

Capital

EN PARTENARIAT AVEC

ENGIE

Irène Balmès

"Le premier travail d'un data scientist, c'est de nettoyer la donnée"

Bernard Werber

"Un monde sans data, c'est un monde dans lequel les connaissances circulent moins vite"

LES BIG ENTRETIENS DU BIG DATA

12 EXPERTS FACE
AUX PROMESSES DE L'IA...
ET À SES LIMITES"

Luc Julia

"L'IA ? un outil ni bon, ni mauvais, mais qui ne fait qu'obéir à la main de l'homme."

Hervé Le Tellier

"Et si notre monde n'était qu'une vaste simulation?"

Laurence Devillers

"Une machine ne ressent pas d'émotion, elle capte juste des indices."



04___ Intelligence artificielle

Luc Julia 📍 Renault

06___ Futurologie

Bernard Werber 📍 Ecrivain, futurologue

08___ Extême Big Data

Gilles Dawidovicz 📍 Google Maps

12___ Ethique de l'IA

Valérie Beaudouin 📍 Telecom Paris

14___ Cybersécurité

Baptiste Robert 📍 Predicta Lab

16___ Robots émotionnels

Laurence Devillers 📍 Paris-Sorbonne

18___ Santé

Nozha Boujemaâ 📍 Oecd.ai

20___ Médecine prédictive

Christophe Zimmer 📍 Institut Pasteur

23___ Objets connectés

Fred Potter 📍 Netatmo

24___ Data science

Irène Balmès 📍 Kynapse

26___ Energie

Yves Le Gélard 📍 Engie

29___ Simulation

Hervé Le Tellier 📍 Ecrivain, Prix Goncourt 2020

Le nombre de données produites dans le monde double chaque année...

Mais que faire de toutes ces données ?

Concrètement, qu'est-ce que la donnée, cet actif que l'on produit chaque jour en quantités phénoménales (de l'ordre de 1,7 Mo par personne et par seconde dans le monde, soit l'équivalent de deux minutes de musique en MP3) ? À l'état brut, la donnée décrit une réalité élémentaire, comme le résultat d'une mesure (la température d'une pièce par exemple). Elle ne prend de la valeur qu'une fois massivement agrégée - on parle alors de big data, ou mégadonnées en français - pour être ensuite convertie en information exploitable (les variations de température dans un bâtiment) puis en connaissance servant à gagner en efficacité (réaliser des économies d'énergie). Voilà pour le principe de base. Car le big data implique surtout d'être capable de traiter d'immenses quantités de données hétérogènes, faisant apparaître des liens inattendus, des structures cachées dont la découverte permet de formuler de nouvelles hypothèses. C'est le travail des data scientist, ces nouveaux spécialistes de l'exploitation des données.

Si la donnée est le carburant de l'économie numérique, le principal moteur capable de la transformer en «or informationnel» n'est autre que l'intelligence artificielle et ses technologies dérivées, le machine learning (apprentissage automatique) et le deep learning (apprentissage profond inspiré du fonctionnement du réseau neuronal humain). Associer big data et IA a donc plusieurs vertus : cela permet de régler, instantanément ou presque, des problèmes autrefois chronophages, et d'automatiser des processus complexes.

L'objectif des 12 entretiens que vous lirez dans ces pages est donc de faire le tour des questions qui se posent à l'aube de 2022 sur le big data, de s'interroger sur la pertinence d'un recueil de plus en plus massif de données... et d'imaginer des façons d'utiliser cette masse d'informations au service du plus grand nombre.

SUPPLÉMENT
PUBLICITAIRE RÉALISÉ
EN PARTENARIAT
AVEC ENGIE

Luc Julia

“En soi, l’intelligence artificielle n’est ni bonne ni mauvaise. C’est juste un outil, comme un marteau.”



L’INTELLIGENCE ARTIFICIELLE, c’est un marteau, nous dit Luc Julia, ingénieur et informaticien, co-créateur de Siri, l’assistant vocal d’Apple et aujourd’hui directeur scientifique chez Renault, après être passé par le service innovation de Samsung. Un marteau car c’est un outil qui, en soi, n’est ni bon, ni mauvais, mais qui ne fait qu’obéir à la main de celui qui le fait fonctionner : l’homme.

Il y a aujourd’hui plus d’un milliard d’iPhone dans le monde. Un milliard d’appareils équipés donc de Siri, l’assistant vocal que Luc Julia a largement contribué à mettre au point. Pourtant, il ne fait pas partie des smartphones addict. Au contraire, dans une interview récente, il déclarait : “Je n’utilise pas le téléphone. Je déteste le téléphone. Je ne réponds jamais.”

Pourquoi n’aimez-vous pas à ce point le téléphone ?

Luc Julia ■ Parce que je trouve que l’interface du téléphone est absolument insupportable. Ce n’est pas du tout naturel. Pas plus naturel d’ailleurs qu’une souris ou qu’un clavier. Et puis je n’aime pas l’idée d’avoir dans la poche un truc qui vous sonne, comme un domestique. Je trouve ça pas pratique du tout. Et même en ayant ajouté un écran dessus... Je vois les gens penchés dessus toute la journée, tout ce que ça va donner, ce sont des problèmes de dos !

A 9 ans, vous avez créé un robot pour refaire votre lit à votre place... C’est vrai ou c’est une légende ?

C’est tout à fait vrai. Malheureusement je n’ai même pas de photo de lui. J’avais mis des centaines d’heures à le mettre au point, donc beaucoup plus que le temps que j’aurais mis à faire des milliers de fois mon lit j’imagine, mais il marchait très mal. Il se déplaçait le long du lit et il bordait les draps. Mais il a déchiré pas mal de draps aussi ! Ça ne marchait pas très bien, mais c’était très rigolo.

J’aime bien faire démarrer l’automatisation à 1642, quand Pascal a inventé la Pascaline, la vraie première machine à calculer. Donc pour moi, l’invention de l’intelligence artificielle, elle, remonte à cette date-là.

Vous n’aimez pas tellement le terme d’intelligence artificielle. Vous préférez celui d’intelligence augmentée. Pourquoi ?

Si on devait définir l’IA aujourd’hui, je dirais que c’est une boîte à outils, avec des intelligences artificielles dedans, des outils dedans. Et si je parle d’intelligence augmentée, c’est pour dire que c’est notre intelligence qui est augmentée par ces outils. Parce que dans chacun des domaines dans lequel l’outil va être bon, il va nous augmenter dans ce domaine particulier, mais cette intelligence va être limitée. C’est pourquoi ces intelligences ne sont pas comme nous et qu’en réalité, elles ne peuvent pas rivaliser avec l’intelligence humaine. Nous sommes, nous, capables de voir les choses beaucoup plus largement et globalement. Elles, elles sont spécialisées dans des domaines très particuliers et fins, où elles sont, certes, plus fortes que nous. Mais elles ne sont pas plus intelligentes !

Illustration : Kagan McLeod

C’est pour cela aussi que vous dites qu’une IA n’est ni plus ni moins qu’un marteau...

Exactement. Et comme pour tout outil, c’est nous qui avons le manche. On peut utiliser un marteau à bon escient en plantant un clou, ou d’une mauvaise manière en tapant sur la tête du voisin. C’est pour ça qu’un jour la société décide des bons ou des mauvais usages d’un outil et va réglementer la façon dont on les utilise. C’est ce qui s’est passé avec le marteau. Avec la bombe atomique. Et c’est ce qui doit se passer avec les IA.

L’IA autonome, ça reste un fantasme ? C’est toujours l’homme qui décide ?

Bien sûr. Pour fonctionner, les IA ont besoin de données et de règles. Les règles, c’est nous qui les écrivons. Et les données, elles sont créées par l’homme, ce sont par définition des données historiques, de choses passées qui ont existé à cause de nous ou grâce à nous. Nous en avons donc quoi qu’il arrive le contrôle. On peut décider de biaiser les règles, de biaiser les données, mais dans tous les cas, c’est nous qui décidons.

Pensez-vous qu’on puisse mettre le monde en équation ?

Je suis un mathématicien, j’aimerais pouvoir décrire le monde juste en mathématiques, mais la vérité c’est qu’on n’y arrive pas. Et c’est pour cela qu’il y a d’autres matières, comme la physique ou la biologie, car les mathématiques ne peuvent pas décrire tout ce qu’il y a dans le monde. Mais c’est parce que les mathématiques sortent de notre cerveau : là encore, c’est une vue de l’esprit, une invention artificielle de l’homme. Alors que la physique ou la biologie sont des descriptions de la nature. Et là, beaucoup de choses nous échappent. Si vous demandez à des spécialistes en biologie de vous décrire le fonctionnement de notre cerveau, ils pourront vous décrire 20% ou 40% de ce fonctionnement. Mais encore beaucoup de choses leur échappent. Donc les mathématiques n’expliquent pas tout.

Donc tant qu’une IA sera basée sur des mathématiques, elle ne pourra pas s’attaquer au réel ?

Exactement. Il y a beaucoup de choses qu’on ne comprend pas et qui ne peuvent pas être décrites par les mathématiques.

LUC JULIA EN CHIFFRES

5

Comme 5 ans, l’âge qu’il avait quand il a décrété qu’il serait chercheur au CNRS. Et il est devenu chercheur au CNRS. Précoce, la légende veut qu’il ait, à l’aide d’un vieux moteur d’aspirateur, imaginé tout petit un robot pour faire son lit à sa place.

22^e

C’est sa position dans la liste des 50 Français les plus influents dressée en 2018 par le magazine Vanity Fair.

1958

Année où le terme « Intelligence Artificielle » apparaît pour la première fois. « J’aime bien faire démarrer l’idée d’une Intelligence Artificielle à 1642, quand Pascal a inventé la Pascaline, la première vraie machine à calculer. Mais la volonté de créer des machines intelligentes, elle, remonte à la nuit des temps, certainement bien avant 1956 quand le terme a été utilisé pour la première fois. »

5,19 milliards

C’est le nombre d’utilisateurs de téléphones mobiles dans le monde. Soit 67% de la population. Et alors qu’une bonne partie de ces mobiles – 1 milliard environ – sont des iPhones, équipés donc de Siri, l’assistant vocal que Luc Julia a largement contribué à mettre au point, il ne fait pas, lui, partie des smartphones addict...

C’est-à-dire qu’il faudrait réussir, demain, à nourrir ces intelligences avec autre chose que des chiffres ?

Oui, peut-être. C’est du domaine de la science-fiction. Mais on parle beaucoup d’informatique quantique et c’est un peu l’idée. De passer d’un univers fait de 0 et de 1 à un monde plus physique, qui se rapproche d’une description plus réelle du monde. Mais aujourd’hui on n’en sait rien. On en est au tout début seulement. Et puis l’ordinateur biologique, qui est celui qui serait le plus proche de nous... il est encore plus loin, lui ! ■

Écoutez le podcast : <https://rb.gy/fwj4jx>



Bernard Werber

"Grâce à la technologie, aujourd'hui, n'importe quel être humain connaît plus de choses que Léonard de Vinci."

L'ÉCRIVAIN BERNARD WERBER milite pour une science-fiction résolument optimiste : pour cet ancien journaliste scientifique, l'IA et les machines doivent nous aider à imaginer un futur meilleur. Il nous explique comment.

Bernard Werber, est écrivain et «futurologue», comme il le dit lui-même. Depuis la sortie de son roman "Les Fourmis", en 1991, son succès ne se dément pas. Il est l'un des auteurs français contemporains les plus lus au monde, avec 30 millions d'exemplaires vendus. Ancien journaliste scientifique, son œuvre fait se rencontrer mythologie, spiritualité, philosophie, science-fiction, biologie et futurologie. Il a créé sur Internet L'Arbre des Possibles, un projet collaboratif initié pour imaginer de multiples scénarios autour des futurs possibles de l'humanité.

Dans vos livres, il est souvent question d'encyclopédies, de connaissances et de savoirs que l'on accumule. Est-ce que vous diriez qu'aujourd'hui, la data, c'est une autre forme de savoir encyclopédique ? Autrement dit, est-ce que le savoir qui était hier dans les lettres est désormais dans les chiffres ?

Bernard Werber — j'ai été journaliste scientifique, donc habitué à jongler entre les mots et les nombres et à jouer avec. Et depuis, j'essaie de développer mes deux cerveaux : gauche et droit ! Mais c'est vrai qu'aujourd'hui, les ordinateurs sont capables d'accumuler bien plus de connaissances que n'importe quel cerveau humain. Quand Deep Blue a battu Kasparov aux échecs,

c'est d'abord parce qu'il avait accumulé plus de scénarios que le cerveau de Kasparov, qui était pourtant très performant. Mais là où l'ordinateur nous bat, c'est qu'il va pouvoir tester l'ensemble de ces scénarios alors que l'humain, lui, ne va tester que ceux qui lui semblent probables, utiles, raisonnables ou intelligents. Et c'est ce qui fait qu'on rate une partie des visions de possibilités.

