Cours PHI8281 | Éthique | UQAM
Jumelé au Cours PHI901B | Séminaire en philosophie pratique I | UQAM : Éthique
Moralité et intelligence artificielle : fondements, implémentation et évaluation
Enseignant : Dominic Martin
Lundi de 14 h à 17 h

Moralité et intelligence artificielle : Fondements, implémentation et évaluation

Un nombre croissant de systèmes d'intelligence artificielle (IA) prennent des décisions complexes sur le plan moral avec beaucoup d'autonomie. Plusieurs villes à l'échelle internationale tentent de déployer des services de taxis autonomes; la plupart des grandes entreprises de technologies développent des modèles génératifs à usage général comme l'agent conversationnel ChatGPT; et beaucoup d'autres organisations essaient d'intégrer des solutions technologiques pour interagir avec leurs employés, le public et automatiser d'autres tâches. La prise de décision morale est aussi une composante essentielle des robots militaires, des robots médicaux et plusieurs autres systèmes.

Le principal objectif de ce séminaire est de réfléchir aux enjeux soulevés par ces nouvelles technologies. Nous essaierons de comprendre, dans un premier temps, la nature d'un agent moral artificiel et comment cette idée est apparue en réponse au problème du contrôle : le risque de perdre le contrôle d'un système d'IA et la difficulté de l'aligner avec les objectifs poursuivis.

Dans un deuxième temps, nous nous pencherons sur les enjeux soulevés par l'implémentation de la moralité artificielle dans un système. Nous essaierons de clarifier les distinctions, les forces et les faiblesses entre les grandes approches pour amener un système d'IA à prendre des décisions plus morales. Nous étudierons le cas de ChatGPT plus spécifiquement et d'autres grands modèles de langage à usage général.

Finalement, on ne peut développer un système d'IA plus moral si on ne peut calibrer ce système et comparer ses performances avec d'autres systèmes. Or, le problème de l'évaluation de l'alignement moral d'un système d'IA soulève des enjeux philosophiques considérables. Nous nous pencherons sur ces enjeux à la fois théoriques et pratiques dans un troisième temps.

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

Les infractions et les sanctions possibles reliées à ces infractions sont précisées aux articles 2 et 3 du Règlement no 18 sur les infractions de nature académique.

Vous pouvez également consulter des capsules vidéos sur le site r18.uqam.ca. Celles-ci vous en apprendront davantage sur l'intégrité académique et le R18, tout en vous orientant vers les ressources mises à votre disposition par l'UQAM pour vous aider à éliminer le plagiat de vos travaux.

INFO SPHÈRE

Infosphère est l'un de ces outils indispensables : un guide méthodologique visant à promouvoir les bonnes pratiques en matière de recherche documentaire et de rédaction de travaux. Cet outil vous accompagnera tout au long de vos études et vous permettra d'éviter les pièges du plagiat. Cliquez sur le logo à gauche pour être redirigé vers le site.

Politique nº 16 visant à prévenir et à combattre le sexisme et les violences à caractère sexuel

Le harcèlement sexuel se définit comme étant un comportement à connotation sexuelle unilatéral et non désiré et consiste en une pression indue exercée sur une personne, soit pour obtenir des faveurs sexuelles, soit pour ridiculiser ses caractéristiques sexuelles.

La Politique nº 16 identifie, notamment, les comportements suivants comme des violences à caractère sexuel :

- la production ou la diffusion d'images ou de vidéos sexuelles explicites et dégradantes, sans motif pédagogique, de recherche, de création ou d'autres fins publiques légitimes;
- les avances verbales ou propositions insistantes à caractère sexuel non désirées;
- la manifestation abusive et non désirée d'intérêt amoureux ou sexuel;
- les commentaires, les allusions, les plaisanteries, les interpellations ou les insultes à caractère sexuel, devant ou en l'absence de la personne visée;
- les actes de voyeurisme ou d'exhibitionnisme;
- le (cyber) harcèlement sexuel;
- la production, la possession ou la diffusion d'images ou de vidéos sexuelles d'une personne sans son consentement;
- les avances non verbales, telles que les avances physiques, les attouchements, les frôlements, les pincements, les baisers non désirés;
- l'agression sexuelle ou la menace d'agression sexuelle;

- l'imposition d'une intimité sexuelle non voulue;
- les promesses de récompense ou les menaces de représailles, implicites ou explicites, liées à la satisfaction ou à la non-satisfaction d'une demande à caractère sexuel.

Pour plus d'information :
https://instances.uqam.ca/wp-content/uploads/sites/47/2019/04/Politique_no_16_2.pdf

Pour obtenir du soutien :

Pour rencontrer une personne ou faire un signalement :
Bureau d'intervention et de prévention en matière de harcèlement
514 987-3000, poste 0886
Pour la liste des services offerts en matière de violence sexuelle à l'UQAM et à l'extérieur de l'UQAM :
harcelement.uqam.ca

CALACS Trêve pour Elles – point de services UQAM :
514 987-0348
calacs@uqam.ca
trevepourelles.org

Service de soutien psychologique (Services à la vie étudiante) :
514 987-3185
Local DS-2110

Service de la prévention et de la sécurité :
514 987-3131

Politique no 44 d'accueil et de soutien des étudiantes, étudiants en situation de handicap

Par sa politique, l'Université reconnait, en toute égalité des chances, sans discrimination ni privilège, aux étudiantes, étudiants en situation de handicap, le droit de bénéficier de l'ensemble des ressources du campus et de la communauté universitaire, afin d'assurer la réussite de leurs projets d'études, et ce, dans les meilleures conditions possibles. L'exercice de ce droit est, par ailleurs, tributaire du cadre réglementaire régissant l'ensemble des activités de l'Université.

Il incombe aux étudiantes, étudiants en situation de handicap de rencontrer les intervenantes, intervenants (conseillères, conseillers à l'accueil et à l'intégration du Service d'accueil et de soutien des étudiantes, étudiants en situation de handicap, professeures, professeurs, chargées de cours, chargés de cours, direction de programmes, associations étudiantes concernées, etc.) qui pourront faciliter leur intégration à la communauté universitaire ou les assister et les soutenir dans la résolution de problèmes particuliers en lien avec les limitations entraînées par leur déficience.

Le Service d'accueil et de soutien aux étudiantes, étudiants en situation de handicap (SASESH) offre des mesures d'aménagement dont peuvent bénéficier certains étudiants. Nous vous recommandons fortement de vous prévaloir des services auxquels vous pourriez avoir droit afin de réussir vos études, sans discrimination. Pour plus d'information, visitez le site de ce service à l'adresse suivante : http://vie-etudiante.uqam.ca/etudiant-situation-handicap/nouvelles-ressources.html et celui de la politique institutionnelle d'accueil et de soutien aux étudiantes, étudiants en situation de handicap :
https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_44.pdf

Vous devez faire connaitre votre situation au SASESH le plus tôt possible :

En personne : 1290, rue Saint-Denis, Pavillon Saint-Denis, local AB-2300
Par téléphone : 514 987-3148
Courriel : situation.handicap@uqam.ca
En ligne : http://vie-etudiante.uqam.ca/