SmartPoop

Lê Nguyên Hoang et Tristan Le Magoarou

Contents

SmartPoop

Bienvenue dans le github du livre *SmartPoop*, co-écrit par Lê Nguyên Hoang et Tristan Le Magoarou, publié sous licence CC BY 4.0 (réutilisable librement à condition de citer les auteurs originaux).
Le livre est disponible en epub et en pdf.
À ce jour, le livre n'est pas dans une version stable.

Sommaire

Pourquoi ce livre ?

SmartPoop est une science-fiction très réaliste, qui vise à sensibiliser au danger des algorithmes et aux défis de les rendres robustement bénéfiques.

Elle raconte l'histoire d'une entreprise du numérique, *SmartPoop*, qui effectue des analyses automatisées d'excréments. Ceci lui permet des diagnostics médicaux ultra-personnalisés. Cependant, dans chaque chapitre, le déploiement de *SmartPoop* conduit à des *effets secondaires imprévus*, avec des conséquences médicales et sociales tragiques. Les co-fondateurs de *SmartPoop*, les héros de l'histoire, doivent alors prendre la responsabilité de leur technologie, en subir les conséquences, et trouver des solutions pour rendre leurs algorithmes sécurisés et éthiques.

À propos des auteurs

Lê Nguyên Hoang

Lê Nguyên Hoang est un chercheur et médiateur scientifique à la faculté d'informatique et communication de l'EPFL. Sa recherche porte sur la sécurité et l'éthique des algorithmes, notamment sur la théorie de l'apprentissage et la conception participative. Lê est aussi YouTubeur, sur les chaînes Science4All, Axiome, Étincelles, EPFL, Wandida, EPFL et ZettaBytes, EPFL. Il est l'auteur de trois livres, intitulés *La formule du savoir*, *Le fabuleux chantier* et *Turing à la plage*. Il est aussi membre du conseil d'éthique d'Orange. Enfin, Lê est Président de l'Association Tournesol, dont la plateforme vise à collecter une grande base de données, fiable et sécurisé, de jugements éthiques humains, pour contribuer à résoudre l'éthique des algorithmes.

Tristan Le Magoarou

Docteur en médecine, spécialiste en santé publique et médecine sociale, Tristan Le Magoarou est médecin d'information médicale et de santé publique dans un centre hospitalier de province. Il est également, depuis 2016, YouTubeur sur sa chaîne Risque Alpha qui possède 26 000 abonnés où il vulgarise l'épidémiologie et les statistiques.

À propos de l'écriture du livre

Ce livre est voué à être constamment mis à jour pour le rendre plus lisible, plus crédible et plus pertinent. Nous invitons nos lecteurs en particulier à proposer des corrections, typographiques ou au niveau de l'histoire. Notez toutefois que les auteurs Lê Nguyên Hoang et Tristan Le Magoarou auront le dernier mot sur la version disponible sur ce github, et qu'ils travaillent dessus en tant que bénévoles. Notez que la licence CC BY 4.0 autorisent n'importe qui à réutiliser et adapter le livre à leur guise, à condition de mentionner la version originale du livre et ses auteurs. Nous encourageons de tels efforts, et nous nous tenons à votre disposition pour tout projet de ce genre.

Bibliographie

1. Le fabuleux chantier : rendre l'intelligence artificielle robustement bénéfique. Lê Nguyên Hoang et El Mahdi El Mhamdi. EDP Sciences (2019).
2. Turing à la plage : l'intelligence artificielle dans un transat. Rachid Guerraoui et Lê Nguyên Hoang. Dunod (2020).
3. Tout ce que vous devez savoir sur le plus tabou des sujets. Michel Lafond. Julien Ménielle (2018).

Chapitre 1 — L'or marron

La bouteille de champagne ouverte, les verres enfin servis, Marc Rofstein demande l'attention des cinq invités de la soirée organisée par Katia Crapinski, dans leur colocation.

> Chers amis, j'aimerais qu'on prenne quelques instants pour féliciter solennellement Katia. Son dernier article de recherche a été accepté pour publication[1], ce qui signifie que Katia a maintenant tout ce qu'il faut pour écrire et valider sa thèse de doctorat en machine learning[2]. Donc si vous le voulez bien, je vous invite à lever vos verres, et à célébrer la bientôt Docteure Katia pour ses accomplissements — même si, contrairement à moi, Katia ne sera pas une docteure qui sauve des vies. À la bientôt Docteure Katia !

Plus tard dans la soirée, après avoir prétendu avoir apprécié la blague de son colocataire, Katia partage toutefois son désir de faire un travail plus altruiste et plus bénéfique pour toute l'humanité. Sa recherche jusque-là se contente d'améliorer les performances d'algorithmes déjà existants[3]. Cependant, Katia est

[1] La production scientifique repose beaucoup sur un système de publication dans des journaux ou des conférences scientifiques, où l'acceptation à publication dépend d'une revue par les pairs. Un article donné est ainsi étudié par deux à cinq relecteurs, qui émettent chacun un avis et exigent des corrections plus ou moins majeures. Les journaux et conférences scientifiques varient beaucoup selon leurs thèmes de prédilection et leur prestige. En informatique, les publications dans les conférences scientifiques sont très prestigieuses. En machine learning en particulier, les conférences les plus prestigieuses sont NeurIPS, ICML et ICLR, avec juste derrière des conférences comme COLT, AISTAT, UAI, et AAAI, parmi d'autres.
:tv: Comment fact-checker une étude scientifique ? Science Étonnante (2019)

[2] Une thèse en informatique est généralement une collection d'articles acceptés à publication. L'auteur de la thèse se doit généralement d'ajouter un résumé, une introduction générale, et une conclusion. Selon les universités, les exigences pour voir sa thèse acceptée varient.

[3] Beaucoup considèrent ainsi que, si le progrès technologique améliore grandement la qualité de vies du plus grand nombre, il conduit aussi au fait que le coût de causer une catastrophe monumentale diminue drastiquement. Si l'on accepte ce postulat, toute recherche sur la performance des algorithmes peut être vue comme la création potentielle de risques nouveaux et mal maîtrisés, et est en cela dangereuse. Pour caractériser ce phénomène, le philosophe Nick Bostrom parle de *l'hypothèse du monde vulnérable*. Il propose ainsi une métaphore, où la recherche scientifique consiste à tirer une boule dans une urne, contenant des boules blanches et des boules noires. Chaque boule blanche améliore le monde, mais tirer une boule noire revient à le mettre en grave danger, à l'instar de la découverte de la réaction nucléaire en

bien consciente que ces progrès alimentent principalement la recommandation de contenus addictifs par les réseaux sociaux et l'optimisation de la publicité ciblée sur ces plateformes[4]. Cette même technologie ne peut-elle pas être davantage utilisée pour sauver des vies ?

Marc, encore étudiant en cancérologie, suggère alors l'utilisation des algorithmes pour le diagnostic précoce. Un cancer diagnostiqué tôt a plus de chance d'être traité avec succès, et avec moins de risques et de complications.

> Ça commence à se faire en plus, précise Marc. L'année dernière des chercheurs américains ont montré qu'une IA était aussi douée qu'une équipe de dermatologues pour détecter des cancers de peau sur des photos[5]. Mais malheureusement, en cette année 2018, beaucoup de cancers sont encore diagnostiqués de manière tardive. À ce stade, des soins risqués et coûteux sont nécessaires.
>
> Il y a clairement un besoin d'algorithmes d'apprentissage en médecine. Mais si on veut diagnostiquer quoi que ce soit, il faut des données. Beaucoup de données[6], fait remarquer Katia.

Malheureusement, la simple collecte de données est une tâche laborieuse, délicate et intrusive pour le grand public, qui préfère ne pas se préoccuper des risques de cancer.

---

chaîne. Selon Bostrom, la quête aveugle de nouvelles connaissances poserait ainsi un risque existentiel, et serait donc immorale. Cela semble d'ailleurs d'autant plus le cas quand il s'agit de la quête d'algorithmes plus performants, dont le déploiement précipité a certainement des effets secondaires difficilement prévisibles. Dès lors, il semble urgent de guider la recherche vers la quête de « boules blanches », voire vers la quête de boules blanches qui nous protègent de boules noires qu'on pourrait tirer à l'avenir.
:tv: How civilization could destroy itself – and 4 ways we could prevent it | Nick Bostrom. TED (2020)
:memo: The Vulnerable World Hypothesis. Nick Bostrom. Global Policy (2019).

[4]De nos jours, les « intelligence artificielles » les plus sophistiquées, celles qui reçoivent des milliards de dollars d'investissements en recherche et développement (si ce n'est plus !), ce sont bien les algorithmes du web, car l'enjeu économique et le besoin d'automatisation y sont monumentaux. Après tout, les chiffres d'affaire de Google, Apple, Facebook, Amazon et Microsoft, entre autres, se comptent en centaines de milliards d'euros. Toute amélioration du service de quelques pourcents représente donc des milliards d'euros. Or ces entreprises doivent gérer des données de milliards d'utilisateurs, qui génèrent chacun sans doute des méga-octets de données par semaine. Ainsi, chaque heure, il y aurait plus de 30 000 heures de nouvelles vidéos mises en ligne sur YouTube. Voilà qui représente des quantités monstrueuses de données, qui ne peuvent être traitées que par des machines. Or les tâches sont de plus en plus complexes, comme détecter des incitations à la haine dans une image ou identifier la désinformation dans des milliards de messages.
YouTube's Blog.

[5]:tv: L'IA sauvera des vies (ft. Primum Non Nocere). Science4All (2018).
:memo: Dermatologist-level classification of skin cancer with deep neural networks. Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun. Nature (2017).

[6]La problématique des données est vraiment critique en machine learning, puisque les algorithmes sont systématiquement conçus pour apprendre des données et pour généraliser à partir des données.

> On n'arrive déjà pas à convaincre les gens de réduire leur consommation de tabac ou d'alcool, note Marc. Si on veut avoir une chance quelconque d'effectuer des diagnostics précoces, il faut se concentrer sur des données très informatives. Mais ce genre de données sont généralement trop invasives à récupérer. Je ne connais pas beaucoup de volontaires pour se faire faire des prises de sang[7] à répétition par exemple...

Katia a le regard songeur. Comment acquérir des données de patients ? Ou, mieux encore, comment acquérir des données de non-patients ? Comment ne pas être « invasif » ? Katia se demande aussi qui est à l'origine du terme « invasif ». Une prise de sang « n'envahit » que rarement le corps de gens.

C'est avec ses réflexions en tête que Katia se dirige vers les toilettes pour faire ses besoins.

Eurêka

Assise sur la cuvette, le clapotis de son urine résonnant au fond de la cuvette frappe alors son esprit. Katia lève la tête, et son visage abattu se transforme tout à coup en un sourire radieux. Archimède avait sa baignoire[8], Katia eut ses WC ! En sortant des toilettes, elle se lance alors dans un discours solennel qui, l'avenir le dira, marquera l'histoire de l'espèce humaine.

> Marc, je viens de trouver l'idée du siècle. Je sais comment résoudre ce problème de diagnostic ! Tu l'as dit. Il nous faut des données. Beaucoup de données très informatives, mais aussi très simples à collecter. Des données qu'on n'aura pas besoin d'extraire violemment des corps humains ; parce que ces données sortent naturellement du corps humains. Ces données ont parcouru tout le corps humain, et contiennent en elles toutes sortes de traces de l'état de ce corps[9]. Ces

[7] Fondée en 2003 par Elizabeth Holmes, l'entreprise Theranos a longtemps prétendu permettre un bilan de santé général à partir d'une prise de sang minime. En 2015, Theranos est évalué à 9 milliards de dollars. En 2018, Theranos est poursuivi pour « fraude massive ». La technologie de Theranos supposément révolutionnaire n'a en fait jamais fonctionné. Les employés de Theranos semble avoir exploité en cachette des techniques classiques d'analyse sanguine, dont la fiabilité était en fait discutable.
:tv: Elizabeth Holmes - la menteuse devenue milliardaire. La chaîne de P.A.U.L. (2020).
:tv: Elizabeth Holmes exposed: the $9 billion medical 'miracle' that never existed. 60 Minutes Australia (2021).

[8] Selon la légende, pour déterminer si une couronne était vraiment en or, Archimède eut l'idée de mesurer son volume, en mesurant la quantité d'eau déplacée dans une baignoire pleine lorsque la couronne était déposée dans la baignoire. En combinant cette mesure à la mesure de la masse de la couronne, Archimède pouvait alors estimer si la couronne avait la densité d'une couronne en or. Quand il s'en rendit compte, selon la légende, Archimède était justement dans sa baignoire. Il s'écria « Eurêka », et sortit tout nu dans la rue pour partager l'enthousiasme de sa découverte.

[9] Les archéologues sont d'ailleurs particulièrement friands des cacas fossilisés, qui leur permettent d'en apprendre sur l'alimentation et la santé des populations passées.
:tv: Le caca, vrai trésor des archéologues - ft. Julien Ménielle, Passé sauvage, Pierre Kerner. C'est une autre histoire (2019)

> données, aujourd'hui, on les jette à chaque fois qu'on tire la chasse. Mais imagine tout ce qu'on pourrait diagnostiquer si, au lieu de les balancer dans les égouts[10], on collectait soigneusement ces données et on prenait le temps de les analyser ! Marc, il faut qu'on analyse le pipi et le caca !

Dans les jours qui suivent, Katia et Marc passent leur temps à discuter de ce projet d'analyse d'excréments. Ils concluent que le produit idéal serait des toilettes intelligentes[11], capables notamment de mesurer et d'analyser les excréments sous toutes leurs coutures avant leur périple dans les eaux usées. Cependant, un tel produit nécessiterait des années de recherche et développement, et donc un grand nombre d'investisseurs. Pour commencer, les deux amis penchent vers un projet moins ambitieux. Après tout, comme Katia le fait remarquer, le meilleur outil de collecte d'information moderne est le téléphone. Pourquoi ne pas l'utiliser pour prendre des photographies des excréments ?

C'est ainsi qu'est conçu petit à petit SmartPoop, destiné à être une application de collecte d'échantillons photographiques de matière fécale et d'analyse automatisée de ces images à l'aide d'algorithmes de machine learning. Pendant les mois qui suivent, Katia écrit sa thèse de doctorat le jour, et programme SmartPoop jusque tard dans la nuit.

Pendant ce temps, Marc s'informe sur la coprologie. Il télécharge des bases de données publiques d'excréments, et prend le soin de vérifier leurs annotations[12]. Certains excréments sont des petites boules solides ; une nette indication de constipation. D'autres sont une bouillie presque liquide ; une bonne grosse diarrhée. Certains étrons ont une forme parfaitement ondulée ; un signe de très bonne santé intestinale[13] ! La couleur donne aussi des informations sur le fonctionnement du foie ou la présence de sang.

---

[10] En fait, il reste possible d'effectuer des diagnostics médicaux à partir d'analyses des égoûts. C'est d'ailleurs ce qui a été fait dans le cadre du suivi épidémiologique du COVID-19. La concentration d'ARN du Sars-CoV-2 dans les égouts permet en effet d'inférer l'incidence du virus dans une population parfois très localisée.
:tv: Aux chiottes le virus ? Science4All (2020)
:memo: How sewage could reveal true scale of coronavirus outbreak. Smriti Mallapaty. Nature (2020)

[11] Il existe bel et bien des projets de toilettes intelligentes, ou *smart toilets*. Notamment une publication dans le prestigieux journal scientifique *Nature* sur un prototype, capable d'effectuer de la reconnaissance... anale !
:memo: A mountable toilet system for personalized health monitoring via the analysis of excreta. Seung-min Park et al. Nature Biomedical Engineering (2020).

