

Allocations doctorales

Espaces Conversationnels Multimodaux

L'université Paris Panthéon Assas recrute en décembre 2022 deux étudiants souhaitant faire une thèse (un en sciences humaines et sociales et l'autre en informatique) pour des allocations doctorales du ministère de l'enseignement supérieur et de la recherche, dans le cadre de son école doctorale EGIC.

Dans le cadre d'une coopération entre le Carism et l'Efrei Research Lab, le projet Espaces Conversationnels Multimodaux (ECM) consiste en la réalisation de deux thèses autour d'un projet commun mené par deux doctorants (PhD Box), l'un issu des sciences humaines et sociales et l'autre des sciences computationnelles. Ce format inédit vise à initier dès le début des carrières des chercheurs une transversalité à la fois des disciplines, des méthodes et des questionnements, le tout encadré par deux laboratoires faisant partie de l'EPEx (établissement public expérimental) Panthéon-Assas et avec le soutien d'un établissement lié par une convention à notre université, l'INA.

Ce projet s'inscrit dans la stratégie de l'université Paris Panthéon-Assas de renforcer l'interdisciplinarité de la recherche en créant de nouvelles synergies au sein de sa communauté académique. Les enseignants-chercheurs encadrant les doctorants se rencontreront autour de ce projet, puis l'élargiront à d'autres collaborations interdisciplinaires. Le projet ECM s'inscrit dans le collège de formation et de recherche (CFR) "Réseaux, médias et information".

1. Présentation du projet

Le débat public a été profondément transformé par l'essor des réseaux sociaux et la possibilité d'échanger, de converser en ligne avec une multitude d'individus, notamment sur des sujets politiques, de société, d'actualité. Pour comprendre ces transformations, de nombreux travaux portant à la fois sur la production et la consommation de contenus sur les plateformes ont été menés depuis une dizaine d'années. Pour autant, il reste de nombreuses zones d'ombre, qui s'expliquent notamment par de forts verrous socio-techniques dans le traitement et l'analyse des bases de données (chiffres, textes, images, vidéo...).

Les travaux étudiant les débats en ligne s'appuient pour l'essentiel soit sur des analyses des productions textuelles (Mercier et Pignard-Cheynel N., 2018), soit sur des ethnographies, s'inspirant notamment de l'analyse des controverses. Ces analyses qualitatives tendent à se concentrer tout particulièrement sur les usagers les plus actifs et les plus visibles, qu'ils s'agissent de professionnels du journalisme (Aubert, 2009; Jeanne-Perrier & Mendès France, 2014), de la communication, de la politique (Boyadjian, 2016; Chibois, 2014) ou simplement d'usagers ayant développé un usage conversationnel hyperactif sur ces plateformes (Gilliotte, 2018). Ceci a pour conséquence de laisser dans l'ombre l'espace des échanges et les réseaux de partage mais aussi la grande majorité des individus qui font circuler des contenus, les commentent, les détournent, en plaisantent, sans en tirer une grande notoriété et tend à exclure les moins visibles, les « immobiles », les « modestes » (Cardon, 2010).

En parallèle, un nombre important de travaux ont profité de l'essor des outils de cartographie pour proposer des analyses de réseaux (et des méthodes d'analyse statistique dérivées) (Azaza et al., 2017; Gérard & Marotte, 2020). Mais là encore, cette perspective se trouve être borgne, car elle met souvent de côté le traitement des conversations et des échanges en ligne. Comment articuler l'étude croisée des messages et des réseaux ? Enfin,

on trouve également de plus en plus de travaux faisant du traitement automatique du langage naturel à partir de larges bases de données (Hellsten & Leydesdorff, 2020; Pak & Paroubek, 2010; Wang et al., 2012). Mais ils se font généralement sans prendre en considération le cadre de l'énonciation, visant notamment à pallier le *context collapse* (Marwick & Boyd, 2010), ou effondrement du contexte d'énonciation, inhérent à ces plateformes. Pourtant, les formes ordinaires et triviales de participation sont une forme de parole politique ou sociale qui ouvre vers des discours non traditionnels ou culturellement moins reconnus, mais qui peuvent être massivement présents, ou représentatifs de positions difficilement visibles. Comment prendre en compte l'ensemble des modalités d'échange et de conversation en ligne ?

