LABORATOIRE SPHERE, SCIENCES, HISTOIRE, PHILOSOPHIE, UMR 7219

Séminaire

PHiS IA : Philosophie, Histoire et Sociologie de l'Intelligence Artificielle

PHiS IA : Philosophy, History & Sociology of Artificial Intelligence

http://www.sphere.univ-paris-diderot.fr/spip.php?article2385

Université de Paris (Diderot), salle 631B, bâtiment Condorcet, 4, rue Elsa Morante, 75013 Paris

2019 – 2020

PRÉSENTATION

L'objet de ce séminaire est de se positionner au centre des transformations scientifiques, épistémologiques, technologiques et sociétales que les développements de l'intelligence artificielle engagent.

L'intelligence artificielle relève d'une histoire qui commence officiellement avec la création de cette dénomination en 1957 à la conférence de Dartmouth ; conférence qui formalise la discipline dans ses objectifs tant épistémologiques que performatifs. Elle relève aussi d'une « préhistoire », dès le XIXe siècle, avec l'apparition de la machine de Babbage et les premiers programmes, jusqu'à l'émergence, après-guerre, des premiers ordinateurs entièrement électroniques, avec mémoire intégrée. L'étude de cette histoire et de cette préhistoire permet de comprendre comment les enjeux et les controverses liés à cette discipline se sont construits : enjeux historiquement scientifiques et méthodologiques dans le cadre de la concurrence entre l'approche connexionniste et l'approche symbolique de l'esprit humain ; enjeux technologiques, à la fois contingents au développement et au déploiement de capacités calculatoires nouvelles, et liés à la capacité à performer des fonctions dévolues à la cognition ; enfin enjeux éthiques dans le dépassement de l'homme dans ses activités cognitives et dans la redéfinition des fonctions humaines. L'IA possède donc ce double attribut d'être à la fois un accessoire de connaissance dans la science de la cognition et une technologie dont la finalité est sans cesse réévaluée à l'aune de ses performances.

L'IA contemporaine, le machine learning avec les réseaux d'apprentissage profond basés sur des modèles simulant les réseaux de neurones biologiques, entre autres exemples, ont transformé profondément cette discipline en ce qu'elle devient un moyen, chaque jour plus efficace, pour réaliser certaines tâches autrefois réservées aux humains. En tant que technologie, elle s'insère de plus en plus au sein des systèmes socioéconomiques, en particulier dans l'organisation des entreprises, et devient un nouvel acteur socioculturel en tant, par exemple, qu'outil du pronostic médical ou support de la gestion des dépenses énergétiques. Cette montée en puissance pose de nombreuses questions, en l'occurrence en termes éthiques, écologiques, politiques, et philosophiques. En tant qu'accessoire de connaissance, elle interroge le statut des théories scientifiques, les modalités épistémologiques nécessaires à maintenir la stabilité de l'édifice de nos connaissances, ainsi que le statut des réseaux de neurones dans le cadre de la modélisation de l'esprit humain et la définition même de ce que nous appelons « intelligence ».

Le séminaire que nous proposons a donc pour objet l'étude de l'IA dans sa dualité technologique/ scientifique, à travers ses aspects historique, philosophique, épistémologique, éthique et sociologique. De façon à comprendre les enjeux modernes de cette discipline, nous l'envisagerons ici comme définie dans son historicité, mais aussi en tant qu'outil scientifique, outil de connaissance et de transformation des connaissances, de compréhension, enfin en tant qu'entité socioculturelle en interaction avec son environnement, que ce dernier soit humain ou non-humain.

Séminaire mensuel proposé par **Benoît Duchemann** (Univ. Paris Diderot, SPHERE), **Jean-Pierre Llored** (chercheur associé à l'UMR SPHERE ; enseignant-chercheur en sciences humaines et sociales à l'Ecole Centrale de Casablanca, Maroc ; visiting teaching fellow au département de philosophie, Université de Bristol, Royaume-Uni ; visiting scholar au Linacre College, Université d'Oxford, Royaume-Uni) et **Jean-Jacques Szczeciniarz** (HPS, Univ. Paris Diderot, SPHERE)

PROGRAMME

Université de Paris (Diderot), salle 631B, bâtiment Condorcet, 4, rue Elsa Morante, 75013 Paris
un jeudi / mois, de 14h à 17h

17/10/2019 SÉANCE INAUGURALE
Jean-Gabriel Ganascia (Pr d'IA, Paris 6 & LIP6) : *Limitations logiques, épistémologiques et éthiques de l'apprentissage machine*

21/11 TRAITEMENT SYMBOLIQUE DU LANGAGE ET NLP CONNEXIONISTE
Laure Soulier (Pr d'IA, Paris 6 & LIP6) : *Le symbolique au service du connexionnisme et vice-versa : apprentissage de représentation augmenté, extraction d'information et bases de connaissances*
Juan Luis Gastaldi (ETH Zürich, SPHERE) : *La revanche connexionniste : le succès des vecteurs des mots dans le traitement automatique du langage*

19/12 HISTOIRE DE L'IA !! ! séance reportée au 16/04/2020, 13:30–16:30 !!

