SÉMINAIRE DE RECHERCHE EN ÉTHIQUE

Éthique de l'intelligence artificielle

Code : PHI8281/PHI9023
Session : Automne 2019
Horaire : Mardi, 14h00-17h00
Heures de bureau : sur rendez-vous

Professeur : Mauro Rossi
Bureau : W-5410
Téléphone : 514-987-3000 Poste 7807
Courriel : rossi.mauro@uqam.ca

DESCRIPTION SELON L'ANNUAIRE

Recherches relatives à un secteur ou un problème particulier de l'éthique. Ce secteur ou ce problème, qui sera déterminé par le professeur, peut concerner soit l'analyse du discours moral (la méta-éthique), soit les fondements de l'éthique normative, soit encore un domaine ou l'autre de l'éthique appliquée (bioéthique, éthique des affaires, éthique environnementale...).

DESCRIPTION DU SÉMINAIRE

Ce séminaire est une introduction à l'éthique de l'intelligence artificielle. Nous considèrerons les questions suivantes: À quelles conditions l'utilisation d'algorithmes « apprenants » dans la prise de décisions politiques et économiques est-elle moralement légitime ? Est-il moralement permissible d'utiliser des armes de guerre autonomes ? Quelle sorte de règles morales devraient suivre les voitures autonomes ? Est-il moralement permissible d'avoir des rapports sexuels avec des robots ? Est-il possible d'avoir des relations amoureuses ou d'amitié avec eux ? Quels enjeux éthiques soulève l'utilisation de robots dans le domaine de la santé ? Les robots peuvent-ils être des agents moraux ? Devraient-ils avoir des droits ? Le développement de l'intelligence artificielle pose-t-il un risque existentiel pour l'humanité ? Comment pouvons-nous aligner les objectifs de systèmes autonomes super-intelligents aux nôtres ?

PLAN PROVISOIRE

1) Introduction (03/09/2019)

- Bostrom, N. and Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. In W.M. Ramsey and K. Frankish (eds.). *The Cambridge Handboom of Artificial Intelligence*, pp. 316-344. Cambridge: Cambridge University Press.
- Müller, V.C. (2019). Ethics of AI and Robotics. Forthcoming in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*.
- Cardon, D. *et al.* (2018). La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux*, 5 (211): 173-219.

2) Algorithmes et prise de décision (10/09/2019)

Lectures obligatoires

- Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy and Technology*, 31 (4): 543-556.
- Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology*, 29 (3): 245-268.

Lectures complémentaires

- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.
- Mittelstadt, B. *et al.* (2016). The Ethics of Algorithms: Mapping the Debate, *Big Data & Society*, 3 (2): 1-21.
- Diakopoulos, N. (2016). Accountability in Algorithmic Decision-Making. *Communications of the ACM*, 59 (2), pp. 56-62. New York: ACM.
- Kroll, J. *et al.* (2017) Accountable Algorithms. *University of Pennsylvania Law Review*, 165: 633-705.

3) L'éthique des armes autonomes (17/09/2019)

Lectures obligatoires

- Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24 (1): 62-77.
- Purves et al. (2015). Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 18 (4): 851-872.

Lectures complémentaires

- Danaher, J. (2016). Robots, Law and the Retribution-Gap. *Ethics and Information Technology*, 18 (4): 299-309.
- Muller, V.C. (2016). Autonomous Killer Robots Are Probably Good News. In E. Di Nucci and F. Santoni de Sio (eds.), *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons*, pp. 67-81. London: Ashgate.
- Simpson, T.W. and Müller, V.C. (2016). Just War and Robot's Killings. *Philosophical Quarterly*, 66 (263): 302-322.
- Burri, S. (2019). What Is the Moral Problem with Killer Robots? Forthcoming in B.J. Strawser, M. Robillard, and R. Jenkins (eds.). Oxford: Oxford University Press.

4) L'éthique des voitures autonomes (24/09/2019)

Lectures obligatoires

- Gogoll, J. and Müller, J.F. (2017). Autonomous Cars: In Favor of a Mandatory Ethics Setting. *Science and Engineering Ethics*, 23 (3): 681-700.
- Hevelke, A. and Nida-Rümelin, J. (2015). Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Science and Engineering Ethics*, 21 (3): 619-630.

- Nyholm, S. (2018). Attributing Agency to Automated Systems: Reflections on Human-Robot Collaboration and Responsibility-Loci. *Science and Engineering Ethics*, 24 (4): 1201-1219.

