

AI and language technology in France

11st LRB meeting — November 25, 2021



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*



National Artificial Intelligence Research Programme (PNRIA)

- The President of the French Republic, Emmanuel Macron, presented an ambitious policy for artificial intelligence (AI) and launched the **national AI strategy** at the closing day of the “AI for Humanity” debate on 29 March 2018.
- This national AI strategy, with a **public budget of EUR 1.5 billion** for five years (2018-2022), is divided into three axes:
 - Achieving the **highest scientific level in AI** by training and attracting the world’s best AI talent
 - **Spreading AI in the economy and society** through startups, public-private partnerships and data sharing
 - Establishing **an ethical framework for AI**

PR[AI]RIE
Paris Artificial Intelligence Research Institute



Objectives of the PNRIA

- Promote **talent attractiveness** and **support programs**
- **Develop AI training** at all levels (initial training, continuing education, and high-level training in master's and doctoral courses)
- Developing state-of-the-art **digital infrastructure**
- Support **AI research** through **dedicated calls** for projects managed by the National Research Agency (ANR)
- Strengthening and encouraging **joint collaboration strategies** at European and international level
- Stimulate **public-private partnerships** and increase the impact of AI research.

Actions carried out in the framework of PNRRIA

- A network of **4 operational Interdisciplinary Artificial Intelligence (3IA) Institutes**: ANITI (Toulouse), 3iA Côte d'Azur (Nice Sophia-Antipolis), MIAI (Grenoble), PRAIRIE (Paris)
- 3 other important centres identified at this stage: SCAI (Sorbonne University), DATA-IA (Paris Saclay), Hi! Paris (Institut Polytechnique de Paris)
- About 190 Chairs in Research and Teaching, **10-15 of which on Language**
- 500 new PhD students in AI per year
- the Jean Zay supercomputer, No. 2 at European level
- + **EUR 100 million** for the ANR AI programmes
- Calls for Projects on International Collaboration with Germany, Japan and Canada
- Industrial and Trusted AI Grand Challenge Program

Language technologies in PNRIA

- Language and language processing technologies are named many times in the official **White Paper** presented at the launch of the programme
- But **they are not a clearly identified** as a strategic axis
- All Interdisciplinary Institutes of Artificial Intelligence (3IA), programme operators, **are led by Inria**, a research operator in the field of technology in France, but these 3IAs do not integrate most of the cutting-edge laboratories in the field of NLP: Inria Rennes, Nancy, Lille and numerous CNRS laboratories
- **Data sharing and reuse** is a particularly well covered topic, for example in the field of health, the Health Data Hub:
You give your blood to help one's neighbour; the same applies to the sharing of your health data" — Gérard Raymond (Health Data Hub vice president)

The 4th “Investments for the future” Programme (PIA 4)

- This programme represents **EUR 20 billion of investment over 5 years in France** in the field of research and innovation
- **Budget doubled** compared to previous programmes: PIA2 (EUR 12 billion) and PIA3 (EUR 10 billion)
- The aim is to “give the economy a medium-term course in order to put it on a virtuous and sustainable path of growth”
- **11 acceleration strategies**, including an “**Artificial Intelligence**” strategy that has not yet been unveiled.
- The acceleration axes are supported by priority research equipment, the first four designated **are led by the CNRS**





The “PIAF” project: For French-speaking AIs



- The ambition of the PIAF project is to build **French-language datasets** for AI in an open and contributory way
- Objective: Produce a **dataset of 25,000 questions and answers in French**
- Given that the data policy is a major focus of the development of artificial intelligence, the Etalab mission (French government) is piloting the establishment of an **interministerial “Lab IA”** (AI Lab), whose mission is to **accelerate the deployment of AI in administrations** via 3 main activities:
 - **Building a core team** to internalise skills and expertise around AI
 - **Supporting AI projects in administrations** through calls for expressions of interest
 - Co-build **shared tools** that can be used as openly as possible
 - More information : <https://piaf.etalab.studio/presentation.html>