30/01/2020 PROGRÈS, IA ET ÉTHIQUE
Vincent C. Müller (Pr de philosophie et d'éthique, Eindhoven Univ. of Technology)
Sonia Desmoulin-Canselier (Laboratoire Droit et Changement Social, UMR 6297 CNRS, Université de Nantes) : *La place de l'humain dans un monde algorithmé : questionnements juridiques sur les conditions du déploiement des systèmes algorithmiques d'aide à la décision*

20/02 PHILOSOPHIE DE L'IA : PERCEPTION, COGNITION ET IA
Gunnar Declerck (Mcf, UTC)
Jean-Michel Roy (Pr, ENS Lyon)

19/03 LOGIQUE ET PREUVE EN IA / IA ET DÉLIBÉRATION
Jean-Jacques Szczeciniarz (Pre, SPHERE, Université de Paris (Diderot))
Jean Sallantin (Pre, LIRMM, Univ. de Montpellier) : *Le débat public : un défi à l'IA*

02/04 IA POUR LA SOCIOLOGIE ET SOCIOLOGIE DE L'IA !! ! 13:30–16:30 !!
Francis Chateauraynaud (Dre, GSPR, EHESS) : *L'IA entre expérimentation cognitive, promesse technologique et problème public. Un regard pragmatiste*
Olessia Kirtchik (Doc en sociologie, EHESS, Paris, & chercheure à l'Institut d'études historiques et théoriques en sciences humaines Poletayev, l'Université Nationale de Recherche "Haute École d'Économie", Moscou) : *L'objet insaisissable pour une sociologie de l'IA : critiques et alternatives*

16/04 HISTOIRE DE L'IA !! 13:30–16:30 !!
Dominique Cardon (Pr de sociologie, Sciences Po & Media Lab) : *La revanche des neurones – la controverse de l'intelligence artificielle*
Frédéric Fürst (Mcf en IA, Université de Picardie & MIS) : *Intelligence artificielle et Astronautique : utopies parallèles*

07/05 IA, VISUALISATION ET MODÈLES
Jean-Daniel Fekete (Dre, Inria) : *Visualisation pour l'analyse de réseaux dynamiques et de données en hautes dimensions*
Franck Varenne (Mcf, ERIAC, Univ. de Rouen, & IHPST)
Une contribution de l'épistémologie des modèles à l'explicabilité des algorithmes

PROGRAMME DÉTAILLÉ

Jeudi 17 octobre 2019
SÉANCE INAUGURALE

Jean-Gabriel Ganascia (Pr d'IA, Université Paris-6 & LIP6)
Limitations logiques, épistémologiques et éthiques de l'apprentissage machine
L'écho grandissant que reçoivent les masses de données (Big Data en anglais) et l'apprentissage profond (Deep Learning en anglais) depuis quelques années, masquent mal leurs limitations. On les pare de toutes les vertus au point de laisser entendre qu'elles donneront accès à une « intelligence artificielle forte » qui transformera l'humanité. Pourtant, si ces techniques apportent et apporteront beaucoup dans un grand nombre de secteurs, par exemple dans le domaine médical, pour aider à diagnostiquer des maladies, ou dans le champ social, pour faire de la prédiction et rationaliser certains choix, elles reposent sur l'induction, c'est-à-dire sur le raisonnement qui va du particulier au général. En conséquence, elles sont soumises aux limitations logiques de toute induction que nous tâcherons de rappeler. De plus, comme nous le montrerons, ces techniques permettent de détecter des corrélations qui ne correspondent pas toutes à des relations de causalités, et qui s'avèrent parfois trompeuses, contrairement à ce qu'affirment les tenants des masses de données. Enfin, nous verrons que l'emploi abusif de procédures de décision fondées sur l'apprentissage machine peut avoir des effets prédateurs sur la société, car loin d'être objectifs, les choix qu'elles engagent comportent des biais générateurs d'injustices.

