==Transcription==

Je ne vais pas avoir besoin de trop motiver parce que beaucoup a déjà été dit, mais on va plutôt parler des données. Comme l’a dit Gilles Dowek, c’est un peu compliqué de séparer données numériques et algorithmes — les données numériques n’ont de sens que de pouvoir permettre l’utilisation d’algorithmes —. Je vais plutôt focaliser mon exposé sur l’utilisation des données et des problèmes que ça peut poser.

Gilles avait un exposé en quatre points. Comme je suis moins fort, j’ai seulement deux points. Donc je vais parler d’extraction de connaissances et de qualité de données.

===Extraction de connaissances des données===

L’extraction de connaissances c’est ce qu’on fait constamment. On a des capteurs, on a des données textuelles, on a des images et on va essayer de trouver des informations dedans, des corrélations, on va essayer de détecter des valeurs aberrantes ; on va réaliser des statistiques pour apprendre quelque chose des données. Donc ça c’est un travail, une technologie qu’on utilise depuis les débuts de l’informatique. Ce n’est pas nouveau, souvent il y a des nouveaux noms, mais en gros c’est toujours la même chose : essayer de comprendre ce qu’il y a dans ces données, essayer d’en extraire des connaissances.

Une fois qu’on utilise ces connaissances, on commence à se dire oui, puisqu’on va utiliser nos connaissances pour faire quelque chose, autant que ce soit des connaissances correctes, autant qu’on les utilise bien. On a tout de suite tendance, et moi le premier, à critiquer des utilisations des algorithmes pour faire ça. Prenons l’exemple de la justice, les humains ne sont pas parfaits. Par exemple, il y a des études qui ont montré que la couleur de votre peau, Gilles en a déjà parlé, pouvait avoir des conséquences importantes pour les décisions des juges. Pour les juges d’application des peines, il y a eu des mesures qui ont été faites, qui ont montré que si vous passez avant le déjeuner ou si vous passez après le déjeuner, il y a une différence. Ça paraît rigolo. Oui, moi je me disais oui, 4-5 % de différence. Non, non ! Je ne sais plus le chiffre, mais c’est plutôt 40 %. Il faut vraiment, si vous devez passer devant un juge d’application des peines, demander à passer après déjeuner. Donc ça ce sont des juges humains et on peut se dire un algorithme, si on arrive à le faire aussi intelligent ou presque aussi intelligent qu’un juge, au moins on pourra lui demander de ne pas avoir de préjugés, on pourra lui demander d’être un peu cohérent : il ne va pas avoir de périodes d’après déjeuner.

Évidemment, on va demander que l’algorithme soit conçu comme ça. Gilles vous a donné l’exemple d’algorithmes qui teste le prénom d’une personne et base ses choix là-dessus ; ça c’est facile à détecter, mais il y a des choses qui sont beaucoup plus complexes comme des études qui ont été faites où vous utilisez l’apprentissage. Vous apprenez, par exemple, à partir de dix ans d’expérience de décisions de juges . Et qu’est-ce que vous allez faire ? Et bien vous allez reproduire, d’une certaine façon, les préjugés des personnes qui ont jugé pendant ces dix ans. Peut-être que si les algorithmes apprennent, au bout de très longtemps, vous allez finir par gommer ces préjugés, mais ce n’est pas clair.

Donc, d’une certaine façon, et c’est exactement de ça dont on parle ici, avec les algorithmes, on peut espérer faire beaucoup mieux sur des sujets particuliers, que des êtres humains. On veut éviter les préjugés ; on veut exiger plus de cohérence. Mais il faut des efforts pour ça.