[12] L'annotation (ou *étiquetage*, *label* en anglais) des données est une phase critique de la conception des bases de données d'entraînement des algorithmes d'apprentissage. Elle consiste à prendre des données brutes (comme des images), et à annoter l'image avec des informations pertinentes, comme la présence ou l'absence de chats, ou comme l'état de santé de la personne qui a produit les excréments photographiés.

[13] :tv: Pourquoi le wombat fait-il des crottes carrées ? La boîte à curiosité (2021).
:tv: L'importance des excréments dans la nature. La boîte à curiosité (2020).
:tv: MinuteEarth Explains: Poop. MinuteEarth (2020).
:books: Tout ce que vous devez savoir sur le plus tabou des sujets. Michel Lafond. Julien Ménielle (2018).

> Tu sais, en fait les selles devraient être incolores ou blanchâtres, explique Marc à Katia lors de l'une de leurs soirées de travail. Elles sont surtout composées d'eau et de fibres après tout. Mais c'est un déchet des globules rouges, la bilirubine, qui les rend foncés. C'est le foie qui s'occupe de la traiter et de la « mettre » dans les excréments.
>
> Je vois... et du coup quand le foie de fonctionne pas on devient marron ?
>
> Presque. Jaune plutôt. C'est ce qu'on appelle l'ictère. Ou « jaunisse » si tu préfères et les selles deviennent blanchâtres. Mais vois-tu...

Katia n'écoutait déjà plus réfléchissant au type d'analyse colorimétrique qu'elle pourrait intégrer à leurs algorithmes. « SmartPoop, ça va être trop cool[14] », se dit-elle intérieurement.

L'application est en ligne !

En mars 2019, Katia défend sa thèse. Une semaine plus tard, l'application SmartPoop est rendue disponible sur l'Apple Store et le Google Play Store[15]. Dans sa version initiale, SmartPoop permet ainsi aux utilisateurs de créer une base de données de leurs déjections solides, que Marc passe chaque soirée à analyser.

> J'ai passé des soirées bien chiantes pendant mes années de médecine, à voir des trucs peu ragoûtants... mais celles-là étaient presque les pires, dira plus tard Marc.

SmartPoop dispose aussi de fonctionnalités d'apprentissage[16]. En particulier, elle apprend ainsi progressivement des données, quotidiennement étiquetées par

[14] Le mot « cool » est souvent présent pour qualifier certaines technologies « dignes d'intérêt ». Il y a certainement une réflexion importante à avoir autour de ce qui rend une technologie ou une idée « cool ». En particulier, ceci peut parfois ne pas être directement relié à la capacité de la technologie ou de l'idée à vraiment rendre le monde meilleur. Typiquement, beaucoup de technologistes semblent trouver « cool » les cryptomonnaies décentralisées ou les performances des algorithmes de machine learning. Il semble toutefois que ceci puisse parfois les amener à ignorer ou à sous-estimer les dangers de ces technologies.

[15] Une brève recherche sur Google Play Store permet de se rendre compte qu'il existe déjà de nombreuses applications pour tracer ses selles, comme Poop Tracker (4,5 étoiles, 100k+ téléchargements), PoopLog (4,2 étoiles, 100k+ téléchargements) et Poopify (4,6 étroiles, 10k+ téléchargements).

[16] La forme la plus développée de machine learning est ce qu'on appelle *l'apprentissage supervisé*. Il s'agit du problème de deviner des propriétés (appelées *étiquettes*) de données brutes comme des images. Typiquement, les images avec un chat peuvent être étiquetées avec l'étiquette « chat ». Dans le cas de SmartPoop, les images de constipation seront étiquetées « constipation ». L'algorithme d'apprentissage va alors cherché à identifier des caractéristiques que les images « constipation » ont, et que les images sans constipation n'ont pas. Si l'algorithme est un succès, il exploitera alors ces caractéristiques pour ensuite généraliser l'étiquetage « constipation » à des images non-étiquetées qui possèdent les caractéristiques des images « constipation ».
Une grande partie de la conception des algorithmes de machine learning par apprentissage supervisé se joue alors dans l'annotation des données. En fait, ce travail est si laborieux que les entreprises du web profitent souvent des utilisateurs de leurs plateformes pour effectuer ce

Marc, et généralise ainsi les annotations de Marc pour prédire les risques de constipation ou de diarrhées chez les utilisateurs à partir d'images que Marc n'a pas eu le temps de visionner[17].

Katia passe alors ces journées à améliorer son application la nuit, et à promouvoir SmartPoop le jour. Elle intervient, en particulier, dans les instituts de recherche, dans les hôpitaux et dans les établissements d'hébergement pour personnes âgées dépendantes, mais aussi dans les réseaux d'entrepreneurs comme la Station F à Paris. Son TEDx à la London School of Economics lui vaut une ovation d'un public conquis par l'opportunité sanitaire.

Cependant, après des mois de promotions, SmartPoop ne décolle pas. Si l'application est téléchargée 1587 fois, elle n'est utilisée quotidiennement que par 75 utilisateurs[18] (dont une vingtaine parmi les proches de Katia et Marc). En juillet 2019, quatre mois après la mise en ligne de l'application, Katia se rend à l'évidence : SmartPoop est un échec.

> SmartPoop deviendrait incontournable si on avait beaucoup plus de données pour entraîner des algorithmes plus sophistiqués, pour ensuite prédire des maladies plus rares et plus dangereuses. Le potentiel sanitaire reste énorme, explique-t-elle à Marc. Mais on a grandement sur-estimé la motivation des utilisateurs à soutenir un tel projet de santé public[19]… Avec le recul, je dirais qu'on a été beaucoup trop bisounours dans notre conception du projet. Les venture capitalists avaient raison. On aurait dû beaucoup plus réfléchir à la croissance, à l'acquisition d'utilisateurs et au marketing du produit. On se serait peut-être rendu compte plus tôt que SmartPoop est une perte de temps. Même si on continue notre travail de promotion pendant des années, on risque de ne jamais dépasser 1000 utilisateurs réguliers. Et ça, ça ne suffira pas pour acquérir assez de données pour avoir un outil de diagnostic utile. SmartPoop, c'est sans espoir.

Katia décide alors de rentrer dans le rang, et accepte un emploi de développement informatique dans une grande entreprise, qu'elle débute en septembre. Cependant, malgré l'environnement de travail exceptionnel qui lui est offert, Katia n'est pas particulièrement enthousiasmée par son travail. Mais au moins, se dit-elle, celui-ci lui permet de payer son loyer.

---

travail à leur place, typiquement via des « CAPTCHA » qui permettent aussi de vérifier que l'utilisateur est humain.

[17] L'un des grands avantages des algorithmes sur l'humain est ce qu'on appelle leur capacité de *mise à l'échelle*. Ainsi, si un algorithme a les mêmes performances de diagnostic qu'un humain à partir d'images, il pourra être utilisé à relativement peu de frais par des milliards d'utilisateurs en même temps. Ceci offre des fantastiques opportunités commerciales, mais aussi médicales ou philanthropiques.

[18] Le grand défi de nombreux produits du web est souvent la rétention des utilisateurs.

[19] D'autres initiatives plus ludiques ont dès aujourd'hui un succès commercial, comme l'application Foodvisor qui permet d'estimer les calories dans un repas à partir d'une photo.

Le ROVID-19

Fin novembre 2019, toutefois, un événement va bouleverser le futur de l'humanité en général, et celui de Katia en particulier. La Kormique[20] déclare une multiplication préoccupante d'une nouvelle pathologie très contagieuse et potentiellement mortelle, qui semble affecter des milliers de kormicains, et qui semble avoir déjà fait des centaines de victimes. Le mois suivant, en décembre 2019, des cas similaires sont observés en Bokistan[21] et, bientôt, dans de plus en plus de pays aux quatre coins du monde. L'Organisation Mondiale de la Santé déclare l'état d'urgence : une nouvelle pandémie est en train de sévir.

On apprend ensuite que le coupable est un nouveau rotavirus, et la terrible maladie qu'il cause est baptisée « ROtaVIrus Disease 2019 », ou ROVID-19[22]. Le ROVID-19 cause de nombreux symptômes dérangeants, comme des maux gastriques, des maux de têtes, une fatigue accrue, des lourdes diarrhées, des grosses fièvres, des vomissements et des tremblements, qui conduisent près de 10% des personnes infectées vers le décès. Cette terrible maladie semble particulièrement mortelle chez les jeunes de moins de 30 ans.

Mais ce qui rend le ROVID-19 extrêmement dangereux, c'est son extrême contagiosité. Pire encore, le traçage des cas d'infections montre que cette contagiosité est particulièrement grande deux ou trois jours avant que les premiers symptômes se déclarent. On parle de contaminations pré-symptomatiques. Les personnes contaminantes ne sont pas encore conscientes d'être malades quand elles transmettent la maladie aux autres. Le taux de reproduction de base[23] du virus, c'est-à-dire le nombre moyen d'individus qu'une personne infectée va contaminer, est estimé à environ 8. La croissance exponentielle[24] de la pandémie

[20] En octobre 2021, Google Translate traduisait « caca » en « korm » en ukrainien. Il se trouve toutefois que korm signifie « nourriture pour animal », et non pas « caca ».

[21] « Bok » signifie « caca » en turque.

[22] Le ROVID-19 est bien sûr une maladie fictive, calquée sur le COVID-19 qui a frappé le monde en novembre-décembre 2019. Une collaboration exceptionnelle de vulgarisateurs scientifiques du web avait d'ailleurs produit une vidéo collective à ce moment là.
:tv: Coronavirus : Chaque JOUR compte (2020)

[23] Le taux de reproduction de base est le nombre moyen d'individus qu'une personne infectée contaminera, en l'absence d'interventions pour contrôler la propagation d'une épidémie. On estime que le COVID-19 avait initialement un taux de reproduction de base autour de 3, et que le variant Delta a un taux bien plus élevé (mais difficile à estimer, vu qu'il est apparu à un moment où de nombreuses interventions étaient déjà en oeuvre). Il s'agit d'une grandeur importante car, si elle est supérieure à 1, alors l'épidémie se propagera exponentiellement en l'absence d'intervention.
Le taux de reproduction effectif est le nombre moyen d'individus qu'une personne infectée contaminera, sachant toutes les mesures sanitaires en place. C'est finalement la grandeur la plus importante. Si elle est supérieure à 1, alors l'épidémie se propagera exponentiellement. Si elle est inférieure à 1, alors l'épidémie disparaîtra exponentiellement vite. En pratique, à cause de relâchements des mesures sanitaires, dans le cas du COVID-19 notamment, ce taux fluctue autour de 1, ce qui fait de l'épidémie une *endémie*, c'est-à-dire une maladie qui persiste dans la population générale.
:tv: Le futur dépend de ce nombre. Science4All (2020).
:tv: Epidemic, Endemic, and Eradication Simulations. Primer (2020).

[24] La croissance exponentielle intervient lorsque le nombre de cas est multiplié chaque

terrifie rapidement toutes les agences sanitaires, tous les hôpitaux et tous les gouvernements du monde.

À partir de début janvier 2020, tous les pays à travers le globe entrent tour à tour dans des périodes de confinement, alors que les tests médicaux se mettent en place petit à petit[25], notamment des tests groupés[26]. Les estimations de janvier sont terrifiantes. Des centaines de milliers de personnes à travers le monde semblent déjà affectées, et des dizaines de milliers de victimes ont déjà succombé.

Une suspicion initiale prend petit à petit de l'ampleur dans la communauté scientifique, notamment suite à l'observation d'une montée fulgurante de cas chez les techniciens de traitement des eaux usées. De plus en plus de données, notamment issues d'analyses chimiques d'excréments des malades, suggèrent que le virus du ROVID-19 se propage principalement via les flatulences des personnes infectées.

« Une très belle étude d'une équipe de Montpellier a aussi montré des clusters importants de malades parmi les gens fréquentant souvent des restaurants spécialisés en cassoulet », précisait souvent Marc afin de détendre l'atmosphère angoissante qui régnait alors. Le port de couches filtrantes[27] est alors conseillé pour toutes les personnes ayant besoin de se déplacer, pour éviter de transmettre le virus.

La quête de données

Et si SmartPoop pouvait aider ? Voilà une question que Marc pose à Katia peu de temps après le début du confinement. Katia répond qu'elle est malheureusement épuisée et débordée par la migration de nombreux produits de son entreprise vers des solutions adaptées au télétravail. Mais alors que Katia est encore en train de résoudre des bugs dans ses codes à 11 heures du soir, Marc insiste.

---

semaine par une constante supérieure à 1. Le danger d'une telle croissance, c'est qu'elle paraît insignifiante les premières semaines, mais devient tout à coup hors de contrôle après quelques semaines ou mois.
:tv: Aller sur la lune avec une feuille de papier et l'échiquier de Sissa. Fabien Olicard (2016).
:tv: Des nombres grands, TRÈS grands. Mickaël Launay (2014)

[25] Il est toujours bon de rappeler qu'un test médical ne permet pas de certifier si une personne a ou n'a pas la maladie en question, car les tests sont toujours imparfaits. En fait, si un individu reçoit un test positif (imparfait) d'une maladie extrêmement rare, alors il reste généralement probable qu'il ne soit en fait pas malade (il conviendra d'analyser alors ses autres symptômes). Les mathématiques des probabilités conditionnelles sont critiques pour bien interpréter ces tests médicaux.
:tv: La loi de Bayes - Argument frappant. Monsieur Phi (2016)
:tv: The medical test paradox, and redesigning Bayes' rule. 3Blue1Brown (2020)

[26] :tv: Les tests groupés : dépister plus avec moins. Science4All (2020).

[27] La contagiosité du pet, mais aussi la capacité des vêtements à filtrer les pets, ont été testées et vérifiées par Dr Karl Kruszelnicki, suite à une expérience de pets dans des boîtes de Petri.
« Notre conclusion finale ? Ne pétez pas nu près de la nourriture. »
:memo: Hot air?. Michael Doyle. The Canberra Times, Reprinted on BMJ (2021).

> On sait que le virus du ROVID-19 est non seulement très présent, mais aussi très actif dans les selles. C'est même très certainement par là qu'il se diffuse le plus. Mais ça veut aussi dire qu'il y laisse des traces. Aujourd'hui, on détecte ces traces en cherchant directement des bouts d'ARN du virus avec des méthodes de biologie classiques[28]. Mais si le virus est si présent, il est possible qu'il laisse une trace visible dans les selles ; qu'il rende les selles visuellement différentes. J'ai vu quelques photos de matières contaminées, et malheureusement, ça ne m'a pas sauté aux yeux. Mais s'il y a ne serait-ce qu'une infime différence, peut-être qu'un algorithme, lui, sera capable de voir la différence. Et si c'est le cas, on pourrait isoler les cas de ROVID-19, et peut-être contrôler la pandémie sans confinement. On sauverait alors des millions de vies, voire des centaines de millions de vies. Sans parler de tous les troubles de santé mentale...

