



**Journée d'étude, le lundi 14 octobre 2024**  
**"Modèles et apprentissages en sciences humaines et sociales"**  
**MASHS 2024**

Maison des Sciences économiques,  
106-112 Boulevard de l'Hôpital  
75013 Paris

L'objectif de la conférence MASHS 2024 est d'explorer le large panorama des collaborations de recherche entre mathématicien.nes/statisticien.nes et chercheur.es de sciences humaines et sociales (historien.nes, psychologues, gestionnaires, géographes, économistes, sociologues, philosophes, linguistes, musicien.nes et artistes...). Au cours de cette journée, les communications envisagées soulèveront en particulier des problèmes d'actualité et ouverts autour des statistiques appliquées, de l'intelligence artificielle, du text mining etc et **viseront à être compréhensibles par l'ensemble des participant.es**. Nous souhaitons explorer de nouvelles perspectives interdisciplinaires et approfondir la compréhension des défis actuels.

Les rencontres MASHS (Modèles et Apprentissages en Sciences Humaines et Sociales), sont organisées par le laboratoire SAMM à l'université Paris 1 Panthéon Sorbonne tous les deux ans depuis 2012 pour avancer dans le dialogue interdisciplinaire entre mathématiques et sciences humaines et sociales.

La journée démarrera par une conférence sur ce que recouvre l'intelligence artificielle sur un plan mathématique avec une leçon qui se veut accessible puis des intervenantes et intervenants invités expliciteront leurs usages de l'IA et du machine learning dans des domaines variés (politiques, traitement de l'information et des deep fakes, archéologie, etc.).

Nous invitons les chercheurs de tous horizons – historiens, psychologues, gestionnaires, géographes, économistes, sociologues, philosophes, linguistes, musiciens, artistes, etc. – à soumettre leurs travaux, sous forme de posters, afin de créer un cadre stimulant et d'interagir avec des spécialistes de différents domaines.

### **Contributions attendues**

Les contributions viseront à donner à voir la partie statistiques/mathématiques vulgarisée et ses conséquences sur les recherches en SHS loin d'une réflexion qui serait purement technique mais aussi éloignée d'une réflexion philosophique.

Ainsi, les usages de données massives avec des techniques récentes pourront être discutées. Par exemple, en sciences politiques, nous observons une montée de l'analyse de ces données depuis l'analyse de l'affaire Cambridge Analytica qui a fait la une des journaux après l'élection du président Trump. Dans les articles récents, on peut lire l'analyse de la cour suprême américaine (Giansiracusa, 2021) ou des changements dans l'analyse de discours (Do and Al., 2022). Les mêmes données massives et l'apparition des "large language models" a permis, en sciences de l'éducation, d'analyser des blogs enseignants (Epstein & Bourgeois 2018) pour en déduire des informations fines sur la manière dont les enseignants perçoivent leur métier. Par exemple, "wikipedia" était classé dans le champ lexical des "logiciels" pour les enseignants de mathématiques et dans celui des "documents" pour les enseignants d'histoire, ce qui reflétait les usages disciplinaires de wikipédia.

En psychologie, le nombre d'expériences rapportées dans des revues universitaires qui utilisent l'apprentissage automatique comme outil analytique pour compléter l'analyse statistique classique est en croissance (Kosinski et coll., 2013 ; Monaro et coll., 2018 ; Pace et coll., 2019).

L'apprentissage automatique a été appliqué avec succès, par exemple, dans l'analyse de données d'imagerie afin de classer les troubles psychiatriques (Orrù et al., 2012 ; Vieira et al., 2017), en génétique (Libbrecht et Noble, 2015 ; Navarin et Costa, 2017), en médecine clinique (Obermeyer et Emanuel, 2016), en sciences médico-légales (Pace et al., 2019), etc.

Par ailleurs, les paradigmes de recherche ont évolué vers des approches multidisciplinaires. Ainsi le comportement et les mécanismes cognitifs sont étudiés d'un point de vue psychologique conjointement à d'autres perspectives disciplinaires telles que l'environnement, la neurologie... Des exemples de types de données de grande dimension pertinents pour la recherche en psychologie comprennent les domaines d'application suivants : données de neuroimagerie, données provenant de jeux et d'outils d'évaluation numériques utilisés en éducation, données textuelles, ou données provenant d'appareils portables qui permettent de suivre les variables physiologiques et la localisation des personnes.