Gilles vous a parlé de cette vente d’agrafeuses sur le Web ; c’est exactement là le sujet. Les gens, qui avaient prévu ça, voulaient que si vous aviez plus à vous déplacer on allait vous faire payer un peu plus cher. Le résultat des courses c’est que si vous étiez dans un quartier défavorisé, vous payiez plus cher. C’est ça qui a fini par soulever, si vous voulez, les embarras de certaines personnes : c’était corrélé avec le niveau social de votre quartier. Pourquoi ? Parce qu’il se trouve que cette marque d’agrafeuses n’était pas vendue dans les quartiers plus défavorisés. Bien ! C‘est ce qu’on veut éviter.

Le sujet aujourd’hui c’est qu’est-ce que peut faire la recherche là-dedans. ? Et donc peut-elle aider à « responsabiliser » tous ces traitements de données ? Et j’ai parlé d’équité, on a déjà parlé de transparence ; il y a la loyauté cette notion très simple, et on en verra un exemple tout à l’heure, qui demande que, si vous annoncez quelque chose, et bien vous vous vous y teniez et qu’on puisse vérifier que c’est le cas. La diversité qui fait que ce ne sont pas uniquement les gens qui sont le plus visibles sur le réseau qui vont avoir tous les clients. Enfin la neutralité, on reviendra là-dessus si j’ai le temps.

On va regarder brièvement quelques techniques, quelques recherches, des fronts dans des sujets de recherche : comment on collecte et analyse les données, notamment comment on trace leur origine. Et puis comment on peut vérifier que ces analyses sont faites de façon responsable et qu’elles ne tiennent pas compte, par exemple, de la couleur de la peau.

Tracer la provenance des données c’est quelque chose qu’on utilise, qui est extrêmement populaire en informatique depuis plusieurs années, et ça, ça ne devrait pas surprendre les scientifiques qui sont dans la salle.

Quand on est scientifique on sait depuis longtemps que des données n’ont pas de valeur si on ne sait pas d’où elles viennent. En gros, quelles sont les données primaires qui ont été utilisées, quelles analyses ont été réalisées ? Si je ne vous donne pas ces informations, je peux vous donner tous les chiffres que je veux, vous n’allez pas me croire ; on est scientifiques, on est incrédules. Ce qu’il faudrait c’est que cette attitude scientifique de vouloir vérifier ce qu’on nous montre, on la retrouve aussi dans la société au moment où les données sont utilisées de façon aussi critique. Donc il y a des méthodes, notamment pour, quand vous avez un <em>workflow</em> de données, suivre comment elles ont été produites, pouvoir être capable de dire d’où elles viennent, qu’est-ce qui a été fait comme calculs dessus. On fait ça en science, mais on fait ça aussi dans l’alimentation : maintenant il y a des grands groupes alimentaires qui sont capables de vous dire où la vache a mangé de l’herbe pour vous donner le lait qui a été utilisé dans ce yaourt. Dans la création vidéo aussi on essaie de voir tout le traçage, très souvent pour des questions de droits d’auteur.

Il y a deux aspects très intéressants, passionnants, mais un peu différents dans la vérification. Le premier c’est l’analyse du code. C’est quoi ? Vous avez un code et vous voulez vérifier que ce code fait « bien » les choses ; ici « bien » est à définir. Je n’en parle pas ici, mais la complexité, c’est déjà de définir des notions aussi simples de d’être équitable ; il y a plein de façons de définir l’équité. Supposons qu’on ait formalisé ce qu’on veut dire par équitable, est-ce que votre logiciel fait bien le boulot ?

Ça, c’est de l’analyse statique du code, c’est très proche de ce qu’on a comme des preuves de théorèmes en mathématiques ; c’est un domaine où Inria a une tradition extrêmement longue et extrêmement forte. De fait, quand on regarde, il y a énormément de travail qui a été fait en analyse de programmes dans différents domaines : la sécurité, la fiabilité, l’optimisation notamment de requêtes, la confidentialité ; encore très peu sur d’autres aspects comme l’équité. Donc il y a tout un front de travaux à faire, de recherches qui sont captivantes.