Jeudi 21 novembre 2019
TRAITEMENT SYMBOLIQUE DU LANGAGE ET NLP CONNEXIONISTE

Laure Soulier (Pr d'IA, Paris 6 & LIP6)
Le symbolique au service du connexionnisme et vice-versa : apprentissage de représentation augmenté, extraction d'information et bases de connaissances
Donner du sens aux mots, aux phrases, aux documents... Que ce soit pour la construction d'un moteur de recherche, pour la traduction automatique, ou encore les systèmes de questions-réponses, la compréhension des textes est un enjeu prépondérant en informatique. Les données textuelles ont en premier lieu été représentées sous la forme d'un "sac-de-mots", représentatif des mots présents/absents et facilement manipulable. Cependant, cette structure perd totalement la notion d'ordre des mots, de structure grammaticale et manque cruellement de sémantique. Une des premières réponses à cette limite a été de considérer les bases de connaissances, structurées en triplet de deux entités et d'une relation, permettant ainsi d'améliorer la compréhension des textes grâce à la "sémantique relationnelle". Plus récemment, les avancées en réseaux de neurones (et plus particulièrement l'apprentissage de représentation a émergé comme une nouvelle solution pour capturer la sémantique des mots grâce à l'hypothèse distributionnelle conditionnant la représentation des mots à être similaire si les mots apparaissent dans un même contexte. Dans cet exposé, nous présenterons dans un premier temps les deux approches de représentation sémantique (relationnelle et distributionnelle), leurs avantages mais aussi leurs limites. En second lieu, nous explorerons la combinaison de ces deux approches ; débouchant d'une part sur de l'apprentissage de représentation augmenté/ancré par rapport aux bases de connaissances, et d'autre part sur l'exploitation des réseaux de neurones pour l'extraction d'information et la construction des bases de connaissances.

Juan Luis Gastaldi (ETH Zürich, SPHere)
La revanche connexionniste : le succès des vecteurs des mots dans le traitement automatique du langage
Si l'efficacité des méthodes symboliques dans le traitement automatique du langage naturel a rendu plus difficile l'arrivée des techniques d'apprentissage profond dans ce domaine, le succès grandissant des modèles basés sur des vecteurs de mots (*word embeddings*) a été l'occasion d'une adoption massive de ces nouvelles techniques connexionistes. Dans cet exposé, il s'agira de revenir sur l'émergence de ces modèles, pour essayer de comprendre les raisons de leur succès au moyen de la conception du langage qu'ils mettent en œuvre.

Jeudi 19 décembre 2019 [séance reportée au 16 avril 2020]
HISTOIRE DE L'IA

Jeudi 30 janvier 2020
PROGRÈS, IA ET ÉTHIQUE

Vincent C. Müller (Pr de philosophie et d'éthique, Eindhoven Univ. of Technology)
Is it time for robot rights ?
I will investigate the suggestions some authors have made, that we should allocate rights to robots, or to artificial moral agents (AMA) that are able to make complex moral decisions and act upon them. I find little reasons to even consider robot rights now, even if one accepts the assumptions made (some of the issue is the confusion of machine ethics with AI ethics). On a more constructive note, I take a look at the notion of 'agency' ('moral status', 'responsibility', 'sentience') in the light of the discussion of embodied and extended cognition. It appears that agency is too easily allocated in the human case, so we should be careful in the case of artificial agents. In AI we should not aim for agency at all, but just for causing the orchestration of intelligent behaviour. If we understand the aims of AI in this way we see that they have no relation to moral status.

Sonia Desmoulin-Canselier (Laboratoire Droit et Changement Social, UMR 6297 CNRS, Université de Nantes)
La place de l'humain dans un monde algorithmé : questionnements juridiques sur les conditions du déploiement des systèmes algorithmiques d'aide à la décision
L'innovation et la décision au XXIe siècle sont placées sous le double signe de la collecte massive de données et du traitement algorithmique. Dans tous les secteurs de la vie économique, juridique et sociale, on attend de la « mine d'or » du big data des pistes de recherches innovantes, des informations inédites, des prévisions et des solutions « individualisées » ou « personnalisées ». De telles perspectives ne peuvent toutefois se réaliser sans le recours à de puissants outils informatiques, pour stocker les données, mais surtout pour les traiter et pour effectuer des calculs. Ces systèmes automatisés permettant de formuler une réponse (ou une proposition de réponse) à une question ou un problème donné sont des systèmes algorithmiques. Dans le secteur médical, les solutions d'aide à la décision se sont multipliées : depuis l'aide au diagnostic (analyse des images médicales et détection des signaux morbides, probabilité diagnostique, examens complémentaires) jusqu'à la prescription médicamenteuse (risque allergique, incompatibilités entre traitements, posologie) en passant par la prédiction (calcul du risque de récidive). A présent, c'est en matière de police (calcul de risque sur une zone géographique ou de risque de récidive pour une personne), d'administration (détermination des affectations dans l'enseignement supérieur, attribution d'une autorisation) et de justice (calcul des prestations compensatoire ou des dommages-intérêts) que les algorithmes se déploient. Les enjeux du recours à de tels outils sont nombreux. Leur complexité et leur fonctionnement peuvent mettre en difficulté l'autonomie du décideur et le respect des droits du destinataire de la décision. A partir de quelques exemples, nous étudierons les réponses du droit en vigueur en Europe et en France et nous nous interrogerons sur les évolutions possibles et nécessaires pour faire face à ces enjeux.

