

---

# Intelligence artificielle, science de la donnée, néo-positivisme numérique et stratégies de recherche

Elisa Grandi\*<sup>1</sup>, Torsten Hiltmann\*<sup>2</sup>, and Dominique Stutzmann\*<sup>3</sup>

<sup>1</sup>Université Paris Cité – Université Paris Cité – France

<sup>2</sup>Humboldt Universität zu Berlin – Allemagne

<sup>3</sup>Institut de recherche et d'histoire des textes – Centre National de la Recherche Scientifique – France

## Résumé

### Pour information

L'argumentaire ci-dessous a été rédigé par les organisateurs de la conférence pour introduire la table ronde et non par les intervenant:e:s listé:e:s ci-dessus. L'argumentaire n'engage donc pas leur responsabilité scientifique.

### Argumentaire

La production massive de données et leur exploitation à grande échelle permet aux historiens et autres humanistes, quel que soit le domaine considéré, un discours scientifique argumenté et fondé sur la "science de la donnée". Faisant primer le document sur la théorie, ou, comme le formule O. Poncet au sujet de l'École des chartes, "se plaç(ant) à la tête d'un combat pour une histoire plus scientifique au contact des sources", la science des données qui propose une vue plus claire des corrélations fait naître ce qui a été appelé un "néo-positivisme numérique" (Mosco 2014). Elle n'est pourtant ni neutre du point de vue axiologique, ni toujours menée de façon correcte pour le domaine historique. Les modélisations et calculs mathématiques oblitèrent une réalité historique plus complexe, des biais de source où les échantillonnages et les corpus de taille restreinte font toujours le risque que "sous l'apparence de rigueur et de scientificité que donnent les calculs de corrélation se trouvent des données chiffrées non probantes, au regard des standards d'administration de la preuve en histoire, y compris en histoire quantitative" tel qu'il a été formulé récemment dans une critique cinglante d'un article de data science (Anglaret et al., 2021).

À chaque étape du traitement des données, depuis la formalisation ("capta" plutôt que "data" selon J. Drucker, 2011) jusqu'à la production des éléments probants et au commentaire de graphiques, le risque est grand de surinterpréter ou de transformer. Quelle place ou espace peuvent prendre les approches théoriques et les réflexions ou sensibilités des chercheur.e.s dans un discours historique fondé sur la donnée de la science et comment formuler l'approche hypothético-déductive dans le discours historique ?

L'intelligence artificielle, et en particulier l'apprentissage automatique, ajoute une couche d'obscurité potentielle dans le traitement des données, en particulier par le biais d'apprentissage difficile à maîtriser. Que faire des données produites par les traitements automatiques quand il est impossible de rendre compte du processus de décision ? Quels changements d'approche

---

\*Intervenant

imposent-ils ou occasionnent-ils ?

Outre l'administration de la preuve, des changements interviennent dans l'ordre du discours et de l'évaluation. Comment les historien.ne.s peuvent apprendre à travailler et à argumenter avec des résultats de traitements dont la performance est évaluée (par exemple, avec un algorithme dont la performance est évaluée sur un autre corpus à 80%) ? Les mesures habituelles du domaine des sciences dures (précision, rappel, F1-score) peuvent-elles aider les historien.ne.s à comprendre les données et la réalité du monde ? Les chercheur.e.s en humanités sont tenté.e.s d'affirmer la primauté de corpus maîtrisés à 100 %, mais peut-être jamais parfaitement saisis, et d'oublier les conséquences des erreurs ou lacunes de dépouillements. Comment considérer l'apport et les risques de l'intelligence artificielle dans ce contexte ?

### **Intervenants**

Elisa Grandi est maîtresse de conférences en histoire économique contemporaine à l'université Paris Cité. Outre ses recherches sur l'histoire économique, les relations internationales et les programmes de développement, ses travaux concernent les méthodes en humanités numériques, notamment la gestion des données massives, l'analyse des réseaux sociaux et le traitement numérique de sources d'archives. Elle participe à l'ANR *Exploitation de Big Data Historiques pour les Humanités Numériques* et au projet européen *Eurhisfirm " Long term data for Europe "* (<https://eurhisfirm.eu/>)

Torsten Hiltmann est professeur de *digital history* à la Humboldt-Universität de Berlin, et aussi spécialiste de l'histoire de la communication visuelle au Moyen Âge. Ses recherches et publications portent sur l'impact de la datafication en histoire et les conséquences épistémologiques des changements de médialité.

Dominique Stutzmann est directeur de recherche au CNRS. Spécialiste de l'histoire de l'écriture au Moyen Âge, et faisant porter ses recherches sur des grands corpus issus de transcription automatique, il explore les conséquences des traitements automatiques et de la massification dans les sciences de l'érudition.