

PROGRAMME INTITUTS ET INITIATIVES
Appel à projet – campagne 2021
Proposition de projet de recherche doctoral (PRD)
SCAI - Sorbonne Center of Artificial Intelligence

Intitulé du projet de recherche doctoral (PRD):

Entrance Patterns and Epistemic Impact of Artificial Intelligence in Medicine

Directeur.rice de thèse porteur.euse du projet (titulaire d'une HDR) :

NOM : DUBOIS **Prénom :** Michel

Titre : Directeur de Recherche ou

e-mail : michel.dubois@cnrs.fr

Adresse professionnelle : 59 - 61 rue Pouchet, 75017, Paris
(site, adresse, bât., bureau)

Unité de Recherche :

Intitulé : GEMASS

Code (ex. UMR xxxx) : UMR8598

École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) : ED433-Concepts et langages

Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1^e inscription et la quotité d'encadrement) : 3 doctorant.e.s (années d'inscription : 2016, 2019, 2020) dont 2 en codirection (codirecteur Patrick Peretti-Watel - quotité 50%, codirectrice Claire Crignon - quotité 25%)

Co-encadrant.e :

NOM : DENIS **Prénom :** Christophe

Titre : Maître de Conférences des Universités ou HDR

e-mail : christophe.denis@lip6.fr

Unité de Recherche :

Intitulé : LIP6

Code (ex. UMR xxxx) : UMR7606

École Doctorale de rattachement : ED130-EDITE
Ou si ED non Alliance SU :



Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1^e inscription et la quotité d'encadrement) : 3 directions de thèse (années d'inscription : 2018, 2020, 2021) avec deux thèses en quotité d'encadrement à 50% et thèse en quotité d'encadrement à 25%

Co-encadrant.e :

NOM : **GARGIULO** Prénom : **Floriana**
Titre : Chargé de Recherche ou HDR
e-mail : Floriana.GARGIULO@cnr.fr

Unité de Recherche :

Intitulé : GEMASS
Code (ex. UMR xxxx) : UMR8598

ED433-Concepts et langages

École Doctorale de rattachement :

Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1^e inscription et la quotité d'encadrement) :

Cotutelle internationale : Non Oui, précisez Pays et Université :

Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?

Non Oui, précisez HuMed - Humanités médicales

Description du projet de recherche doctoral (en français ou en anglais) :

Ce texte sera diffusé en ligne : il ne doit pas excéder 3 pages et est écrit en interligne simple.

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.

Préciser le profil d'étudiant(e) recherché.

1-Context:

In the history of science, several technical and/or methodological paradigms have been able to generate cascades of scientific innovations in apparently disconnected disciplines: it is the case, for example, of the introduction of calculus in mathematics, at the middle of the 17th century, that allowed to formulate the fundamental laws of me-chanics or the developments in electromagnetism at the beginning of the 20th century. More recently, it is also the case for Artificial Intelligence (AI). Since the beginning of 2000s, AI, with its various sub-domains, has experienced an extremely rapid growth in terms of publications. Moreover, AI research technology is getting outside its traditional academic framework, in computer science and mathematics, impacting other independent research fields. In the last decade, in particular, AI provided new innovative tools to face the complexity of medical data, impacting the methodologies for imaging, diagnosis, treatment and predicting outcomes in several clinical situations(1;2;3;4). With the augmentation, in terms of mass and complexity, of clinical data applications to the medical domain attracted several computer scientists, creating a florid interdisciplinary domain named 'Artificial Intelligence in Medicine' (AIM). In 1984, Clancey and Shortliffe provided the following definition (5): 'Medical artificial intelligence is primarily concerned with the construction of AI programs that perform diagnosis and make therapy recommendations. Unlike medical applications based on other programming methods, such as purely statistical and probabilistic methods, medical AI programs are based on symbolic models of disease entities and their relationship to patient factors and clinical manifestations'. The number of publications concerning AI in medical imaging has exponentially increased in several medical domains suggesting a transversely progressive affirmation of such interdisciplinary approach.