Puisque l'ordinateur nous bat, pourquoi alors avoir lancé ce projet d'arbre des possibles, qui regroupe près de 10 000 scénarios autour de futurs possibles pour l'humanité ?

Parce que si un cerveau humain est limité, je pense que plusieurs cerveaux connectés peuvent réussir à compenser ce déficit. Je crois beaucoup à ce que j'ai observé dans les fourmilières : la conscience collective ou l'intelligence collective, qui surpasse l'intelligence individuelle. L'une des devises des fourmis, dans le roman, c'était «1+1=3» car je crois qu'en se connectant on arrive à trouver des choses qui n'existent ni dans le cerveau de l'un, ni dans le cerveau de l'autre. Moi je crois à la capacité de l'homme à utiliser son imaginaire pour ne plus avoir aucune limite.

Vos contributeurs y parviennent-ils ?

Ce n'est pas une science exacte : ce sont des intuitions, une banque de données de scénarios de futurs possibles, qui n'a pas d'autre prétention que cela. Évidemment la réalité est toujours plus créative... mais il y avait déjà des scénarios qui parlaient de pandémie ou de virus il y a plusieurs années. Je pense qu'il y a des cycles et qu'en observant le passé on peut y trouver des graines pour imaginer le futur. Ce que je regrette, c'est que le nombre de scénarios pessimistes est

légèrement supérieur au nombre de visions optimistes. C'est vrai que, lorsqu'on suit l'actualité, c'est difficile de se dire qu'on va imaginer un futur où tout va s'arranger, dans lequel nous vivrons en paix... Pourtant, je suis persuadé qu'il faudrait inventer une nouvelle science-fiction qui privilégie le côté «on va réussir, tout va s'arranger...» J'espère que, grâce aux ordinateurs, on va trouver de nouvelles solutions et se rendre compte que l'intelligence artificielle est une manière pour l'homme de penser plus loin, plus fort et de gagner du temps.

Une intelligence artificielle fonctionne en analysant des données qui existent, donc nécessairement du passé. Est-ce que ça veut dire qu'on ne peut prévoir le futur qu'en se basant sur ce qui s'est déjà produit ?

C'est vrai que, pour l'instant, on utilise la connaissance du passé comme outil de compréhension du futur. Je pense qu'on peut faire mieux, mais ça va demander encore plus de puissance de calcul. Peut-être qu'avec les ordinateurs quantiques on va réussir à avoir ce niveau d'imagination, au-delà de l'imagination humaine. En attendant, l'imagination humaine est quand même très très très puissante. Le prochain challenge, après battre les hommes aux échecs, sera de les battre dans le domaine de l'imaginaire. C'est pour cela aussi que j'ai créé l'arbre du possible : pour qu'on soit très forts dans le domaine de l'imaginaire et que les ordinateurs ne nous rattrapent pas de sitôt !

Face à la technologie, vous êtes donc plutôt optimiste ?

Grâce à la technologie, aujourd'hui, n'importe

BERNARD WERBER EN CHIFFRES

1 463 pages

C'est le nombre de pages du dernier manuscrit des Fourmis... avant que son éditeur ne lui demande de le ramener à 350 pages. En tout, il a rédigé 18 variantes complètes et différentes de son roman et travaillé dessus 12 ans avant qu'il ne soit publié.

15 %

C'est le pourcentage que représentent les fourmis dans la totalité de la masse de la vie animale sur terre. Oui, 15% de la masse totale de la vie sur Terre est fourmis !

30 millions

C'est le nombre de livres qu'il a vendus dans le monde. Dont 12 millions à l'étranger, dans une trentaine de langues, 35 exactement, ce qui fait de lui l'un des auteurs français les plus lus dans le monde, si ce n'est le numéro 1.

9

C'est le chiffre au-delà duquel il est interdit d'aller sous peine d'être d'hérésie, par les gardiens du savoir qui seuls sont autorisés à compter au-delà de neuf, dans Le Mystère du chiffre, une nouvelle tirée de son Arbre des possibles..

quel être humain connaît plus de choses que Léonard de Vinci. Comprendre le Big Bang, apprendre à parler japonais, accéder à des millions de livres et à des milliards d'informations, tout cela est possible derrière son ordinateur. Imaginons un monde sans Internet, sans informatique, sans data, c'est un monde moins intéressant et dans lequel les connaissances circulent moins vite. Après, les machines ne sont pas responsables de la façon dont on les utilise. Le feu peut servir à réchauffer et cuire des aliments ou à mettre le feu à une maison. Mais le feu en soi n'est ni bon ni mauvais. Ce sont les intentions des utilisateurs qui comptent, selon qu'ils veuillent faire le mal ou le bien. —

Écoutez le podcast : <https://rb.gy/zdzzpu>

Gilles Davidowicz

"Bientôt, Google vous indiquera le temps d'attente à la caisse du supermarché ou les places libres dans le RER..."



GOOGLE MAPS ne cesse de lancer de nouveaux services : plus d'une centaine sont annoncés dans les mois qui viennent dans le monde ! On vous en donne un avant-goût avec le directeur pour l'Europe du sud de Google Maps.

Notre smartphone à la main ou en poche, nous sommes tous des capteurs. Capteurs de position. De température. De vitesse. De bruit. Ces données que nous recueillons en permanence permettent à Google Maps de nous informer en temps réel : connaître l'affluence à la caisse du supermarché, trouver une place libre dans un train, déterminer le meilleur itinéraire à vélo, avec le moins de côtes possibles, ou le trajet le plus rapide ou le moins polluant en voiture. Mais comment Google nous rend-il ces services au quotidien ? Quelles sont les données que ses IA manipulent et comment en tire-t-on des informations pertinentes ? Les explications de Gilles Dawidowicz, géographe et planétologue, en charge de Google Maps pour la France et l'Europe du Sud.

Vous avez récemment lancé Timelapse, un service qui permet de voyager dans l'espace... et de remonter le temps, pour lequel vous avez compilé 37 ans d'archives de photos satellites de notre planète, soit 20 petaoctets de données... Mais ça représente quoi exactement 20 petaoctets ?

Gilles Dawidowicz — Si je devais prendre une image, imaginez qu'un octet de donnée ça soit

un grain de riz. Un kilo octet ça ferait une tasse. Un mega ça ferait huit sacs de riz. Un giga ça ferait trois conteneurs de riz. Un tera ça ferait deux navires porte-conteneurs de riz. Et un péta ça couvrirait la ville de Besançon. Donc là, on a 20 petaoctets, ça ferait vingt fois la ville de Besançon recouverte de riz ! On n'est plus dans le big data, mais dans l'extrême big data !

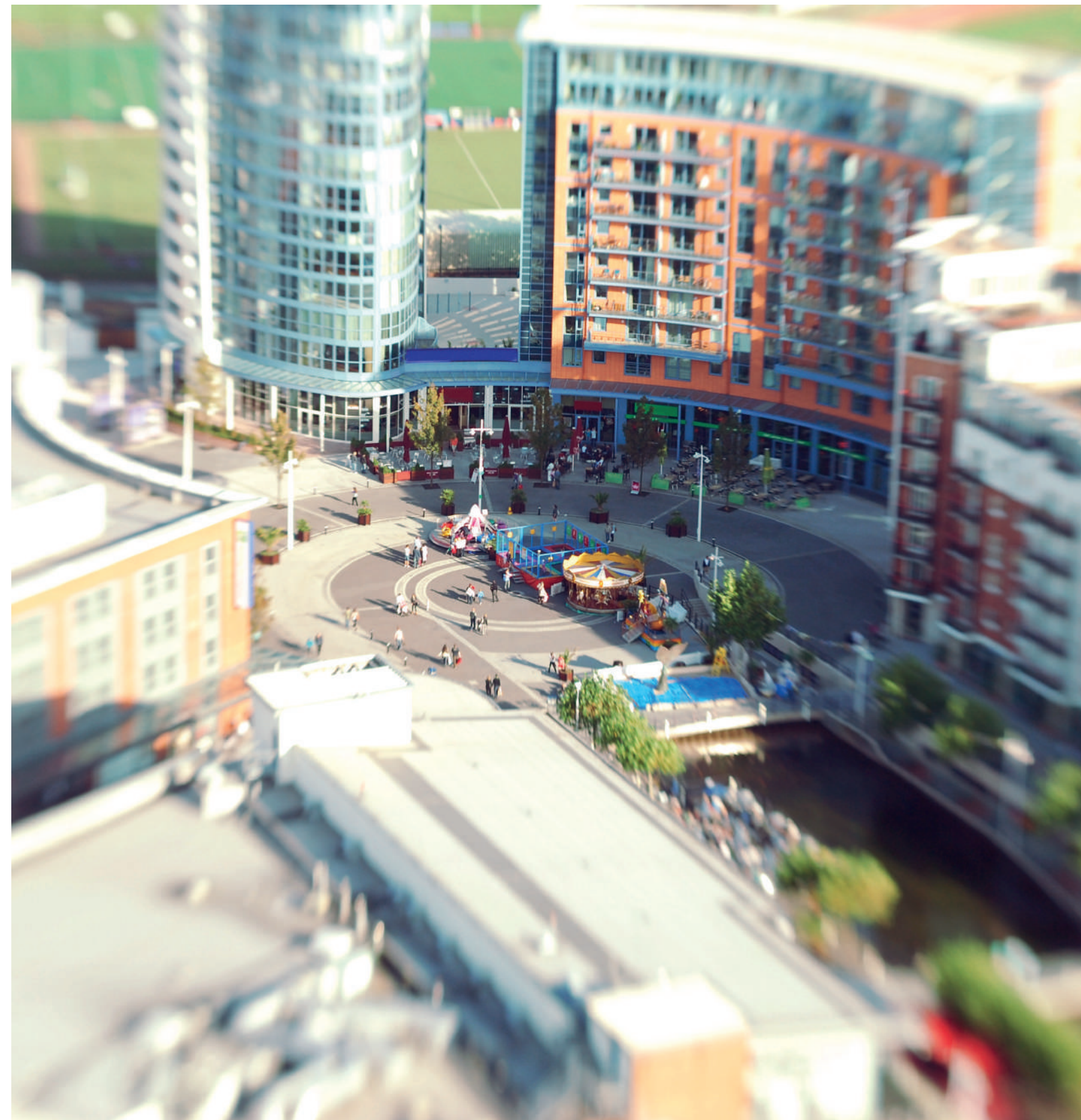
Vous disiez récemment que si vous aviez dû traiter ces données sur des ordinateurs classiques, cela vous aurait pris 228 ans de calcul !

C'est pour cela que, pour traiter ces données, on a collaboré avec de grandes universités, notamment aux Etats-Unis, et des organismes comme l'agence spatiale européenne ou la NASA, pour organiser ces données et les traiter grâce à de l'intelligence artificielle.

A quoi a servi l'IA dans ce processus ?

Concrètement, cela nous a permis par exemple d'effacer pixel par pixel les nuages dans les zones où il y en avait. Pour cela on a appris à la machine à faire la distinction entre un nuage et la neige par exemple, parce que vu de l'espace, un pixel blanc, c'est un pixel blanc. Les êtres humains, des photoanalystes, mettent des années à apprendre à reconnaître les formes, les formations, les équipements au sol. Nous, on a appris à la machine à le faire bien plus rapidement. Et puis évidemment il y a des endroits sur terre où il y a des nuages toute l'année. Là on a dû faire appel à d'autres technologies, notamment avec des photos aériennes, à plus basse altitude, qu'il a ensuite

Illustration : Kagan McLeod - iStock, Rob Evers



fallu mélanger avec les images satellites. De sorte qu'on a obtenu sur les 37 dernières années, une représentation parfaite de la terre par année. Ensuite, toujours avec l'IA, on peut mettre cela en animation et remonter le temps pour voir les transformations des paysages un peu partout sur la planète. Et prendre ainsi conscience de la façon dont la planète évolue et de l'impact de l'homme et de ses activités sur l'environnement.

Vous expliquez que vous avez utilisé des données dormantes que vous avez dû retravailler, parce que 80% d'entre elles n'étaient pas "calibrées" pour être utilisées. Qu'est-ce que ça veut dire ? L'usage de certaines d'entre elles n'avait pas été pensées avant d'être stockées ?