L'intégration des données de grande dimension dans la recherche en psychologie constitue un défi dans la mesure où peu de théories sont disponibles pour constituer un cadre pour les recherches.

En histoire, l'article d'Oliveira et al (2018) propose une méthode générique basée sur l'apprentissage profond pour la segmentation automatique des documents historiques, utilisant des réseaux convolutifs pour identifier avec précision les différentes parties telles que les titres, paragraphes, images, etc. En parallèle, l'outil "docExtractor" (Monnier, T., & Aubry, M., 2020) se concentre sur l'extraction automatique des éléments clés des documents historiques. Ces approches montrent comment l'intelligence artificielle, en particulier les techniques de deep learning, peut révolutionner le traitement et l'analyse de documents historiques variés. En facilitant la recherche en codicologie, ces technologies ouvrent de nouvelles perspectives interdisciplinaires, permettant aux historiens, archivistes et chercheurs d'autres domaines d'accéder plus facilement à des informations cruciales et de mener des analyses plus approfondies.

Pour faire progresser la recherche transdisciplinaire à l'ère de cette évolution, nous invitons à soumettre des communications centrées sur une application de techniques numériques sur des données de grande dimension (dans le cadre des SHS), en particulier celles utilisant l'intelligence artificielle. Nous encourageons tout particulièrement les jeunes chercheurs à soumettre leurs recherches ou leurs projets de recherche liés à ce sujet.

### **Calendrier et informations pratiques**

- Remise des résumés des posters en moins de 300 mots pour le 15 juillet
- Retours aux contributeurs et contributrices pour le 28 août
- Posters le 14 octobre

La journée se déroulera le lundi 14 octobre à la maison des sciences économiques, 106-112 Boulevard de l'Hôpital, 75013 Paris.

Vous pourrez vous inscrire à la journée via <https://mashs2024.sciencesconf.org/>. La journée sera gratuite. Le nombre de places étant limité, en cas de sur-inscription, nous privilégierons les personnes ayant proposé des travaux.

### **Présentation des organisateurs**

Le SAMM est une unité de recherche (UR 4543 de l'Université Paris 1 Panthéon Sorbonne et FR 2036 CNRS) ayant un positionnement transversal en mathématiques appliquées, couvrant des sujets d'apprentissage statistique et d'apprentissage machine, d'analyse, d'optimisation, de contrôle, de modélisation et analyse stochastique. Le SAMM souhaite avoir un rôle de moteur dans le développement des interactions avec les SHS. De plus, l'unité se positionne comme référent interne à Paris 1 pour l'expertise en Intelligence Artificielle (IA), notamment dans le cadre de la mise en place de l'Observatoire de l'IA qui regroupe également des chercheurs issus des SHS et des disciplines juridiques.

### **Le comité d'organisation est composé de**

Jean-Marc Bardet, SAMM, Paris 1 Panthéon-Sorbonne

Marie Cottrell, SAMM, Paris 1 Panthéon-Sorbonne

Muriel Epstein, EMA, CYU, associée au SAMM, Paris 1 Panthéon-Sorbonne

Dominique Haughton, Bentley University, associée au SAMM, Paris 1 Panthéon-Sorbonne

Jean-Charles Lamirel, Loria, Université de Strasbourg, associé au SAMM Paris 1 Panthéon-Sorbonne

Fabien Navarro, SAMM, Paris 1 Panthéon-Sorbonne

### **Le comité scientifique est composé de**

Frédéric Aschehoug, Institut de Psychologie Université Paris Cité

Muriel Epstein, CY Cergy Paris Université

Dominique Haughton, Bentley University

Stéphane Lamassé, Paris 1 Panthéon-Sorbonne

Jean-Charles Lamirel, Université de Strasbourg

Fabien Navarro, Paris 1 Panthéon-Sorbonne

Etienne Ollion, CNRS, école polytechnique

Alexandre Steyer, Paris 1 Panthéon-Sorbonne

## **REFERENCES BIBLIOGRAPHIQUES**

Alerini J., M. Olteanu, J. Ridgway, (2017) "Markov and the Duchy of Savoy : segmenting a century with regime-switching models", 2017, Journal de la Société Française de Statistique, Vol. 158(2), p.89-117.

Bhaduri, M., Dery, J., Haughton, D., Li, T., Liu, P., & Nan, W. (2021). Dynamics of Reward-Based Crowdfunding: Kickstarter Rock Music Projects. *Chance*, 34(1), 4-13.