Katia fait alors le lien avec un appel d'offres qu'elle a vu passer à son travail, et qui cherchait activement des projets d'informatique liés au ROVID-19. Sûrement, se dit-elle, les investisseurs seraient intéressés par un projet comme SmartPoop, si SmartPoop promet de résoudre la crise du ROVID-19. Mais pour les convaincre, il faudra d'abord avoir un *Proof of Concept*, ou PoC comme on le dit dans le jargon[29]. Autrement dit, il faudra une première version de SmartPoop, pas encore tout à fait fonctionnelle, mais suffisamment convaincante pour attirer ces investisseurs. Mais pour cela, Katia sait qu'il manque surtout à SmartPoop des données.

> Il nous faut des données de patients malades. Beaucoup de données de beaucoup de patients malades, s'exclame Katia. Tu penses que tu peux nous avoir ça ?

Marc passe alors ses journées à contacter tous les collègues dans son carnet d'adresse, puis tous les médecins qu'il connaît sur Facebook, puis tous les médecins qu'il trouve sur Twitter, pour les supplier de photographier les déjections des malades dans les hôpitaux. La plupart ne répondent pas[30]. Certains

[28] La méthode la plus standard pour détecter des bouts d'ARN est le test qPCR. Elle consiste à répliquer un morceau d'ARN cible un grand nombre de fois, via à une réaction polymérase en chaîne, et en insérant un signal fluorescent dans les copies. Ceci permet d'accroître exponentiellement le nombre de ces morceaux d'ARN, s'ils étaient initialement présents dans l'échantillon analysé, ce qui les rend plus faciles à détecter ensuite.
:tv: Le principe de La PCR et leur différents étapes. Physiologie Santé (2020).
:tv: Comment fonctionnent les tests de dépistage du Covid19 ? ENS Paris-Saclay (2020).

[29] En innovation, et en particulier en informatique, la conception d'un nouveau produit débute généralement avec le PoC, qui permet d'attirer des investissements, en temps et en argent, pour ensuite développer un produit fonctionnel. La seconde étape est ensuite la conception d'un *Minimum Viable Product*, ou MVP. Ce produit se veut minimaliste (pour être fabriqué rapidement), mais parfaitement fonctionnel (pour être vendu aux premiers clients). De nouvelles fonctionnalités sont ensuite ajoutées au MVP au fur et à mesure.

[30] La problématique des données des hôpitaux est en fait horriblement plus complexe, notamment à cause de considérations de protection des données sensibles des patients. L'infrastructure informatique des hôpitaux est d'ailleurs souvent soumise à des attaques de hackers malveillants, qui exploitent souvent le manque d'investissement dans la sécurité de cette infrastructure

rétorquent avec des insultes. « Vous avez déjà été dans un hôpital ? On est déjà débordé pour sauver des vies. On n'est pas là pour alimenter un compte Instagram de merdes », commentent les plus agressifs. « Ils abusent... On ne peut mettre aucun filtre, et tu ne peux pas en faire de story », ironise Marc.

Katia adapte alors le site web SmartPoop.com pour appeler les médecins, mais aussi le grand public à contribuer. Elle y demande aux visiteurs d'utiliser SmartPoop pour prendre des photos de leurs excréments, et de renseigner sur la plateforme leur état de santé jour après jour. Elle supplie également les médecins d'encourager leurs patients à utiliser SmartPoop.

Katia contacte aussi des médiateurs scientifiques, sur Twitter et sur YouTube, pour qu'ils communiquent sur ce projet. Science4Alpha[31], une vidéaste scientifique avec 200 000 abonnés accepte de vulgariser le projet de Katia. La vidéo très pédagogique de Science4Alpha récolte 100 000 vues dès la première semaine, et conduit à l'adoption de SmartPoop par des dizaines de milliers d'utilisateurs[32].

En mai 2020, alors que les nombres de cas diminuent lentement, mais demeurent encore très élevés, SmartPoop récolte quotidiennement des centaines de milliers de photographies. Au total, SmartPoop dispose alors de dizaines de millions de photos d'excréments. Malheureusement, si les algorithmes de Katia distinguent aisément les diarrhées glairosanglantes des cas avancés, ils échouent encore à détecter une quelconque différence entre les cas infectés pré-symptomatiques[33] et les cas sains. Katia est frustrée.

Ça semble peine perdue, s'exaspère-t-elle auprès de Marc.

Tu regrettes encore nos efforts ?

---

pour paralyser des services et exiger des rançons. Dans le cadre du roman, toutefois, on peut imaginer que la situation sanitaire est peut-être là suffisamment catastrophique pour justifier (éthiquement) des collaborations (illégales) avec SmartPoop.

[31] Le nom Science4Alpha est bien sûr un clin d'oeil aux chaînes Science4All et Risque Alpha des deux auteurs de ce roman.

[32] « Sans nos médiateurs scientifiques, qui informent, expliquent, enseignent, décodent, combattent la désinformation et débattent des questions scientifiques, beaucoup resteraient dans un espace où ils ne disposent pas des informations dont ils ont besoin, ce qui les conduirait à faire de mauvais choix à des moments vraiment cruciaux, » affirmait Jacinda Arden, Première Ministre de la Nouvelle-Zélande, en juillet 2020. Malheureusement, les collaborations avec le monde de la vulgarisation, pendant la crise du COVID-19 et autour de sujets comme le changement climatique ou l'éthique des algorithmes, ont sans doute été très déficientes à travers le monde, avec plus généralement un très probable manque d'investissements dans la communication scientifique. Elles sont aussi souvent très délicates, pour les autorités et pour les vulgarisateurs, surtout dans le climat de défiance actuel. Par exemple, la vidéo de Science4All sur les vaccins, en collaboration avec le Ministère de la Santé en France, a causé un lever de bouclier (sans doute accentué par une campagne de désinformation organisée), recevant plus de dislikes que de likes sur YouTube (ce qui est extrêmement rare !).
:tv: Un vaccin pour permettre aux étudiants de retrouver leur vie d'avant (ft. Prof. Fischer). Science4All (2021).

[33] Dans le cas du COVID-19 (et donc du ROVID-19), le grand défi du contrôle de la pandémie était justement le fait d'éviter les contaminations pré-symptomatiques, c'est-à-dire par des sujets infectés avant que ceux-ci ne développent des symptômes (et donc avant qu'ils se rendent compte qu'ils sont infectés).
:tv: Contenir la pandémie sans confinement ? Science4All (2020).

> Non... Je pense que, cette fois, le pari était bon[34]. Il y avait peu de chances que SmartPoop résolve la crise du ROVID-19. Mais si c'était le cas, on aurait sauvé l'humanité. Bon, ce n'est pas le cas. Mais, cette fois, je pense qu'on a bien fait d'essayer. Ça n'en reste pas moins frustrant...

Déçu, Marc reconnaît que les confinements vont probablement s'éterniser, probablement pendant des années, le temps qu'un vaccin efficace soit développé, testé et déployé à très grande échelle — si tant est qu'il voit le jour un jour[35]. Pendant ce temps, le ROVID-19 ne cessera de se diffuser.

En entendant ces mots, Katia se lève, agitant son index droit qui illustre alors le bouillonnement intellectuel qui anime ses neurones.

> Se diffuser... Mais oui ! C'est ça.

Katia se jette alors sur son ordinateur, et se met à coder. Marc la suit et demande des explications. Quelle est la nouvelle idée de Katia ?

SmartPoop versus ROVID-19

> Le virus du ROVID-19 se diffuse particulièrement bien à travers les flatulences, explique Katia. Mais de ce que je comprends, d'habitude, les flatulences ne diffusent pas très loin, car nous portons des culottes et des pantalons. Bon aussi des jupes et des robes... Mais oublions les jupes et les robes. Elles sont interdites depuis des mois maintenant ! Pour que le ROVID-19 se diffuse vraiment bien, il doit probablement affecter la manière dont les gaz sont produits et se diffusent. Mais ce gaz, il doit aussi être présent dans les excréments ! Et on devrait le voir. Mais pas dans une photo. Pour le voir, il faut une vidéo !

C'est ainsi que SmartPoop propose désormais, non pas de photographier les excréments, mais de les filmer ! Quelques jours plus tards, des centaines de milliers de vidéos de quelques secondes sont collectées par SmartPoop. Katia, qui n'a pas dormi entre temps, est alors sur le point d'achever la conception des nouveaux algorithmes de SmartPoop, désormais adaptés à l'analyse de vidéos.

---

[34] Katia insiste là sur la différence entre le *pari* et le résultat. Comme en parle très bien l'ancienne joueuse de poker Annie Duke, nous autres humains avons malheureusement trop tendance à juger une décision en fonction de son résultat, quand bien même ce résultat était imprévisible au moment de la décision, notamment sachant les informations dont on disposait alors. Pour progresser, selon Duke, il est critique de juger la prise de décision à partir de l'état de connaissance au moment de cette prise de décision, qui correspondait alors nécessairement à un pari, car le futur était incertain.
:tv: Jugez le pari. Pas le résultat. Science4All (2020).
:books: Thinking in Bets: Making Smarter Decisions When You Don't Have All the Facts. Penguin. Annie Duke (2019).

[35] Il est intéressant de se rappeler qu'au début de la pandémie COVID-19, il n'était pas clair que les vaccins auraient l'efficacité qu'ils ont finalement eue. Et il est aussi bon de rappeler que certains vaccins ont permis d'éradiquer certaines maladies terribles, comme la variole.
:tv: Le plus grand triomphe de l'humanité. Science4All (2020).

À 4 heures du matin, Katia rentre dans la chambre de Marc pour le réveiller. « J'ai fini l'algorithme. Il faut que tu vois ça. » Marc se réveille en sursaut, court chercher une bouteille de champagne et rejoint Katia dans le salon. Katia explique qu'elle a entraîné son algorithme avec 90% de la base de données de SmartPoop, et qu'elle s'apprête à tester les performances de l'algorithme sur les 10% restants[36]. Katia explique que ces 10% restants ont été tirés au hasard, avec la simple contrainte qu'ils contiennent autant d'excréments infectés pré-symptomatiques que d'excréments sains[37]. Si l'algorithme échoue, alors il aura un taux de reconnaissance d'excréments de 50%. S'il est parfait, sa précision sera de 100%.

Il ne reste plus qu'à lancer le test de l'algorithme pour connaître sa performance. Katia et Marc se lancent dans un décompte. Cinq. Quatre. Trois. Deux. Un. Le test est lancé[38].

Dix secondes plus tard, 10% du test est effectué. Il faudra donc attendre encore une minute et demie pour avoir les résultats. Pendant cette longue minute et demie, Katia et Marc ont le souffle coupé. Enfin, le résultat s'affiche. Le verdict : 52,4%.

Tête baissée, Marc se lève, et part remettre le champagne au réfrigérateur. Katia, elle, s'affale dans le canapé. Quand Marc revient dans le salon, Katia dort déjà. Il va lui chercher une couverture pour lui éviter de prendre froid. Puis il va se coucher à son tour. Le lendemain, Marc se réveille à 15 heures. Katia dort encore. En fait, Katia dormira vingt heures de suite.

Le réveil de SmartPoop

C'est encore au beau milieu de la nuit que Katia réveille tout à coup Marc.

[36] Ce qui est décrit là est la séparation du jeu de données en un jeu de données d'entraînement (*training set*) et un jeu de données de test (*test set*), qui est une technique classique en machine learning pour valider un algorithme après apprentissage. On parle alors de *validation croisée*, qui peut prendre des formes plus sophistiquées.
:tv: La validation croisée | Intelligence Artificielle 13 (ft.@La statistique expliquée à mon chat). Science4All (2018)

[37] Notez que le taux de succès de prédiction d'un algorithme dépend fortement du taux de choses à détecter dans les données (ici, le taux d'excréments infectés pré-symptomatiques). En effet, si 99% des données ne sont pas des excréments infectés pré-symptomatiques, alors un algorithme idiot qui systématiquement prédit « cet excrément n'est pas infecté pré-symptomatique » permet d'avoir un 99% de précision ! De façon plus générale, pour estimer le succès d'un algorithme prédictif dans une tâche de prédiction binaire (infecté versus pas infecté), il est nécessaire de préciser deux statistiques (par exemple le taux de base et la précision, ou par exemple le taux de faux positif et le taux de faux négatif).
:tv: Les grands scientifiques veulent se tromper. Science4All (2019)

[38] Cet instant est bien sûr très romantisé. En pratique, les datascientistes doivent souvent lancer et relancer les calculs un grand nombre de fois, en essayant d'ajuster les paramètres de l'apprentissage pour trouver une configuration qui marche bien. Par exemple, dans ce cas, Katia pourrait tester plusieurs architectures de réseaux de neurones, des astuces de « normalisation de batchs », différents optimiseurs (SGD, Adam...) avec différentes paramétrisations, et ainsi de suite. Le datascientiste finit souvent par progresser petit à petit, ou par se résigner. Mais ce travail laborieux est aussi bien sûr moins spectaculaire.

> Cinq écarts-types, cinq écart-types[39], répète-t-elle ! Le test n'a pas échoué. Il est en fait assez nettement au-dessus de 50%.
>
> Mais un taux de succès de 52,4% ne nous aidera absolument pas à arrêter le ROVID-19.

Katia explique alors que, en effet, l'algorithme actuel est très largement insuffisant. Cependant, la supériorité de 52,4% par rapport à 50%, cela suffit à suggérer que SmartPoop est bel et bien en train de relever un signal distinctif des excréments infectés.

> Si SmartPoop ne détectait absolument rien, alors on s'attendrait à un taux d'erreur de 50%, explique Katia. Mais pas exactement de 50%, à cause des fluctuations statistiques. Sachant que le test a été effectué sur des dizaines de milliers de vidéos d'excréments, on s'attendrait à obtenir 50% plus ou moins un erreur de l'ordre de 0,5%. Or, là, on est à 52,4%, soit 2,4% de plus que 50%. Un écart de 2,4%, c'est donc presque 5 fois la fluctuation de 0,5%[40]. C'est beaucoup. Et ça veut dire que la distinction existe très probablement ! SmartPoop n'est simplement pas encore capable de l'identifier !

Marc demande à Katia ce qu'il manque pour discerner ce signal. Katia répond :

> Si on veut diagnostiquer quoi que ce soit, il faut des données. Beaucoup de données. Et il va nous falloir aussi beaucoup de machines pour analyser toutes ces données. Mais maintenant qu'on a cinq écarts-types, je suis sûre qu'on va pouvoir trouver des investisseurs pour nous y aider ! On tient notre PoC !

Katia et Marc décident alors de se lancer à corps perdu dans le développement de SmartPoop. Katia démissionne de son entreprise, et passe désormais jour et nuit à améliorer les algorithmes de SmartPoop, à promouvoir l'application et à chercher des investisseurs. Elle loue alors plus de puissances de calculs encore sur les serveurs d'Amazon Web Service, et appelle aussi ses anciens camarades de thèse, pour qu'ils l'aident dans le développement de SmartPoop.

---

[39] En sciences, et notamment en physique en particulier, on parle parfois de « 5 sigmas ». Il s'agit du signal considéré suffisamment marquant pour être parfois qualifié de « découverte scientifique », quoique son interprétation exacte est en fait complexe, voire très trompeuse. En particulier, l'utilisation de tels signaux est très critiqués, notamment par les statistiques dites *bayésiennes*.
:tv: La plus grosse confusion des sciences : la p-value !! :hot_pepper: Science4All (2019).