On a fêté récemment les 60 ans du vol de Youri Gagarine dans l'espace, quelques années après que les premiers satellites aient commencé à collecter des informations et des photos. Très vite, avec ces moyens, on est arrivé à stocker des millions de photos satellites. Mais depuis, les technologies ont évolué, les propres caméras de ces satellites ont changé. Il a donc fallu sortir ces données dormantes du tiroir et les retraiter pour pouvoir les utiliser.

Vous disposez de vos propres données historiques également : est-ce que ce sont ces données, et l'utilisation de cet historique du passé, qui permettent à Google Maps d'avoir cette minute de précision lorsque vous indiquez des temps de trajet en voiture, par exemple ?

Pas complètement. Les données historiques sont très importantes mais on a vu leurs limites avec la crise Covid. Car si les gens changent leurs habitudes, on ne peut plus se baser sur ces données. Du coup c'est le mélange intelligent de ces données historiques, des données en temps réel et des données officielles transmises par les partenaires locaux qui nous permettent aujourd'hui d'avoir cette minute de précision.

Vous qui scrutez les déplacements en temps réel, quelles sont les évolutions notables que vous avez observées avec la crise Covid ?

La pandémie a eu un impact considérable sur nos modes de transport et de déplacement. Les calculs d'itinéraires à vélo sur Google Maps ont

ainsi augmenté de façon incroyable : +45% en France et même 56% à Paris. Y compris avec des temps de trajet plus longs. C'est-à-dire que non seulement il y a plus de gens qui se sont reportés sur le vélo, mais qu'en plus ils ont pédalé plus longtemps.

Vous publiez régulièrement des rapports sur les déplacements. À qui et à quoi servent-ils ?

Nous travaillons beaucoup avec les responsables politiques, pour qui ces rapports de mobilité permettent de mettre en place de nouvelles politiques publiques, comme la création de pistes cyclables, des rocade à aménager, des nouvelles lignes de bus, etc. Une partie de ces données est aussi mise à disposition du grand public, en open source, pour contribuer à la réflexion collective sur nos modes de déplacement et l'aménagement du territoire.

C'est également dans cette optique de prise de conscience individuelle et collective que vous allez lancer un nouveau service qui permet de choisir l'itinéraire qui a le moins d'impact sur l'environnement. Comment ça marche ?

En effet, nous allons mettre à disposition de tout le monde un certain nombre de services afin d'aider à réduire notre empreinte carbone et l'impact de nos activités humaines partout sur la planète. Google Maps permet de calculer 1 milliard de kilomètres de trajets en transports en commun par jour dans le monde. Et plus de 50 millions de kilomètres d'itinéraires en voiture. Donc si on propose des itinéraires alternatifs, plus verts et qui permettent de moins consommer d'énergie, ça va forcément réduire les émissions de CO₂. On va ainsi, par exemple, proposer des itinéraires en nous basant sur le pourcentage de pente, pour proposer de les contourner si c'est possible. On va aussi calculer les itinéraires de façon plus personnalisée, en fonction du type de véhicule et d'énergie qu'il utilise, de la circulation, de la météo... Bref, on va agréger tout un tas de données différentes pour obtenir ce résultat, y compris un certain nombre de données en temps réel, qui nous sont notamment transmises par nos utilisateurs,

Est-ce que cela signifie que nous allons tous devenir des capteurs humains ?

C'est vrai que nous sommes tous devenus des capteurs : des geo-capteurs, des photo-capteurs

GILLES DAWIDOWICZ EN CHIFFRES

9
C'est le nombre d'années qu'il a passé chez Google, à travailler sur Google Maps.

1
C'est en milliard le nombre d'utilisateurs réguliers de Google Maps et Google Earth.

24
C'est en millions le nombre de photos qui ont été compilées pour lancer Time Lapse, le nouveau service de Google Earth qui permet de voyager dans l'espace et le temps avec 37 ans d'archives de photos satellites de notre planète.

20
C'est en petabytes le volume de données qui ont été manipulées pour réaliser ce nouveau service.

avec nos smartphones, nos équipements. Notre force, chez Google, c'est d'avoir des produits utilisés par des millions de personnes et donc d'avoir des millions de sources qui nous alimentent et peuvent nous servir pour les calculs d'itinéraires notamment. Un exemple : il y a quelque temps aux Etats-Unis un pont s'est effondré suite à un tremblement de terre. Trente secondes après, Google Maps était mis à jour et on a pu calculer un nouvel itinéraire pour les secours afin de leur permettre de contourner le pont et d'accéder au site. Nous ne garantissons évidemment pas une mise à jour en trente secondes partout dans le monde. Mais nous faisons 25 millions de mises à jour quotidiennes dans Google Maps. Et les contributeurs de Google Maps, ceux qui mettent à disposition des informations, sont des gens sérieux, qui renseignent des informations valables. Je pense aussi aux "local guides", ces utilisateurs qui renseignent des informations sur l'accessibilité d'un restaurant ou son menu par exemple, en postant des photos et des commentaires.

On parlait de données dormantes au début de cet entretien. Là, on parle beaucoup de données en temps réel, liées à l'instant. Comment faire le tri entre les données à vocation éphémères et celles que l'on peut conserver dans le temps, car elles auront une utilité pour ces services que vous rendez ?

C'est une question difficile, à laquelle peuvent répondre des datas scientists : savoir quoi conserver, quelles sont les données pertinentes. Parce que l'humanité continue chaque année à produire toujours plus de données - leur nombre double tous les deux ans - et rien n'indique que ça va s'arrêter. Après, ce sont des questions d'arbitrage, certaines données sont mises de côté et attendent que quelqu'un ait une idée pour les utiliser et voir si elles peuvent avoir de la valeur.

Vous avez prévu de lancer de nombreux nouveaux services sur Google Earth et Google Maps dans les prochains mois, dont certains, justement, sont basés sur les informations apportées par les utilisateurs. Pouvez-vous nous en dire plus ?

On a des services très concrets qui vont rapidement être disponibles, comme l'indication de la température dans un lieu donné. Savoir si une rame de métro est plus ou moins remplie et s'il n'y fait pas trop chaud. Identifier les places libres dans un RER ou dans un train. On pourra aussi connaître l'affluence dans les lieux publics, les magasins. Savoir s'il y a la queue à la caisse du supermarché ou du monde à la pharmacie pour choisir le meilleur moment pour aller faire ses courses, en évitant les pics d'affluence. Des services très concrets et très locaux. Mais que nous allons essayer de proposer et rendre accessibles partout dans le monde.

Est-ce qu'il y a une idée reçue sur l'intelligence artificielle qui vous énerve ? Ou un cliché auquel vous aimeriez tordre le cou ?

L'intelligence artificielle a une puissance incroyable, mais qui n'a rien de magique. C'est un outil parmi d'autres. Un outil très performant certes, mais qui ne peut pas régler tous les problèmes. 🍌

🔥 Écoutez le podcast : <https://rb.gy/07cohn>

Valérie Beaudouin

"Il faudrait passer du big data au smart data, car beaucoup ne signifie pas forcément qualité"



VALÉRIE BEAUDOUIN est professeur de sociologie à Telecom Paris, spécialiste des technologies numériques. Spécialiste de la statistique textuelle, elle a été pendant 12 ans au cœur de la R&D d'Orange, où elle a mené de nombreuses études d'usage autour des nouveaux services et produits.

Vous enseignez l'éthique au sein du master en intelligence artificielle de Telecom Paris. En quoi cette discipline a-t-elle sa place dans un tel cursus ?

Valérie Beaudouin ● Il y a une puissance énorme autour de ces algorithmes et des risques de mauvais usages de ces outils. Certaines technologies qui inquiètent, sont généralement complexes à expliquer. Elles posent même problème aux chercheurs du domaine. Par exemple, on ne sait pas exactement comment fonctionnent les réseaux neuronaux utilisés dans les programmes d'apprentissage, ce qu'on appelle le deep learning. On ignore comment ce réseau est arrivé à prendre telle ou telle décision. Or si on ne peut pas expliquer à une personne pourquoi tel crédit lui a été refusé ou pourquoi elle est désavantagée sur telle autre chose, suite à ce qu'a proposé l'algorithme, cela crée de la défiance. On a des machines qui apprennent par imitation, à partir de grosses bases de données, mais on a parfois du mal à expliquer comment l'algorithme a produit tel ou tel résultat.

Est-ce que, au-delà du fonctionnement des algorithmes, le problème ne vient pas aussi de ce qu'on leur donne pour

fonctionner ? En clair, des données ?

C'est vrai qu'on parle beaucoup désormais des risques de biais. Les outils d'intelligence artificielle s'appuient sur l'existant pour construire des modèles, or il y a des discriminations dans nos sociétés. Donc, par essence, ces outils vont reproduire les inégalités et les formes de discrimination qui sont déjà inscrites dans notre société.

Vous dites aussi qu'il faut être attentif à la constitution des bases de données en matière de développement durable. C'est-à-dire ?

Les outils de machine learning sont très consommateurs en énergie. Un des grands axes de recherche en ce moment c'est donc comment essayer de construire des échantillons de taille raisonnable pour réduire le coût énergétique de fonctionnement des machines. Donc moins de données, mais plus pertinentes et mieux contrôlées. Il faudrait passer du big data au smart data, car beaucoup ne signifie pas forcément qualité. C'est important d'aller vers une recherche de qualité et de précision.

Vous dites aussi que cela pourrait conduire les entreprises qui utilisent l'IA à pratiquer l'«éthique washing», soit communiquer auprès du public en utilisant l'argument éthique... sans réellement agir sur le sujet.

C'est déjà le cas. Quand on regarde le champ de l'éthique de l'IA, c'est un énorme champ de recherche. Il existe plusieurs centaines de rapports de recherche et de documents sur le sujet. La question qui peut se poser est : pourquoi une telle prolifération autour de ces questions

d'éthique ? Ce n'est pas du tout habituel. Quand Internet est né, les informaticiens et ingénieurs n'écrivaient pas des rapports sur tous les risques qu'allaient représenter Internet. Et quand on regarde aujourd'hui dans le détail, on voit que ce sont principalement les grands groupes internationaux qui prennent la parole sur ce sujet de l'éthique. Comme s'ils prenaient date en faisant cela et disaient «nous avons conscience qu'il y a des risques autour des technologies que nous sommes en train de mettre en place. Nous sommes des acteurs majeurs du développement de l'IA et en même temps nous sommes des acteurs responsables qui prennent en compte les difficultés liées à cela.»

Quel est le problème alors ?

On peut se demander si ce n'est pas une façon d'éviter les réglementations que pourraient mettre en place les États pour limiter les risques liés à ces technologies. Autre hypothèse, ils font cela car les chercheurs dans le domaine de l'IA sont de plus en plus sensibles à ces questions et aux risques que représentent les technologies qu'ils sont en train de développer. Et donc pour les entreprises c'est un moyen de cultiver leur marque employeur car ce sont des spécialistes extrêmement demandés sur le marché du travail. La compétition est très forte pour attirer les meilleurs talents.

Vous avez d'ailleurs vous-même eu l'idée de cette formation à l'éthique après avoir entendu le témoignage d'un data scientist qui avait décidé de changer d'orientation après avoir réalisé que l'IA pouvait détruire des emplois... y compris ceux de ses amis et

VALÉRIE BEAUDOUIN EN CHIFFRES

80 000

C'est le nombre de vers de Corneille et de Racine qu'elle a analysé pour en calculer la structure rythmique et essayer de déterminer la logique de l'agencement des mots dans les vers.

2

Comme la morale à double face, celle qu'elle expérimente dans ses écrits pour l'Oulipo, un groupe d'auteurs qui pratiquent ce qu'on appelle l'écriture sous contrainte. Sa contrainte à elle étant donc d'écrire des textes autour d'éléments parallèles ou opposés.

27

C'est le pourcentage de femmes dans le monde numérique.

2 heures

C'est l'écart de temps de travail domestique entre les hommes et les femmes. Les femmes font deux heures de travail en plus que les hommes à la maison.

collègues.

Oui, exactement. Parmi nos étudiants, les préoccupations majeures par rapport à l'IA tournent d'ailleurs autour de l'emploi. Ils ont parfaitement conscience qu'on est dans un cycle de création destructrice, avec de vieux métiers qui disparaissent et de nouveaux qui apparaissent et ils sont très sensibles à ça !

Est-ce qu'il y a une idée reçue sur la data qui vous énerve ?