Jeudi 20 février
PHILOSOPHIE DE L'IA : PERCEPTION, COGNITION ET IA

Gunnar Declerck (Mcf, UTC) : *Formaliser et mécaniser la cognition. Quelques critiques et comment les désamorcer*

Jean-Michel Roy (Pr, ENS Lyon)

Jeudi 19 mars
LOGIQUE ET PREUVE EN IA / IA ET DÉLIBÉRATION

J.-J. Szczeciniarz (Pre, SPHERE, Université Paris-Diderot)

Jean Sallantin (Pre, LIRMM, Univ. de Montpellier)
Le débat public : un défi à l'IA
Nous sommes à un moment critique de l'histoire où les modèles qui fondent la gouvernance de notre société sont contestés à toutes les échelles depuis la commune jusqu'aux règles des échanges commerciaux mondialisés de la planète.
En effet, les citoyens - peut-être parce qu'ils sont plus instruits que par le passé et parce qu'ils ont accès plus facilement à davantage d'informations - revendiquent un pouvoir de décision politique et pas seulement, comme c'est le cas actuellement, de choix de représentants. Voter pour élire des représentants ne satisfait pas leur aspiration légitime à participer aux choix de sociétés dans lesquelles ils vont vivre.
L'argument souvent mis en avant pour justifier l'absence de parole directe donnée aux citoyens est l'impossibilité matérielle de faire discuter ensemble un grand nombre de personnes. En effet, comment mettre en relation conversationnelle des individus qui se trouvent dans des lieux différents et parfois fort éloignés et qui sont disposés à discuter à des moments différents ?
Si l'on admet qu'il existe une volonté politique de donner aux citoyens le rôle qu'ils revendiquent (ce que montrent le « grand débat » organisé en 2019 ou les lois relatives à l'organisation de référendum populaires), le problème semble

donc être essentiellement de nature technique. Comment l'intelligence artificielle peut-elle permettre des débats citoyens sur les axes de la société de demain ?

Jeudi 2 avril !! 13:30–16:30 !!

IA POUR LA SOCIOLOGIE ET SOCIOLOGIE DE L'IA

Francis Chateauraynaud (Dre, GSPR, EHESS)

L'IA entre expérimentation cognitive, promesse technologique et problème public. Un regard pragmatiste

Olessia Kirtchik (Doc en sociologie, EHESS, Paris, & chercheure à l'Institut d'études historiques et théoriques en sciences humaines Poletayev, l'Université Nationale de Recherche "Haute École d'Économie", Moscou)

Le débat public : un défi à l'IA

Le terme de l'"IA" est extrêmement flou en recouvrant une multiplicité d'applications et d'approches (systèmes experts basés sur des règles, diverses méthodes de l'apprentissage statistique, *data mining*…) qui possèdent des généalogies et des enjeux très divers. En outre, des efforts appliqués à la résolution de problèmes spécifiques (NLP, robotique, etc.) sont souvent menés à l'intersection de différentes disciplines telles que computer science, statistique, l'ingénierie des systèmes, sciences cognitives, linguistique et bien d'autres… Pour ces raisons, il n'est pas facile de définir l'"IA" en tant qu'objet de recherche en sciences sociales de manière cohérente. Dans cette intervention, je reviendrai sur des critiques de la notion de l'IA et proposerai quelques stratégies et questions de recherche, proprement sociologiques, que peut susciter cet ensemble techno-scientifique.