[40] Notez que Katia fait bien attention à parler d'intervalle de *fluctuation*, et non pas d'intervalle de *confiance*. Ces deux notions sont souvent confondues à tort, alors qu'elles décrivent des objets assez distincts. Dans le premier cas, il s'agit d'une incertitude sur les données à observer, alors que, dans le second cas, il s'agit d'un intervalle qui estime les valeurs d'un paramètre d'un modèle, à partir des données observées. Cependant, cet intervalle de confiance doit aussi ne pas être confondu avec un troisième type d'intervalles, appelé intervalle de *crédence* (ou de *crédibilité*). Contrairement à l'intervalle de confiance, l'intervalle de crédence prend aussi en compte l'état global des connaissances scientifiques avant d'avoir observé les données collectées.
:tv: Peut-on faire confiance aux intervalles de confiance ? :hot_pepper: Science4All (2019)

Marc, lui, passe son temps à tester SmartPoop, et à suggérer des améliorations de l'interface pour rendre son utilisation plus facile et compréhensible pour tous ces utilisateurs[41]. Marc contacte également régulièrement différents médias, et les appelle à promouvoir SmartPoop pour recueillir davantage d'utilisateurs et de données. Science4Alpha parle régulièrement des progrès de SmartPoop dans ses vidéos YouTube, et encourage ses collègues du web à en faire de même.

Jour après jour, la performance de SmartPoop s'améliore. En juillet 2020, elle passe à 55%. En août, elle passe à 60%. Katia et Marc sont désormais invités sur les plateaux de télévision, pour parler de SmartPoop. Les journaux nationaux titrent : « Filmez vos excréments pour sortir du confinement ! »

C'est alors qu'un investisseur, appelé Luc, décide d'investir 10 millions d'euros pour 10% de SmartPoop, dont Katia et Marc deviennent alors co-fondateurs. SmartPoop embauche ainsi ses premiers développeurs, chargés d'améliorer l'application et les algorithmes de SmartPoop, ainsi que des commerciaux pour encourager l'adoption massive de l'application. Cet argent permet aussi de payer les factures de plus en plus importantes des serveurs de calculs.

Le triomphe de SmartPoop

En décembre 2020, SmartPoop possède désormais près de 100 millions d'utilisateurs réguliers, et plusieurs milliards de vidéos d'excréments atteignant un total de 320 années de vidéos[42]. Mais surtout, les performances de SmartPoop atteignent alors 90%. L'application est alors auditée et approuvée par les autorités sanitaires, qui encouragent désormais son adoption massive. Après une année complète de confinement, en janvier 2021, celui-ci est enfin levé, et la population retrouve une vie plus normale.

> C'est super, explique Marc, invité à paraître sur Science4Alpha. Le taux de reproduction de base du ROVID-19 est autour de 8. Si on suppose que dès qu'un individu est diagnostiqué positif par SmartPoop, alors lui et ses colocataires s'isolent chez eux, sachant que le taux d'erreur est de 10%, on devrait ainsi diviser le taux de reproduction par 10, ce qui théoriquement nous ramène à 0,8. Comme 0,8 est en dessous de 1, cela nous donne une chance de contenir le ROVID-19, sans requérir de confinement global. Mais bien sûr, il ne s'agit là que d'estimations. Il reste crucial que l'on prenne encore soin de la distanciation physique, des gestes barrières et du port de couche, et de surveiller constamment le taux de reproduction effectif qui détermine comment la maladie se propage.

Petit à petit, tous les pays au monde adoptent SmartPoop, désormais utilisé

[41] En informatique, on parle de UX/UI design, pour expérience utilisateur et interface utilisateur.

[42] Chaque utilisateur régulier met en ligne une vidéo par jour, depuis plusieurs mois, d'où les milliards de vidéos. Chaque vidéo fait quelques secondes, ce qui représente environ 10 milliards de secondes de vidéos, soit environ 320 années de vidéos.

par 3 milliards d'humains sur terre. Le ROVID-19 est alors contenu à quelques milliers de cas seulement par pays. Fin 2021, l'Organisation Mondiale de la Santé l'annonce publiquement. Grâce à SmartPoop, dont la précision atteint désormais 99%, le ROVID-19 est désormais déclaré contenu.

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

Chapitre 2 — Filtrer les données fécales

Le couple sort tremblant de l'hôpital. Quelques mots épars prononcés par le cancérologue[43] quelques minutes plus tôt résonnent encore dans leurs têtes. « Échappement thérapeutique », « métastases osseuses », « traitements expérimentaux ». Lucile Polmon, qui accompagnait son mari pour la consultation, connaît bien ces termes. Elle en a déjà lu les définitions sur internet, plusieurs fois et elle sait bien à quoi cela correspond pour celui qu'elle aime.

> Ça va aller, dit-il en retenant ses larmes... Ça va aller.

Malheureusement, l'optimisme de façade du mari de Lucile ne suffira pas. Deux ans plus tard, malgré des séances de chimiothérapie intensive[44], le mari décède, à seulement 42 ans. Il abandonne ainsi Lucile avec leur fille unique.

Depuis lors, Lucile, profondément traumatisée par le cancer de son mari, dévore toutes les informations de santé qu'elle peut trouver. Le jour où SmartPoop était sorti, elle avait aussitôt téléchargé l'application. Elle était l'une des premières à adopter l'application quotidiennement. Très rapidement, elle a forcé sa fille de douze ans, Jeanne, à utiliser SmartPoop elle aussi. Tous les soirs, désormais, elle analyse ses propres données SmartPoop et celles de sa fille, à la recherche de toute anomalie.

> Maman, je vais bien. Ce n'est pas la peine de surveiller constamment ton truc là, s'énerve Jeanne. T'es en train d'en faire une obsession malsaine. Depuis qu'on utilise SmartPoop, tu es tout le temps sur les nerfs, tu ne manges quasiment plus et je suis sûre que tu ne dors pas assez.
>
> Le jour où tu auras un cancer, tu riras moins.
>
> Personne n'en ri, maman. Je me fais juste du souci pour toi.

[43]:tv: Le cancer. Science Étonnante (2017).

[44][:tv: Chimiothérapie. Geneva University Hospitals (2019).])https://www.youtube.com/watch?v=eZT8Hgsyh1Q)

Mais l'obsession de Lucile ne se limite toutefois pas à SmartPoop. Elle passe ses journées à lire toutes sortes de blogs de santé du web, à regarder des vidéos YouTube sur de la médecine alternative et à fréquenter des groupes Facebook où des inconnus partagent leurs expériences personnelles[45].

La radioactivité des bananes

Un soir, sur son téléphone, Lucile tombe sur une vidéo YouTube sur les impacts sanitaires de la radioactivité, et qui affirme que le potassium est radioactif. Même si la vidéo affirme que cette radioactivité n'est pas préoccupante[46], l'algorithme de YouTube recommande[47] alors une autre vidéo intitulée « Les bananes, un OGM conçu pour exterminer la population ». Curieuse d'un tel titre, Lucile clique sur la vidéo. Elle y apprend que la banane moderne n'a rien de naturel. Selon l'auteur de la vidéo, la banane résulte de manipulations par des groupes agro-alimentaires[48]. La vidéo fait aussi intervenir un témoin masqué.

> Le potassium a été injecté dans les bananes naturelles pour en retirer les noyaux et pour qu'elles se vendent ainsi, affirme-t-il. Ça a rapporté des milliards d'euros aux exploitations industrielles. Quand j'ai démontré que le potassium était radioactif, j'ai reçu des menaces. Et quand j'ai insisté, j'ai été licencié[49].

[45] Quelques études scientifiques suggèrent un lien étroit entre détresse mentale et croyances en la médecine alternative.
:memo: Are modern health worries associated with medical conspiracy theories? Journal of Psychosomatic Research. Y Lahrach and A Furnham (2017).
:memo: Belief in a COVID-19 Conspiracy Theory as a Predictor of Mental Health and Well-Being of Health Care Workers in Ecuador: Cross-Sectional Survey Study. Xi Chen, Stephen X Zhang, Asghar Afshar Jahanshahi, Aldo Alvarez-Risco, Huiyang Dai, Jizhen Li et Verónica García Ibarra. JMIR Public Health and Surveillance (2020).

[46] La radioactivité est en fait omniprésente. Ce qui la rend dangereux n'est pas sa présence, mais la dose de la radioactivité. Bien sûr, celle de la banane est beaucoup trop faible pour être préoccupante pour la santé.
:tv: La radioactivité et notre exposition aux rayonnements ionisants.. Le Réveilleur (2019).

[47] Cet algorithme de recommandation est devenu l'une des entités les plus influentes au monde. Pour s'en rendre compte, il est utile de constater quelques statistiques. Depuis 2016, il y a plus de vues sur YouTube que de recherches sur Google. En 2019 (avant le COVID-19 !), les vues sur YouTube représentaient un milliard d'heures de visionnage pour deux milliards d'humains sur terre, soit une moyenne d'une demi-heure par jour. Or l'algorithme de recommandation de YouTube est responsable de 2 vues sur 3. Après tout, à chaque fois qu'un utilisateur va sur YouTube.com ou clique sur l'application YouTube sur son téléphone, c'est cet algorithme qui décide quelles vidéos seront proposées à l'utilisateur, sans parler de l'auto-play ou de la barre de proposition à droite sur le site. Dans leur livre Le Fabuleux Chantier, El Mahdi El Mhamdi et Lê Nguyên Hoang affirment que ceci fait de l'algorithme de YouTube l'entité la plus puissante au monde, car il est capable d'enfermer certaines populations dans leurs convictions, ou de réduire au silence certaines informations en ne les recommandant jamais.
:tv: L'IA nous gouverne déjà. Science4All (2018).
:books: Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique. Lê Nguyên Hoang et El Mahdi El Mhamdi. EDP Sciences (2019)
:computer: YouTube (Wiki Tournesol).

[48] :tv: 4 Histoires Terrifiantes sur les BANANES. Trash (2020).

[49] :tv: Les réseaux sociaux sont dangereux. Très dangereux. Science4All (2021).

Deux heures plus tard, Lucile est encore sur son téléphone[50]. Elle découvre désormais des groupes Facebook qui dénoncent le « scandale du potassium », et qui exigent l'interdiction de tous les produits contenant du potassium, à commencer par la banane. Ces sites expliquent également que la consommation de gélules de magnésium réduit le taux de potassium. « Le magnésium détruit les particules de potassium », affirment certains messages.

Trois heures plus tard, au milieu de la nuit, Lucile découvre maintenant un blog qui présente le magnésium comme un traitement miracle, non seulement contre les particules radioactives, mais aussi contre les tempêtes solaires et ses effets cancérigènes[51]. « Nous l'avons testés et approuvés, sur les milliers de membres de notre communauté », affirme le blog.

Il est quatre heures du matin quand Lucile décide enfin de se coucher. Avant de rejoindre son lit, elle ouvre toutefois l'application SmartPoop. Elle y découvre que SmartPoop y fournit une estimation de la teneur en magnésium et en potassium des excréments. « Excellent », se dit-elle. « Je vais pouvoir surveiller ma santé, et celle de ma fille ».

C'est alors avec une ferme intention qu'elle ferme les yeux pour s'endormir. Dès son réveil, elle achètera des compléments alimentaires de magnésium — ça tombe bien le blog proposait un lien vers un site qui en vendait[52]. Tout ça pour sa santé, et plus encore pour celle de sa fille.

Tout pour sa fille

Le lendemain soir, avant de manger, Lucile dit à Jeanne : « Je veux que tu prennes trois doses de magnésium par jour. Un le matin, un le midi et un le soir. »

[50] Les plateformes du web ont des intérêts économiques énormes à rendre leurs produits addictifs pour que les utilisateurs restent sur leurs plateformes. On parle parfois de *l'économie de l'attention*. En 2021, les *Facebook files*, ainsi que la lanceuse d'alerte Frances Haugen, ont révélé le fait que, depuis plusieurs années, Facebook a sciemment et systématiquement privilégié la rétention de l'attention de ses utilisateurs aux risques que ceci endurait sur leur santé mentale, sur la désinformation médicale et sur les tensions géopolitiques.
:tv: L'économie de l'attention : le commencement. Stupid Economics (2018).
:tv: Big Tech's Battle For Our Attention. BrainCraft (2018).
:tv: The Social Dilemma. Netflix (2020)
:computer: The Facebook Files. Wall Street Journal (2021)
:tv: Facebook Whistleblower Frances Haugen: The 60 Minutes Interview (2021).

[51] Les tempêtes solaires posent en fait de sérieux risques pour les systèmes électriques et électroniques, et peut être très cancérigène pour un astronaute dans l'espace, voire pour un équipage dans un avion. Cependant, au niveau de la mer, le champ magnétique terrestre nous protège largement de ses effets cancérigènes.
:tv: Éruption solaire et risque d'apocalypse électrique... - Astro'Stylé #07 - String Theory (2018)
:tv: Sur le soleil, il pleut aussi | Miho Janvier | TEDxSaclay (2018)

[52] Beaucoup de sites de désinformation médicale sont très lucratifs, soit parce qu'ils vendent directement de la médecine alternative, soit parce qu'ils vendent de la publicité ciblée pour un public vulnérable, qu'ils pourront tarrifier à haut coût.
:computer: Facebook 'still making money from anti-vax sites'. The Guardian (2021)

> T'es sûre que c'est une bonne idée ?
>
> Si tu veux éviter d'avoir un cancer comme papa, il te faut ces trois doses de magnésium.
>
> Mais la boîte dit de ne pas prendre plus d'une dose par jour.
>
> Oublie ce que dit la boîte et crois-moi. Il te faut trois doses de magnésium par jour. Et tiens-toi éloignée des bananes aussi. Ce sont des OGM pleins de potassium radioactif, produits par des entreprises capitalistes.
>
> Hein ? Mais tu es complètement folle !
>
> Comment oses-tu parler comme cela à ta mère ? Excuse-toi !
>
> Calme-toi et je m'excuserai.
>
> Prends tes trois doses et va te coucher.
>
> Je vais me coucher, mais je ne prendrai pas tes trois doses. Tu délires complètement.

Lucile prend alors une boîte de magnésium et la lance sur sa fille qui reçoit la boîte dans l'œil. S'en rendant compte, Lucile court vers Jeanne pour s'excuser. Mais Jeanne se lève alors et court vers sa chambre en pleurant. Au moment de quitter la salle à manger, elle se retourne vers Lucile et lui dit : « Tu devrais avoir honte de la manière dont tu traites ta fille. »

Seule et au bord des larmes, impuissante et se sentant détestée, Lucile trouve refuge dans les groupes Facebook qu'elle a fréquentés la veille[53]. Elle lit le témoignage d'autres internautes qui partagent une expérience similaire : « Ma famille est aveugle et naïve ». « Ma femme ne veut pas me croire ». « J'ai décidé de rajouter le magnésium directement dans la sauce que je leur sers ». Cette dernière citation inspire Lucile. Pour sauver sa fille et la protéger de tout risque de cancer, Lucile décide désormais d'insérer directement le magnésium dans la nourriture qu'elle prépare.