Bourgeois N., Cottrell M., Lamasse S., Olteanu M., (2015) Search for meaning through the study of co-occurrences in texts, in I. Rojas, G. Joya and A. Catala (eds.), *Advances in Computational Intelligence, IWANN 2015, Part II*, Lecture Notes in Computer Science, 9095, p. 578-591, Springer-Verlag, Majorque, Spain, June 10-12, 2015.

Deichmann, J. I., Haughton, D., Mingfei, L. I., & Heyao, W. A. N. G. (2022). Does European Union Membership Result In Quality-Of-Life Convergence?. *Regional Science Inquiry*, 14(1), 31-46.

Do, S., Ollion, E., & Shen, R. (2022, August 1). The Augmented Social Scientist. Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. <https://doi.org/10.31235/osf.io/3fkzc>

Epstein M. et Bourgeois N. (2018) , «A statistical analysis of French teachers' blogs: beyond institutional perspectives », *French Journal For Media Research* [en ligne], Le web 2.0 : lieux de perception des transformations des sociétés. <https://www.frenchjournalformediaresearch.com:443/lodel-1.0/main/index.php?id=1642>.

Giansiracusa, N., & Giansiracusa, N. (2021). Deepfake Deception: What to Trust When Seeing Is No Longer Believing. *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More*, 41-66.

Giansiracusa, N. (2021). An evolutionary view of the US supreme court. *Mathematical and Computational Applications*, 26(2), 37.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. United States* 110, 5802–5805.

Libbrecht, M. W. et Noble, W. S. (2015). Applications de l'apprentissage automatique en génétique et génomique. *Révérend Genet.* 16, 321–332. DOI : 10.1038/NRG3920

Monaro, M., Galante, C., Spolaor, R., Li, Q. Q., Gamberini, L., Conti, M., et al. (2018). Détection de mensonges cachés à l'aide de la dynamique du clavier. *Rapports scientifiques* 8 :1976

Monnier, T., & Aubry, M. (2020). docExtractor: An off-the-shelf historical document element extraction. In 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR).

- Navarin, N. et Costa, F. (2017). Une méthode efficace de noyau graphique pour la prédiction fonctionnelle de l'ARN non codant. *Bioinformatique* 33, 2642 à 2650.
- Obermeyer, Z., et Emanuel, E. J. (2016). Prédire l'avenir : mégadonnées, apprentissage automatique et médecine clinique. *N. Engl. J. Med.* 375, 1216–1219.
- Oliveira, S. A., Seguin, B., & Kaplan, F. (2018). dhSegment: A generic deep-learning approach for document segmentation. In 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR).
- Olteanu M., Hazan A., Cottrell M., Randon-Furling J.. (2019) Multidimensional urban segregation: toward a neural network measure. *Neural Comput & Applic.* <https://doi.org/10.1007/s00521-019-04199-5>
- Olteanu M., J. Alerini, (2019)“Quelques réflexions sur la périodisation en histoire”, 2019, In G. Bonnot, S. Lamassé (Ed.), Dans les dédales du web : Historiens en territoires numériques, Editions de la Sorbonne,p.57-85.
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using a support vector machine to identify imaging biomarkers of neurological and psychiatric diseases: a critical review. *Neurosci. Biobehav. Revelation* 36, 1140-1152.
- Pace, G., Orrù, G., Merylin, M., Francesca, G., Roberta, V., Boone, K. B., et al. (2019). La simulation de la détection des troubles cognitifs avec le test B est renforcée par l'apprentissage automatique. *Front. Psychol.* 10 :1650.
- Soleiman, S., Randon-Furling, J., Cottrell, M. (2022). Machine Learning and Data-Driven Approaches in Spatial Statistics: A Case Study of Housing Price Estimation. In: Faigl, J., Olteanu, M., Drchal, J. (eds) *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization. WSOM+ 2022. Lecture Notes in Networks and Systems*, vol 533. P.31-40 Springer, Cham. [https://doi.org/10.1007/978-3-031-15444-7\\_4](https://doi.org/10.1007/978-3-031-15444-7_4)
- Vieira, S., Pinaya, H. et Mechelli, A. (2017). Utilisation de l'apprentissage profond pour étudier les corrélats de neuroimagerie des troubles psychiatriques et neurologiques : méthodes et applications. *Neurosci. Biobehav. Rev.* 74 (partie A), 58-75.