Jeudi 16 avril [séance initialement prévue le 12 décembre 2019]

HISTOIRE DE L'IA

Dominique Cardon (Pr de sociologie, Sciences Po & Media Lab)

La revanche des neurones – la controverse de l'intelligence artificielle

Profitant des réussites spectaculaires des techniques de *deep learning*, les promesses de l'Intelligence artificielle (IA) sont revenues occuper le débat public. Dans cette conférence, on propose de retracer quelques aspects de l'histoire de l'Intelligence artificielle afin de comprendre les défis présents. Le « troisième printemps de l'IA » se caractérise, en effet, par le retour du paradigme connexionniste qui avait été marginalisé lors du « second printemps » de l'Intelligence artificielle dans les années quatre-vingt. La nouvelle vague de promesses de l'IA a pris forme grâce aux opportunités nouvelles offertes par les données massives et aux nouvelles capacités de calcul des ordinateurs. Mais, elle propose surtout une autre manière de rendre les machines « intelligentes ». Il ne s'agit plus de demander aux calculateurs de raisonner, mais bien d'apprendre des données et de former des modèles de prédiction à partir des données. Ce changement de paradigme de l'IA a de nombreuses conséquences : il propose une représentation différentes de la société qui s'appuie sur le comportement des individus plutôt que sur leurs catégories d'appartenance, il ouvre des possibilités d'automatisation du travail, il présente des opportunités nouvelles mais aussi des risques de biais dans les calculs. Cette conférence propose d'examiner les défis éthiques et politiques que pose les développements de l'IA.

> Les travaux de Dominique Cardon portent sur les usages d'Internet et les transformations de l'espace public numérique. Ses recherches récentes portent sur les réseaux sociaux de l'Internet, les formes d'identité en ligne, l'autoproduction amateur et l'analyse des formes de coopération et de gouvernance dans les grands collectifs en ligne. Il conduit aujourd'hui une analyse sociologique des algorithmes permettant d'organiser l'information sur le web.
> Il a publié *La démocratie Internet*, Paris, Seuil/La République des idées, 2010, avec Fabien Granjon, *Mediactivistes*, Paris, Presses de Science po', 2010 (2e éd. enrichie : 2013), avec Antonio Casilli, *Qu'est-ce que le digital labor ?*, Paris, Ina Éditions, 2015, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Seuil/République des idées, 2015 et, avec Jean-Philippe Heurtin, *Chorégraphier la générosité. Le Téléthon, le don et la critique*, Paris, Economica, 2016.

Frédéric Fürst (Mcf en IA, Université de Picardie & MIS)

Intelligence artificielle et Astronautique : utopies parallèles

L'intelligence artificielle est un projet lancé en 1956 visant à réaliser des machines dotées d'une intelligence similaire à celle des humains. Les initiateurs du projet, et de nombreux chercheurs en IA depuis cette époque, ont oeuvré avec la conviction que cette ambition était rationnelle et réaliste. Mais au fil des décennies, l'objectif initial, qu'on appelle IA forte, est apparu de plus en plus utopique et a fini par se diluer et se morceler en un ensemble d'objectifs bien plus modestes, ce qu'on nomme l'IA faible. Bien que critiqué dans sa faisabilité et pour ses implications sociales et éthiques, l'intelligence artificielle n'en demeure pas moins un champ des sciences et techniques dynamique et qui suscite toujours autant d'enthousiasme. Les années 1950 ont également vu la naissance d'un autre domaine scientifique et technique fantasmatique : l'astronautique. En retraçant l'histoire de l'IA en parallèle avec celle de l'astronautique, on mettra en évidence des analogies entre les deux, au-delà de leur aspect utopique et de leur corrélation chronologique.

Jeudi 7 mai

IA, VISUALISATION ET MODÈLES

Jean-Daniel Fekete (Dre, Inria)

Visualisation pour l'analyse de réseaux dynamiques et de données en hautes dimensions

Franck Varenne (Mcf, ERIAC, Univ. de Rouen, & IHPST)

Une contribution de l'épistémologie des modèles à l'explicabilité des algorithmes

Il y a aujourd'hui une demande croissante d'explicabilité pour les algorithmes assurant des fonctions de modèles de prédiction ou de décision (médicale, juridique, etc.) et qui utilisent pour ce faire des techniques formelles d'apprentissage machine : régression, classification bayésienne, arbres de décision, forêts aléatoires, machine à vecteurs de supports, réseaux de neurones et leur perfectionnement, le deep learning. Comme on s'en doute, cette demande d'explicabilité, très discutée actuellement dans la littérature, est riche et n'est pas dépourvue d'ambiguïtés ainsi que plusieurs auteurs l'ont déjà signalé (Herman, 2017 ; Gilpin et al. 2018). Il convient d'abord de distinguer les différents types de sujets humains, cibles de cette explicabilité : ce peut être le modélisateur du problème, le programmeur, l'utilisateur final ou le bénéficiaire de la décision. Il convient ensuite de distinguer ce qui, de l'algorithme, doit être explicable : l'algorithme proprement dit, ou bien ses données et ses représentations (les ontologies), ses processus ou encore son seul résultat. Il convient enfin de distinguer ce que, dans ce contexte, cette demande d'explicabilité entend par « explication », à la différence d'« interprétation » et de « compréhension ». Cet exposé se penchera sur ce troisième axe de questionnement. Il présentera les premiers résultats d'un travail de collaboration que je mène actuellement avec Christophe Denis (LIP6, ACASA). Il tâchera de montrer en quoi une épistémologie qui se concentre sur la pluralité des fonctions épistémiques des modèles scientifiques, sur leurs critères distinctifs, tout en mettant l'accent sur leurs articulations et leurs déterminations réciproques peut contribuer à clarifier cette demande d'explicabilité.

