

## CONFÉRENCE D'OUVERTURE

### **STATISTIQUE ET INTELLIGENCE ARTIFICIELLE : ET SI L'ON REMETTAIT LA DÉCISION AU CENTRE ?**

*Christine CHOIRAT (\*)*

*(\*) Data Science Competence Center (cheffe de la section Data Science & Artificial Intelligence), Office fédéral de la statistique (Suisse)  
Professeure invitée au Global Studies Institute (GSI) de l'Université de Genève.*

---

L'intelligence artificielle (IA) et l'apprentissage automatique transforment progressivement la statistique publique, en facilitant certaines tâches, en améliorant les prédictions et en ouvrant de nouvelles perspectives pour l'évaluation et le pilotage des politiques publiques.

Cependant, au-delà de ces avancées technologiques, une question centrale demeure : comment garantir que ces outils soutiennent efficacement la prise de décision, plutôt que de l'obscurcir ou de la biaiser ?

Cette présentation propose de replacer la décision au cœur de la réflexion sur l'usage de l'IA en statistique publique. Nous explorerons les composantes essentielles des systèmes d'IA : l'importance des données en entrée, le rôle des modèles dans la production des résultats (prédictions, recommandations, décisions automatisées), et les implications de ces systèmes sur la gouvernance des données et des infrastructures. Nous aborderons les défis clés auxquels font face les instituts de statistique :

- la qualité et la représentativité des données, ainsi que l'enrichissement par les métadonnées,
- le choix entre solutions internes ou externes (*open source* versus boîtes noires propriétaires),
- la gestion des risques de confidentialité et de divulgation,
- les enjeux de souveraineté liés aux infrastructures numériques (ex. initiatives comme Onyxia),
- l'évaluation des modèles IA : comment garantir leur performance, leur équité et leur robustesse avant leur mise en production, et comment éviter la reproduction ou l'amplification des biais existants dans les données ?

À travers des exemples concrets, notamment l'utilisation des grands modèles de langage (LLM) et une exploration approfondie de Lomas, notre plateforme open-source pour l'analyse confidentielle des données, nous illustrerons comment ces défis peuvent être relevés. Nous montrerons notamment comment les nouvelles méthodes de contrôle du risque de divulgation permettent d'exploiter le potentiel de l'IA tout en garantissant une utilisation responsable et équitable des données.