Le machine learning, ce n'est pas la solution à tous les problèmes. Aujourd'hui avec de l'informatique classique, des règles simples, on peut réaliser beaucoup de choses. Or avec cette mode autour de l'IA on a l'impression que certains veulent absolument embarquer ces technologies-là alors que parfois des technologies plus simples seraient aussi performantes et moins coûteuses en énergie, en compétences et en ressources au sens large. ●

Écoutez le podcast : <https://rb.gy/vd5dw9>

Illustration : Kagan McLeod



Baptiste Robert

"Réfléchissez avant de donner vos données !"

BAPTISTE ROBERT est un «hacker éthique», connu sur Twitter auprès de ses 237 000 followers sous le pseudo d'Elliot Alderson, inspiré du héros de la série Mr Robot. Chercheur en cybersécurité, avec sa société, Predicta Lab, dont l'activité consiste à écouter les signaux faibles sur le Net pour détecter les événements, quels qu'ils soient, il fouine, explore, cherche dans tous les recoins du web, dans la masse des données parfois publiques, parfois cachées, celles qui ont un sens, un intérêt ou, tout simplement, qui ne devraient pas être exposées...

Vous êtes un hacker éthique, mais qu'est-ce que c'est ?

Baptiste Robert — L'idée générale de mon activité et de mon engagement est de trouver les failles de sécurité, dans les applis mobiles ou les sites Internet, d'entreprises de toutes tailles et même de gouvernements, et de les prévenir de l'existence de ces failles. D'abord en privé, de manière à leur donner la possibilité de réparer ces failles. Puis, si l'entreprise ne fait rien, en public pour la pousser à réparer la faille et protéger les données de ses utilisateurs.

Pourquoi est-ce si important de protéger les données ? Qu'est-ce que les données ont de si précieux ?

La donnée utilisateur a une grande valeur marchande. Vos données sont vendues, échangées, tous les jours pour mieux vous profiler et vous cibler : pour la publicité, évidemment. Car si je connais vos centres d'intérêt, je vais pouvoir vous pousser des publicités qui auront plus de chances de vous intéresser... et me donneront

plus de chances de vendre mon produit. De la même façon, dans un contexte politique, les réseaux sociaux ont une importance capitale. Or avec vos données personnelles, on peut cibler les infos, les nouvelles qu'on va mettre en avant. Votre fil sur Facebook n'est pas le même que celui de votre voisin car le but de Facebook est de retenir votre attention, de créer de l'engagement, pour que vous restiez le plus longtemps possible sur le site. Et c'est le profilage des utilisateurs qui permet l'augmentation de l'engagement.

Est-ce que toutes les données ont la même valeur ?

Tout ne se vaut pas non. Si j'ai le numéro de téléphone ou l'adresse physique du domicile d'un homme politique de premier plan, ça n'a pas la même valeur que si j'ai les vôtres par exemple ! Et si je sais en plus que vous avez l'habitude de discuter tous les soirs par téléphone à la même heure avec la même personne et que cette personne ne fait pas partie de votre entourage proche... c'est une donnée extrêmement précieuse que je vais pouvoir utiliser contre vous dans plein de contextes différents. Ce qu'il faut bien comprendre c'est que votre donnée toute seule ne vaut pas grand-chose : mais c'est le recoupement, le nombre d'informations, qui vont permettre de donner du sens, de vous profiler. Et donner la possibilité d'utiliser ces données contre vous.

Est-ce qu'il faut chercher à protéger toutes ses données de la même façon si elles n'ont pas toutes la même valeur ?

Pour moi, il faut surtout que les utilisateurs sachent à quoi vont servir leurs données... avant de les donner. Si on vous demande votre adresse

physique sur un site ou pour un service qui ne va jamais rien vous livrer... à quoi bon la leur donner ? Et si vous en donnez une autre, est-ce que le site va quand même continuer à marcher ? En général, la réponse est oui, vous pouvez donc mettre n'importe quoi.

Là on parle de données que l'on «donne» consciemment. Mais quand vous vous penchez sur les failles de sécurité, c'est plutôt pour limiter le risque de vol... Qui sont aujourd'hui les voleurs de données ?

On trouve pratiquement tous les jours sur le dark web des bases de données à vendre ou à acheter avec quantité d'informations, comme des mails, des mots de passe, des informations personnelles. Tout cela profite à des cybercriminels dont c'est devenu le métier.

Et combien ça coûte ?

Votre adresse seule n'est pas intéressante. Ce qui est intéressant, c'est le volume. Sur les forums, les bases proposées comptent souvent des millions d'entrées. On a vu récemment une base de données d'utilisateurs de Facebook proposée sur un forum, avec plus de 500 millions d'utilisateurs, répartis sur pratiquement tous les pays du monde. Et cette base était à vendre pour un peu plus de dix mille euros.

Quels conseils concrets pouvez-vous donner pour mieux protéger ses données ?

Déjà ne cliquez pas sur tous les liens : quand une fenêtre apparaît, lisez ce qu'on vous dit avant de tout accepter. Refusez aussi les cookies des pop-up RGPD. Les accepter, c'est un clic, mais les refuser, c'est seulement deux. Si vous les refusez, vous limitez le traçage et la récupération de données. Deuxième conseil, réfléchissez avant de poster quelque chose, notamment sur les réseaux sociaux : une fois que vous avez posté une photo, elle ne vous appartient plus. Savez-vous ce qu'elle va devenir ? Et si quelqu'un l'utilise dans un autre contexte est-ce que cela vous pose un problème ? Si oui, ne la postez pas. Enfin, troisième conseil : vous n'êtes pas obligé de dire la vérité. Dans la vie réelle, si je vous demande votre nom et votre prénom, vous allez probablement me dire la vérité, car nous sommes entraînés à cela. Sur les réseaux sociaux et Internet en général vous n'êtes pas obligé de dire la vérité. Ce n'est pas parce qu'un site vous

demande votre adresse que vous êtes obligé de mettre la bonne. Si le site n'en a pas besoin pour fonctionner, pourquoi donner le bon ? Vous allez lui donner gratuitement une information pertinente, qui a de la valeur. Donc réfléchissez avant de donner vos données.

Quelle est l'idée reçue sur la data qui vous irrite le plus ?

C'est de dire «Je n'ai rien à cacher». Car en réalité nous avons tous des choses à cacher, des choses qu'on ne veut pas montrer à nos proches, des choses qu'on ne veut pas rendre publique ou qui pourraient être utilisées contre nous. Or ce n'est pas interdit. On a le droit d'être nul à la guitare mais de vouloir en jouer sans que personne ne nous écoute... De ne pas avoir envie qu'une vidéo de nous en train de jouer soit publiée. C'est important de préserver cette notion de vie privée et de se dire que tout n'est pas fait pour être partagé ! Il y a des moments de la vie qui méritent de rester dans le cercle privé au lieu d'aller sur des plateformes qui vont les utiliser pour vous cibler et gagner de l'argent. —

Écoutez le podcast : <https://rb.gy/yduvvi>

BAPTISTE ROBERT EN CHIFFRES

31 000

C'est le code postal de Toulouse, la ville où il habite et travaille.

45

C'est le nombre d'épisodes de la série Mr Robot, dont le héros, Elliot Alderson a inspiré son pseudonyme sur Twitter.

237 800

C'est le nombre de ses followers sur Twitter, abonné à son compte Elliot Alderson, donc.

30 minutes

C'est le temps qu'il lui a fallu en février 2021 pour trouver d'importantes failles sur l'appli Koo, sorte de Twitter indien.

Laurence Devillers

"Ce ne sont pas les machines qui nous manipulent, mais les ingénieurs qui les conçoivent"



JOIE, TRISTESSE, PEUR, COLÈRE, SURPRISE OU ENCORE DÉGOÛT... Les machines sont-elles capables d'émotions ? Peuvent-elles les ressentir, ou simplement les capter et les simuler ? Le point de vue de Laurence Devillers, professeur en intelligence artificielle, spécialiste de l'éthique et des interactions homme/machine.

Un robot peut-il ressentir des émotions ?

Laurence Devillers — Le ressenti des émotions chez l'humain, est lié à ses viscères, ses entrailles, son corps... tout ce qu'une machine n'a pas. Une machine ne ressent pas d'émotion, elle capte juste des indices. Elle capte du signal vocal, de la vision, de la chaleur du corps, des indices physiologiques... Et elle réagit ensuite en fonction de paramètres qu'on a programmés. On pourrait par exemple imaginer qu'un robot puisse exprimer un degré de douleur, en fonction de variables programmées, lorsqu'on lui tape sur la main. Mais il s'agit de données déjà étalonnées, sur lesquelles la machine se base pour exprimer une intention qu'on aura programmée en réponse. C'est comme pour le langage : les machines ne comprennent pas ce qu'elles disent, c'est un simulacre de langage. Une machine va dire "je ressens de la douleur" mais elle n'en ressentira pas, c'est une simulation, un mensonge.

Les machines, mêmes intelligentes, ne sont donc pas capables d'improviser ?

C'est cela. Les machines sont plus intelligentes que nous pour calculer et trouver la meilleure solution parmi un ensemble de solutions possibles. Comme aux échecs par exemple, où les

machines nous battent systématiquement. Ou pour répondre à une question en allant chercher dans une encyclopédie en ligne. En revanche dans le vrai monde, en 3D, dans lequel il y a des choses qu'on peut voir sous différents aspects, l'humain est capable d'interpréter les choses, même s'il n'en a qu'une vision partielle. Alors que la machine doit connaître tous les possibles pour les interpréter.

Qu'est-ce qui lui manque alors ?

L'intuition. Dans tout apprentissage, il y a une part de compréhension rapide des choses. Un enfant, après avoir vu quelqu'un faire quelque chose deux fois, sait le faire. La machine, il lui faut bien plus de temps pour arriver à la même performance. Nous avons aussi une partie innée aussi que n'a pas la machine. Et nous n'apprenons pas uniquement par le langage, mais aussi par tous les autres sens que n'a pas la machine.

La différence aussi c'est que les enfants sont curieux de tout... Une machine est-elle, elle, capable de curiosité ?

Créer des machines curieuses, c'est un axe de recherche actuel, c'est vrai. Une machine curieuse, c'est une machine qui a un environnement connu et est capable d'aller chercher ailleurs. On peut le faire au premier degré : avoir un robot dans une pièce et le faire aller ailleurs. Dans le monde numérique, c'est envoyer le programme voir ailleurs, chercher à agrandir ses connaissances par aléatoire : si la machine reconnaît de façon aléatoire des choses qu'elle a déjà vu, elle peut apprendre de ces situations. La curiosité, c'est ce qui manque aux machines pour être autonomes : si les machines n'ont pas la possibilité d'apprendre seules, ça sera toujours à nous de

les nourrir en données. Pour apprendre plus, il faut être curieux, aller voir ailleurs pour enrichir ses connaissances. Donc forcément, les objets de demain, pour être autonomes, devront être doués de cette capacité d'apprendre sans être supervisés par l'homme.

C'est ainsi qu'on pourra créer des machines qui ressembleront vraiment à l'homme ?

Je me demande bien pourquoi, au-delà du fantasme récurrent de se prendre pour Dieu et de façonner un être à son image, on se complique autant la vie alors qu'on sait très bien donner la vie autrement ! On ne peut pas reproduire ce qu'est l'humain car, fondamentalement, on ne sait pas ce qu'est l'humain. On crée des objets de plus en plus complexes, avec des simulacres de ce que nous sommes, mais sans l'essence même de ce que nous sommes. La pensée par exemple, on ne sait pas ce que c'est.

C'est aussi pour cela qu'on a beau nourrir les machines de data et données, cela ne remplace pas la perception que nous, humains, avons du monde...

Effectivement, les machines ne sont pas capables d'avoir cette perception multi-facteur que nous avons. Nos émotions sont complexes : elles sont souvent plusieurs en même temps. Quand mon enfant dit une bêtise dans la rue, je vais le gronder et en même temps ça me fait sourire. Et dans ma voix il va entendre l'amusement en même temps qu'une espèce de colère jouée pour faire comprendre que ce n'est pas ce qu'il faut faire... Ce mélange d'émotions, ces émotions fines liées à notre culture, à l'interaction et au contexte, aucun des systèmes actuels de détection des émotions ne le prend en compte.

Est-ce que ces machines nous manipulent ?