Oui, j’ai zappé un truc c’est l’<em>open source</em>. L’<em>open source</em> est quand même très important dans ce contexte-là. C’est-à-dire que si vous n’avez pas accès au logiciel ça va être difficile de vérifier qu’il fait bien les choses ; on peut faire des audits, etc., mais ça veut dire que la vérification sera relativement homéopathique.

L’autre versant c’est tester les effets. Donc si le premier s’approchait de vérifier des preuves de théorèmes, de vérifier des théorèmes, là ça s’approche plutôt de l’étude de phénomènes physiques comme le climat ou du cœur humain. L’idée est de développer des modèles mathématiques et de faire des analyses statistiques.

Par exemple un truc qui a été fait à Chicago : la municipalité a décidé un changement assez considérable des transports urbains et il y a des gens qui ont commencé à dire « non, non, c’est n’importe quoi, votre transformation est très négative pour les minorités ethniques. » À partir de là c’était difficile à vérifier parce qu’il y avaient dans des quartiers différents des gens qui étaient avantagés, d’autres qui ne l’étaient pas. Donc il y a des gens qui ont dit : « On peut formaliser ça, on peut trouver un modèle mathématique et on peut essayer de vérifier ». En l’occurrence ils ont montré que, de fait, le changement ne défavorisait pas ces minorités. Cela aurait pu être une conséquence malheureuse de choix politiques ; en l’occurrence ce n’était pas vrai. Une étude de data journalisme a permis de le montrer.

Je vais maintenant prendre un autre exemple qui est la vérification de la transparence dans Google Ads ; ça nous permettra de parler de transparence.

A quoi sert Google Ads Settings : vous recevez de la pub, vous trouvez que c’est intrusif. Qu’est-ce qu’on connaît de vous ? Google va vous dire quelles informations sur vous ont été utilisées pour vous envoyer cette publicité. Je ne fais pas du <em>bashing</em> de Google ; ce sont pratiquement les seuls à essayer de faire un peu de transparence, donc on peut quand même apprécier leurs efforts. Sur l’écran, on voit les données d’une personne que j’ai anonymisées, à qui il est écrit qu’elle aime les chats, le fitness, etc. Première chose les résultats sont quand même assez médiocres, parce cette personne m’a avoué qu’elle n’aimait pas les chats et qu’elle ne faisait pas de fitness non plus. Donc ces résultats sont encore un peu à améliorer pour Google. Mais regardons ce qui se passe du point de vue de la transparence. Il y a une équipe de recherche qui a développé un logiciel qui s’appelle Adfisher, qui a testé cette transparence.

Première chose qu’ils ont testé : vous pouvez dire « je ne veux pas qu’on m’envoie de pubs sur les chats ». Si vous dites ça vous n’allez plus recevoir de pubs sur les chats ; de ce point de vue-là ça marche.

Deuxième chose sur la transparence, qui marche moins bien, c’est qu’on vous dit : « Voilà les paramètres qui sont utilisés ». Et bien, dans la réalité, ce n’est pas si simple que ça : il y a d’autres paramètres qui sont utilisés, qu’on ne vous dit pas, en particulier il y a le sexe. Ça, c’est gênant, parce que normalement ils ne devraient pas garder ce genre d’informations dans des certains pays parce que c’est un attribut protégé. Il se trouve qu’Adfisher a mis en évidence que les hommes recevaient des pubs pour des boulots qui étaient bien mieux payés que pour celles que les femmes recevaient, statistiquement, de manière significative. On pourra se poser après la question de savoir pourquoi les hommes recevaient des pubs pour des jobs mieux payés, mais ce que je voulais souligner ici, c’est la transparence : on vous a dit qu’on ne tenait compte que de ces paramètres, mais les publicités de Google en utilisent d’autres. Une autre propriété éthique qu’on peut vérifier.

**10’ 30**

**Évaluer la qualité des données**

Deuxième partie de cet exposé : l’évaluation de la qualité des données.