Les jours qui suivent, malgré les excuses répétées de Lucile, les repas de famille sont tendus et silencieux. À chaque fin de repas, toutefois, Lucile regarde les données SmartPoop de sa fille. À sa grande satisfaction, le taux de magnésium augmente régulièrement. En revanche, c'est avec frustration que Lucile constate que le taux de potassium ne diminue pas. Après une semaine, Lucile reste inquiète. « Le potassium est radioactif, et la radioactivité cause des cancers »,

[53] Selon Christian Picciolini, ancien extrêmiste et désormais actif dans les mouvements de dé-radicalisation, les victimes de la radicalisation sont souvent des personnes qui se sentent elles-mêmes tristes, impuissantes et détestées, après avoir subi ce qu'il appelle des « potholes » (des coups durs).
:tv: My descent into America's neo-Nazi movement & how I got out | Christian Picciolini | TEDxMileHigh (2017). :books: Breaking Hate Confronting the New Culture of Extremism. Christian Picciolini. Hachette books (2020).

ne cesse-t-elle de lire sur Facebook. Tant que ce taux de potassium ne sera pas réduit à zéro[54], sa fille sera à risque de cancer. Lucile trouve cela inacceptable.

Le groupe Facebook fétiche de Lucile explique que chaque individu est unique et que les dosages de magnésium peuvent beaucoup varier entre les individus. De fait, si le potassium ne baisse pas, il faut augmenter les doses de magnésium. Le groupe ajoute que les diarrhées qui peuvent en découler sont le signe que le traitement commence à fonctionner[55]. Convaincue par ces explications, Lucile décide alors d'augmenter les doses de magnésium, tout en prenant soin que le taux de magnésium dans les excréments ne dépasse jamais le seuil considéré dangereux par les sites web qu'elle fréquente. Semaine après semaine, ces doses augmentent. Elles passent à deux doses par jour par personne, puis à quatre, puis à huit. Rien n'y fait pourtant. L'application indique que Jeanne a constamment la diarrhée mais les taux de potassium ne diminuent pas. Ceci dit, dans le même temps, les taux de magnésium n'atteignent pas les seuils jugés dangereux.

Un soir, toutefois, après avoir fini sa soupe, Jeanne devient pâle. Elle se lève de sa chaise, fait deux pas, puis s'écroule. « Jeanne, Jeanne », s'écrit Lucile. Elle appelle les secours, qui emmènent Jeanne en unités de soins intensifs. Lucile est priée de patienter dans la salle d'attente, pendant que Jeanne est suivie de près. Le médecin revient finalement vers Lucile, lui expliquant que Jeanne est malheureusement décédée d'une overdose de magnésium.

> Ce n'est pas possible. Je surveille quotidiennement son taux de magnésium, et il est très en dessous des taux dangereux pour la santé, s'exclame Lucile.
>
> Je ne sais pas quel outil de mesure vous utilisez, mais je peux vous assurer que son taux de magnésium n'avait rien de normal et de sain.

L'erreur des barres d'erreurs

C'est toute énervée que Katia entre alors dans le bureau de Marc. « On a un énorme problème ». Elle lui explique que SmartPoop vient de recevoir une lettre de l'avocat de Lucile, qui explique l'histoire du décès de Jeanne.

> La mère nous poursuit maintenant en justice, ajoute Katia. Elle reproche à SmartPoop d'avoir menti sur la quantité de magnésium dans des excréments. En sous-estimant le risque d'overdose, elle estime que SmartPoop l'a encouragée à augmenter les doses de magnésium données à sa fille. Selon elle, SmartPoop est co-auteur de l'homicide involontaire de sa fille[56]. C'est n'importe quoi !

[54] En fait, une grave déficience en potassium, appelée hypokaliémie, est aussi dangereuse pour la santé. Tout comme un excès de potassium, appelé hyperkaliémie.

[55] L'excès de magnésium conduit effectivement à des diarrhées.
:memo: Fecal Excretion of Soluble Magnesium by Humans. David Saunders and Hugh Wiggins. Western Journal of Medicine (1983).

[56] D'après l'article 221-6 du code pénal français, « le fait de causer, dans les conditions et selon les distinctions prévues à l'article 121-3, par maladresse, imprudence, inattention, négligence

> Et c'est vrai ? SmartPoop a-t-il sous-estimé le taux de magnésium ?
>
> Je ne sais pas. Je n'ai pas vérifié les données.

Après un silence, Katia rajoute toutefois :

> « Mais oui ça me paraît probable. On n'a jamais entraîné nos algorithmes avec des données d'excréments avec de telles doses aberrantes de magnésium. Les algorithmes n'ont sans doute pas réussi à estimer ces doses adéquatement[57].
>
> Tu peux vérifier les chiffres maintenant ?

Katia va alors dans son bureau, suivie de Marc. Elle ouvre son ordinateur, et tapote sur le clavier de manière frénétique. Après quelques minutes, elle trouve enfin l'information.

> J'ai trouvé. Oui en effet, la dose est sous-estimée d'un facteur dix. Cependant, on a bien une barre d'erreur immense[58] qui contient la valeur trouvée par les médecins. L'algorithme est juste. On n'est pas en tort.
>
> Certe mais Katia, on ne peut pas s'attendre à ce que les utilisateurs interprètent correctement les barres d'erreur[59]. La plupart des

---

ou manquement à une obligation de prudence ou de sécurité imposée par la loi ou le règlement, la mort d'autrui constitue un homicide involontaire puni de trois ans d'emprisonnement et de 45 000 euros d'amende.
En cas de violation manifestement délibérée d'une obligation particulière de prudence ou de sécurité imposée par la loi ou le règlement, les peines encourues sont portées à cinq ans d'emprisonnement et à 75 000 euros d'amende. »

[57] Katia souligne là l'un des grands défis de la sécurité des algorithmes, à savoir la validation de leurs performances sur des cas critiques et rares (parfois appelés *edge cases*). D'un côté, la rareté de ces cas fait qu'on dispose de très peu de données, voire parfois d'aucune donnée, pour permettre de tester la sécurité de l'algorithme face à ces cas. De l'autre, le fait que ces cas sont critiques fait que la sécurité de l'algorithme pour ces cas précis peut être une question de vie ou de mort. Toutefois, même si chaque cas rare et critique est très rare, l'ensemble de tous les cas rares et critiques peut rester important et survenir avec une probabilité non-négligeable, surtout lorsque l'algorithme est utilisé des milliards de fois par jour. Malheureusement, comme en parle cette publication dans le cas des voitures autonomes, il n'existe pas beaucoup de méthodes efficaces de validation d'un algorithme face aux cas rares et critiques. L'article en propose justement une nouvelle.
:memo: Efficient statistical validation with edge cases to evaluate Highly Automated Vehicles. Dhanoop Karunakaran, Stewart Worrall et Eduardo Nebot. ITSC (2020).

[58] Lors de collectes ou d'analyses de données, il est généralement très utile de préciser l'incertitude sur les données, ou sur les estimations inférées des données. La manière la plus simple d'y parvenir, c'est de rapporter un intervalle comme « entre 5 et 200 milligrammes par litre de sang », qui idéalement correspond à un intervalle de crédence (qui diffère d'un « intervalle de confiance »). Un tel intervalle devrait alors être interpréter comme suit : « selon le modèle utilisé par SmartPoop et sachant les données collectées par SmartPoop, SmartPoop estime que, avec crédence 95%, la concentration de magnésium dans le sang du patient est entre 5 et 200 milligrammes par litre ».
:tv: Peut-on faire confiance aux intervalles de confiance ? :red_pepper: Science4All (2019).

[59] La précision des tests ADN par les entreprises spécialisées dans leur séquençage est en fait étonnamment mauvaise. Dans un reportage pour CBC News, deux jumelles ont reçu des estimations différentes de leurs origines ethniques par la même entreprise, après avoir envoyé

utilisateurs ne comprennent pas du tout ces barres d'erreur. Ils ne comprennent pas que si on estime à 20 milligrammes par litre, avec une barre d'erreur entre 5 et 200, ça veut dire que ça pourrait très bien être 5 milligrammes par litre ou 200 milligrammes par litre. La plupart des gens vont juste retenir le chiffre de 20 milligrammes par litre[60].

Ouais. Enfin... Ce n'est pas notre faute si certaines personnes sont débiles[61].

Voyons, Katia...

En tout cas, on est dans le caca. La mère demande dix millions d'euros de réparation.

C'est beaucoup ! Mais bon, il suffit de payer, non ?

Ce n'est pas si simple, Marc. Si ça, c'est arrivé à une personne...

Ça va arriver à plein d'autres personnes aussi !

On a maintenant 3 milliards d'utilisateurs. Si 0,0001% d'entre eux ont une histoire comme cela, on va se retrouver avec 3000 procès[62]. On est dans un gros pétrin...

---

des échantillons de leurs salives au même moment ! Et ces estimations étaient très différentes d'une entreprise de séquençage ADN à l'autre. De façon intrigante, l'une des entreprises, 23AndMe, avait une option sur la crédence des résultats de l'analyse, qui était placée par défaut à 50%. Lorsque cette crédence était déplacée à 90%, l'analyse fournissait alors des origines géographiques très vagues, comme « quelque part en Europe ». La réglementation semble dangereusement laxiste vis-à-vis de ce qui semble être clairement une mésinformation potentiellement dangereuse, vendue par des entreprises privées.
:tv: Twins get 'mystifying' DNA ancestry test results (Marketplace). CBC News (2019).

[60] Ce que Marc décrit là correspond à la différence entre raisonner avec l'étendue de l'ignorance et raisonner avec uniquement le modèle (déterministe) qu'on juge le plus probable. En bayésianisme, on parle de la différence entre le *multivers bayésien* (qui décrit tous les scénarios crédibles) et le *maximum a posteriori*. Face à l'incertitude, peut-être parce que cette incertitude nous fait peur ou est trop complexe, nous avons souvent trop tendance à raisonner avec le second plutôt qu'avec le premier.
Le multivers bayésien. Science4All (2021).

[61] Rejeter la faute sur les utilisateurs (et leurs erreurs) est une excuse récurrente des entreprises du web pour ne pas prendre la responsabilité d'une mauvaise conception de leur expérience utilisateur. Cependant, il est bon de rappeler que la loi encadre en grande partie cela. Si un fournisseur vend un produit à un client en sachant pertinemment que l'utilisation de ce produit par ce client mettra en danger le client ou d'autres personnes, alors il a alors une responsabilité légale dans cette mise en danger d'autrui (voir l'article 221-6 du code pénal français, dans la note de bas de page [^homicide]).

[62] C'est ce que Lê Nguyên Hoang appelle l'effet « N epsilon ». Si un risque très faible (de probabilité epsilon) affecte un très grand nombre d'individus, alors notre intuition va généralement être très inadéquate pour déterminer si le risque à l'échelle de la population est important ou négligeable, car notre intuition a souvent une très mauvaise intuition des très grands nombres et des très petits nombres. Il convient alors d'essayer de pauser le calcul, pour mieux estimer les risques.
:tv: Un million de milliards de dilemmes. Science4All (2021).

La responsabilité de SmartPoop

Après un long silence, Marc demande alors : « Katia, tu penses qu'on est responsable de ce décès. Est-ce qu'on est vraiment coupable d'un homicide involontaire ? »

> On a sauvé des millions, peut-être même des milliards de vies, en diagnostiquant le ROVID-19[63]. SmartPoop, c'est un produit cool. C'est complètement injuste de penser que SmartPoop a quelque chose à voir avec un homicide !
>
> Oui mais est-ce que tuer une jeune fille peut être justifié par en avoir sauvé un million d'autres[64] ?
>
> Marc, on ne l'a pas tué cette jeune fille. Arrête de déconner. Notre algorithme n'a même pas commis d'erreur ! C'est surtout sa mère qui est suffisamment cinglée pour la bourrer de magnésium ! On n'y est pour rien si elle n'est pas foutue d'interpréter correctement des barres d'erreurs.
>
> Certes. Mais d'un certain point de vue on l'a aidée à le faire. C'est un peu comme si quelqu'un voulait se suicider, et qu'on lui tendait un pistolet pour y arriver[65].
>
> Doucement Marc. Et surtout, ne va pas dire ça au tribunal.
>
> Ça ne te dérange pas Katia que cette jeune fille soit morte, et que SmartPoop semble avoir une part de responsabilité dans ce décès ?
>
> Comment ça, « une part de responsabilité » ? On n'a rien fait d'incorrect ! Et puis, il y a plein d'autres causes dans cette affaire, comme les théories du complot que cette mère a gobées. Comparativement, nous, on n'y est pour rien[66]. On ne fait que donner des

[63] Cette remarque fait appel à une notion de crédit moral, l'idée selon laquelle le fait d'avoir agi très moralement dans le passé justifie des actions moins morale à l'avenir.
:tv: Moral Licensing. Mind Field S3E2. VSauce (2018).

[64] Marc pose là une question classique (et en fait un peu trop caricaturale) de l'opposition entre la déontologie et l'utilitarisme.
:tv: Jusqu'où serez-vous utilitariste ? (Ft. Science4All). Monsieur Phi (2017).
:tv: Encore plus utilitariste ? Monsieur Phi (2017).

[65] Il est intéressant de noter que le port et même la vente d'armes est interdit pour le grand public dans la plupart des démocraties à travers le monde, probablement parce que ces technologies représentent un danger pour autrui, voire pour soi-même si ces technologies sont mal utilisées. On pourrait faire la remarque que, de façon similaire, beaucoup de technologies de l'information sont aujourd'hui mal utilisées et représentent un danger pour autrui, notamment lorsqu'elles promeuvent massivement des discours de haine, du cyber-harcèlement ou de la désinformation médicale qui, en temps de COVID-19, peut conduire à des épidémies et à des mesures sanitaires dangereuses et contraignantes pour tous.

[66] Katia souligne là l'aspect *multifactoriel* du décès de Jeanne. De façon plus générale, l'idée d'identifier *une* cause ou *un* coupable (ou au moins *un* responsable) semble limitée à des contextes monofactoriels. Cependant, la complexité des flux de l'information modernes et le fait qu'il intègre de nombreuses entités rendent le raisonnement monocausal bancal pour analyser ce qui devrait être fait à l'avenir pour éviter de telles situations tragiques. Une

statistiques parfaitement objectives[67].

Je veux dire, « contrafactuellement[68] ». Dans un monde où cette mère n'avait pas accès à SmartPoop, elle n'aurait sûrement pas osé donner autant de magnésium à sa fille. Sans SmartPoop, cette jeune fille serait probablement encore en vie.

Déjà tu n'en sais rien. Je suis sûr que même sans SmartPoop les gens feraient n'importe quoi avec les compléments alimentaires. Et à ce jeu-là, Facebook et YouTube me semblent largement plus responsables que nous[69]. La mésinformation partagée sur ces plateformes, à l'origine, c'est surtout ça qui a causé toute cette histoire[70].

Katia, je pense que tu es beaucoup trop orientée sur le business ou la protection de SmartPoop. En fin de compte, l'objectif de SmartPoop, ce n'est pas de faire de l'argent ou de protéger notre image, c'est de sauver des vies, et d'éviter de causer des torts[71].

---

approche plus systémique semble nécessaire.

[67] Les algorithmes sont parfois considérés « objectifs ». À bien y réfléchir, ce qui les rend plus « fiables », c'est plutôt leur transparence (s'ils sont Open Source), ou au moins la reproductibilité de leurs calculs. Cependant, il peut être trompeur d'y voir quelque chose d'objectif, dans la mesure où un algorithme alternatif aurait pu conduire à des conclusions distinctes. Au moins peut-on dire que les statistiques calculées et transmises par l'algorithme sont sujettes au choix de l'algorithme utilisé pour effectuer les estimations statistiques (et justement, une grosse partie de l'éthique de l'information consiste à déterminer quels algorithmes sont préférables à déployer à grande échelle). Quoi qu'il en soit, même si une statistique est objective, elle n'est pas nécessairement pour autant *désirable* à communiquer, notamment car les statistiques peuvent être extrêmement trompeuses, et finissent par guider un grand nombre de décisions.
:tv: Prebunking : les vaccinés hospitalisés. Science4All (2021).
:books: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Cathy O'Neil. Penguin (2017).