L'objectif est de se pencher sur les très nombreux participants à bas bruit du débat en ligne, de comprendre les réseaux dans lesquels ils s'inscrivent, les relations qu'ils établissent et leurs formes de participation. Il s'agira de reconstituer, pour une étude de cas donnée, la construction du débat public en ligne à partir de l'étude d'un ample corpus numérique (corpus de tweets, de commentaires, des pages Facebook, etc.), en considérant les médiations par des dispositifs sociotechniques et l'impact de ces derniers sur la production des messages et la circulation des informations.

L'ampleur des données à recueillir, leur difficile caractérisation, la complexe identification des réseaux et de leurs liens font que ces questions nécessitent d'être explorées du point de vue de l'analyse computationnelle mais aussi des sciences de l'information et de la communication, avec la nécessité de s'appuyer sur différentes méthodes, à la fois qualitatives, quantitatives, computationnelles et mixtes. Il faudra pour les deux doctorants recrutés répondre aux enjeux liés à la structuration des données et à l'analyse d'un important corpus conversationnel, grâce à des méthodes à définir, mais pouvant notamment tenir des techniques de traitement du langage naturel (NLP), de l'analyse automatisée d'image et de vidéo, de l'analyse de réseau, etc.

Ce projet se fera en mode agile, favorisant les allers-retours entre les deux doctorants, confrontant et enrichissant mutuellement les besoins d'analyse exprimés par le doctorant du Carism et les capacités de l'Intelligence Artificielle proposés par le doctorant de l'Efrei Research Lab.

2. Partenaires

Le projet implique trois équipes de l'EPEX : le Carism (centre d'analyse et de recherche interdisciplinaires sur les médias), l'équipe "Données & IA" de l'Efrei Research Lab ainsi que la contribution de l'INA.

Carism

La thématique du débat public et ses transformations liées au développement d'internet sont au cœur des préoccupations du Carism, avec en particulier les travaux reconnus de Romain Badouard, Tristan Mattelart, Cécile Méadel, Arnaud Mercier, parmi d'autres, sur les débats en ligne et les usages du numérique ; par ailleurs le laboratoire est structurellement multidisciplinaire (sociologie, histoire, géopolitique, sciences politiques, sémiologie et économie dialoguent autour des sciences de la communication), ce qui le rend particulièrement apte à mettre en place et accompagner ce PhD Box.

Efrei Research Lab

Les chercheurs de l'équipe "Données & IA" de l'Efrei Research Lab mènent des travaux pour faire face à la transformation digitale, au développement accéléré des publications et la diffusion de l'information sur le Web, les réseaux sociaux et les multivers. Les travaux visent à répondre à des questions comme : où est saisie l'information ? Comment la trouver, l'organiser et la classifier ? Comment vérifier sa valeur et sa qualité ? Comment valider le contenu pour ne traiter que l'information utile ? Où et comment la stocker ? Comment créer de la valeur à partir de données ? Pour répondre à toutes ces questions, l'équipe "Données & IA" est constituée de chercheurs dans les domaines de compétences qui font appel à l'informatique, aux mathématiques et statistiques, à l'analyse et à la modélisation des contenus numériques (texte, image, vidéo, son).

INA

L'INA, qui collabore déjà activement avec le Carism sur des projets de recherches interdisciplinaires, travaille depuis longtemps sur l'extraction et l'analyse de réseaux sociaux, par exemple sur la propagation de nouvelles à

travers les différents médias. Il apportera au projet ECM son expertise ainsi que l'état de ses données et outils. Par la suite, il pourra participer à la valorisation des résultats des recherches.

3. Modalités d'organisation du PhD Box

Chaque laboratoire de recherche assurera dans le cadre de l'école doctorale EGIC le recrutement d'un doctorant. Sera alors constitué un groupe de recherche (ECM) composé par les deux directeurs de thèse, les deux doctorants, le titulaire de la chaire de professeur junior MédiaM, au moins un chercheur de l'INA et tout chercheur désigné par les directeurs et pouvant contribuer au projet.

