



Ce qui nous distingue d'ores et déjà, c'est une volonté de voir 10 ans en avance et une indépendance d'esprit.

machines à apprendre par elles-mêmes derrière la magie de Google ou de Facebook –, deux ont été engagés par la grande entreprise, et le dernier est professeur à l'Université de Montréal. Cela nous donne le droit de rêver, de viser la production de valeur sociale autant qu'économique.

PLAN : Qui dit données massives dit financement massif. Montréal est-elle aussi un endroit stratégique en ce qui a trait à l'accès au financement et à la concentration d'intervenants suffisamment gros ?

V.B. : Beaucoup de capital-risque dort dans la Province, mais les investisseurs québécois sont prêts à soutenir un produit, rarement une idée. Par conséquent, Montréal a vraiment moins de moyens que les autres pôles pour traverser cette « vallée de la mort » qui précède la commercialisation. Beaucoup d'entrepreneurs d'ici hésitent aussi à avoir recours au capital-risque par peur de perdre le contrôle de leur compagnie.

PLAN : Un frein que soulèvent les acteurs du domaine des données massives, c'est le manque criant de cerveaux compétents en la matière. Est-ce que l'IVADO compte répondre à ce problème, et avec quelle stratégie ?

V.B. : C'est une de nos motivations premières.

ées massives a besoin s comme des docteurs ants. Les universités ont être en mesure de former ces experts, des experts susceptibles de se destiner autant à l'entreprise qu'à la recherche. En plus d'analystes, l'industrie a besoin de gestionnaires capables d'orienter l'utilisation qui doit être faite des données.

Nous dirigeons nos efforts vers les étudiants de l'étranger pour attirer les meilleurs profils et nous travaillons en parallèle à retenir les cerveaux locaux, car les emplois les plus intéressants restent à l'étranger : certains ingénieurs formés à Montréal travaillent maintenant pour Google.

PLAN : Les données n'ont pas de frontière. Quels problèmes cela pose-t-il en droit international et de la personne ? Faut-il réfléchir à une législation particulière aux données massives en parallèle des aspects technologiques, et est-ce un axe de recherche de l'IVADO ?

V.B. : Une technologie de rupture comme l'utilisation des données massives apporte immanquablement des défis pour la société. Le droit et l'éthique se rapportant à ces questions restent de fait très flous, alors que ce sont des éléments cruciaux. L'IVADO envisage de jouer un rôle consultatif en la matière, mais pas de développer un axe de recherche spécifique. Nous souhaitons toutefois attirer des éthiciens et des sociologues pour les intégrer à nos équipes. Avant tout, je crois qu'une grosse part passe par l'éducation du public qui ne devrait pas diffuser d'information sans en comprendre les conséquences.

PLAN : Vous venez d'une université italienne, vous travaillez aujourd'hui au Québec. Faut-il en tirer des conclusions sur la géopolitique des données massives ?

Andrea Lodi : L'utilisation des données massives aujourd'hui, ça se passe essentiellement aux États-Unis dans des entreprises comme Google, Facebook ou IBM. Mais la création de la chaire que je dirige montre que le gouvernement du Canada a aussi compris l'importance stratégique du sujet du point de vue économique et social.

PLAN : Quels sont les principaux enjeux scientifiques et techniques de l'utilisation des données massives ?

A. L. : Depuis 10 ou 15 ans, les efforts dans le domaine ont essentiellement porté sur le développement d'un stockage efficace et bon marché des données d'une part,

et sur leur acquisition d'autre part, grâce aux capteurs que l'on a installés partout, sur les téléphones intelligents notamment. Les décisions étaient prises en différé, sur la base d'analyses ponctuelles faites *a posteriori* sur des données non dynamiques. Aujourd'hui, les décisions doivent se prendre en temps réel sur un flux continu d'informations, pour, par exemple, être capables de renseigner en permanence les automobilistes sur le meilleur itinéraire possible. C'est un changement radical de paradigme qui amène les chercheurs à modifier complètement leur façon d'aborder la question de l'utilisation des données massives et de l'accès rapide et fiable aux informations dans les bases de données.

PLAN : On parle toujours de quantité mais rarement de qualité dans le domaine de l'utilisation





des données massives. Comment peut-on valider le traitement de milliards d'informations issues de milliards de sources différentes ?

A. L. : La question de la qualité des données massives est effectivement fondamentale. Mais la problématique est finalement la même qu'avec les données statiques en

Aujourd'hui les décisions doivent se prendre en temps réel sur un flux continu d'informations.

quantités plus faibles. Nous disposons de méthodes comme l'exploration, la fusion et le nettoyage de données qui visent à l'optimisation de l'extraction des connaissances à partir de quantités d'informations issues de sources différentes; nous avons la possibilité de procéder à des analyses statistiques pour nous assurer de la fiabilité des décisions prises à partir des données; dans le cas des variables aléatoires, on peut faire des prédictions sur la base d'un historique. Mais le test ultime de fiabilité en matière d'utilisation des données, c'est la réponse du marché et de la société civile : pensez-vous que si Google ne faisait pas de bonnes prévisions, autant de personnes et d'entreprises les utiliseraient ?

PLAN : De plus en plus de décisions, et de plus en plus importantes, vont être prises sur la base de l'utilisation des données massives. Comment assurer la résilience des systèmes et des processus qui accompliront ces tâches ?

A. L. : Là encore, je ne pense pas que la problématique soit différente de n'importe quelle autre industrie. On fait confiance à Hydro-Québec pour sa capacité à fournir du courant parce que cette société d'État entretient bien le réseau hydroélectrique et les centrales, et qu'elle les exploite efficacement. Dans le cas des données massives, la quantité d'informations et la diversité des sources sert plutôt la résilience qu'elle ne la dessert : si vous produisez votre électricité à partir d'une combinaison d'éoliennes, de panneaux solaires, de barrages et de centrales thermiques, la probabilité que vous puissiez faire face à toutes les situations est bien plus grande que si vous utilisez seulement du charbon, par exemple. De même, si un téléphone intelligent est en panne, il y en a 100 autres pour prendre le relais. C'est comme un écosystème : plus les éléments qui le composent sont variés et nombreux, plus il est résistant.

PLAN : En cas de défaillance, qui devra être tenu responsable ? Le producteur des données, celui qui gère le réseau qui les achemine, celui qui les traite ?

A. L. : Je m'intéresse au sujet en tant que citoyen. En tant que chercheur par contre, je n'ai pas plus à dire sur cette question qu'un ingénieur doté d'une culture générale en données massives. Ce n'est plus mon champ d'expertise, je suis très occupé par le mien et une journée n'a que 24 heures ! Je pense que la science peut jouer un rôle consultatif en cas de litige, mais qu'il est aussi important qu'elle garde en tout temps son indépendance par rapport au débat sociétal. ◀