[68] Au moment de juger une décision X, le raisonnement contrafactuel consiste à comparer les conséquences probables de faire X, aux conséquences probables de ne pas faire X (ou de faire une alternative à X). Il s'agit de l'approche souvent privilégiée par la théorie de la décision, surtout en présence d'incertitude.
:tv: Fallait-il confiner ? La pensée contrafactuelle. Science4All (2020).
:books: How to Decide: Simple Tools for Making Better Choices. Annie Duke. Portfolio (2020).

[69] La régulation des plateformes de diffusions de l'information, notamment vis-à-vis des fausses informations, demeure un flou juridique délicat.
:tv: Fake news, tous coupables ? | Angle Droit | Le Vortex (2021).

[70] Le rôle central de ces plateformes, et en particulier des algorithmes de ces plateformes, dans la crise de la mésinformation actuelle est le sujet central d'un livre précédent de Lê Nguyên Hoang, co-écrit avec le chercheur El Mahdi El Mhamdi, et d'articles publiés en sociologie et en philosophie.
:books: Le Fabuleux Chantier : Rendre l'Intelligence Artificielle Robustement Bénéfique. Lê Nguyên Hoang & El Mahdi El Mhamdi. EDP Sciencess (2019).
:memo: Science Communication Desperately Needs More Aligned Recommendation Algorithms. Lê Nguyên Hoang. Frontiers in Communication (2020).
:memo: Recommendation Algorithms, a Neglected Opportunity for Public Health. Lê Nguyên Hoang, Louis Faucon & El-Mahdi El-Mhamdi. Revue Médecine et Philosophie (2021).

[71] Cette remarque fait écho aux *facebook files*, qui révèlent que, à de nombreuses reprises, la direction de Facebook a privilégié les profits et l'image de l'entreprise aux tragédies que sa plateforme aggravait.

> Marc, tu raisonnes trop sur le court terme. Si tu veux que SmartPoop sauve des vies, il faut que SmartPoop continue de vivre. Le jour où SmartPoop est démantelé ou fait faillite, s'il y a une nouvelle pandémie de ROVID-19, ce n'est pas la vie d'une fille d'une cinglée qui va y passer ; ce seront des millions, voire des milliards de vies[72].

Visiblement tous deux frustrés, Katia et Marc marquent un silence dans leur discussion.

> En attendant, je persiste à penser que Facebook et YouTube devraient davantage être poursuivis que nous, ajoute Katia.

Marc reste pensif. Après quelques secondes, il reprend la parole.

> C'est peut-être une défense envisageable. On peut accepter quelques dommages, mais à condition que les poursuites soient aussi engagées à l'encontre de Facebook et YouTube, ou tout autre réseau social utilisé par la mère[73].

L'engagement de SmartPoop

Katia et Marc organisent alors une réunion avec leurs avocats pour préparer un plan d'action selon cette ligne directrice. Ils rajoutent qu'en réponse au procès, et pour éviter d'autres procès futurs, SmartPoop s'engage à être un acteur majeur dans la lutte contre la mésinformation.

> L'objectif étant que les prochaines plaintes soient dirigées davantage vers Facebook et YouTube que vers nous, explique Katia.
>
> Ça ne va pas être évident, prévient toutefois l'avocat. La mère ne semble pas considérer que les informations qu'elle a lues sur le potassium et le magnésium soient de fausses informations. Je ne suis pas sûr qu'elle sera prête à porter plainte contre eux. Mais on peut essayer de trouver d'autres victimes de la mésinformation et déposer

:computer: The facebook files. Wall Street Journal (2021).

[72] Alors que certains appellent au démantèlement des grandes entreprises du numérique, d'autres soulignent le fait que ce démantèlement court le risque de conduire à des effets secondaires imprévus beaucoup plus préoccupants encore, comme l'impossibilité de modérer le flux de l'information, y compris les informations dangereuses et compromettantes, comme les images pédo-pornographiques, les appels à la haine raciale ou le cyber-harcèlement. C'est le cas par exemple de Sinan Aral, professeur du MIT, qui a beaucoup étudié la désinformation sur les réseaux sociaux.

:books: The Hype Machine : How Social Media Disrupts Our Elections, Our Economy, And Our Health–and How We Must Adapt. Sinan Aral. Penguin (2021).

[73] Ce que suggèrent là Katia pourrait être une solution de mieux aligner les intérêts des employés des entreprises pour promouvoir l'éthique. Au lieu de leur demander de souligner les dilemmes éthiques de leurs propres entreprises (ce qui semble néanmoins important et désirable, mais est en pratique affecté par *l'ethics washing*), ces employés pourraient être payés pour souligner les problèmes éthiques avec les produits de leurs concurrents. En pratique, un frein énorme à cela est l'accès aux données des concurrents, sans lesquels l'identification des problèmes éthiques est souvent impossible.

> une plainte collective. Ceci peut convaincre Lucile de poursuivre ces autres entreprises aussi.

Des mois plus tard, le procès a lieu. Le verdict tombe. SmartPoop est reconnu co-auteur de l'homicide involontaire de Jeanne. Cependant, les co-auteurs sont nombreux, puisqu'ils incluent de nombreux réseaux sociaux qui ont propagé d'énormes quantités de mésinformations médicales dangereuses. Une compensation d'un million d'euros est exigée de la part de SmartPoop. De plus, le tribunal demande à SmartPoop de prendre la responsabilité des mésinformations publiées dans l'application, mais aussi des risques de mécompréhension par un public qui n'est pas formé à comprendre les données médicales.

En réponse à la demande du tribunal, Katia et Mark s'engagent alors à embaucher Lucile en tant que consultante, et de travailler avec des médecins, des psychologues et des infographistes, pour améliorer le design de l'application SmartPoop et pour y proposer des bilans de santé simples, sécurisés et aisément compréhensibles[74]. Après des semaines de discussions parfois tendues, et de travail laborieux, un nouveau design de SmartPoop est établi. Science4Alpha invite alors Marc à présenter le nouveau de design de SmartPoop sur sa chaîne YouTube.

> SmartPoop se transforme en une application d'alerte d'anomalie, et de recommandations médicales sécurisées, explique Marc. Plutôt que de noyer l'utilisateur dans un océan de données dans lequel il pourrait se perdre, l'application va maintenant accompagner les utilisateurs. En cas de risques très bien identifiés, comme dans le cas des constipations et des diarrhées, ou pour certaines carences comme les carences en vitamine A qu'on arrive maintenant très bien à détecter, SmartPoop va notifier l'utilisateur[75] et lui fournir une recommandation médicale, conformément à ce qui a été décidé avec l'Ordre des médecins. Typiquement, ce sera une recommandation de la forme « veuillez boire davantage d'eau » ou « pensez à manger des carottes ».
>
> Que se passe-t-il si une anomalie mal identifiée est repérée ?

[74] On peut insister sur ce que cette phrase implique. En pratique, le fait qu'une information soit une mésinformation ou non est une question très complexe, qui nécessite de comprendre la psychologie du public, et en particulier ses erreurs d'interprétation probables. Une entreprise qui souhaite y parvenir se doit alors d'engager des équipes sérieuses capables de répondre à cette tâche délicate. Cependant, on peut noter que les entreprises du numériques investissent souvent déjà beaucoup pour optimiser l'expérience utilisateur, voire l'addiction des utilisateurs à leurs produits.
:tv: How a handful of tech companies control billions of minds every day | Tristan Harris. TED (2017)
:tv: The Social Dilemma. Netflix (2020)

[75] En 2013, la *Food and Drug Admnistration* (FDA) des États-Unis a interdit l'entreprise de séquençage d'ADN 23AndMe de communiquer les résultats de leurs tests génétiques avec leurs patients, faute d'études convaincantes sur la fiabilité de ces tests.
:computer: FDA bans 23andme personal genetic tests. BBC (2013).

Excellente question ! Dans ce cas, SmartPoop recommandera simplement de consulter un médecin[76]. Ce qu'on veut avant tout, c'est ne pas causer de mal. Primum non nocere, dit-on en médecine. Autrement dit, la priorité, c'est de ne pas nuire. SmartPoop implémente désormais ce principe avec beaucoup de rigueur.

Les données brutes de SmartPoop sont-elles encore accessibles ?

Oui. Conformément au Règlement Générale pour la Protection des Données, le fameux RGPD, toutes les données collectées par SmartPoop demeurent accessibles à l'utilisateur. Mais on a décidé de faire très attention à la manière dont elles sont présentées, pour que les utilisateurs se méfient plus, en particulier des erreurs de mesure inévitables de notre dispositif. Nos algorithmes d'apprentissage sont encore en train d'améliorer leurs estimations des propriétés physico-chimiques des excréments à partir de photographies, et nous sommes bien conscients que, sur de nombreuses tâches, la fiabilité de ces algorithmes n'est pas encore au rendez-vous. C'est pour cela que nous faisons très attention à ce que de telles données soient présentées avec beaucoup de barres d'erreurs. Nous préférons que les utilisateurs prennent conscience de l'incertitude de nos modèles avant qu'ils ne lisent les résultats de l'analyse de ces modèles[77].

Marc, tu dis aussi prendre des mesures contre la mésinformation ?

Oui tout à fait. Nous avons aussi travaillé étroitement avec des médecins, des psychologues, des infographistes, mais aussi avec des patients, pour accompagner SmartPoop d'informations médicales de qualité, soit directement dans l'application, soit via des liens vers des sites web de confiance. L'affaire de Lucile nous a montré qu'il s'agit d'un problème de santé public majeur.

---

[76] Sur son site web, l'entreprise 23AndMe insiste désormais beaucoup sur la nécessité de consulter un médecin avec les résultats de leur analyse, et même avant de décider de séquencer son ADN. En octobre 2021, ils écrivaient : « **Ces rapports ne remplacent pas les visites chez un professionnel de la santé**. Consultez un professionnel de la santé pour vous aider à interpréter et à utiliser les résultats génétiques. Les résultats ne doivent **pas** être utilisés pour prendre des décisions médicales. » Et rajoutaient : « Nous vous encourageons à parler à un conseiller en génétique. »
:computer: 23andMe Genetic Health Risk Reports: What you should know.

[77] De façon plus générale, la recherche en sécurité des algorithmes insiste beaucoup sur l'importance de permettre aux algorithmes de mesurer leur propre incertitude sur les résultats qu'ils calculent.
:memo: Benchmarking Uncertainty Estimation Methods for Deep Learning With Safety-Related Metrics. Maximilian Henne, Adrian Schwaiger, Karsten Roscher, Gereon Weiss. SafeAI@AAAI (2020)
:books: Human Compatible: Artificial Intelligence and the Problem of Control. Stuart Russell. Penguin (2019)
Cette remarque semble aussi primordiale pour les humains, y compris les experts qui pêchent souvent par excès de confiance.
:tv: L'excès de confiance tue. Science4All (2020).

Mais qu'est-ce que la mésinformation ? Est-ce que dire que les bananes ne sont pas naturelles et qu'elles sont radioactives, est une mésinformation ?

Ce qu'on a découvert avec l'histoire de Lucile, c'est que la notion de mésinformation est en fait très délicate[78]. D'un côté, non, ce n'est pas tout à fait une fausse information, ou une fake news comme diraient certains. Mais d'un autre côté, dite dans un mauvais contexte, cette phrase peut être trompeuse pour certains publics, si ceux-ci sont ensuite amenés à croire que les bananes sont du coup dangereuses, ou qu'il vaut mieux s'en éloigner[79]. Pour être clair, oui, les bananes ont été modifiées par l'agriculture. Cela fait 7000 ans que l'on sélectionne les variétés de bananes qui nous arrangent, ce n'est pas naturel du tout. Mais c'est le cas de quasiment toute la nourriture qu'on mange, du blé au bœuf, en passant par les pommes, les oranges et oui, les bananes aussi. Et oui aussi, la banane est radioactive. Mais la radioactivité d'une banane est deux cents fois moindre que ce que vous subissez si vous prenez l'avion pour un trajet de 6 heures, juste parce que l'avion vole un peu haut dans le ciel. Si vous avez peur de la radioactivité de la banane, vous devriez être terrifié par l'avion. D'ailleurs il existe une unité, la DEB, pour « dose équivalente en banane », qui sert de façon didactique à présenter certaines exposition très faible à de la radioactivité[80].

Malheureusement, en tant que YouTubeur éducatif, je peux te certifier qu'expliquer tout cela prend du temps[81]...

Oui et l'attention du public est très limitée. C'est pourquoi, plutôt que d'inonder les utilisateurs de SmartPoop d'informations complexes, nous avons opté pour leur communiquer uniquement des informations très simples et très fiables, tout en ajoutant des liens vers des informations plus complètes.

On en vient à la grande annonce de cette vidéo. Marc, tu es maintenant mon boss.

Oui en effet. Science4Alpha, tu es une pédagogue exceptionnelle, qui nous a supporté très tôt dans notre démarche, et qui est toi-même très préoccupée par les enjeux de santé publique. Nous avons

[78]:books: The Reality Game How the Next Wave of Technology Will Break the Truth. Samuel Woolley. PublicAffairs (2020).

[79]:tv: Les Fake News, un Fake Problem ? Science4All (2021).

[80]:tv: The Most Radioactive Places on Earth. Veritasium (2014)

[81]On parle parfois de la « loi de Brandolini », ou de le principe d'asymétrie des idioties, selon quoi le coût de déconstruire une croyance erronée est beaucoup plus grand que celui de la défendre. Voilà qui rend la rectification de croyances erronées extrêmement ardue, lorsqu'elle est confrontée au « mille-feuille argumentaire », qui correspond à enchaîner des arguments bancals défendant une croyance.
:tv: Comment réagir au bullshit ? Philoxime

décidé de soutenir officiellement ton travail en te garantissant un revenu stable. Et nous avons conclu des accords similaires avec neuf autres YouTubeurs scientifiques. Nous sommes très excités par ces partenariats car nous pensons que l'information de qualité est une priorité pour la santé publique[82].

Mais, Marc, comme certains sont d'ailleurs probablement en train de l'écrire dans les commentaires, n'y a-t-il pas des risques de conflits d'intérêt ?

Toujours. Nous avons fait de notre mieux pour trouver un système qui te permette de nous surveiller de près et de vivre confortablement, tout en te garantissant la liberté qu'exige le travail de communication scientifique, en particulier sur des sujets sensibles comme la santé publique. En particulier, nous nous sommes engagés à garantir un an de revenu en cas de rupture de contrat avec toi ou tes collègues[83], et nous n'avons aucun droit de regard sur les contenus que vous publiez[84].

Vous entendez cela, chers vieweurs ? Promis, je resterai libre dans ce que je dis. Si SmartPoop chie son produit ou son interface, je serai le premier à le signaler.