La définition du périmètre de l'étude de cas choisie sera la première étape de la coordination, le terrain devant à la fois être en cohérence avec les questions posées par les sciences de l'information et de la communication et pertinentes par rapport aux approches informatiques ; aucune discipline ne devant être l'outil de l'autre (l'informatique avec ses outils, l'infocom avec ses représentations des usages). Lors de cette première étape seront définis les besoins des deux côtés, les formats de représentation de données ainsi que les interfaces d'échange de données. Les deux doctorants devront ensuite mener leurs travaux selon les normes de leur propre discipline et en participant aux conférences et publications reconnues de celles-ci.

Le groupe ECM devra mettre en place des réunions au moins semestrielles permettant de présenter les avancées des travaux. Pour assurer le bon fonctionnement en continu et la bonne compréhension des besoins dans les deux disciplines, les deux doctorants devront travailler en mode agile, avec des points réguliers et fréquents au démarrage. Après le recrutement des deux doctorants (temps n), les étapes conjointes prévues sont :

- une communication commune à n+18 mois ;
- un séminaire croisé organisé par ECM sur le thème à n+24 mois
- une publication commune à n+30 mois

Par ailleurs les deux doctorants participeront au projet qui doit être déposé conjointement par le Carism et l'Efrei Research Lab lors de l'appel à projet ANR de 2023.

4. Profil des candidats

1. Profil du candidat en sciences humaines et sociales

Le ou la candidat.e devra disposer d'un diplôme de niveau Master 2 dans une discipline des sciences humaines et sociales, en particulier les sciences de l'information et de la communication, la sociologie, les sciences politiques... avec un goût prononcé pour l'enquête de terrain et l'articulation de différentes techniques d'enquêtes (à la fois entretiens qualitatifs, traitement statistique, analyse de contenu...). Il ou elle pourra s'appuyer sur les compétences et les connaissances d'un autre doctorant (côté Efrei Research Lab) spécialisé dans les sciences computationnelles, de façon à produire conjointement un travail interdisciplinaire de pointe. La connaissance de certains langages de programmation serait un plus, mais n'est pas une condition.

Le dossier de candidature sera composé d'un CV, d'une lettre de motivation, du mémoire de M2, d'un relevé de note des deux années de Master et d'un projet de recherche.

Le projet, de trois pages maximum, sera structuré ainsi :

- Contexte, enjeux de la recherche
- Question de recherche
- Dispositif méthodologique
- Hypothèses

Le ou la candidat.e proposera un terrain de recherche autour d'une thématique définie (événement médiatique, politique, de société) , mais cela peut n'être qu'exploratoire, le sujet proposé pouvant être amené à évoluer en fonction des réflexions des doctorants et des données disponibles

Seront convoqués en audition les candidat.e.s préalablement sélectionnés sur dossier par l'école doctorale EGIC.

La thèse sera dirigée par Cécile Méadel (Professeure des universités, IFP, Université Paris-Panthéon-Assas) et Quentin Gilliotte, titulaire de la chaire de professeur junior MédiaM.

2. Profil du candidat en informatique

- Diplôme : Master 2 (ou équivalent) en informatique.
- Compétences :
 - Solides compétences en informatique : programmation (Python), programmation web, XML, base de données, apprentissage automatique,
 - TALN : base de NLP, outils d'extraction de données, d'exploration et d'annotations de corpus,
 - Des connaissances en analyse d'images/vidéos.
- Langues : bon niveau en anglais,
- Qualités personnelles : bon relationnel pour le travail en équipe, rigueur scientifique, autonomie et esprit d'initiative.

Pour postuler, merci de joindre à votre message un CV et une lettre de motivation, les relevés de notes des deux dernières années d'études. Les lettres de recommandation seraient un plus.

La thèse sera codirigée par Katarzyna Wegrzyn-Wolska et Faten Chaieb-Chakchouk.