2-Objectives:

This project aims to analyze the penetration patterns of artificial intelligence, as a new epistemic approach, in the medical field. Going beyond the previous global approaches regarding both AI and Medicine as a whole, the focus of this project is to realize an in-depth analysis of the internal innovation process generated by the combination of different AI methods on selected fields of medicine. Considering AI sub-domains as exogenous novelties (introduced for the first time from another domain), we will analyze how their introduction in a medical discipline reshapes what Kauffman calls the disciplinary "adjacent possible" (6;7): namely the innovations that could be reached from the combination of the actual knowledge structures. The project focuses on three



separated case studies: neurology, a discipline in which medical imaging plays a central role; cardiology that has an important tradition in time series analysis ; and finally epidemiology, a domain traditionally open to interdisciplinary “contaminations”.

3-Methods:

The project will be conducted with a computational social science approach based on the interplay between quantitative scientometric analysis, modelling and qualitative field work in medical research unit. The project requires large textual database analysis, modelling but also interviews with researchers, at different stages of their career, in three Sorbonne institutes or research units:

- Institut du Cerveau et de la Moelle épinière (UM75)
- Unité de recherche sur les maladies cardiovasculaires, le métabolisme et la nutrition (UMRS 1166)
- Institut Pierre Louis D'épidémiologie Et Santé Publique (UMRS 1136)

Different publication metadata sources will be mobilized for the project. In particular the PubMed APIs will be used to extract the information on all the publications, in the last 20 years. A list of keywords related to AI will be created from the Wikipedia glossary pages, in order to identify the papers using AI tools. This list will be extended and validated through experts' interviews. Further information on the citation patterns of the papers using AI will be retrieved using the Scopus APIs.

4- Project development:

4.1 Entrance Patterns: The first part of the project will be focused on the entrance patterns of AI in the selected case studies with a top-down approach going from a general observation of the diffusion phenomenon to the in-depth analysis of the most important AI tools mobilized for each case study. This first part is organized in three different tasks.

— Task1: Data collection and first survey : The first months of the project will be devoted to the collection and the preparation of scientometric data. At the same time the student will take the first contacts with the selected laboratories and collect the information on the AI tools the researchers and their collaborators use more frequently.

— Task2: Innovation and adoption : After the initial data collection, within a general science study frame-work (8;9;10;11), we will analyze the early adopters of IA tools and their academic and institutional profiles. Following the subsequent diffusion from the early adopters we will try to make a solid hypothesis on the mechanisms driving the adoption of AI methods to the actual widespread of AI in the different domains. We intend to measure the influence of multiple variables : the role of the collaboration networks, of the researchers reputation, of the geographic positioning of the institutions, of the stage of the career, etc.

— Task3: Dynamics : With a sociologically informed approach, mixing text mining techniques and qualitative information from the field work, we will use the citation structure to reconstruct the AI dynamics (12;13) in each medicine domain, namely the network of directed thematic relationships between papers and their AI contents. This representation will allow us to reconstruct a dynamical picture of which sub-domains of AI have been mobilized in each domain.

4.2 Epistemic Impact:

The second part of the project will focus on the impact of AI methods on the internal knowledge structure of the case study disciplines, on the working practices in the laboratories, and on the training of the next generations of researchers.

— Task 4: Field work : Interviews will be conducted with researchers to understand how these new methods are challenging their preexisting practices. We will select researchers at different stages of their careers, to get possibly different feedbacks on the perception of the new methodologies, in terms of potential perspectives, limits and of controversial issues. We will also investigate how the students are exposed to AI in their training. This empirical part is the core of the project: it allows, on one side to give a solid interpretation to the data driven analyses, on the other it will allow to reconstruct, in a detailed way, the local French landscape of AI research in Medicine.

—Task 5: Research strands dynamics: Going back to the metadata analysis driven by the expertise collected on the fieldwork, we will analyze if, thanks to AI methodologies, new research strands start to develop or restart from “sleeping beauties” (14). Making the hypothesis that the AI novelty has largely impacted the “possible adjacent” of the field knowledge, we will investigate how AI themes mix up with the pre-existing semantic structures associated to the field. We will use this complex bipartite structure (AI and medical themes) to identify the emerging strands and their life cycles. We will analyze how these strands enter in competition with the pre-existing ones (15) developing controversial scientific patterns. From these results, compared with the expertise from the fieldwork, we will assess the evolution of the role of AI research strands in the selected medical fields. Considering that the AI innovation enters in competition with the pre-existing approaches, a modelling approach mixing innovation and competition dynamics, will be used to test if some simple mechanisms (hypothesized from the data and the fieldwork expertise) could be identified to as the driving forces reproducing the actual knowledge frameworks.