J'espère bien, oui. Notre relation avec toi va beaucoup s'appuyer sur une relation de confiance. Nous avons confiance en toi, et en

[82]La pertinence de telles collaborations, avec le privé ou avec des institutions gouvernementales, est un dilemme permanent pour la communication scientifique, surtout sachant le peu de financement qu'elle reçoit pour l'instant, et la difficulté d'accéder aux informations internes à des grandes entreprises. On peut citer l'exemple de cette collaboration entre le YouTubeur scientifique SmarterEveryDay et 23AndMe, qui est un cadre de collaboration jugé satisfaisant par le YouTubeur scientifique.
:tv: DNA Testing and Privacy (Behind the scenes at the 23andMe Lab) - Smarter Every Day (2017).
En particulier, sur des sujets controversés, de telles collaborations peuvent avoir des effets contreproductifs sur la confiance envers les communicateurs scientifiques.
:memo: Trust in scientists in times of pandemic: Panel evidence from 12 countries. Yann Algan, Daniel Cohen, Eva Davoine, Martial Foucault & Stefanie Stantcheva. PNAS (2021)
Lê Nguyên Hoang a par exemple produit une vidéo en collaboration avec le Ministère de la Santé en France, qui aura reçu plus de dislikes que de likes, questionnant ainsi la pertinence d'une telle collaboration.
:tv: Un vaccin pour permettre aux étudiants de retrouver leur vie d'avant (ft. Prof. Fischer). Science4All (2021).

[83]Les employés de l'équipe d'éthique des IA de Google n'a pas eu ce luxe. En particulier, Timnit Gebru a été licencié pendant ses vacances, sans préavis, peu de temps après celle-ci a critiqué dans un article de recherche les algorithmes de langage dans lesquels Google avait beaucoup investi.
:tv: L'éthique des algorithmes en sérieux danger. Science4All (2020).

[84]Notez que cela n'est pas le cas des chercheurs en éthiques des IA de Google, dont les publications sont soumis à une approbation en interne par des managers haut placés de Google. Comme on l'a vu suite aux licenciements de Timnit Gebru et Margarett Mitchell, ce manque d'indépendance entre l'éthique et la direction de Google est un sérieux risque pour l'intégrité de la recherche de Google.
:tv: Google démantèle son éthique (et tout le monde s'en fout...). Science4All (2021).

ta volonté de prioriser la santé publique dans tes vidéos. Et nous espérons que toi, Science4Alpha, ainsi que le grand public et nos différents partenaires, feront confiance en notre engagement pour la santé publique avant tout[85].

On se fait un petit selfie avec un check pour finir cette vidéo ? Et oui on évite de se serrer les mains, parce qu'on n'a pas encore la garantie que le ROVID-19 est complètement parti. Voilà, merci Marc, et à vous vieweurs, j'espère que vous êtes excité comme moi à l'idée d'avoir une vue de l'intérieur de SmartPoop. Je ne pourrai bien sûr pas tout vous dire, car certains sujets sont sensibles, et il y a même des histoires de délits d'initiés que je préfère éviter[86]. Mais promis, on ne sera pas tendre avec Marc et Katia, surtout s'ils dérapent.

On essaiera de ne pas trop déraper alors !

Pour ne pas manquer les futures vidéos de santé publique et de défis algorithmiques, pensez à vous abonner, à mettre la cloche et n'hésitez pas à partager cette vidéo. Et je vous dis à très bientôt, sur Science4Alpha.

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

[85] L'une des grandes difficultés d'un tel partenariat, notamment pour les entreprises ou les gouvernements, c'est la confiance en le communicateur scientifique. Celui-ci pourrait être payé par des concurrents pour produire une désinformations nocives. Malheureusement, mettre en place de tels partenariats, c'est compliqué...

[86] La transparence, c'est un sujet compliqué aussi...

Chapitre 3 — Le biais étroniste

> Le vrai confort, ce n'est pas seulement de savoir que vous pouvez faire vos besoins en toute sécurité ; c'est surtout de savoir que votre santé sera prise en charge à tout instant grâce à l'intelligence de vos toilettes. SmartToilets[87] de SmartPoop. Laissez-nous prendre soin de vous.

C'est la cinquième fois déjà aujourd'hui qu'Issa Gueye voit passer cette publicité sur son téléphone. Ironiquement, cette fois-ci, c'est assis sur la cuvette[88], entre deux vidéos d'humoristes, qu'il y est exposé. Quelques minutes plus tard, alors qu'il ouvre justement l'application SmartPoop pour filmer ses excréments, il se rend alors compte que l'installation des SmartToilets lui éviterait de faire lui-même ce travail de tournage des étrons. Même si cela fait maintenant trois ans qu'il l'effectue presque quotidiennement, il persiste à trouver cet effort laborieux et repoussant.

Issa, ce grand sportif qui aime suivre son état de forme, clique alors sur l'offre spéciale proposée par son application. En tant qu'utilisateur régulier depuis les débuts de SmartPoop, il a en effet une réduction de 30% sur l'achat d'un SmartToilet de base, pour un prix total de 12 000 euros. Cela coûte cher, mais ce trader en salles de marché peut se le permettre. En fait, Issa découvre alors l'existence d'une version SmartToilet Deluxe. Cette version permet ainsi un pré-chauffage de la cuvette, une filtration des odeurs par filtre céramique alvéolé, un nettoyage par jet d'eau fin et un séchage par air chaud, le tout optimisé

[87] Les SmartToilets sont déjà en développement ! Ils ont même conduit à une publication académique, avec un prototype capable de reconnaissance... rectale !
:memo: A mountable toilet system for personalized health monitoring via the analysis of excreta. Seung-min Park et al. Nature Biomedical Engineering (2020).

[88] L'invention des toilettes, et en particulier des circuits d'eaux usées, a en fait été l'une grandes avancées de l'histoire des civilisation, car elle a permis de combattre efficacement la diffusion de nombreuses maladies comme le choléra. D'ailleurs, plus de 2 milliards d'humains n'y ont toujours pas accès. On estime que, à cause de cela, environ 800 000 enfants meurent chaque année de diarrhées. :tv: How The Toilet Changed History. It's Okay to Be Smart (2017).

par des algorithmes d'intelligence artificielle pour maximiser le bien-être de l'utilisateur[89]. Intrigué, Issa finit par opter pour ce produit, malgré son prix exorbitant de 25 000 euros.

Poo

Quelques jours plus tard, les SmartToilet Deluxe de SmartPoop sont livrées et installées chez Issa. Le soir même, c'est avec un grand sourire, et quelques hésitations, qu'Issa va alors tester ses nouvelles toilettes. Alors qu'il se rapproche de celles-ci, le couvercle s'ouvre automatiquement. Issa s'assoit ensuite sur la cuvette. Une étrange sensation de chaleur accompagne alors son contact avec la cuvette. « C'est comme si j'utilisais les toilettes juste après quelqu'un d'autre... Pas très rassurant. Mais je dois sans doute juste m'y habituer », se dit-il.

> Bonjour Issa, dit alors une voix masculine.

Issa sursaute. Les toilettes viennent de lui parler !

> Détendez-vous et profitez du confort des SmartToilet Deluxe, explique-t-elle. Je suis Poo, votre assistant vocale. Est-ce que vous préféreriez que j'adopte une voix féminine[90] ?

Issa est très perturbé. Après quelques secondes, il répond : « Oui, je préfère une voix féminine. »

> N'hésitez pas à me dire ce que vous ressentez pour que j'optimise votre confort, dit Poo avec désormais une voix féminine[91].

Perturbé, Issa reste silencieux pendant quelques autres secondes. Il finit par répondre : « On peut peut-être commencer par se tutoyer ».

> Avec plaisir, Issa. Je suis à ton service.

Le téléphone d'Issa vibre tout à coup, Issa prend son téléphone, qui lui annonce alors « Félicitations ! Vous utilisez désormais nos SmartToilet Deluxe. Nous

[89] Au Japon, les toilettes ont souvent plusieurs de ces fonctionnalités.
:tv: Why You Need to Try a High-Tech Japanese Toilet. Lifehacker (2019).

[90] Suite à de nombreuses critiques sur le renforcement du biais de genre, Apple a décidé de donner à Siri une voix masculine par défaut.
:computer: How AI bots and voice assistants reinforce gender bias. Caitlin Chin and Mishaela Robison. Brookings (2020).
:computer: Apple's Siri is no longer a woman by default, but is this really a win for feminism?. Eleonore Fournier-Tombs. The Conversation (2021).

[91] Cet exemple souligne la tension entre le contrôle des utilisateurs sur les technologies qu'ils utilisent et les conséquences que ce contrôle peut avoir à l'échelle sociale. Ainsi, en permettant à l'utilisateur de customiser ses technologies, il y a un risque que celui-ci le fasse en aggravant, consciemment ou non, des biais de genre (dans ce cas), la diffusion des appels à la haine ou encore celle de la désinformation. Cette tension réside finalement dans le principe d'absence de tort de John Stuart Mill, ou dit plus simplement dans le principe « les libertés des uns s'arrêtent là où commencent celles des autres ».
:tv: Ce principe sur lequel tout le monde s'entend. Monsieur Phi (2021).

vous souhaitons une agréable expérience ». Encore tendu, Issa a alors du mal à se relâcher.

> Tu me sembles tendu. C'est normal. La première fois, ce n'est jamais simple. Mais détends-toi. Essaie de profiter de ton nouveau confort quotidien.
>
> Je n'ai pas l'habitude de parler avec un assistant vocal dans les chiottes !
>
> Je comprends. Moi non plus, je n'ai pas encore l'habitude de parler aux utilisateurs dans les chiottes !

Petit à petit, toutefois, Issa perçoit de plus en plus ses SmartToilets comme un nouveau confort important[92]. Après tout, mis bout à bout, les humains passent un temps faramineux assis sur des toilettes. Autant en faire un bon moment !

Enfin, Issa avance dans sa tâche.

> Euh... dit Issa, avec une certaine hésitation. Est-ce que tu... l'analyses ?
>
> Oui Issa, je suis en train de filmer tes « données » sous différents angles, et de mettre les vidéos sur ton compte en ligne. J'effectue aussi d'autres prélèvements pour une analyse plus fine. Est-ce que tu veux en savoir plus ?
>
> Oui, dis-en moi plus.
>
> Avec plaisir. L'eau dans laquelle baignent les excréments est aussi constamment prélevée et analysée en spectroscopie. Autrement dit, on étudie la manière dont cette eau absorbe différentes couleurs du spectre lumineux, de manière très précise. Ceci nous permet d'identifier certaines molécules à l'intérieur de tes excréments, ce qui va nous permettre un bilan de santé beaucoup plus précis[93]. On effectue aussi de la spectrométrie de masse[94], et des mesures chimiques, par exemple de l'acidité de cette eau[95].

[92] Des milliards d'humains sur terre n'ont pas accès à ce confort. Comme on l'a vu, ceci pose de sérieux problèmes sanitaires.
:tv: 6 Toilets From History, and What They Taught Us. SciShow (2021). Mais la recherche est en train de développer des solutions pour rendre ce confort accessible à plus de monde, en innovant notamment dans le traitement des excréments.
:tv: 3 Groundbreaking New Toilets. SciShow (2018).

[93] :tv: Spectrophotometry and Beer's Law. Professor Dave Explains (2019).

[94] La spectrométrie de masse consiste à découper une molécule à analyser en sous-molécules chargées, par exemple à l'aide d'une ionisation par électrons, et à mesurer la capacité d'un champ magnétique à dévier la trajectoire de sous-molécules. Puisque les sous-molécules plus lourdes sont plus dures à dévier, ceci nous renseigne sur la masse de ces sous-molécules, et donc sur la masse et la composition de la molécule à analyser.
:tv: IR Spectroscopy and Mass Spectrometry: Crash Course Organic Chemistry #5 (2020).

[95] L'acidité d'un liquide est mesurée par son pH, via par exemple des mesures électriques.
:tv: How a pH meter works! pH Professor (2017).

> Quand est-ce que j'aurais des résultats ?
>
> Je te tiendrai au courant, aux toilettes ou dans ton application, si jamais il y a des mesures préoccupantes.

« Très préoccupant »

Quelques jours plus tard, alors qu'il vient de rentrer chez lui après une dure journée de travail. Issa reçoit une notification de SmartPoop. Issa demande alors : « Poo, que se passe-t-il ? »

> J'ai fait des analyses très préoccupantes. Issa, je pense que tu devrais aller en urgence à l'hôpital.

Terrifié, Issa ressort de chez lui, prend la voiture et va directement à l'hôpital. Arrivé aux urgences, il demande alors de consulter un médecin. « Pour quelles raisons ? », lui demande-t-on. « J'ai reçu une alerte de SmartPoop, qui m'a juste dit de venir aux urgence », explique Issa.

Après une minute ou deux, Issa est pris en charge.

> Bonjour Issa, je suis Dr. Paola Marta. Sachez que nous avons l'habitude de gérer des urgences SmartPoop. Depuis deux ans, la plupart de nos patients sont venus suite à une alerte SmartPoop, et le rapport qu'ils nous envoient nous aide toujours beaucoup à soigner nos patients[96]. Comment vous sentez-vous ?
>
> Pour l'instant, je n'ai que des maux gastriques. Mais je suis assez terrifié. SmartPoop dit que mon état de santé est « très préoccupant ».
>
> Très préoccupant, vous avez dit ?
>
> Oui. Docteure, est-ce que vous savez ce que j'ai et si ça va aller ?

Dr. Marta fuit alors Issa du regard. Clairement gênée, Dr. Marta paraît terrifiée par le diagnostic de SmartPoop. Les 27 patients qu'elle a eus, et dont l'état de

[96] D'après une étude récente, la plupart des gens semblent faire en effet davantage confiance aux jugements d'un algorithme que d'un humain, au moins pour certaines prédictions comme le succès de Tesla, des sanctions européennes contre la cyber-guerre ou le futur du Brexit. De façon toutefois perturbante, les professionnels expérimentés, eux, font moins confiance aux algorithmes, ce qui les conduit à de moins bons jugements que les non-professionnels !
:memo: Algorithm appreciation: People prefer algorithmic to human judgment. Jennifer Logg, Julia Minsona & Don Moore. Organizational Behavior and Human Decision Processes (2019).
Cependant, la recherche en psychologie suggère également que ce qui nous est familier nous paraît plus crédible, et ce de manière parfois troublante. On parle de *biais de familiarité*. Par exemple, un sujet exposé à de multiples reprises à la phrase « la température corporelle d'une poule » a une plus grande probabilité de juger la phrase « la température corporelle d'une poule est de 34°C » vraie. Le fait que Dr. Paola Marta interagit quotidiennement avec des rapports SmartPoop, qui en plus sont efficaces pour la guider dans le traitement de ses patients, pourrait alors expliquer pourquoi la docteur Paola Marta fait confiance à SmartPoop.
:tv: The Illusion of Truth. Veritasium (2016).

santé avait été jugé « très préoccupant » par SmartPoop, sont tous décédés[97]. Après un silence interminable, Dr. Marta finit par répondre, avec visiblement un effort pour paraître aussi rassurante que possible, malgré les circonstances.

Je vous promets qu'on fera de notre mieux.

Après une prise de sang, Issa se retrouve enfin allongé dans un lit d'hôpital, seul, abandonné à sa propre imagination. Il craint le pire, et pense à tout ce qu'il aurait aimé avoir fait. Issa nourrit des regrets, et envisage sa postérité. Toute sa vie, Issa n'a fait que suivre l'appât du gain. On lui a dit que réussir en finance, c'est rouler sur l'or et triompher dans la vie. C'est ce qu'il a fait. Cela fait-il de lui une mauvaise personne[98] ?

