html>

Jean-François Bonnefon, Iyad Rahwan, and Azim Shariff. *The social dilemma of autonomous vehicles*. Science (2016), 352 (6293). p. 1573-1576.J.

Antoine Cully, Jeff Clune, Danesh Tarapore & Jean-Baptiste Mouret. *Robots that can adapt like animals*. Nature Vol 521 503-507. 2015.

Ethics Commission of the Federal Ministry of Transport and Digital Infrastructure of Germany, Automated and connected driving report, 2017

Christian Gerdes, Sarah M. Thornton. *Implementable Ethics for Autonomous Vehicles*. Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte. Springer, Berlin. 2015.

Pierre-Yves Oudeyer. *Developmental Robotics*. Encyclopaedia of the Sciences of Learning, N.M. Seel ed., Springer References Series, Springer. 2012.

### *AI and cognition*

Stanislas Dehaene, Apprendre !: Les talents du cerveau, le défi des machines (French). Odile Jacob sciences, 2018



Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes and Adrien Baranes. *Information-seeking, curiosity, and attention: computational and neural Mechanisms*. Trends in Cognitive Science (2013) 1-9. 2013.

Douglas Hofstadter & Emmanuel Sander. *L'analogie, cœur de la pensée*. Ed. Odile Jacob, 2013.

Daniel Kahneman. *Thinking, Fast And Slow*. New York : Farrar, Straus And Giroux, 2011

Luc Steels. *Self-organization and selection in cultural language evolution*. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*, 1 – 37. Amsterdam: John Benjamins. 2012.

### *Natural language, speech, audio*

Daniel Adiwardana et al., Towards a Human-like Open-Domain Chatbot, arXiv:2001.09977v1, 2020

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, et al.. CamemBERT: a Tasty French Language Model. 2019.

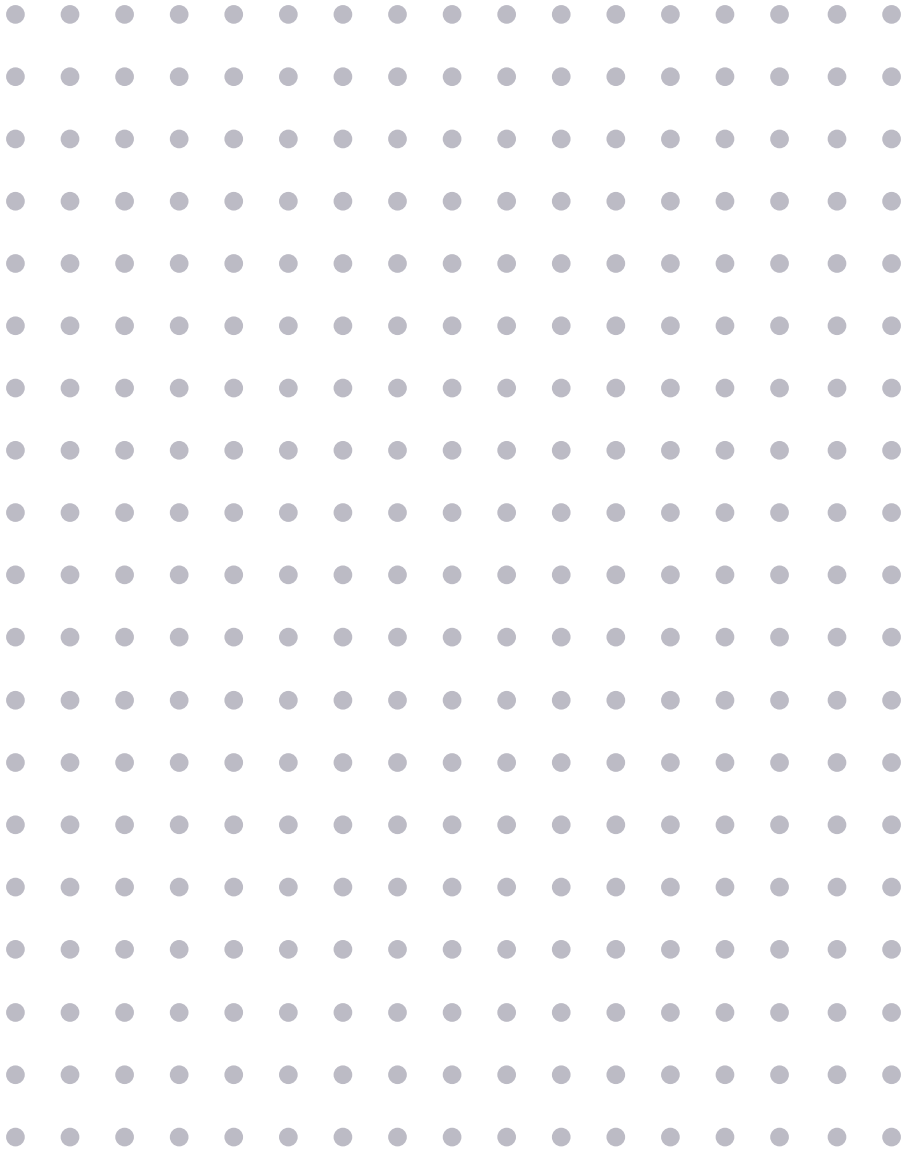
Kenneth Church. *A Pendulum Swung Too Far*. Linguistic Issues in Language Technology – LiLT. Volume 2, Issue 4. 2007

G. Hinton, L. Deng, D. Yu, G.E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, B. Kingsbury, *Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups*. IEEE Signal Processing Magazine, 29(6):82-97, 2012.

Alec Radford, Karthik Narasimhan, Tim Salimans and Ilya Sutskever. *Improving Language Understanding by Generative Pre-Training*. OpenAI, 2018. [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)

Stephen Roller et al., Recipes for building an open-domain chatbot, arXiv:2004.13637v2, 2020

Ashish Vaswani et al., Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, arXiv:1706.03762v5, 2017



*Inria*

Domaine de Voluceau, Rocquencourt BP 105  
78153 Le Chesnay Cedex, France  
Tél. : +33 (0)1 39 63 55 11  
[www.inria.fr](http://www.inria.fr)