Je ne vais pas vous faire la liste de tous les dérives contre lesquelles on voudrait lutter sur le Web, il y en a trop, mais ce que je voulais dire c’est que ça peut avoir des influences sur notre vie quotidienne.

Prenons un exemple connu : en 2013, il y a des gens qui sont arrivés à pirater le compte Twitter de Associated Press et qui ont annoncé que la Maison Banche avait été bombardée. C’est sympa, le Dow Jones a chuté de 100 points en deux minutes. Pourquoi ? Parce que l’information paraissait crédible — elle venait de Associated Press — et donc là on a une information qui est une information fausse qui peut avoir des conséquences dans la vie de tous les jours.

Des informations fausses, il y en a beaucoup qui ont circulé au cours des dernières élections aux États-Unis. À partir du moment où elles sont là, où elles peuvent biaiser les décisions des gens, qu’est-ce qu’on fait à ce sujet ?

C’est évidemment un sujet de société ; ce n’est pas un sujet d’informatique, mais ça pose des questions aux informaticiens. Une fois que des politiques auront décidé de déployer des solutions, se seront mis d’accord — ce n’est pas facile, c’est peut-être eux qui ont la partie la plus dure —, nous en tant qu’informaticiens avons le job le plus facile : c’est arriver à mettre au point des outils qui fassent ça.

Hou là ! Quelle est la question que nous avons à résoudre ?

**Public :** C’est peut-être la véracité.

**Serge Abiteboul :** C’est la véracité, oui. Donc sur cette véracité, je voulais dire que les difficultés qui sont dues à la véracité sont des difficultés classiques dans le traitement de données : le volume, la variété, la véracité, mais ce sont d’autres difficultés qui sont dues très précisément au fait que les données sont des données générées par des humains.

Les humains sont très sympas, mais quand ils publient des informations ou quand ils donnent des informations, ils font plein de fautes ; ils omettent volontairement ou pas des données ; ils sont incohérents, ils sont incomplets, etc. Évidemment, quand on va essayer de raisonner sur tout ça et puis de comprendre ces données, ça va être compliqué. En plus, il y aussi des affaires de goûts : il n’y a pas juste la vérité, il y a aussi « j’aime bien, je n’aime pas ». Donc c’est compliqué, ça demande du travail de recherche. Par exemple, sur le fait que les données soient incohérentes, ça veut dire qu’on ne peut pas utiliser des logiques standards, parce qu’en logique standard si vous avez impliqué deux faits contradictoires, vous êtes foutu. Donc là il va falloir plutôt, par exemple, utiliser une logique probabiliste et il y a beaucoup de travaux dans cette direction.

Ça continue avec d’autres problèmes. Je voulais mentionner un autre aspect de ces propriétés de responsabilité, c’est la propriété de neutralité.

Alors neutralité des réseaux, vous avez tous déjà peut-être entendu parler de cette notion-là. C’est qu’un réseau devrait transporter les données sans biais, le réseau, Internet par exemple, devrait transporter les données sans biais depuis la source jusqu’à la destination, sans tenir compte de cette source, de cette destination et du contenu. C’est une espèce de justice. On ne va pas vous privilégier parce vous êtes, je ne sais pas, Google ou Microsoft. On ne va pas vous pénaliser parce que vous êtes une petite entreprise ou on ne va pas vous pénalisez parce que vous êtes un opposant au gouvernement. Le réseau ne voit pas tout ça, c’est une espèce de justice du réseau.