Références :

[1] Denis, C., Varenne, F., « Interprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique », Conférence Nationale en Intelligence Artificielle 2019, *Actes de la CNIA@PFIA 2019*, J. Lang (dir.), Toulouse, pp. 60-69, https://hal.sorbonne-universite.fr/hal-02184519. Actes de la CNIA@PFIA 2019 : https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_CNIA_PFIA2019.pdf

[2] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A. Specter, M., Kagal, L., "Explaining Explanations : An Approach to Evaluating Interpretability of Machine Learning", https://arxiv.org/abs/1806.00069, 2018.

[3] Herman, B., "The Promise and Peril of Human Evaluation for Model Interpretability", *Thirsty-first Conference on Neural Information Processing Systems*, 2017, https://arxiv.org/abs/1711.07414

[4] Varenne, F., *From Models to Simulations*, London, Routledge, 2018.

Jeudi 30 janvier 2020

PROGRÈS, IA ET ÉTHIQUE

Vincent C. Müller (Pr de philosophie et d'éthique, Eindhoven Univ. of Technology)

Sonia Desmoulin-Canselier (Laboratoire Droit et Changement Social, UMR 6297 CNRS, Université de Nantes)
La place de l'humain dans un monde algorithmé : questionnements juridiques sur les conditions du déploiement des systèmes algorithmiques d'aide à la décision

L'innovation et la décision au XXIe siècle sont placées sous le double signe de la collecte massive de données et du traitement algorithmique. Dans tous les secteurs de la vie économique, juridique et sociale, on attend de la « mine d'or » du big data des pistes de recherches innovantes, des informations inédites, des prévisions et des solutions « individualisées » ou « personnalisées ». De telles perspectives ne peuvent toutefois se réaliser sans le recours à de puissants outils informatiques, pour stocker les données, mais surtout pour les traiter et pour effectuer des calculs. Ces systèmes automatisés permettant de formuler une réponse (ou une proposition de réponse) à une question ou un problème donné sont des systèmes algorithmiques. Dans le secteur médical, les solutions d'aide à la décision se sont multipliées : depuis l'aide au diagnostic (analyse des images médicales et détection des signaux morbides, probabilité diagnostique, examens complémentaires) jusqu'à la prescription médicamenteuse (risque allergique, incompatibilités entre traitements, posologie) en passant par la prédiction (calcul du risque de récidive). A présent, c'est en matière de police (calcul de risque sur une zone géographique ou de risque de récidive pour une personne), d'administration (détermination des affectations dans l'enseignement supérieur, attribution d'une autorisation) et de justice (calcul des prestations compensatoire ou des dommages-intérêts) que les algorithmes se déploient. Les enjeux du recours à de tels outils sont nombreux. Leur complexité et leur fonctionnement peuvent mettre en difficulté l'autonomie du décideur et le respect des droits du destinataire de la décision. A partir de quelques exemples, nous étudierons les réponses du droit en vigueur en Europe et en France et nous nous interrogerons sur les évolutions possibles et nécessaires pour faire face à ces enjeux.

Jeudi 20 février

PHILOSOPHIE DE L'IA : PERCEPTION, COGNITION ET IA

Gunnar Declerck (Mcf, UTC)

Jean-Michel Roy (Pr, ENS Lyon)

Jeudi 19 mars

LOGIQUE ET PREUVE EN IA / IA ET DÉLIBÉRATION

J.-J. Szczeciniarz (Pre, SPHERE, Université Paris-Diderot)

Jean Sallantin (Pre, LIRMM, Univ. de Montpellier)
Le débat public : un défi à l'IA

Nous sommes à un moment critique de l'histoire où les modèles qui fondent la gouvernance de notre société sont contestés à toutes les échelles depuis la commune jusqu'aux règles des échanges commerciaux mondialisés de la planète.