Lectures complémentaires

- Nyholm, S. (2018). The Ethics of Crashes with Self-Driving Cars: A Roadmap I-II. *Philosophy Compass*, 13 (7): e12507.
- Nyholm, S. and Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory and Moral Practice*, 19 (5): 1275-1289.
- Sparrow, R. and Howard, M. (2017). When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport, *Transportation Research Part* C, 80: 206-215.
- Himmelreich, J. (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory and Moral Practice*, 21 (3): 669-684.
- Nyholm, S. and Smids, J. (2019). Automated Cars Meet Human Drivers: Responsible Human-Robot Coordination and the Ethics of Mixed Traffic. Forthcoming in *Ethics and Information Technology*.

5) Les robots et les soins de santé (01/10/2019)

Lectures obligatoires

- Sharkey A. (2014). Robots and Human Dignity: A Consideration of the Effects of Robot Care on the Dignity of Older People. *Ethics and Information Technology*, 16 (1): 63-75.
- Sparrow, R. (2015). Robots in Aged Care: A Dystopian Future? *AI and Society*, 31 (4): 445-454.

Lectures complémentaires

- Sparrow, R. and Sparrow, L. (2006). In the Hands of Machines? The Future of Aged Care. *Minds and Machines*, 16 (2): 141-161.
- Coeckelbergh, M. (2012). "How I Learned to Love the Robot": Capabilities, Information Technologies, and Elderly Care. In I. Oosterlaken & J. van den Hoven (eds.), *The Capability Approach, Technology and Design*, pp. 77-86. Dordrecht: Springer.
- Sharkey A. and Sharkey N. (2012). Granny and the Robots: Ethical Issues in Robot Care for the Elderly. *Ethics and Information Technology*, 14 (1): 27-40.
- Sorell, T. and Draper, H. (2014) Robot Carers, Ethics, and Older People. *Ethics and Information Technology*, 16 (3): 183-195.
- van Wynsberghe, A. (2016). *Healthcare Robots: Ethics, Design and Implementation*. London: Routledge.

6) Sexe, amour et amitié avec les robots (08/10/2019)

Lectures obligatoires

- Danaher, J. (2017). The Symbolic-Consequences Argument in the Sex Robots Debate. In J. Danaher and N. McArthur (eds.), *Robot Sex: Social and Ethical Implications*, pp. 103-131. Cambridge, MA: MIT Press.

- Danaher, J. (2019). The Philosophical Case for Robot Friendship. Forthcoming in *Journal of Posthuman Studies*.

Lectures complémentaires

- Turkle, S. (2017), *Alone Together* (2nd ed.). London: Basic Books.
- Nyholm, S. and Frank, L.E. (2017). From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible? In J. Danaher and N. McArthur (eds.), *Robot Sex: Social and Ethical Implications*, pp. 219-243. Cambridge, MA: MIT Press.
- Eskens, R. (2017). Is Sex With Robots Rape? *Journal of Practical Ethics*, 5 (2): 62-76.
- Devlin, K. (2018). *Turned On: Science, Sex and Robots*. London: Bloomsbury Press.
- Danaher, J. (2019). Building Better Sex Robots: Lessons from Feminist Pornography. In Yuefang Zhou and Martin Fischer (eds.), *AI Love You - Developments on Human-Robot Intimate Relations*, pp. 133-148. Dordrecht: Springer.

7) Les droits des robots (15/10/2019)

Lectures obligatoires

- Sparrow, R. (2004). The Turing Triage Test. *Ethics and Information Technology*, 6 (4): 203-213.
- Schwitzgebel, E. and Garza, M. (2015). A defense of the Rights of Artificial Intelligences. *Midwest Studies in Philosophy*, 39 (1): 89-119.

Lectures complémentaires

- Bryson, J. J. (2010). Robots should be Slaves. In Y. Wilks (ed.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, pp. 63-74. Amsterdam: John Benjamins Publishing Company.
- Neely, E. L. (2014). Machines and the Moral Community. *Philosophy & Technology*, 27 (1): 97-111.
- Gunkel, D. (2018). *Robot Rights*. Cambridge, MA: MIT Press.
- Turner, J. (2019). Chapitre 4 de *Robot Rules: Regulating Artificial Intelligence*. Berlin: Springer.
- Danaher, J. (2019). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. Forthcoming in *Science and Engineering Ethics*.