Au Conseil national du numérique on a travaillé là-dessus. On a aussi beaucoup travaillé sur une autre neutralité qui est celle des plateformes. En général les gens n’utilisent pas que le réseau, ils utilisent des plateformes de l’Internet comme Google, Facebook, des grosses plateformes et, d’une certaine façon, leurs informations ils les ont par ces plateformes-là. Alors à quoi servirait que le réseau soit neutre si l’information qu’on vous donne est complètement biaisée par la plateforme à laquelle vous avez accès ? La critique, quand on commence à dire on pourrait exiger de ces plateformes d’être neutres, la critique c’est toujours « mais attendez quand vous utilisez le site du *Monde* ou de Carrefour, rien n’est neutre. Un site comme Google c’est à la fois le site du *Monde*, c’est à la fois Carrefour. *Le Monde* a le droit d’éditorialiser son contenu. Carrefour a le droit de mettre en avant ses produits par rapport aux produits d’autres marques.

Pourquoi ces différences pour ces grandes plateformes ? Je pense que c’est une question fondamentale, c’est que ces plateformes sont devenues tellement importantes que, d’une certaine façon, elles sont devenues incontournables. Quand vous avez autour de 90 % des Européens qui cherchent sur le Web avec Google, ou un même pourcentage d’Européens qui est sur Facebook, quand vous êtes d’une certaine façon dans un « service public ». Et donc, à cette puissance considérable, vous devez pouvoir demander aussi des exigences particulières, des responsabilités considérables.

Le problème c’est qu’il y a quelques compagnies qui sont en train de concentrer toutes les données du monde et une énorme partie des puissances de calcul et que ça menace la vraie concurrence entre les entreprises. Ça menace aussi nos libertés.

Là c’était pour montrer quelque chose qui est du niveau du domaine public. Ça tient du procès antitrust. Mais je vais vous donner juste un exemple un peu personnel. J’aime beaucoup Google, j’ai toujours utilisé Google et quand on me disait : « Google biaise les résultats », je disais : « Non, non, ces mecs ne font pas ça ; ils ont toujours dit qu’ils mettaient un mur entre le moteur de recherche et leur business. Non, non, ce sont des gens bien, ils ne font pas ce genre de trucs. » La question c’est comment pouvez-vous démontrer qu’il y a un biais ou qu’il n’y en a pas ? Le problème n’est pas uniquement un problème de société. La société peut très bien dire « oui, oui, on ne veut pas de biais », mais comment vous le vérifiez ça ? Eh bien ce n’est pas facile. On peut utiliser des tests sur les résultats des moteurs de recherche, mais, ces moteurs sont très complexes, ils tiennent compte d’où vous êtes, ils tiennent compte de plein d’autres paramètres. Montrer que systématiquement ils vont favoriser un service par rapport à un autre ce n’est pas simple. Dans la réalité, vous pouvez vous convaincre. C’est-à-dire que vous avez vu un copain qui avait un produit qui était dans la première page des réponses de Google sur une question particulière et puis un produit concurrent de Google arrive et deux mois après il est dans la troisième page. Mais la question c’est comment vous le démontrez.

**Conclusion pour l’Académie des sciences**

Je vais avoir deux diapos de conclusion pour l’Académie des sciences.

La première c’est sur la recherche dans ce domaine, la recherche autour des données, la gestion de données, maintenant on aime bien le mot *datascience*.

Mon expérience, c’est vraiment une opinion, là, ce n’est pas un théorème, mon expérience, c‘est que dans le domaine, par le passé on a été tirés par les données des entreprises. J’ai travaillé pendant des dizaines d’années avant qu’on me dise : « Tiens, tu pourrais regarder aussi les données de Facebook ». Il n’y avait pas Facebook. D’accord ? On travaillait sur les données des entreprises. Typiquement ce qu’on faisait, on faisait des modèles de données et puis on faisait des systèmes et on visait deux choses : la performance, et la performance. D’accord ? Et après on nous a dit : « C’est bien, ça marche, ça marche vite, mais ce n’est pas assez fiable ». Alors on a bissé aussi sur la fiabilité. En gros notre but, on voulait être capables de faire des systèmes capables de gérer des téraoctets de données pendant des mois sans se casser la gueule. Maintenant on sait faire. Je ne dis pas que la recherche est finie, il y a plein de trucs encore à regarder, il y a plein de recherches passionnantes. Mais, maintenant qu’on sait faire ça, on va essayer de faire mieux sur deux fronts. Le premier front ce sont les données personnelles et les données sociales, et j’ai insisté sur le fait que les données humaines sont différentes des données des entreprises en ce qu’elles ont des erreurs, des incohérences, etc. ; et puis sur les questions éthiques, c’est-à-dire le sujet aujourd’hui. Le second, je n’en ai pas parlé, c’est la sécurité, c’est protéger les données.