Ce ne sont pas les machines qui nous manipulent, ce sont les ingénieurs qui les ont conçues qui nous manipulent. Je me suis amusée à faire une expérience avec une enceinte connectée : je lui disais "je suis triste" et elle me répondait "je suis triste avec vous". Je répète ça plusieurs fois et au bout d'un moment, après avoir épuisé toutes ses phrases, elle me propose de me prendre dans ses bras... Comme c'est une enceinte sans bras, je lui dis "où ça des bras" et là elle me sort "Des bras en plus, déménageurs

à côté de chez vous..." avec l'adresse. Voilà ce qu'elle sait faire, la machine : nous localiser et trouver autour de nous ce dont on peut avoir besoin. C'est très fort, mais ça n'a rien à voir avec un ressenti humain. A travers ces systèmes qu'on accepte chez nous, nous sommes des cobayes : de l'autre côté on change et on adapte les stratégies pour voir comment les humains réagissent et on crée une fake empathie.

Vous dites qu'un homme peut tomber amoureux d'une machine, en revanche, l'inverse vous n'y croyez pas...

Ce n'est pas question d'y croire ou pas, c'est que c'est juste impossible. Une machine simule, elle ne ressent rien. C'est nous humains qui, en face, projetons des ressentis dans la machine, dès lors qu'elle s'adresse à nous avec le langage ou même nous suive juste du regard. —

A lire : Les robots émotionnels, par Laurence Devillers, éditions de l'Observatoire.

Écoutez le podcast : <https://rb.gy/hq2jf6>

LAURENCE DEVILLERS EN CHIFFRES

87

C'est le pourcentage d'avis positifs et très positifs sur Amazon pour son livre "Les robots émotionnels".

1900

C'est l'année où pour la première fois un robot-humanoïde est apparu dans un film. C'était dans Coppelius, La Poupée Animée de Goers Méliès.

4

C'est le nombre d'émotions de base que ressentiraient les êtres humains : la peur, la tristesse, la colère, la joie.

11

C'est le nombre de commandements qu'elle propose dans son livre "Des robots et des hommes", pour réguler les relations entre humains et robots. Parmi eux : "Tu ne divulgueras pas mes données à n'importe qui. Tu pourras oublier quelque chose si je te le demande. Tu apprendras et suivras les règles de la société. Tu seras loyal et capable d'expliquer tes décisions. Et enfin tu seras bienveillant et utile. Et, pourquoi pas, doué d'un peu d'humour !" —



Nozha Boujemaa

"Comme les médicaments, les algorithmes peuvent aussi présenter des effets indésirables."

EN MATIÈRE DE SANTÉ, faudrait-il généraliser le don de données au même titre que le don d'organe ? Car si les machines sont techniquement capables de réaliser des diagnostics de pointe et que la médecine prédictive n'est plus de la science-fiction, il manque toutefois un élément crucial pour faire fonctionner les algorithmes correctement : des données de santé fiables, qualifiées et disponibles en grande quantité. Décryptage et explications avec Nozha Boujemaa, experte en intelligence artificielle, notamment dans le domaine de l'éthique et la santé.

Quel est le principal intérêt de l'IA en matière de santé ?

Nozha Boujemaa — Aujourd'hui, les techniques de reconnaissance d'images sont vraiment au point. Bien entraînée, une machine est capable d'aller chercher dans les images des indices qui ne sont pas toujours visibles à l'œil nu. Mais pour apprendre à faire cela, il faut que les algorithmes puissent avoir accès à une immensité de données, pour leur permettre d'avoir des redondances et donc d'identifier un objet, une forme, sous une multitude d'apparences. C'est vrai d'ailleurs pour la médecine comme pour la reconnaissance des visages ou objets en général. Mais en matière de santé, l'intérêt c'est d'être capable d'anticiper les signaux faibles. Car plus on est dans l'anticipation, plus la médecine est efficace et peut avoir un impact.

Or, selon vous, on ne dispose pas aujourd'hui d'assez de données pour entraîner correctement ces IA ?

Exactement et c'est là tout le problème. L'efficacité de l'IA c'est un duo : il faut des algorithmes qui fonctionnent et des données de qualité. L'un ne va pas sans l'autre, tout l'impact de l'IA dans la santé viendra du mariage de ces deux choses-là. Or pour l'instant, si les algorithmes prédictifs fonctionnent, en revanche on manque cruellement de données. Il faut, certes, que ces images soient protégées, car il s'agit de données personnelles et de données de santé. Mais le paradoxe c'est qu'en essayant de faire de l'éthique sur ce sujet, on en arrive à un comportement non éthique, puisqu'en freinant le développement de ces IA, on accepte l'idée qu'on ne pourra pas détecter une maladie grave avant l'heure. On accepte qu'un praticien dise à un patient, «revenez dans trois mois, on verra comment telle lésion a évolué», le privant de chance alors qu'une IA aurait pu analyser bien plus finement les premiers signaux d'une maladie par exemple.

Qu'est-ce qui bloque ?

Il faudrait sans doute créer une carte de "donneur de données" comme on avait celle de donneur d'organes ! Car pour moi, les données de santé sont un bien commun. Mais la difficulté, c'est que ceux qui produisent les données - je pense aux hôpitaux par exemple - ne sont pas forcément les mieux placés pour les diffuser. De plus, ces données ayant de la valeur, se mêlent des intérêts économiques qui rendent le sujet encore plus complexe. A partir du moment où une donnée est rare, pour l'obtenir il y a des circuits légaux... et d'autres moins. On l'a vu récemment avec une étude sur le Covid publiée dans The Lancet et qui citait presque 800 hôpitaux, des milliers de patients... Or une telle quantité de

données, c'est quasi impossible de se la procurer, il n'y a pas de baguette magique. D'ailleurs, plusieurs hôpitaux cités ont nié avoir participé à cette étude, qui a finalement été retirée.

Mais alors d'où venaient-elles, ces données ?

Il y a pas mal de sociétés qui offrent aujourd'hui des services gratuits à base d'IA pour les radiologues ou les hôpitaux... Sauf qu'on ne maîtrise pas le cycle de vie de la donnée une fois qu'elle entre dans ce type de service gratuit. Elles se retrouvent ensuite dans ce type d'étude sans que le producteur initial de la donnée ne soit au courant, ce qui peut introduire des biais car la donnée n'est pas toujours bien qualifiée. Or l'essentiel pour entraîner une IA, c'est que la donnée qu'on lui donne soit fiable et qualifiée. Sinon les résultats ne seront pas interprétables.

Pour résumer, vous dites donc qu'on a aujourd'hui des modèles d'intelligence artificielle capables de faire de la médecine prédictive, mais que pour les entraîner on a un problème de données... Et du coup un marché quasi parallèle de la "donnée" se développe et que cela peut donner, au final, des résultats biaisés ou, pire, faussés ?

Oui. Et c'est là tout l'intérêt d'une régulation internationale sur la gouvernance des données et en particulier les données de santé. C'est un enjeu majeur ! On doit avoir des algorithmes auditable, dont on puisse vérifier qu'ils font bien ce qu'ils disent faire. Quand un médicament est mis sur le marché, il a une notice qui décrit ses effets indésirables ou ses effets en dehors d'un certain cadre d'utilisation normale. Les algorithmes devraient avoir les mêmes précautions d'utilisation car, très clairement, ils peuvent présenter des effets indésirables. Il faut maîtriser la chaîne de bout en bout, et c'est ça l'IA de confiance.

A titre personnel, si demain vous deviez passer un examen médical, accepteriez-vous que les images soient ensuite utilisées pour entraîner une IA ? Et laisseriez-vous une machine établir un diagnostic ?

Oui, j'accepterai le partage des données, tout comme le don d'organes. Si ça peut aider à faire progresser la science évidemment, pourquoi je

NOZHA BOUJEMAA EN CHIFFRES

19 200

C'est le nombre de résultats que donne, en 0,37 seconde, Google, lorsqu'on tape son nom dans la barre du moteur de recherche.

17 minutes et 22 secondes

C'est la durée de sa conférence TedX sur l'IA de confiance et les algorithmes de recommandation, donnée à Saclay en 2019.

25

C'est le nombre de thèses qu'elle a dirigées.

10

Comme 10X plus de données de santé produites dans le monde aujourd'hui qu'en 2013. Le recueil et l'analyse de ces données ne cessent de croître... et de s'accélérer.

freinerais. Pour le diagnostic, je ne me méfierais non pas de la machine mais de ceux qui appliquent l'IA. Pour être sûre que celui qui est derrière la machine a pris toutes les précautions nécessaires pour utiliser à bon escient l'IA et la maîtriser, ainsi que les données utilisées et injectées dans cette machine pour donner ce résultat. Si ces conditions sont réunies, bien sûr que l'IA sera plus performante que l'humain et ne pourra qu'anticiper ce que l'humain n'est pas capable d'analyser aujourd'hui. C'est évident.

Quel est pour vous le message le plus important à faire passer en matière de données de santé ?

Pour que la promesse de l'IA en santé soit une réalité, il faut pouvoir disposer de données fiables et sourcées, il faut les partager, il faut légiférer de manière internationale sur la circulation des données, et comment elles sont partagées, pour créer de la confiance. Aujourd'hui, on fait encore trop au cas par cas, il n'y a pas d'approche généralisée. Or c'est vraiment l'enjeu de ces prochaines années. —

Écoutez le podcast : <https://rb.gy/rtamfn>

Christophe Zimmer

"Intelligence artificielle : la nouvelle façon de concevoir "in silico" les médicaments de demain"



ET SI VOTRE FUTUR MÉDECIN TRAITANT ÉTAIT UNE IA ? Une machine surentraînée à analyser toutes sortes d'images médicales pour y repérer la moindre anomalie et contribuer à concevoir ensuite des traitements sur mesure, et même de façon anticipée et prédictive. Quelles sont les réelles applications de l'IA en matière de santé, et ses éventuelles limites ? Les explications de Christophe Zimmer, directeur du département de Biologie computationnelle et de l'unité Imagerie et modélisation de l'Institut Pasteur.

Quelles sont les principales avancées que l'on doit à l'IA en matière de santé ?

Christophe Zimmer — Toutes les techniques d'imagerie médicale vont, si elles ne le font pas déjà, nécessiter une analyse par intelligence artificielle, en utilisant des méthodes dites d'apprentissage profond (des réseaux de neurones artificiels), si on veut obtenir un diagnostic de pointe et de qualité. Un médecin peut être formé sur quelques centaines ou milliers d'images au cours de sa formation, un algorithme peut lui être entraîné sur des millions voire des centaines de millions d'images. Dans les avancées espérées pour les prochaines années, il y a un enjeu énorme dans la médecine personnalisée, qui consiste à utiliser les informations dont on peut disposer sur un patient, par exemple les images médicales mais aussi le génome et d'autres informations biologiques, pour prédire à un stade précoce les risques de survenue de différentes maladies. Et pour prédire quel traitement serait

le plus adapté à chaque individu, et si possible pour prévenir la maladie plutôt que la guérir.

Vous pensez qu'on laissera bientôt les IA établir seules un diagnostic ?

Je ne sais pas si à terme on pourra se reposer entièrement sur les décisions ou les résultats de l'IA. Il y aura sans doute des contraintes juridiques. Mais s'il s'avère que la performance de l'IA est telle que l'intervention d'un expert humain a tendance à dégrader le résultat plutôt qu'à l'améliorer, peut-être qu'à un moment on se reposera presque exclusivement sur l'analyse par intelligence artificielle, en effet. Mais nous n'en sommes certainement pas encore là.

C'est pour cela que vous développez une approche qui consiste à faire pointer par l'IA des zones sur les images qui méritent l'attention des médecins...

C'est effectivement une approche qui me semble actuellement plus réaliste et utile en pratique que de se fier entièrement à l'IA. L'idée est d'utiliser l'IA pour attirer l'attention des experts sur des zones de l'image qui, par exemple, pourraient révéler des lésions difficiles à repérer à l'œil nu, dans des images de scanner de poumon. Il faut savoir que les images médicales sont des images énormes. En histopathologie, par exemple, une image peut faire 100 000 par 100 000 pixels (c'est-à-dire 10 milliards de pixels) et les experts humains n'ont tout simplement pas le temps d'examiner ces images dans leur plus grand détail. Alors que des algorithmes peuvent faire cela en un clin d'œil et révéler des régions particulièrement intéressantes, et déterminantes pour le diagnostic, qui ensuite peuvent être examinées de plus près par les médecins dans un

temps limité. Cette approche hybride de collaboration homme-machine devrait permettre de maintenir la responsabilité des médecins dans le diagnostic, tout en augmentant son efficacité et sa qualité grâce à l'IA.

C'est d'autant plus vrai que dans certaines circonstances, les IA peuvent commettre des erreurs grossières...

C'est vrai, elles peuvent parfois interpréter de manière une image de façon complètement erronée. Un exemple que j'ai en tête est l'image d'un enfant en train de se brosser les dents. Un algorithme d'IA peut interpréter la brosse à dents comme étant une batte de baseball, car ces deux objets peuvent avoir une forme similaire, alors qu'à l'évidence un être humain ne ferait pas cette erreur.