5- Project outputs

Beside the thesis, for each task (except the data collection) we expect to produce one or two scientific papers, to be published in interdisciplinary journals (like PLoS one or Scientific Reports) or disciplinary journals in Sociology or Medicine. The figure represents the project timeline with the scheduled periods for each task.

6- Institutional aspect and doctoral student profile

This project will be conducted under the supervision of Michel Dubois, Christophe Denis and Floriana Gargiulo. Michel Dubois is a sociologist of science, CNRS Senior Research Fellow, director of the Gemass (UMR8598) and former Deputy director of the interdisciplinary research unit, Epigenetics, Data and Politics (EpiDaPo, UMI 2006). Christophe DENIS is Associate Professor (HDR) at LIP6 (Sorbonne Université) and AI Expert Research Engineer at EDF R&D. His current research activities deal on the explainability and on the ethic of AI based applications (mainly for health care applications). The doctoral student will also work closely with Floriana Gargiulo, Gemass researcher specialized in data mining and social network analysis (16 ; 17; 18). For the position, we will look for a doctoral student with a good computational background, above all concerning data extraction and large-scale network analysis. The PhD will be in sociology and therefore the candidate should be willing to be trained at the interface between qualitative and quantitative approaches.

[1] Ramesh, A. N., et al. “Artificial intelligence in medicine.” *Annals of the Royal College of Surgeons of England* 86.5 (2004): 334.

[2] Szolovits, Peter, ed. *Artificial intelligence in medicine*. Routledge, 2019.

[3] Patel, Vimla L., et al. “The coming of age of artificial intelligence in medicine.” *Artificial intelligence in medicine* 46.1 (2009): 5-17.

[4] Hamet, Pavel, and Johanne Tremblay. “Artificial intelligence in medicine.” *Metabolism* 69 (2017): S36-S40.

[5] Clancey, William J., and Edward H. Shortliffe, eds. *Readings in medical artificial intelligence: the first decade*. Addison-Wesley Longman Publishing Co., Inc., 1984.

[6] Kauffman, S. A. *Investigations* (Oxford University Press, New York/Oxford, 2000).

[7] Tria, F., Loreto, V., Servedio, V. et al. The dynamics of correlated novelties. *Sci Rep* 4, 5890 (2015).

[8] Dubois M, *La nouvelle sociologie des sciences*, Paris, Presses Universitaires de France, 1999

[9] Dubois M., *Introduction à la sociologie des sciences*, Paris, Presses Universitaires de France, 1998

[10] Dubois M., *Social Dynamics of Biomedical Research. An Actionist Perspective on the Sociology of Science*, Oxford, The Bardwell Press, 2012, 216p.

[11] Fortunato, Santo, et al. “Science of science.” *Science* 359.6379 (2018): eaao0185.



- [12] Chavalarias, David, and Jean-Philippe Cointet. "Phylogenetic patterns in science evolution—the rise and fall of scientific fields." *PloS one* 8.2 (2013). 3
- [13] Chinazzi, M., Gonçalves, B., Zhang, Q., Vespignani, A. (2019). Mapping the physics research space: a machine learning approach. *EPJ Data Science*, 8(1), 33.
- [14] Ke, Q., Ferrara, E., Radicchi, F., Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, 112(24), 7426-7431.
- [15] Aleta, A., Meloni, S., Perra, N., Moreno, Y. (2019). Explore with caution: mapping the evolution of scientific interest in physics. *EPJ Data Science*, 8(1), 1-15.
- [16] Gargiulo, Floriana, and Timoteo Carletti. "Driving forces of researchers mobility." *Scientific reports* 4 (2014):4860.
- [17] Floriana Gargiulo, Auguste Caen, Renaud Lambiotte, Timoteo Carletti. "The classical origin of modern mathematics". *EPJ Data Science*, 2016, Volume 5, Number 1
- [18] Gargiulo, Floriana, et al. "Asymmetric participation of defenders and critics of vaccines to debates on french-speaking twitter." *Scientific reports* 10.1 (2020): 1-12.

**Merci d'enregistrer votre fichier au format PDF et de le nommer :
«ACRONYME de l'initiative/institut – AAP 2021 – NOM Porteur.euse Projet »**

***Fichier envoyer simultanément par e-mail à l'ED de rattachement et au programme :
cd_instituts_et_initiatives@listes.upmc.fr avant le 20 février.***