5. Planning

- 12 décembre 2022 : fermeture de l'appel à candidature
- janvier 2023 : auditions des candidats retenus et sélection des deux doctorants
- 1er février 2023 : début du contrat

6. Nature du contrat et rémunération

Les deux postes sont des contrats doctoraux signés pour 3 ans, sur la base des allocations doctorales du ministère de l'enseignement supérieur et de la recherche. Les deux doctorants dépendront de l'EGIC (l'école doctorale du Carism et de l'Efrei Research Lab) et devront en respecter [la charte](#).

7. Dépôt des dossiers de candidature

Les dossiers doivent être envoyés (complets) en indiquant en objet du mail "Candidature PhD Box ECM - [Nom du candidat]" aux adresses suivantes :

- Pour le doctorant en SHS (Carism) : cecile.meadel@u-paris2.fr
- Pour le doctorant en Informatique (Efrei Research Lab) : etienne.pernot@efrei.fr

8. Contact et demande d'informations complémentaires

Pour tout renseignement, merci d'envoyer mail en fonction des laboratoires de rattachement :

Contact Carism :

Cécile Méadel : cecile.meadel@u-paris2.fr

Quentin Gilliotte : quentin.gilliotte@gmail.com

Contact Efrei Research Lab :

Etienne Pernot : etienne.pernot@efrei.fr

Katarzyna Wegrzyn-Wolska : katarzyna.wegrzyn@efrei.fr

Faten Chaieb-Chakchouk : faten.chakchouk@efrei.fr

9. Références bibliographiques

- Aubert, A. (2009). Le paradoxe du journalisme participatif. Motivations, compétences et engagements des rédacteurs des nouveaux médias (enquête). *Terrains & travaux*, 15, 171-190.
- Azaza, L., Savonnet, M., Leclercq, É., Kirgizov, S., & Faiz, R. (2017). Évaluation de l'influence polarisée dans un réseau multirelationnel. Application à Twitter. *Document numérique*, 20(1), 67-100. <https://doi.org/10.3166/dn.2017.00003>
- Boyadjian, J. (2016). Les usages politiques différenciés de Twitter. Esquisse d'une typologie des « twittos politiques ». *Politiques de communication*, 6(1), 31-58. <https://doi.org/10.3917/pdc.006.0031>
- Cardon, D. (2010). *La démocratie Internet*. Paris, Seuil.
- Chibois, J. (2014). Twitter et les relations de séduction entre députés et journalistes. La salle des Quatre Colonnes à l'ère des sociabilités numériques. *Réseaux*, 188(6), 201-228. <https://doi.org/10.3917/res.188.0201>
- Gérard, C., & Marotte, G. (2020). #AffaireBenalla : Déconstruction d'une polémique sur le rôle de la communauté Twitter « russophile » dans le débat politique français. *Hérodote*, 177-178(2-3), 125-147. <https://doi.org/10.3917/her.177.0125>
- Gilliotte, Q. (2018). Je tweete donc je suis ? In V. Sacriste (Éd.), *Nos vies, nos objets : Enquêtes sur la vie quotidienne*. Presses universitaires du Septentrion, 169-184. <http://books.openedition.org/septentrion/21002>
- Hellsten, I., & Leydesdorff, L. (2020). Automated analysis of actor–topic networks on twitter : New approaches to the analysis of socio-semantic networks. *Journal of the Association for Information Science and Technology*, 71(1), 3-15. <https://doi.org/10.1002/asi.24207>
- Jeanne-Perrier, V., & Mendès France, T. (2014). Les particularités journalistiques du traitement judiciaire via Twitter. Le cas du procès d'Anders Breivik en Norvège, 2012. *Les Cahiers de la Justice*, 1(1), 131-149. <https://doi.org/10.3917/cdlj.1401.0131>
- Marwick, A. E., & Boyd, D. (2010). I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society*. <https://doi.org/10.1177/1461444810365313>
- Mercier A., Pignard-Cheynel N. (dir.), *#info. Commenter et partager l'actualité sur Twitter et Facebook*, Paris, Éditions de la Maison des sciences de l'homme, 2018.
- Pak, A., & Paroubek, P. (2010, mai). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). LREC 2010, Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. *Proceedings of the ACL 2012 System Demonstrations*, 115-120. <https://aclanthology.org/P12-3020>