

Cursus	Sem.	Type
Sciences humaines et sociales	MA2	Obl.
UNIL - Autres facultés	E	Obl.
UNIL - HEC	E	Obl.

Langue d'enseignement	français
Crédits	3
Retrait	Non autorisé
Session	Eté
Semestre	Printemps
Examen	Pendant le semestre
Charge	90h
Semaines	14
Heures	3 hebdo
Projet	3 hebdo
Nombre de places	30

It is not allowed to withdraw from this subject after the registration deadline.

Résumé

Le cours "Critical Data Studies" s'inscrit dans la nouvelle offre d'enseignements UNIL/EPFL qui propose de croiser des savoirs provenant des SHS et des sciences de l'ingénieur afin d'aborder des thématiques complexes qui nécessitent une méthodologie interdisciplinaire.

Contenu

Une donnée n'est jamais neutre. Elle s'inscrit toujours dans un « assemblage de données » qui dépend d'une multitude de facteurs contextuels, d'agents humains, d'institutions, ou encore de déterminations techniques. Et pourtant, les données, aujourd'hui traitées massivement dans le domaine social, culturel, autant que dans celui de la recherche, sont souvent perçues comme des entités objectives, permettant de limiter les biais de l'interprétation subjective. Cela dit, les débats contemporains autour de la reconnaissance faciale ou de l'intelligence artificielle nous montrent que les données massives peuvent elles-mêmes produire des résultats problématiques, qui dépendent souvent de leur usage en société et des intentions des acteurs qui les emploient, les étudient, ou les commercialisent. Dès lors, quels sont les a priori à partir desquels nous tirons des informations et construisons des savoirs à partir de jeux de données ? Comment comprendre cette notion englobante qui qualifie aujourd'hui des informations provenant de l'astrophysique, d'interactions sociales, autant que de sources textuelles ?

Le cours UNIL/EPFL « Critical Data Studies » s'attache à mettre en lumière les défis méthodologiques autant que l'ensemble des biais qui peuvent découler de l'étude de données considérées, a priori, comme « objectives » et naturelles.

Durant la première partie du semestre, les étudiants se familiarisent avec un écosystème numérique permettant le traitement, l'analyse et le partage de données massives (environnement Renku) tout en découvrant les travaux issus du champ contemporain des « critical data studies » qui portent sur « les défis culturels, éthiques et critiques que posent les données massives » (Iliadis et Russo 2016).

Le cours, organisé autour de projets interdisciplinaires, proposera de retracer la chaîne des données telle qu'elle apparaît dans les dispositifs numériques contemporains (acquisition, transformation, modélisation, visualisation, etc.), afin de questionner chacune de ces étapes à partir d'outils critiques et réflexifs. Qu'est-ce que le traitement quantitatif fait aux données ? Comment sont-elles transformées lors de leur collecte ? Comment concilier la prise en compte du contexte et le traitement automatisé de larges ensembles de données ? A partir de quels présupposés leur donne-t-on du sens ? Quelle est l'histoire de la notion de "données" ?

Compétences requises

Cours prérequis indicatifs

Une formation préalable dans le langage de programmation Python ou R est recommandée. Une formation

courte (mise à niveau) sera disponible pendant les séances d'atelier qui suivent le cours, chaque semaine.

Des connaissances de base en programmation (Python ou R) et/ou en statistiques sont bienvenues.

Acquis de formation

A la fin de ce cours l'étudiant doit être capable de:

- Appliquer des outils de data science
- Analyser une étude de cas
- Critiquer les hypothèses des modèles

Compétences transversales

- Accéder aux sources d'informations appropriées et les évaluer.
- Ecrire un rapport scientifique ou technique.
- Recueillir des données.
- Planifier des actions et les mener à bien de façon à faire un usage optimal du temps et des ressources à disposition.
- Dialoguer avec des professionnels d'autres disciplines.
- Faire preuve d'esprit critique
- Recevoir du feedback (une critique) et y répondre de manière appropriée.

Méthode d'enseignement

Le deuxième semestre sera dédié à la réalisation de projets interdisciplinaires menés par les groupes mixtes d'étudiants (UNIL/EPFL). A l'exception de la première semaine du semestre et de trois séances « milestone », le travail sur le projet s'effectuera hors-classe. Chaque équipe sera accompagnée par un superviseur qui se tiendra à leur disposition pour répondre à leurs questions et les guider dans leur travail.

Méthode d'évaluation

L'évaluation du cours porte sur les éléments suivants :

- Soumission d'un "data acquisition plan" en début de deuxième semestre (20%)
- Note finale du projet de groupe (deuxième semestre) 80%

Encadrement

Office hours	Oui
Assistants	Oui
Forum électronique	Oui

Ressources

Bibliographie

O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. College & Research Libraries (Vol. 78). <https://doi.org/10.5860/crl.78.3.403>

Salganik, M. J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press. Open review edition.

Gold, M., & Klein, L. (Eds.). (2019). Debates in the Digital Humanities 2019. Minneapolis; London: University of Minnesota Press. doi:10.5749/j.ctvg251hk

Strasser, B. (2019) Collecting Experiments: Making Big Data Biology, The university of Chicago Press.

Leonelli, S. (2019) La recherche scientifique à l'ère des Big Data: cinq façons dont les données massive nuisent à la science, et comment la sauver. Éditions Mimésis, pp. 122. ISBN: 978-8869761843.

Ressources en bibliothèque

- [Salganik, M. J. \(2017\). Bit by Bit](#)
- [Strasser, B. \(2019\) Collecting Experiments](#)
- [O'Neil, C. \(2017\). Weapons of Math Destruction](#)
- [Leonelli, S. \(2019\) La recherche scientifique à l'ère des Big Data](#)
- [Gold, M., & Klein, L. \(Eds.\). \(2019\). Debates in the Digital Humanities 2019](#)

Sites web

- <https://renkulab.io/cdc-2020>