Example of a public achievement: the Chatbot Factory

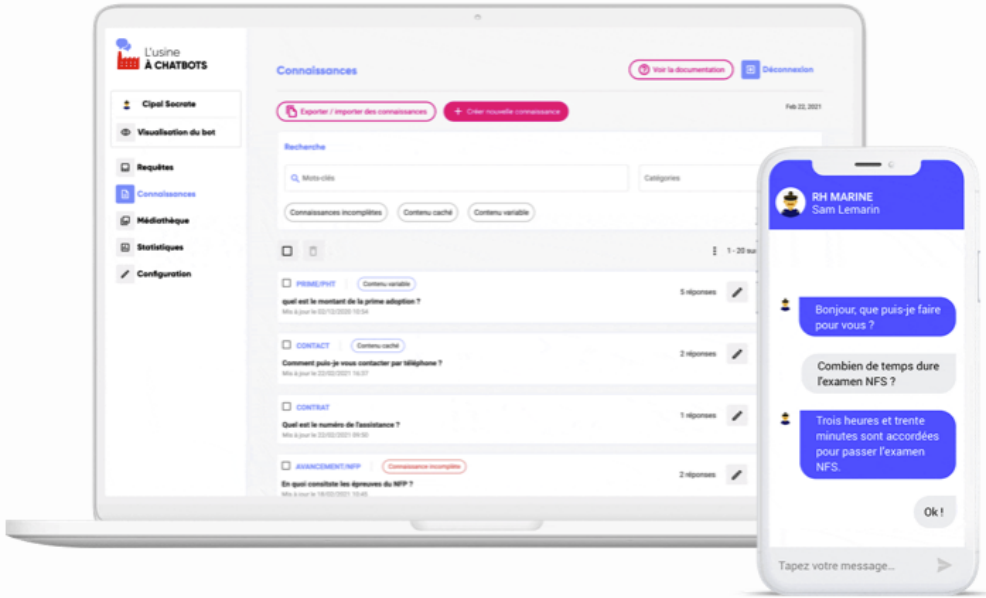
 

[Voir la documentation](#)

Créer facilement votre chatbot

Outils la diffusion d'informations à destination des agents du Ministère.

[Contactez-nous](#) [Créer votre chatbot](#)



This tool is proposed by the digital factory of the Ministry of Armies



Other AI programmes in the field of language

- Big science: develop an open source multilingual language model with huge computing power in just one year (2021-2022)
- Consortium of some **100 researchers** including research laboratories, companies, startups (including Hugging Face as founder) and universities
- Guaranteed access to 5 million hours of calculation on the **Jean Zay supercomputer** (its capacity will double in early 2022, with a total of 3152 nVidia GPUs)
- Objective: offer a service comparable to that provided by OpenAI (GPT-3) or Google (Switch-C) within a year, but with **more open design** and **easy reuse**



Other AI programmes in the field of language

- [The Voice Lab](#): building a **common data ecosystem and standards to enable France and Europe to remain competitive in the global voice market**
- A consortium of more **than 40 companies and startups** from NLP sector and **several research laboratories**
- Objectives: establish a **data marketplace for voice services** and promote the **sharing of voice data** at the French level
- Financed by the Investments for the future (Investissements d'avenir - PIA) and BPI France programmes (**EUR 4.7 million**)

LeVOICELab

**Inventons la
voix française!**

Association des acteurs Français
et Européens de la voice tech.



Towards a new strategy for language in AI in France?

- **Very promising context** for NLP sector, with heavy investments (PNRIA, PIA) and very structuring programmes (Big Science, VoiceLab...)
- Appointment of a **national AI coordinator** in 2020, Mr Renaud Vedel, attached to the Directorate-General for Enterprise, and ongoing work on a **national AI roadmap for language and NLP**
- Opening in 2022 of the **Cité internationale de la langue française et de la Francophonie** in Villers-Cotterêts, with the ambition to be “a laboratory around linguistic technologies” (source: President of the Republic)
- Proposal to set up an **International Centre for Digital and Language Technologies**, that could be the core of AI’s national language strategy, which could open on Villers-Cotterêts site in 2023.



2022 is coming very soon...



- French Presidency of the European Union in the first half of 2022: **multilingualism will be one of the key topics of the Presidency.**
- A major event of the French EU Presidency will be devoted to **multilingualism, technology and innovation**, which will take place in Lille (France) **from 7 to 9 February**: save the date!
(Under Non-disclosure rules until 1st december 2021)
- 2022 is also a year of elections in France :
 - Presidency of the Republic (April)
 - Parliament (June)
- **Changes in strategic directions are therefore possible/likely**

Thank you for your attention



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

Thibault Grouas

ELRC NAP for administrations in France

Thibault.grouas@culture.gouv.fr

Language and digital mission

General delegation for the French language and languages of France