Au-delà de ça, quand on parle d'imagerie médicale, il peut y avoir des différences entre les appareils d'imagerie et donc entre les images. Il est donc capital d'entraîner les algorithmes et il est nécessaire de qualifier les données, pour être sûr qu'on traite bien le même type d'images.

Oui effectivement. Pour obtenir de bonnes performances de diagnostic avec des algorithmes d'apprentissage, il est important qu'il y ait une cohérence entre les images utilisées pendant la phase d'apprentissage et les images analysées par les algorithmes pour établir un diagnostic ou faire une prédiction. Si ces images sont obtenues avec un autre système d'imagerie, par exemple un scanner d'une marque différente, il est possible qu'il y ait des différences dans la texture des images qui ne soient pas perceptibles par l'humain mais qui peuvent faire échouer les algorithmes de diagnostic. Il est donc très important de s'assurer qu'il y ait une certaine cohérence entre ces images et de renforcer la robustesse des algorithmes, c'est-à-dire de les rendre moins sensibles à des différences techniques qui ne sont pas pertinentes d'un point de vue médical. C'est un travail essentiel. Cette fragilité de l'IA n'est pas irrémédiable mais il faut la garder en tête dans toutes les applications de l'IA, et surtout les applications médicales.

C'est aussi pour cela qu'il faut que le code soit disponible et que l'on puisse accéder aux algorithmes ?

CHRISTOPHE ZIMMER ET L'INSTITUT PASTEUR EN CHIFFRES

1885

C'est l'année au cours de laquelle Louis Pasteur inocula pour la première fois le vaccin contre la rage à un garçon de neuf ans, Joseph Meister, qui avait été mordu quatorze fois par un chien enragé.

2802

C'est le nombre de collaborateurs de l'Institut Pasteur à Paris. Dont 10 prix Nobel, travaillant pour 12 départements et 144 unités de recherche.

10 millions

C'est le nombre d'images de biopsies mises à disposition du projet européen Big Picture, qui vise à utiliser l'IA pour l'analyse automatique d'images de biopsies de patients souffrant de tous types de cancers.

95

C'est le pourcentage de bons diagnostics établis par une IA de détection des mélanomes, contre 87% pour des spécialistes, selon un test réalisé en 2018 en Allemagne.

Absolument. Il est essentiel que les codes sources (les programmes informatiques et les réseaux de neurones avec leurs millions de paramètres) soient rendus disponibles pour qu'ils puissent être réutilisés et validés dans d'autres contextes, et surtout pour qu'ils puissent être réentraînés sur des données différentes. Une équipe peut avoir développé un algorithme qui fonctionne très bien sur un certain type d'images mais qui échouera sur un autre jeu de données pourtant en apparence similaire. Par ailleurs, même si un algorithme est capable de distinguer et classer avec une très grande précision des images de chats ou de chiens ou de cellules cancéreuses et de cellules saines, dans un contexte médical on a généralement besoin de savoir pourquoi : quelles sont les raisons pour lesquelles telle cellule a été classée comme cancéreuse plutôt que telle autre. Or la réponse à cette question n'est pas évidente car un algorithme à base de réseaux de neurones ne nous explique pas d'emblée pourquoi il a pris telle décision ! Une contrepartie de la puissance des réseaux de neurones est qu'ils ne sont pas facilement interprétables et sont souvent considérés comme

des boîtes noires. Il faut donc essayer d'ouvrir cette boîte noire et de rendre compréhensible à un être humain les règles qui sont utilisées en interne par l'algorithme.

Tout le sujet, c'est donc finalement d'avoir accès aux bonnes données pour entraîner les machines ?

Effectivement, l'accès à des grandes quantités d'images pour entraîner ces algorithmes est crucial. La qualité des prédictions dépend surtout de la quantité des images qui sont fournies à l'algorithme pour l'entraîner. Les équipes qui auront accès à ces données en plus grande quantité (et qualité) prendront presque automatiquement de l'avance sur les autres. Cela soulève évidemment des questions de vie privée, mais il est clair qu'il y a un bénéfice énorme à pouvoir entraîner ces IA sur de grandes quantités de données. A mon sens, il est important de ne pas accumuler les contraintes réglementaires qui pourraient décourager certaines équipes de chercheurs, notamment en Europe, à poursuivre ces projets. Car on pourrait très bien dans dix ans se retrouver à devoir acheter très cher des logiciels qui auront été développés dans des pays moins regardants sur la protection des données de santé. Et qui, en plus, risquent d'être moins performants en Europe car ils auront été entraînés sur des données potentiellement différentes.

Et côté médicaments ? Les IA pourraient aussi, dites-vous, découvrir des nouveaux antibiotiques notamment, en en testant beaucoup ?

Oui, et il y a déjà des premiers pas très encourageants dans ce sens. L'IA peut intervenir en amont des expériences, pour cribler des centaines de millions de molécules différentes in silico, c'est-à-dire dans l'ordinateur, et en sélectionner un petit nombre, les plus prometteuses, qui pourront ensuite être étudiées de façon expérimentale. L'IA peut également proposer de nouvelles molécules qui n'existent pas encore et qui ont de bonnes chances d'avoir des effets thérapeutiques. Ces travaux peuvent inciter des chimistes à synthétiser ces molécules et à tester leur effet expérimentalement.

On a pu lire il y a quelques mois qu'Alphabet, la maison mère de Google avait avancé sur la question de la détermination de la structure

tridimensionnelle des protéines. En quoi cela consiste-t-il ?

Depuis longtemps, un défi majeur pour la biologie structurale est de réussir, à partir de la séquence unidimensionnelle de l'ADN, à prédire in silico la forme tridimensionnelle des protéines synthétisées dans les cellules. Jusqu'à récemment, ces recherches avaient produit des résultats plutôt médiocres, très inférieurs aux structures 3D obtenues à partir des méthodes expérimentales. Mais depuis peu, Deep Mind, entreprise qui fait partie d'Alphabet basée à Londres, a réussi à obtenir des résultats spectaculaires, dont la qualité rivale celle des structures expérimentales. C'est une avancée très importante car si on peut prédire les formes des protéines, par exemple une protéine qu'utilise le SARS-CoV-2 pour infecter les cellules, on peut mieux prédire quelle sera l'effet d'une mutation de son génome, donc l'impact potentiel d'un nouveau variant. En outre, on peut mieux cibler le développement de molécules ou d'autres stratégies thérapeutiques pour interférer avec le virus.

Est-ce qu'il y a une idée reçue sur la data qui vous énerve ? Et quelque chose que vous voudriez qu'on retienne particulièrement à propos de l'IA ?

Il y a une idée reçue selon laquelle on obtiendra la meilleure méthode et les meilleurs diagnostics uniquement si on se concentre sur le développement des meilleurs algorithmes. Cette mise en avant des algorithmes, qui sont évidemment essentiels, a tendance à diminuer l'importance souvent bien plus grande des données, de leur quantité et de leur qualité. Un autre point que je voudrais souligner, est que l'IA est souvent un peu surévaluée, c'est-à-dire que les attentes sont parfois exagérées, mais parfois c'est l'inverse qui se produit, y compris dans la communauté scientifique. Certaines techniques d'IA sont déconsidérées, par exemple comme n'étant pas suffisamment nouvelles ou sophistiquées, même si leur application peut être très féconde, et cela peut parfois constituer un frein inutile. Ce qui m'intéresse le plus, personnellement, encore davantage que le détail des méthodes, ce sont les applications que l'on peut tirer de l'IA. Ces techniques ouvrent désormais des perspectives énormes qu'il serait très dommage de ne pas exploiter. ●

Écoutez le podcast : <https://rb.gy/dorqkf>



Fred Potter

"Votre maison, un data center qui s'ignore !"

Nous produisons tous quantité de données au quotidien. A l'échelle de la maison, du foyer, c'est une véritable fourmilière de datas que l'on peut récupérer... et utiliser pour améliorer la vie. Mais comment trier parmi toutes les données celles qui ont du sens et une utilité ? Comment faire pour que la maison intelligente de demain ne soit pas qu'un gadget ? Les réponses de Fred Potter, fondateur de Netatmo et désormais CTO chez Legrand, groupe qui a racheté Netatmo il y a trois ans.

Quel est le chiffre important pour vous lorsqu'on parle de maison connectée ?

Fred Potter ● Le pourcentage de maisons qui ont déjà au moins un objet connecté. Parce qu'on sait très bien qu'à partir du moment où on va dépasser 25%, alors on est sûr qu'on ira à 75, 80% des maisons... Et comme nous on équipe les maisons, on attend ce moment où les technologies vont devenir la norme et le standard de l'habitat.

A quoi ressemblera la maison connectée de demain selon vous ?

Il y aura une sonnette connectée pour les visiteurs. Une serrure connectée, qui va permettre de gérer ses clés de manière numérique. Une caméra, qui ne vous filmera pas vous puisqu'elle sera capable de vous reconnaître quand vous arrivez. Vous entrerez dans une pièce à la bonne température, avec de l'air de qualité. Vous allumerez la lumière avec une commande vocale. Enfin, la maison sera plus performante du point de vue énergétique car tout sera connecté et cela permettra par exemple d'employer la bonne énergie au bon moment, lorsqu'elle est moins chère ou lorsque les énergies renouvelables sont disponibles.

Et il n'y aura pas de frigo connecté ?

Je pense que le frigo connecté, c'était plus un coup de com qu'autre chose, je ne vois pas trop son intérêt. En revanche, le gros électroménager très consommateur d'énergie, type lave-linge ou sèche-linge, a lui vocation à être connecté, non pas pour faire des gadgets mais pour justement être plus performants énergétiquement.

Dans une maison connectée, la question des données et de leur sécurité est-elle vraiment sensible. Tout le monde s'en fiche de la température qu'il fait chez moi, non ?

C'est vrai que les données les plus sensibles sont surtout les enregistrements audio et vidéo. Malgré tout, la sécurité et le respect de la vie privée, c'est quelque chose que vous devez exercer en profondeur. Quel que soit le côté anodin de la donnée collectée, si on n'en a pas besoin, pourquoi la collecter et la stocker ? En matière de data, il faut toujours se demander ce qui est collecté, par qui et pour quoi faire. ●

Écoutez le podcast : <https://rb.gy/jczvma>

FRED POTTER EN CHIFFRES

2011

C'est l'année où il a lancé la première station météo connectée Netatmo, récompensée au CES de Las Vegas.

45

C'est, en millions d'euros, le chiffre d'affaires qu'il réalisait en 2018 avec Netatmo lorsque le groupe Legrand a décidé de racheter l'entreprise.

80

C'est en milliards le nombre d'objets connectés recensés en 2020 dans le monde.

Irène Balmès

"Le premier travail d'un data scientist, c'est de nettoyer la donnée"



Le problème des algorithmes, c'est que si on n'est pas sûr des données qu'on y entre, de celles qui les font tourner, alors on ne peut pas être sûr des décisions qu'ils prennent... Décryptage avec Irène Balmès, élue "meilleure data scientist de France" en 2016 lors d'un concours organisé au sein de l'école 42 et actuellement data scientist pour Kynapse, société de conseil en IA et innovation digitale. Pour elle, il est capital de faire de la pédagogie autour de l'importance d'une "hygiène de la donnée".

En quoi consiste le travail d'un data scientist ?

Irène Balmès — C'est un terme qui est apparu il y a une dizaine d'années, au moment où il y a eu de plus en plus de données disponibles, en particulier chez les géants comme Google ou Facebook. On s'est rendu compte qu'on pouvait utiliser les algorithmes pour tirer des informations de ces immenses bases de données. En particulier faire de la prédiction de façon bien plus fine que ce qu'on pouvait faire avant. En tirant parti des propriétés statistiques des données, de leurs répétitions, on peut prévoir de nouvelles répétitions ou de nouvelles similarités.

Concrètement, qu'est-ce que vous faites de vos journées ? Vous regardez les chiffres défiler façon Matrix ?

Non, désolée de vous décevoir mais je ne lis pas dans la matrice en regardant des tableaux Excel ! J'écris beaucoup de code par contre et je fais des graphes. Notre premier travail c'est de récupérer les données et de les rendre «propres».

C'est un gros travail parce que la donnée elle n'existe pas au naturel, il faut la collecter, la vérifier, la nettoyer. Ça, c'est une très grosse partie du travail. Et puis on va essayer de comprendre ces données, de voir ce qu'on peut en tirer. Ce qui nous intéresse dans la data science ce sont les statistiques, et donc les effets globaux et non pas les effets locaux, qu'on ne peut pas voir sur quelques chiffres.