La dernière diapo. Quand on regarde ce genre de problème on se demande de quels moyens on dispose, parce que ces problèmes sont compliqués, parce qu’il faut aller bille en tête contre des grosses plateformes, qu’il faut aller bille en tête contre des habitudes, contre des nouveaux outils qui arrivent sans arrêt, ça change très vite. Quels sont les moyens à notre disposition ?

Évidemment les outils c’est d’abord la loi et la réglementation. Ça ne nous concerne pas tellement à l’Académie, mais on peut quand même donner notre avis.

Un autre outil, ce sont les associations de consommateurs. Alors là, les associations de consommateurs, on est tous des consommateurs et on a, à mon avis, dans ces domaines-là, un pouvoir considérable parce que, d’une certaine façon, si on décide qu’on ne veut plus utiliser Facebook, Facebook n’existe plus ! Et d’ailleurs, ces entreprises sont extrêmement soucieuses de leur image de marque parce que, d’une certaine façon, toutes ces grandes entreprises du Web si vous partez, si elles perdent leurs clients, elles disparaissent. Un des cas où elles ont reculé, par exemple, c’est Instagram qui avait changé sa politique sur les droits des photos. Il y a eu des levées de boucliers dans les réseaux sociaux ; des gens ont commencé à quitter Instagram par dizaines de milliers et puis Facebook, qui avait acquis Instagram a sorti un communiqué en disant « non, non, mais vous n’aviez rien compris ; ce n’est pas qu’on voulait. On enlève ce contrat. On veut respecter votre propriété sur les données. »

Les gouvernements et les associations qui doivent s’appuyer sur des citoyens éclairés. Donc là on retrouve un domaine sur lequel l’Académie a un rôle considérable à jouer : c’est pousser pour que les citoyens soient beaucoup plus éduqués. C’est un sujet considérable. C’est le sujet de l’enseignement de l’informatique, l’enseignement de ces questions éthiques autour de l’informatique à tous les citoyens. Ça c’est un projet sur lequel l’Académie a une responsabilité considérable.

Et le dernier point, évidemment, c’est celui qui, en tant que chercheur, me passionne le plus, c’est le travail sur des nouveaux outils, qui permettront, par exemple, de faire de la recherche, des analyses de données plus correctes, de vérifier que les analyses de données sont faites de manière correcte.

Je vous remercie de votre attention.

[Applaudissements]

**Organisateur :** Très bien, on va passer aux questions. Odile Macchi. Odile vas-y.

**21’ 20**

**Odile Macchi :** Est-ce qu’il y a des moteurs de recherche « éthiques » entre guillemets ? Par exemple il y a Qwant, qui est un moteur de recherche français, est-ce qu’il est éthique ?

**Serge Abiteboul :** C’est une excellente question. Personnellement j’ai arrêté d’utiliser Google et j’utilise Qwant ; je vous conseille d’en faire autant d’abord parce qu’ils payent leurs impôts en France, etc. Mais non ! Est-ce qu’il est éthique ? C’est une vraie question. Qwant dit, par exemple, qu’il ne garde aucune trace de vos données sur cet aspect-là de vos recherches. De ce point de vue-là, il ne vous espionne pas et vous gardez plus votre *data privacy*. Il y a des aspects qui le sont moins : par exemple j’ai demandé au Eric Léandri de Qwant quel était l’algorithme qui permettait de classer leurs données. J’ai posé la même question à des responsables chez Google et à chaque fois on me dit : « Non, non on ne veut pas publier ça, parce que si on publie ça les gens vont s’en servir pour les contourner. » C’est aussi ce qu’a dit, évidemment, Léandri. Mais la suite de sa réponse m’a plu. Il m’a dit : « On est en train de travailler sur des techniques qui nous permettront de les publier plus tard. » Donc ils font de la recherche autour, notamment, des blockchains, je ne sais pas les détails, pour espérer un jour pouvoir publier l’algorithme de classement. Pour moi c’est fondamental. C’est la transparence.

