PHI 1961 - Éthique, responsabilité sociale et IA
Hiver 2027

Un algorithme peut-il être injuste ? Quels éléments de l'actualité de l'IA méritent une analyse éthique? Faut-il craindre une superintelligence, ou plutôt ce que les systèmes actuels font déjà à nos emplois, à notre vie privée et à nos démocraties ? Et devons-nous quelque chose aux machines elles-mêmes ?

Ce cours est une introduction aux enjeux éthiques et sociaux des systèmes d'intelligence artificielle : algorithmes de recommandation, de prédiction ou de conduite, et IA génératives comme ChatGPT, Gemini ou Claude. Il ne suppose aucune formation préalable en philosophie ou informatique.

Les premières séances fournissent les outils : notions de base en éthique et en métaéthique, types d'IA, problème de l'alignement. Nous examinons ensuite, cas concrets à l'appui, une série de domaines d'éthique appliquée à l'IA: enjeux démocratiques et internationaux, discriminations, surveillance et vie privée, remplacement du travail humain, éthique des GML, statut moral des IA, environnement, animaux.

La lecture de départ est le livre de Martin Gibert *Faire la morale aux robots* ; les autres textes et vidéos sont déposés sur StudiUM. Le fil conducteur de la session est un projet personnel : un rapport d'éthique de 1000 à 1500 mots sur un cas de votre choix. Une première version est relue et commentée par vos pairs, réunis en comités, avant la remise de la version finale. L'évaluation repose sur un examen oral de mi-session (35 %), la relecture par les pairs (10 %), le rapport final (30 %) et un examen oral final (25 %).