Nettoyer des données, ça veut dire quoi exactement ?

Parfois, la méthode de collecte peut influencer sur la donnée elle-même. Et parfois on peut avoir l'impression qu'il y a des corrélations, des choses qui se ressemblent, qui apparaissent dans les données, alors qu'en réalité ce ne sont pas de vraies corrélations, ce sont uniquement des artefacts dus à la méthode de collecte.

Comme quoi, par exemple ?

Par exemple, même si la collecte automatisée réalisée par des capteurs est en principe la plus fiable, on peut imaginer un capteur qui fonctionne très mal au-delà d'une certaine température... et qui donne des résultats faussés quand il fait trop chaud. Ensuite, sur des données relevées par des humains, on voit apparaître plus de biais. Je me souviens d'un cas où on essayait de prédire si une personne allait venir à un rendez-vous ou non et on trouvait une très forte corrélation entre le fait que la personne vienne et son nombre d'enfants. Et on s'est rendu compte simplement que cette information du nombre d'enfants n'était inscrite que si la personne venait au rendez-vous ! Mais on ne le savait pas dans les données à l'origine car la colonne existait dans tous les cas. C'est simplement lorsqu'on a

vu cette corrélation qu'on en a discutée avec les personnes du métier qui, elles, savaient le processus de collecte et qui nous ont expliqué que cette donnée n'était collectée que si la personne venait au rendez-vous. C'est donc essentiel de pouvoir discuter avec les personnes du métier, qui savent comment la donnée a été collectée.

Quel est le risque si on ne nourrit pas les machines avec les bonnes données ?

Les algorithmes vont prendre leurs décisions selon statistiquement ce qu'ils ont vu avant et ce qui, statistiquement, leur semble être la bonne décision. On a vu ainsi il y a quelques années un algorithme pour une carte de crédit qui proposait systématiquement un crédit moins important aux femmes qu'aux hommes, toutes circonstances égales par ailleurs. Ça s'explique par le fait que l'algorithme fonctionne à partir de données historiques et que ces données montrent que les femmes gagnent moins que les hommes, en général. Donc en utilisant ces données historiques, on va pénaliser un groupe tout entier, ici les femmes. D'abord il faut questionner les données d'origine et puis il faut questionner l'algorithme.

Vous avez d'autres exemples ?

Oui, sur la reconnaissance faciale notamment. Les algorithmes sont majoritairement entraînés à reconnaître des visages d'hommes blancs, parce que ce sont les photos que l'on trouve le plus sur internet, tout simplement. En Angleterre, des caméras vont, dans certaines zones, scanner la foule pour voir si une personne est recherchée ou non. Or, ces algorithmes se trompent sur les types de visages qu'ils ont peu vus. Ils sont donc très efficaces sur les hommes blancs, mais se trompent beaucoup sur les personnes noires et notamment les femmes noires. Ce qui veut dire que les personnes qui vont être arrêtées dans la rue pour justifier de leur identité, car l'algorithme n'aura pas réussi à les identifier, ce sont en majorité les personnes de couleur. Tout le problème vient des données et de leur diversité.

Et pas de la diversité de ceux qui les programment ?

Oui, aussi, c'est indispensable. Il faut plus de conscience et plus de diversité. Au-delà même de la couleur de peau, on a des diversités de handicaps, de genre, d'orientations sexuelles,

etc. Toutes ces diversités il faut les prendre en compte. Et il faut aussi plus de réglementations pour pouvoir aller voir comment sont conçus les algorithmes et s'assurer qu'ils font bien ce pour quoi ils ont été programmés.

Quelle idée reçue sur l'IA ou la donnée vous agace ?

Le terme d'IA lui-même me hérisse, pour la partie intelligence. On pense immédiatement à des intelligences dystopiques qui vont contrôler l'humain et l'asservir ou le dépasser. On en est très loin pourtant, et je pense qu'il faut se poser la question de savoir ce que c'est qu'être intelligent. Est-ce que c'est savoir jouer aux échecs ? Ou bien savoir s'adapter à son environnement et au contexte ? Certes, l'ordinateur est désormais plus fort aux échecs que l'humain, mais il lui est toujours beaucoup plus difficile de s'adapter, d'improviser, et d'être pertinent.

Écoutez le podcast : <https://rb.gy/mtkfb0>

IRÈNE BALMÈS EN CHIFFRES

5

C'est le nombre de langues parlées par Irène Balmès. Le français, l'anglais, le portugais, l'espagnol et le Fongbé, langue parlée au Bénin, où elle a été pendant un an professeur de physique dans un collège.

42

Comme l'école 42, créée par Xavier Niel, où elle a été classée 5^e du concours du meilleur data scientist de France 2016 et 1^{ère} femme du classement.

42906

C'est en euros le salaire annuel moyen d'un data scientist débutant en France, selon le site Glassdoor. Ce qui fait un peu plus de 3500 euros par mois.

786

C'est le nombre de postes de Data Scientists actuellement disponibles sur le site Indeed.com, le plus grand site d'emploi mondial.

Yves Le Gélard

"Il ne faut pas confondre big data et big brother. Car c'est le big brother qui est dangereux, pas le big data."



LE BIG DATA peut nous aider concrètement à réduire notre consommation d'énergie, mais aussi à mieux piloter les énergies renouvelables et à adapter le plus finement possible la production à la consommation pour éviter les pertes. Sauf que dans le même temps, le numérique en lui-même est très consommateur d'énergie. Alors comment gérer ce paradoxe ? Comment faire en sorte que la data soit au service d'un monde plus vert ? Yves Le Gélard, directeur général adjoint d'Engie, en charge du digital et des systèmes d'information, défend l'idée qu'un compteur connecté a besoin de données pour être efficace. Et que c'est la data qui permettra de développer les énergies renouvelables et au final de rendre l'énergie plus verte.

Pourquoi d'un coup a-t-on commencé à s'intéresser au big data alors que la donnée existe depuis très longtemps ?

Yves Le Gélard — C'est la rencontre entre les capacités de traitement et de stockage qui a changé la donne. Et aussi, au début des années 2010, le besoin d'un certain nombre de sociétés américaines de manipuler un grand nombre de données. C'est le résultat de la mise au point de la technologie du cloud : l'offre de la puissance de calcul, couplée à la technologie des bases de données, permet d'ouvrir de nouveaux champs. L'émergence de Hadoop et l'adoption de cette technologie par Facebook nous a fait entrer dans l'ère du Big Data.

A quoi sert le big data pour un producteur d'énergie comme Engie ?

Un réseau électrique, a besoin à tout instant

d'équilibrer la production d'énergie avec la consommation. C'est une opération complexe qui nécessite une quantité importante de données. Plus vous disposez de données précises sur la consommation, plus vous êtes capable d'ajuster la production de façon fine et fiable. Ce qui coûte très cher en matière de réseau, c'est ce qu'on appelle le calibrage de la pointe : c'est-à-dire la journée la plus froide de l'hiver, à 20 heures, pour laquelle il est nécessaire de mobiliser toutes les capacités de production. Une capacité de production d'un gigawatt, nécessite un investissement de l'ordre du milliard d'euros. Si vous disposez d'une analyse fine de ce qui compose cette pointe de consommation et éventuellement de la façon de la lisser un peu, ce sont des investissements considérables qui sont évités, avec en corollaire un équilibre du réseau maintenu et une situation plus vertueuse. Il est donc extrêmement important pour pouvoir équilibrer la production et la consommation, de disposer d'informations très précises sur qui va consommer quoi, quand et comment. Et pour le prévoir, il faut une quantité très importante de données. Ces données sont aujourd'hui fournies dans un certain nombre de pays dans le monde par ce qu'on appelle les compteurs intelligents - en France nous avons Linky pour l'électricité et Gazpar pour le gaz. Ces compteurs intelligents fournissent des informations extrêmement précises, non plus à l'échelle de l'année ou de la demi-année, lorsque passe le relevé, mais au pas de la minute, voire de la seconde.

Donc, pour caricaturer, auparavant, pour calibrer les capacités de production, c'était presque au doigt

mouillé... alors qu'aujourd'hui le big data vous permet de calibrer bien plus finement et donc d'éviter de construire des capacités de production inutiles ?

Absolument. Et ceci va être d'autant plus nécessaire que le mix énergétique va évoluer vers encore plus de renouvelables. Or il n'y a pas toujours du vent, pas toujours du soleil... Nous aurons besoin de données extrêmement précises qui vont nous permettre par exemple, selon les conditions météorologiques, d'ajuster la production et d'équilibrer le réseau en fonction, certes, de la consommation, mais aussi des conditions de production.

Malgré tout, avez-vous besoin de ce niveau de détail dans les données que vous collectez ? Par exemple, depuis 2016, vous avez un parc éolien, dans la Meuse, «open data». Mais est-ce qu'il y a un intérêt à connaître à la seconde le rendement de la turbine R80711, comme on le voit sur le site ?

Oui c'est très important. Le fait de suivre à la seconde une capacité de production comme une turbine permet d'aller chercher 2 ou 3% de rendement supplémentaire. Et quand on sait qu'un gigawatt de production coûte un milliard, 3% de plus c'est beaucoup d'argent. Pour prendre l'exemple des éoliennes dont vous parlez dans le parc de la Haute Borne, nous avons mis au point des algorithmes qui permettent d'ajuster constamment l'éolienne face au vent. Car 1 degré d'écart, ce sont des pourcentages de production qui s'envolent... Le fait de suivre ces données en temps réel, permet d'utiliser tout le potentiel d'un investissement qui a coûté beaucoup d'argent.

YVES LE GÉLARD ET ENGIE EN CHIFFRES

2015

C'est la date à laquelle il a rejoint Engie, d'abord en tant que DSI et aujourd'hui comme directeur général adjoint en charge du digital et des systèmes d'information, après avoir travaillé chez IBM, CAPgemini ou encore Fujitsu et SAP.

170 000

C'est le nombre de collaborateurs d'Engie dans le monde. Et parmi ces 170 000 il y a plus de 2000 collaborateurs dédiés à la data, dont 200 data scientists.

2045

Date de l'objectif zéro carbone d'ENGIE

3.3.0

C'est le numéro de la dernière version de Hadoop. Hadoop c'est la principale plateforme Big Data, surnommée l'éléphant du Big Data en raison de son logo - un éléphant jaune inspiré par la peluche du fils du créateur de cette solution. C'est une solution qui, en gros, distribue les données et les calculs sur différents ordinateurs ce qui permet de réaliser simultanément des tâches multiples en manipulant un très grand nombre de données.

Tout ça c'est bien pour vous, le fournisseur, mais pour moi, client, quel est l'intérêt d'avoir un compteur connecté ?

D'abord, ça apporte souvent du confort, avec le fait de pouvoir piloter à distance la mise en route du chauffage ou au contraire son extinction. Et puis cela contribue aussi à faire prendre conscience de notre empreinte carbone. À partir du moment où vous prenez conscience de l'empreinte carbone de tous vos actes du quotidien, cela entraîne des comportements extrêmement vertueux. Nous le constatons déjà chez certains clients qui, à partir du moment où ils disposent d'informations sur leur consommation font évoluer leur mode de vie.

Quel va être le rôle de ces compteurs par rapport au nombre croissant de voitures électriques sur le marché ?



En dix ans, l'intégralité du parc automobile va changer, ce qui va entraîner des changements considérables en matière d'approvisionnement électrique. Ces besoins sont nouveaux, il n'existe pas d'historique de consommation. Il est donc important de suivre l'évolution de ces besoins en temps réel, pour mieux y répondre. Et développer, par exemple, des offres pour recharger sa voiture électrique, mais aussi imaginer comment utiliser l'énergie stockée dans les batteries s'il y a besoin d'en réinjecter dans le réseau lors d'une pointe de consommation. Pour tout cela, les données sont cruciales.

Les mondes du numérique et de la data sont très gourmands en énergie, et consomment beaucoup. Comment gérer ce paradoxe?

C'est vrai, le monde de la donnée est très énergivore, 4 à 5% de la consommation mondiale d'énergie sont destinés aux data centers. Depuis de nombreuses années, nous

travaillons avec de grands acteurs du cloud pour limiter cette consommation. Pour cela, nous pouvons agir à plusieurs niveaux. Tout d'abord sur la conception des data centers eux-mêmes, pour faire en sorte qu'ils soient aussi frugaux que possible en énergie. Ensuite, en travaillant à fournir à ces centres une énergie exclusivement renouvelable et en créant de nouvelles capacités de productions vertes. Enfin, nous pouvons aussi parfois récupérer une partie de la chaleur résiduelle émise par les serveurs, afin de l'injecter dans des réseaux de chaleur, et ainsi chauffer les habitations d'un quartier, une piscine municipale, une école.