**Odile Macchi :** Est-ce qu’ils ont un marché, parce qu’ils sont tellement petits ? Personne ne sait ce que c’est quand on demande autour de nous.

**Serge Abiteboul :** Ils ont une personne de plus : je suis passé chez eux. Ils ont exactement le *business model* qu’avait Google dans le début des années 2000, qui a fait la richesse de Google. C’est-à-dire qu’ils vendent de la publicité autour de leur moteur de recherche et c’est un excellent *business model*. Leandri m’a dit qu’avec 15-20 % du marché français, ça lui suffirait à être rentable. Vous aviez une question.

**Public :** Oui. Est-ce que vous pourriez commenter cette remarque qu’en fait les qualités d’un système de données dépend de ce qu’on veut en faire, par exemple la provenance des données. Si on veut faire une base de données parapharmaceutiques, qu’on veut être sûr de la qualité des médicaments, on va exiger la provenance des données parce qu’on veut en être sûr. Mais si on veut faire un système pour la recherche ou pour des systèmes de renseignement, on pourra admettre que, comme les journalistes, on ne donne pas les sources. On aura comme ça des informations peut-être moins fiables, mais plus surprenantes ou plus intéressantes.

**Serge Abiteboul :** Il peut y avoir plein de raisons pour ne pas donner l’origine de ses données. Celles que vous avez données, par exemple : je ne veux pas donner des secrets. Ça peut être aussi pour des raisons commerciales, le secret commercial. Ça peut être aussi parce que franchement ça coûte de l’argent de regarder la provenance et, peut-être, pour ce qu’on fait, vous vous moquez un peu de la qualité. Donc oui, il y a plein de raisons. Mais il y a plein d’endroits où les données et leur provenance et les calculs qui ont été faits dessus me permettent d’avoir confiance, ou non, dans ce qu’elles font et c’est dans ce cadre-là que c’est nécessaire qu’elles soient publiées.

**Public :** Ce que je veux dire surtout c’est que la qualité d’un système de données n’est pas intrinsèque ; elle dépend de l’utilisateur.

**Serge Abiteboul :** Mais la confiance aussi. C’est aussi la confiance que vous voulez donner aux utilisateurs. Bien sûr. C’est complètement dépendant du contexte. Je ne dis pas qu’il faut de la provenance sur tout, mais encore une fois, il semble que ce soit une cause de confiance très fréquemment.

**Organisateur :** Nicolas Ayache.

**Nicolas Ayache :** Oui, Serge. Il y a une autre menace qui est évoquée dans votre livre *Le temps des algorithmes* et qui m’a frappée, cette extrême concentration des données menace aussi un certain nombre de professions. Vous avez parlé du juge, mais il y a, par exemple, les chauffeurs de taxi avec les grandes bases de données géographiques, l’automatisation des véhicules ; les connaissances médicales et la connaissance médicale pourrait également menacer le métier de médecin ; la robotisation de la chirurgie, le chirurgien. Il y a des problèmes éthiques aussi qui sont liés à la menace de ces emplois. Est-ce que vous y avez pensé ?