8) Semaine de lecture (22/10/2019)

9) IA et agentivité morale (29/10/2019)

Lectures obligatoires

- Savulescu, J. and Maslen, H. (2015). Moral Enhancement and Artificial Intelligence: Moral AI? *Beyond Artificial Intelligence*, pp. 79-95. Dordrecht: Springer.
- Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology*, 28 (1): 107-124.

Lectures complémentaires

- Giubilini, A. and Savulescu, J. (2018). The Artificial Moral Advisor. *Philosophy & Technology*, 31 (2): 169-188.

- Lara, F. and Deckers, J. (2019). Artificial Intelligence as a Socratic Assistant for Moral Enhancement. Forthcoming in *Neuroethics*.
- Danaher, J. (2019). The Rise of the Robots and the Crisis of Moral Patiency. *AI & Society*, 34 (1): 129-136.
- Scheutz, M. (2016). The Need for Moral Competency in Autonomous Agent Architectures. In V.C. Müller (ed.), *Fundamental Issues in Artificial Intelligence*, pp. 517-527. Dordrecht: Springer
- van Wynsberghe, A. and Robbins, S. (2019). Critiquing the Reasons for Making Artificial Moral Agents. Forthcoming in *Science and Engineering Ethics*.

10) Les conséquences de l'automation sur le travail (05/11/2019)

Lectures obligatoires

- Celentano, D. (2019). Automation, Labour Justice, and Equality. *Ethics and Social Welfare*, 13 (1): 33-50.
- Danaher, J. (2017). Will Life Be Worth Living in a World Without Work? Technological Unemployment and the Meaning of Life. *Science and Engineering Ethics*, 23 (1): 41-64.

Lectures complémentaires

- Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton and Co.
- Ford, M. (2015). *The Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.
- Autor, D. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29 (3): 3-30.
- Loi, M. (2015). Technological Unemployment and Human Disenhancement. *Ethics and Information Technology*, 17 (3): 201-210.
- Korinek, A. and Stiglitz, J. (2017). Artificial Intelligence and its Implications for Income Distribution and Unemployment. National Bureau of Economic Research, Inc. http://www.nber.org/papers/w24174

11) Pas de cours (12/11/2019)

12) La singularité (19/11/2019)

- Chalmers, D. (2010). The Singularity: A Philosophical Analysis. *The Journal of Consciousness Studies*, 17 (9-1): 7-65.

13) Les pouvoirs et les risques des super-intelligences (26/11/2019)

- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press, chapters 6-8.

14) Le problème de l'alignement des valeurs (03/12/2019)

- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press, chapters 12-14.

15) Séance de discussion et de rattrapage (10/12/2019)

LECTURES

Toutes les lectures obligatoires sont disponibles sur la page moodle du cours.

ÉVALUATION

Modalités

La note finale du séminaire se basera sur :

1. La participation (présence et interventions) au séminaire (10%) ;
2. Deux présentations en classe (40%) ;
3. Une dissertation finale de 5000 mots (50%), à remettre sur moodle au plus tard le **mardi 17 décembre 2019** à 23h59.

Critères

1. Compréhension des textes de référence : exactitude de l'identification des thèses, arguments et concepts dans les textes étudiés. (25%)
2. Sélection des éléments de discussion : pertinence et suffisance des éléments retenus. (25%)
3. Partie critique : présentation et discussion des critiques que l'étudiant juge pertinentes. (20%)
4. Formulation : maitrise de l'usage des catégories d'analyse, clarté et précision terminologique. (15%)
5. Organisation du texte: structure, enchainement des raisonnements et cohérence. (15%)

Retards

Pour des raisons d'équité, aucun retard non justifié ne sera toléré. Une pénalité de -2% par jour (ouvrable) de retard sera appliquée.

PLAGIAT

Règlement no 18 sur les infractions de nature académique

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par un.e étudiant.e, 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.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes ;
- l'utilisation totale ou partielle du texte d'autrui en le faisant passer pour sien ou sans indication de référence ;
- la transmission d'un travail aux fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis aux fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant ;
- l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée ;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé ;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle ;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances ;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-création, un rapport de stage ou un rapport de recherche.

Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18 sur les infractions de nature académique.

Vous pouvez consulter sur le site r18.uqam.ca des capsules vidéos qui 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.

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.

INFO SPHÈRE

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/