En effet, les citoyens - peut-être parce qu'ils sont plus instruits que par le passé et parce qu'ils ont accès plus facilement à davantage d'informations - revendiquent un pouvoir de décision politique et pas seulement, comme c'est le cas actuellement, de choix de représentants. Voter pour élire des représentants ne satisfait pas leur aspiration légitime à participer aux choix de sociétés dans lesquelles ils vont vivre.

L'argument souvent mis en avant pour justifier l'absence de parole directe donnée aux citoyens est l'impossibilité matérielle de faire discuter ensemble un grand nombre de personnes. En effet, comment mettre en relation conversationnelle des individus qui se trouvent dans des lieux différents et parfois fort éloignés et qui sont disposés à discuter à des moments différents ?

Si l'on admet qu'il existe une volonté politique de donner aux citoyens le rôle qu'ils revendiquent (ce que montrent le « grand débat » organisé en 2019 ou les lois relatives à l'organisation de référendum populaires), le problème semble donc être essentiellement de nature technique. Comment l'intelligence artificielle peut-elle permettre des débats citoyens sur les axes de la société de demain ?

Jeudi 2 avril !! 13:30–16:30 !!

IA POUR LA SOCIOLOGIE ET SOCIOLOGIE DE L'IA

Francis Chateauraynaud (Dre, GSPR, EHESS)

L'IA entre expérimentation cognitive, promesse technologique et problème public. Un regard pragmatiste

Olessia Kirtchik (Doc en sociologie, EHESS, Paris, & chercheure à l'Institut d'études historiques et théoriques en sciences humaines Poletayev, l'Université Nationale de Recherche "Haute École d'Économie", Moscou)

Le débat public : un défi à l'IA

Le terme de l'"IA" est extrêmement flou en recouvrant une multiplicité d'applications et d'approches (systèmes experts basés sur des règles, diverses méthodes de l'apprentissage statistique, *data mining*…) qui possèdent des généalogies et des enjeux très divers. En outre, des efforts appliqués à la résolution de problèmes spécifiques (NLP, robotique, etc.) sont souvent menés à l'intersection de différentes disciplines telles que computer science, statistique, l'ingénierie des systèmes, sciences cognitives, linguistique et bien d'autres… Pour ces raisons, il n'est pas facile de définir l'"IA" en tant qu'objet de recherche en sciences sociales de manière cohérente. Dans cette intervention, je reviendrai sur des critiques de la notion de l'IA et proposerai quelques stratégies et questions de recherche, proprement sociologiques, que peut susciter cet ensemble techno-scientifique.

Jeudi 16 avril [séance initialement prévue le 12 décembre 2019]

HISTOIRE DE L'IA

Dominique Cardon (Pr de sociologie, Sciences Po & Media Lab)

La revanche des neurones – la controverse de l'intelligence artificielle

Profitant des réussites spectaculaires des techniques de *deep learning*, les promesses de l'Intelligence artificielle (IA) sont revenues occuper le débat public. Dans cette conférence, on propose de retracer quelques aspects de l'histoire de l'Intelligence artificielle afin de comprendre les défis présents. Le « troisième printemps de l'IA » se caractérise, en effet, par le retour du paradigme connexionniste qui avait été marginalisé lors du « second printemps » de l'Intelligence artificielle dans les années quatre-vingt. La nouvelle vague de promesses de l'IA a pris forme grâce aux opportunités nouvelles offertes par les données massives et aux nouvelles capacités de calcul des ordinateurs. Mais, elle propose surtout une autre manière de rendre les machines « intelligentes ». Il ne s'agit plus de demander aux calculateurs de raisonner, mais bien d'apprendre des données et de former des modèles de prédiction à partir des données. Ce changement de paradigme de l'IA a de nombreuses conséquences : il propose une représentation différentes de la société qui s'appuie sur le comportement des individus plutôt que sur leurs catégories d'appartenance, il ouvre des possibilités d'automatisation du travail, il présente des opportunités nouvelles mais aussi des risques de biais dans les calculs. Cette conférence propose d'examiner les défis éthiques et politiques que pose les développements de l'IA.

> Les travaux de Dominique Cardon portent sur les usages d'Internet et les transformations de l'espace public numérique. Ses recherches récentes portent sur les réseaux sociaux de l'Internet, les formes d'identité en ligne, l'autoproduction amateur et l'analyse des formes de coopération et de gouvernance dans les grands collectifs en ligne. Il conduit aujourd'hui une analyse sociologique des algorithmes permettant d'organiser l'information sur le web.
> Il a publié *La démocratie Internet*, Paris, Seuil/La République des idées, 2010, avec Fabien Granjon, *Mediactivistes*, Paris, Presses de Science po', 2010 (2e éd. enrichie : 2013), avec Antonio Casilli, *Qu'est-ce que le digital labor ?*, Paris, Ina Éditions, 2015, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Seuil/République des idées, 2015 et, avec Jean-Philippe Heurtin, *Chorégraphier la générosité. Le Téléthon, le don et la critique*, Paris, Économica, 2016.