**Serge Abiteboul :** Oui, bien sûr. Il y a un chapitre du livre qui s’appelle, « La fin du travail ». Ce serait long de répondre, mais pour faire une réponse rapide, le problème d’éthique n’est pas tellement le fait que les gens n’ont plus de travail, le problème éthique c’est : le revenu des machines et des algorithmes qui remplacent ces personnes, comment est-il partagé ? C’est ça la vraie question. Ce n’est pas tellement le fait que les caissières de supermarchés ne travaillent plus qui est un problème. C’est le fait que quand elles sont remplacées par des robots, l’argent des robots ne va pas aux caissières de supermarchés. C’est là qu’est le vrai problème.

**Organisateur :** Dernière question là-bas, présentez-vous madame.

**Public :** Bonjour. Merci. Je suis chercheuse à la chaire humanisme numérique de Sorbonne Universités. Je m’intéresse beaucoup aux questions éthiques et numériques et la question que je voudrais vous poser, monsieur Abiteboul, c’est concernant la globalisation de ces questions éthiques et en particulier par rapport à l’exemple que vous citiez sur la neutralité des réseaux et la neutralité des contenus. Wikipédia, enfin Wikimedia a connu, en 2016, un véritable scandale au Nigéria puisqu’ils se sont aperçus qu’en ayant promu des plans ce qu’on appellerait data0, un peu comme ce que fait Facebook avec un projet Facebook 0 ou internet.org, avaient donc permis, en faisant un partenariat avec le Telco local, de donner accès gratuitement à Wikipédia pour la plupart des Nigérians sur leur téléphone. Et ils se sont aperçus que les usages de Wikipédia correspondaient à des usages de piratage et qu’on se servait de Wikipédia pour télécharger, partager des fichiers piratés. La question que je pose donc est la suivante, est-ce que, puisqu’on a vu aussi grâce à l’exposé de Milad Douheihi, qu’il s’agit bien d’une problématique également culturelle, de rappeler que le numérique est une culture, puisque ces plateformes sont globales comment est-ce que la communauté scientifique peut-elle s’emparer de ces questions éthiques en tenant compte de leur dimension globale ? Merci.

**Serge Abiteboul :** Vous avez absolument raison sur la globalité de la question et je pense que la réponse n’est pas une réponse de la communauté scientifique, mais c’est une réponse de la société en général. C’est-à-dire que c’est à la société, en général, de décider, de mettre des *guidelines* ; ça peut devenir des traités internationaux, mais c’est loin dans le futur. Ce n’est pas une question d’informatique. C’est-à-dire qu’en tant qu’informaticien moi, ce que je peux être amené à faire, je peux être amené à développer des algorithmes qui vérifient différentes propriétés, par exemple d’équité. Mais ce n’est pas à moi de fixer la règle de l’équité. La règle de l’équité c’est une règle qui doit être sociale, qui doit être décidée par tout le monde. Mon rôle, en tant que scientifique, est d’expliquer aux gens, d’éduquer les gens et de les aider à décider quelles règles ils veulent. Là où vous avez tout à fait raison, c’est qu’il n’y a absolument aucune raison pour que ces règles soient les mêmes dans tous les pays. Pour prendre juste un exemple qui est les robots soignants, les robots soignants sont beaucoup plus facilement adoptés en Asie qu’en Europe. Donc là on rentre dans des questions de société qui vont prendre beaucoup de temps à se définir.

**Public :** Est-ce que je peux juste intervenir une dernière une petite seconde. En particulier puisqu’on a quand même vu que la question est dans le code aussi, donc dans le fait de penser une certaine pragmatique du code donc de ses domaines d’applicabilité et de son extension, en tout cas jusqu’où peut aller le code sur les cultures locales, en fait. C’est pour ça aussi que je pense que ça reste quand même une question scientifique, mais peut-être avez-vous un autre avis là-dessus.

**Serge Abiteboul :** C’est une question scientifique dans le sens où l’adoption de ces approches transforme les cultures locales. Je ne sais pas si c’est ça sur lequel vous vouliez arriver, mais je pense que cette transformation ne concerne pas que les scientifiques.