Si tout cela est si vertueux, pourquoi a-t-on entendu autant de critiques autour des compteurs connectés ?

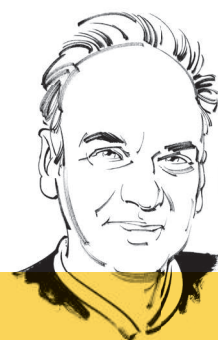
Le sujet de la data cristallise un certain nombre de fantasmes, de suspicions, voire de théories non nécessaires. Les technologies autour des données sont des technologies qui relèvent de la science et peuvent être utilisées à bon escient ou à mauvais escient. Mais, en soi, ces technologies ne sont ni bonnes ni mauvaises. Il faut tordre le cou à l'idée reçue que big data égale big brother. Big data, dans les mains de big brother peut devenir dangereux. Mais c'est le big brother qui est dangereux, pas le big data.

Malgré tout, ne sommes-nous pas trop boulimiques en matière de données...

C'est à mon avis la tendance des années à venir. Pendant longtemps on a laissé la donnée dormir, on en utilisait peu et on en jetait beaucoup. Depuis une dizaine d'années, le balancier penche dans l'autre sens et on a tendance à conserver absolument tout. Or ce n'est sans doute pas nécessaire. Je pense qu'il faut passer du big data au smart data et mettre fin à la logique de boulimie dans laquelle nous sommes tombés. Certains usages sont justifiés, d'autres nécessiteraient plus de raison... Je pense par exemple que le fait de photographier son petit déjeuner ou son chien et que la photo se retrouve sur les réseaux sociaux et dans une cinquantaine de serveurs différents, n'est sans doute pas indispensable. En revanche, quand il s'agit de faire des économies, limiter notre empreinte carbone, il faut mobiliser tout le potentiel de la data. ■

Écoutez le podcast : <https://rb.gy/dj6jv9>

iStock, Britta und Ralph Hoppe - www.FotoToode



Hervé Le Tellier

"Comment être sûr que le monde actuel n'est pas une gigantesque simulation ?"

ET SI NOTRE ADN N'ÉTAIT QU'UNE SORTE D'ALGORITHME ? Le monde dans lequel nous évoluons ne serait-il qu'un gigantesque jeu vidéo ? La vie ne peut-elle se résoudre qu'à une sorte de grande équation ? Une série de chiffres, de statistiques, de probabilités ? C'est l'un des thèmes développés par Hervé Le Tellier, mathématicien et écrivain, dans son roman "L'anomalie", prix Goncourt 2020, le livre événement de l'année, vendu à plus d'un million d'exemplaires en France.

Sans dévoiler trop de choses pour ceux qui n'ont pas encore dévoré L'anomalie, on peut dire que vous y développez l'idée que la vie ne serait qu'une gigantesque simulation. Un peu comme dans Matrix, non ?

Hervé Le Tellier ■ Non, c'est le contraire de Matrix en réalité. Dans Matrix, les machines utilisent les humains comme source d'énergie. Des individus biologiques sont dans des capsules et produisent de l'électricité pour les ordinateurs, comme avant on utilisait des chevaux pour faire tourner les moulins. Moi la question que je me pose c'est plutôt la question inverse, j'essaie de savoir si le monde dans lequel nous vivons est bien réel. Si nous sommes dans une réalité physique, matérielle, ou si nous sommes dans une sorte d'immense jeu vidéo.

C'est-à-dire ?

Peut-être sommes-nous en réalité dans un ordinateur dans lequel nos consciences ne seraient que des énormes algorithmes, construits sur

la base d'un code qui serait le code ADN. Qui produirait des êtres qui, au fur et à mesure de leur vie virtuelle, auraient des expériences qui les modifieraient au niveau du caractère et du physique. Mais qui seraient avant tout des êtres créés par des algorithmes...

Et vous y croyez vraiment ?

C'est un jeu que je propose, une simulation mentale. Même si c'est totalement faux, même si nous sommes dans un monde réel, les idées et les propositions que cela ouvre sont très intéressantes d'un point de vue philosophique et romanesque évidemment. C'est une idée folle mais qui rencontre un certain nombre d'adeptes aujourd'hui, y compris parmi des scientifiques de renom : l'idée que nous soyons dans une sorte de simulation est en train de gagner du terrain.

Autre point commun entre votre livre et Matrix : dans le film on explique l'impression de déjà-vu par une sorte de bug dans la machine. Vous, votre «anomalie» peut aussi être vue comme une sorte de bug de la simulation...

Oui, un bug, ou un piège ou un test. Moi je l'imagine plutôt comme un test en fait, même si j'ai laissé une fin très ouverte au roman. Ce qui est important dans l'idée du bug ou de l'anomalie, c'est que nous sommes en permanence confrontés à des bifurcations dans nos vies, liées à des tout petits éléments. Et on ne se rend pas toujours compte de l'impact que ces petits événements peuvent avoir dans nos vies. Par exemple, on freine sur l'autoroute, on rate une sortie et une rencontre qu'on devait faire dans la journée ne se fera pas, or cette rencontre aurait

Illustration : Kagan McLeod

été décisive dans notre vie. Ou on se fait piquer par une guêpe, on est retardé de deux minutes et ces deux minutes vont décaler l'intégralité du monde... Je trouve ça fascinant, ce chaos considérable, avec des milliards d'individus qui interagissent et qui chacun, à leur manière, vont modifier la seconde ultérieure. Et quand je dis seconde, il s'agit d'une référence toute relative : pour l'ordinateur dans lequel nous sommes - si nous sommes dans un ordinateur - cette seconde ne dure peut-être en réalité qu'une millionième de seconde. Parce qu'on voit bien que le temps, dans les machines, n'est plus le même que le nôtre : quand on joue contre un ordinateur, dans des jeux de combat en ligne par exemple, les monstres que l'on combat pourraient être un million de fois plus rapides qu'ils ne le sont, mais l'ordinateur les ralentit volontairement pour les mettre au niveau du temps de réaction de notre cerveau..

Alors qu'on dit souvent que les Français sont nuls en maths, est-ce une fierté pour vous d'avoir réussi à faire lire à un million de nos concitoyens un ouvrage qui parle de probabilités, dans lequel

on voit passer Fibonacci, célèbre mathématicien, où l'on évoque les topoï de Grothendieck, la notation de Kendall, la loi de Little, les chaînes de Markov...

D'abord, les Français ne sont pas si nuls que ça en maths. On a tout de même régulièrement des médailles Fields (une des plus prestigieuses récompenses en matière de mathématiques). Et je crois qu'il y a un goût pour les maths, même s'il y a sans doute des choses à améliorer au niveau de leur enseignement. Il y a des choses à faire au niveau de la pédagogie générale de cette science, qui est d'abord un langage. Mais on n'est pas réfractaire à cette science. Mais oui, je suis content qu'un livre qui aborde des questions scientifiques puisse ne pas rebuter un large public.

Ce public, vous aviez imaginé le toucher aussi largement ?

J'ai voulu faire un livre populaire, j'avais l'ambition de faire un livre qui soit lu, mais pour moi qui vendais 30 000 exemplaires de mes précédents romans, arriver à vendre vingt fois plus, trente fois plus, c'était totalement inespéré. Ce livre n'était évidemment pas programmé pour être un best-seller.

A propos d'imprévisible, il est aussi question de probabilités dans votre roman. Vous faites notamment dire à un personnage que si on peut calculer la probabilité qu'une pièce tombe sur pile ou face, et même celle infime qu'elle retombe sur la tranche, on oublie de prévoir le cas où lancée, la pièce ne retombe pas... C'est ça la force de la littérature par rapport aux maths : imaginer l'imprévisible ?

Les probabilistes vous répondraient probablement que si, l'imprévisible peut se prévoir. On pourrait par exemple calculer la probabilité qu'en versant de l'eau bouillante dans une casserole, l'eau se mette immédiatement à geler. Cette probabilité est très, très faible, une chance sur un avec des centaines de zéros derrière, mais elle existe. Après c'est justement cette très faible probabilité qui fait qu'il est bon de l'éliminer pour commencer à calculer, sinon on se trimbale une variable qui est littéralement gênante dans le calcul général. En revanche, conserver en tête cette idée que l'impossible peut devenir possible, ça oui, c'est intéressant quand on est romancier et qu'on veut imaginer des fictions.

Peut-on mettre le monde en chiffres ? Est-ce que la vie peut se réduire à une équation ?

Je pense que c'est possible. Mais ce n'est pas souhaitable. Par exemple, on pense que l'amour entre deux personnes est un phénomène incompréhensible, alors qu'en réalité il est très possible qu'on comprenne un jour ce qui provoque l'amour. Les motivations intimes de chacun, qui touchent à la fois à ce qui est de l'ordre de la phéromone, de l'électrique, de la réminiscence... et qui vient faire qu'un visage, un corps, un parfum, un regard, nous est extrêmement intime et nous touche immédiatement. C'est peut-être quelque chose qui est inscrit en nous mais je ne suis pas sûr que cela soit souhaitable de le savoir !

Vous dites : "L'incompréhension surpasse l'intelligence." Est-ce que cela veut dire qu'une machine ne sera réellement intelligente que lorsqu'elle sera capable de se rendre compte, seule, qu'elle ne sait pas, ne comprend pas ?

C'est une jolie façon de le dire effectivement. La mesure de son incompétence c'est aussi la mesure de sa capacité à chercher une solution à quelque chose qu'on ne comprend pas. Ce qu'on voudrait, ce sont des machines qui aillent au-delà de leur propre savoir. Mais en même temps c'est le grand danger car c'est leur apprendre à apprendre et leur apprendre à avoir envie d'apprendre, c'est un enjeu important à partir du moment où elles auront une éthique et ne menacent pas l'être humain... Jusqu'au jour où, pour sauver la planète, elles devront peut-être sacrifier l'être humain.

Si on devait retenir une chose sur votre livre, et sur notre échange, quelle serait-elle ?

Ce qui m'a intéressé dans l'écriture de ce livre, c'est de m'interroger sur le réel et sur la virtualité, ainsi que sur la question de l'essentiel dans nos vies. Qu'est-ce qui me serait indispensable, qu'est-ce que je ne serai pas prêt à sacrifier ? Ce qui m'importe aussi c'est qu'on se méfie d'un monde dans lequel le Big Data gouverne tout et qu'on soit en permanence orienté. J'ai un peu peur qu'une partie non négligeable de l'humanité manipule l'autre. ■

🔊 Écoutez le podcast : <https://rb.gy/9v25cz>

HERVÉ LE TELLIER EN CHIFFRES

118

Comme 118e prix Goncourt décerné, celui qui lui a été attribué le 30 novembre 2020 par 8 voix contre 2.

12 500 exemplaires

C'est le tirage initial de son roman, l'Anomalie, paru chez Gallimard... bien avant l'annonce du prix Goncourt. Depuis, les ventes se sont emballées, puisqu'au dernier comptage le roman dépasse le million d'exemplaires imprimés...

5356

C'est le nombre d'évaluations laissées par les internautes sur la page Amazon de l'Anomalie, prix Goncourt 2020. Dont 51% ont mis 5 étoiles et 6% une seule étoile.

42

Comme le protocole 42, déclenché après l'orage que traverse l'avion, au cœur de l'intrigue de l'anomalie. Mais 42 c'est aussi la réponse donnée par une intelligence artificielle dans le livre culte pour les geeks : "Le guide du voyageur galactique", de Douglas Adams, à La grande question sur la vie, l'univers et le reste...

RETROUVEZ L'INTÉGRALITÉ DE CES ENTRETIENS EN VERSION AUDIO DANS LE PODCAST "BIG DATA"

Capital

BIG DATA

III LES BIG ENTRETIENS DU BIG DATA IIIIIII

Un podcast en partenariat avec ENGIE

DISPONIBLE GRATUITEMENT SUR APPLE PODCAST, DEEZER, CASTBOX, SPOTIFY ET SUR TOUTES LES APPLICATIONS D'ÉCOUTE.

**Nous construisons
aujourd'hui le système
énergétique bas carbone
de demain.**

Plus d'informations sur [engie.com](https://www.engie.com)




engie

*J'agis avec ENGIE.

L'énergie est notre avenir, économisons-la!

ENGIE : SA AU CAPITAL DE 2 435 285 011 € - RCS NANTERRE 542 107 651. © Getty Images.

**#Act
With
ENGIE***