Frédéric Fürst (Mcf en IA, Université de Picardie & MIS)

Intelligence artificielle et Astronautique : utopies parallèles

L'intelligence artificielle est un projet lancé en 1956 visant à réaliser des machines dotées d'une intelligence similaire à celle des humains. Les initiateurs du projet, et de nombreux chercheurs en IA depuis cette époque, ont oeuvré avec la conviction que cette ambition était rationnelle et réaliste. Mais au fil des décennies, l'objectif initial, qu'on appelle IA forte, est apparu de plus en plus utopique et a fini par se diluer et se morceler en un ensemble d'objectifs bien plus modestes, ce qu'on nomme l'IA faible. Bien que critiqué dans sa faisabilité et pour ses implications sociales et éthiques, l'intelligence artificielle n'en demeure pas moins un champ des sciences et techniques dynamique et qui suscite toujours autant d'enthousiasme. Les années 1950 ont également vu la naissance d'un autre domaine scientifique et technique fantasmatique : l'astronautique. En retraçant l'histoire de l'IA en parallèle avec celle de l'astronautique, on mettra en évidence des analogies entre les deux, au-delà de leur aspect utopique et de leur corrélation chronologique.

Jeudi 7 mai

IA, VISUALISATION ET MODÈLES

Jean-Daniel Fekete (Dre, Inria)

Visualisation pour l'analyse de réseaux dynamiques et de données en hautes dimensions

Franck Varenne (Mcf, ERIAC, Univ. de Rouen, & IHPST)

Une contribution de l'épistémologie des modèles à l'explicabilité des algorithmes

Il y a aujourd'hui une demande croissante d'explicabilité pour les algorithmes assurant des fonctions de modèles de prédiction ou de décision (médicale, juridique, etc.) et qui utilisent pour ce faire des techniques formelles d'apprentissage machine : régression, classification bayésienne, arbres de décision, forêts aléatoires, machine à vecteurs de supports, réseaux de neurones et leur perfectionnement, le deep learning. Comme on s'en doute, cette demande d'explicabilité, très discutée actuellement dans la littérature, est riche et n'est pas dépourvue d'ambiguïtés ainsi que plusieurs auteurs l'ont déjà signalé (Herman, 2017 ; Gilpin et al. 2018). Il convient d'abord de distinguer les différents types de sujets humains, cibles de cette explicabilité : ce peut être le modélisateur du problème, le programmeur, l'utilisateur final ou le bénéficiaire de la décision. Il convient ensuite de distinguer ce qui, de l'algorithme, doit être explicable : l'algorithme proprement dit, ou bien ses données et ses représentations (les ontologies), ses processus ou encore son seul résultat. Il convient enfin de distinguer ce que, dans ce contexte, cette demande d'explicabilité entend par « explication », à la différence d'« interprétation » et de « compréhension ». Cet exposé se penchera sur ce troisième axe de questionnement. Il présentera les premiers résultats d'un travail de collaboration que je mène actuellement avec Christophe Denis (LIP6, ACASA). Il tâchera de montrer en quoi une épistémologie qui se concentre sur la pluralité des fonctions épistémiques des modèles scientifiques, sur leurs critères distinctifs, tout en mettant l'accent sur leurs articulations et leurs déterminations réciproques peut contribuer à clarifier cette demande d'explicabilité.

Références :

[1] Denis, C., Varenne, F., « Interprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique », Conférence Nationale en Intelligence Artificielle 2019, *Actes de la CNIA@PFIA 2019*, J. Lang (dir.), Toulouse, pp. 60-69, https://hal.sorbonne-universite.fr/hal-02184519. Actes de la CNIA@PFIA 2019 : https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_CNIA_PFIA2019.pdf

[2] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A. Specter, M., Kagal, L., "Explaining Explanations : An Approach to Evaluating Interpretability of Machine Learning", https://arxiv.org/abs/1806.00069, 2018.

[3] Herman, B., "The Promise and Peril of Human Evaluation for Model Interpretability", *Thirsty-first Conference on Neural Information Processing Systems*, 2017, https://arxiv.org/abs/1711.07414

[4] Varenne, F., *From Models to Simulations*, London, Routledge, 2018.