Certes, sa carrière lui a au moins permis de s'offrir toutes sortes de matériel de luxe, comme les SmartToilets Deluxes. Preuve ultime de la réussite sociale. Mais Issa aura-t-il vraiment accompli quelque chose dans sa vie ? Qu'en est-il de toutes ces nombreuses autres personnes qui ont souffert à cause de lui ? De son ex-femme qu'il a délaissée ? De ses enfants avec qui il a passé si peu de temps[99] ?

Dr. Marta revient alors vers Issa. « Enfin », s'exclame intérieurement Issa, qui a l'impression d'avoir été abandonné depuis plusieurs heures. Pourtant, cela ne fait que dix minutes que Dr. Marta s'était absentée[100]. Elle demande alors à Issa de l'accompagner pour passer une radiologie, et lui annonce qu'une coloscopie est aussi envisagée. En l'espace de quelques heures, Issa subira en fait toutes sortes

[97] La loi de succession de Laplace suggère que, si on avait a priori une incertitude initiale totale sur le fait qu'un diagnostic « très préoccupant » de SmartPoop conduisait à un danger de mort, alors, sachant que les 27 premiers patients avec ce diagnostic sont tous morts, la probabilité qu'Issa meurt à son tour serait alors de 28/29. Voilà qui justifie pleinement la peur de Dr. Marta.
:tv: Binomial distributions | Probabilities of probabilities, part 1. 3Blue1Brown (2020).

[98] La psychologie empirique suggère que s'auto-qualifier avec des adjectifs peut donner une impression de permanence de certaines de nos propriétés, qui peut ensuite être très néfaste à notre développement personnel. On parle de *fixed mindset* (mode d'esprit fixe). Par opposition, les personnes qui pensent pouvoir progresser (*growth mindset*, mode d'esprit de croissance) semblent beaucoup plus épanouie. Dès lors, plutôt que de se qualifier de « mauvaise personne », il semblerait plus judicieux de se qualifier de « personne ayant potentiellement fait certaines mauvaises choses », voire de « personne qui aspire à faire de meilleures actions ».
:tv: The power of believing that you can improve | Carol Dweck. TED (2014).
:books: Mindset: The New Psychology of Success. carol Dweck. Penguin (2007).

[99] Ces questions renvoient à ce que certains appellent le *goal factoring* (reconstruction des objectifs) ou le *self alignment* (auto-alignement), qui consistent à se demander si les buts que nous nous sommes fixés sont vraiment les buts que nous devrions nous fixer, voire que nous voudrions vraiment nous fixer.
:computer: Goal Factoring. LessWrong (2018).
Il semble en effet souvent être le cas que ces buts sont des *croyances orphelines*, c'est-à-dire des buts qu'on s'est fixé suite à des motivations fondamentales, et qu'on persiste à se fixer alors même que ces motivations fondamentales ont disparu. Typiquement, on pouvait vouloir plaire à ses parents, à une époque où la fierté de nos parents nous importait énormément ; mais en y réfléchissant, on a fini par se rendre compte que d'autres motivations fondamentales prévalaient sur cela.
:tv: La croyance orpheline d'Einstein. Science4All (2019).

[100] :tv: How Your Brain Makes Time Pass Fast or Slow. It's Okay To Be Smart (2020).

d'analyses. En fin de journée, Dr. Marta lui dit qu'il restera en observation pendant la nuit. Elle l'encourage à essayer de manger le repas qui lui est servi, et à essayer de trouver le sommeil pour se reposer.

Le lendemain matin, Dr. Marta arrive enfin dans la chambre d'Issa. Issa est épuisé. Terrifié, il n'a pas dormi de la nuit.

> Quelles sont les nouvelles, docteure ?
>
> Vous avez clairement une fatigue très avancée et un haut niveau de stress, avec visiblement des troubles digestifs et des troubles du sommeil. Quelque chose ne va pas. Cependant, toutes nos mesures ne parviennent toujours pas à identifier l'origine de vos troubles. Nous avons appelé l'entreprise SmartPoop, dont le rapport de santé n'est pas vraiment en phase avec notre examen, et ils sont en train d'analyser manuellement vos données. Nous espérons qu'ensemble nous arriverons à cerner le problème. Mais ceci va nous prendre du temps. J'en suis désolée. Vraiment désolée.

Alerte chez SmartPoop

En fin de matinée, c'est au tour de Marc d'entrer dans le bureau de Katia.

> On a un problème, annonce-t-il.
>
> Que se passe-t-il, Marc ?
>
> Mon assistante vient de me dire qu'une médecin, Dr. Marta, n'a cessé d'essayer de nous joindre hier soir et ce matin. Elle a un patient à qui SmartPoop a dit d'aller aux urgences. Ils ont lancé plein de tests sur le patient. Mais ils n'arrivent pas à identifier la cause du problème.
>
> T'as lancé l'équipe de diagnostics personnalisés sur le problème ?
>
> Oui. Ils bossent depuis 9h ce matin, et ne trouvent toujours pas le problème. Ils m'ont demandé de te demander de jeter un œil aux données.
>
> T'es sûr que c'est à moi de regarder ? Je suis très occupée. Je dois préparer SmartPoopCon 2024 en fin de semaine.
>
> Ils ont une idée du problème, mais ils veulent ton expertise pour confirmer leur intuition...
>
> Tu peux m'envoyer les références du cas ?
>
> C'est déjà fait.

Katia ouvre alors son client email, qui lui dit qu'elle a 25 251 messages non-lus. Elle trie ses messages par destinataire et trouve celui de Marc[101]. Katia

[101] Le philosophe Michel Serres aimait insister sur l'impact des technologies de l'information,

copie-colle les références du cas, et lance des requêtes à la base de données de SmartPoop. Elle récupère alors les données d'Issa, et analyse les statistiques de ses excréments. Katia exécute quelques commandes, qui génèrent alors toutes sortes de graphes. Après avoir vu une quinzaine de graphes, Katia s'exclame : « Oh non ! Issa est hors distribution[102] ».

> Qu'est-ce que tu veux dire ?
>
> Les données vidéos d'Issa sont parfaitement normales. Ce sont ses données spectrographiques qui sont signalées « très préoccupantes ». Mais je crains que ce soit parce qu'Issa est un utilisateur de SmartToilet Deluxe statistiquement très différent des autres utilisateurs de SmartToilet Deluxe.
>
> Il est sénégalais.
>
> Oh non... Je crois qu'on a un gros problème. Tu as le numéro de la médecin ? Il faut l'appeler tout de suite.

L'effet nocebo

Marc compose le numéro de Dr. Marta.

> Bonjour Dr. Marta, ici Marc. Vous êtes en ce moment en speakerphone avec Katia et moi, Présidente et Vice-Président de SmartPoop. On a du nouveau sur le cas d'Issa.
>
> Bonjour Dr. Marta, ici Katia. J'ai analysé le cas d'Issa, et je suis à peu près sûre qu'il s'agit d'un faux positif. Autrement dit, SmartPoop s'est trompé en signalant un problème avec les données d'Issa.

---

comme le papier, l'imprimerie ou les ordinateurs, sur *l'externalisation* de notre cognition. Il est ainsi remarquable de constater à quel point nos boîtes emails sont parvenues à externaliser une grosse partie de notre mémoire. En un sens, nos boîtes emails nous connaissent beaucoup mieux que nous nous connaissons nous-mêmes, non pas car elles sont « intelligentes », mais simplement car leur mémoire est beaucoup plus fiable que la mémoire humaine, et car rechercher dans ces boîtes emails est beaucoup plus efficace que rechercher dans ntore mémoire.
:tv: Michel Serres - Les nouvelles technologies : révolution culturelle et cognitive. I Moved to Diaspora (2012).

[102] Les données dites « hors distribution » sont des données très distinctes de l'ensemble des autres données. Elles sont souvent considérées erronées, voire adversariales, si bien que beaucoup d'algorithmes d'apprentissage cherchent à les éliminer.
:memo: Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui & Julien Stainer. NIPS (2017)
:memo: A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. Kimin Lee, Kibok Lee, Honglak Lee & Jinwoo Shin. NeurIPS (2018).
Les algorithmes qui prévilégient la sécurité sont alors conduits à ignorer les données hors distribution, qui ignorent donc minorités. En fait, comme en parle l'article suivant, il existe une tension fondamentale entre inclusion et sécurité. Pour résoudre cette tension, il est critique de mieux comprendre la distribution, et de mieux sécuriser et authemtifier les sources des données, ou d'être beaucoup plus modeste dans la conception des algorithmes.
:memo: Collaborative learning in the jungle. El-Mahdi El-Mhamdi, Sadegh Farhadkhani, Rachid Guerraoui, Arsany Guirguis, Lê Nguyên Hoang & Sébastien Rouault. NeurIPS (2021).

Bonjour Marc et Katia, répond Dr. Marta. Vous êtes sûrs de ce que vous dites ? Issa a des symptômes préoccupants, notamment au niveau digestif et en termes de fatigue.

C'est étrange. Les données que j'ai ne me semblent pas suggérer des symptômes préoccupants, explique Katia.

C'est peut-être un effet nocebo, intervient alors Marc.

Un effet nocebo ? Oui ça pourrait être ça, en effet, s'exclame Dr. Marta.

Un effet nocebo, demande Katia, qu'est-ce que c'est ?

Lorsqu'un patient pense qu'il lui arrive quelque chose d'horrible, il peut arriver qu'il développe les symptômes qu'il craint, explique Marc. Des symptômes digestifs par exemple, c'est tout à fait probable. Issa a sans doute une confiance telle en SmartPoop que, lorsque SmartPoop lui a dit que son état de santé était très préoccupant, son état de santé est devenu très préoccupant[103].

Mais expliquez-moi, demande Dr. Marta, pourquoi y a-t-il eu ce cas de faux positif ? Jusque là, à chaque fois que SmartPoop a déclaré un cas « très préoccupant », dans notre hôpital, ça avait toujours terminé en unité de soin intensif, puis en décès. C'est pour ça que j'ai moi-même était paniquée pour Issa — ce qui a sans doute malheureusement contribué à son nocebo.

On vient de déployer les SmartToilets et les SmartToilets Deluxe, explique Katia. Je pense que les SmartToilets ont des diagnostics relativement fiables. Mais la version Deluxe a des capteurs encore plus avancés, qui n'ont pas été aussi utilisés que ceux des SmartToilets de base. Et c'est pour cette raison qu'elle n'est pas aussi fiable.

Je comprends, merci, réponds Dr. Marta. Je vais en informer Issa. Avec un peu de chance, ça aidera à soigner son nocebo. Est-ce que je peux vous demander de manuellement corriger son SmartPoop ?

C'est... fait, dit Katia, après avoir entré quelques commandes sur son ordinateur.

Merci beaucoup ! Bonne journée, conclut Dr. Marta, avant de raccrocher.

163 faux positifs

Katia raccroche deux fois, pour vérifier que Dr. Marta n'est plus connectée. Katia tape alors plusieurs nouvelles commandes, qui permettent d'afficher trois nouveaux graphes. Elle se retourne vers Marc.

[103]:tv: This Video Will Hurt. CGP Grey (2013).

Bien joué Katia. Toujours là pour nous secourir, s'exclame Marc.

Marc, je pense que tu ne réalises pas encore à quel point on est dans la merde.

Je commence à ne plus supporter cette expression.. Le problème n'est pas résolu ?

Je vais faire les mises à jour cet après-midi pour éviter d'autres alertes de faux positifs. Mais le mal est déjà fait. SmartPoop a lancé 163 alertes de faux positifs pour des utilisateurs de SmartToilet Deluxe.

Ça nous fait 163 appels téléphoniques à effectuer. Ce n'est pas la mer à boire !

Marc, ce n'est pas ça le coeur du problème.

Quel est le problème ?

Le problème, c'est que les 163 faux positifs sont tous d'origine africaine ; et que ce n'est pas un hasard.

Que se passe-t-il ?

SmartPoop est... raciste[104], affirme Katia avec ton grave.

Raciste ? Que racontes-tu ? C'est un algorithme, pas un humain ?

Ce que je veux dire, c'est qu'il est dangereux pour les noirs africains. Et du coup, on va subir une énorme shitstorm dans les semaines à venir, pour avoir déployé une technologie raciste avant de la tester suffisamment[105].

Attends. Est-ce que tu suspectes un de nos développeurs d'être raciste et d'avoir fait faire cela à notre algorithme ?

Non. Ce n'est pas cela. Aujourd'hui, avec le machine learning, les algorithmes apprennent beaucoup plus des données que des développeurs. Donc leurs performances dépendent de la qualité et de la quantité des données à disposition[106]. Sauf que les données

[104]:tv: L'IA est raciste (mais vous aussi !). Science4All (2018).

[105]Ce fut le cas des technologies de reconnaissance faciale, déployées de manière précipitée, avant que des audits externes ne découvrent que ces technologies ont un taux d'erreur inacceptable pour les minorités, ce qui pouvait parfois les empêcher d'accéder à leurs propres bâtiments, lorsque cette entrée était permise par de tels algorithmes. Ces technologies ont fini par être bannies par le Sénat américain, ce qui aura conduit à une rétraction des produits développés par IBM, Amazon et Microsoft.
:tv: Coded Bias. Netflix (2020)
Un article récent affirme que des mesures similaires doivent être urgemment prises pour les algorithmes de traitement de langage, dont les vulnérabilités mal comprises et sous-auditées risquent fortement de conduire à des catastrophes.
Citation à venir

[106]L'exemple qui montre cela mieux que tout autre est sans doute l'histoire de Tay et Xiaoice. Ces deux algorithmes conversationnels de Microsoft ont été conçus selon les mêmes principes.

dont on dispose via les SmartToilet Deluxe, ce sont uniquement des données des utilisateurs des SmartToilet Deluxe...

Or ces utilisateurs sont quasiment exclusivement des blancs et des asiatiques, complète Marc.

Oui... Ce produit est un produit de luxe. Il est donc acheté uniquement par des riches, qui se trouvent être majoritairement des blancs et des asiatiques.

Et donc, faute de données sur les noirs africains, SmartPoop est devenu mauvais pour diagnostiquer leurs excréments.

C'est pire que ça. En médecine, quand quelque chose est inhabituel, ou anormal[107]...

On le considère préoccupant. D'où l'alerte...

Devine combien il y a d'utilisateurs de SmartToilet Deluxe d'origine africaine.

Non... 163 ? Non...

On vient de classer tous les utilisateurs de SmartToilet Deluxe d'origine africaine dans la catégorie « malade ». On a automatisé le racisme !

Mais on ne l'a pas fait intentionnellement[108].

Va expliquer ça aux médias ! Je crains le pire. On va avoir des titres sensationnalistes du genre « SmartPoop est raciste », et on risque probablement une poursuite en justice. On risque de perdre

---

Cependant, Tay fut lancée sur Twitter, où elle fut déraillée par les données de trolls qui l'encourageait à exprimer des propos sexistes et racistes. Tay alla même jusqu'à appeler au génocides de certaines populations. Cependant, de façon moins connue en Europe et en Amérique du Nord, Xiaoice, elle, fut lancée 2 ans auparavant sur les réseaux sociaux chinois (WeChat en particulier), et est devenue adorable. À tel point que Xiaoice est aujourd'hui utilisée par 600 millions de chinois, avec des histoires d'hommes séduits romantiquement par Xiaoice. Bref, Tay est devenue horrible ; Xiaoice est devenue adorable. Pourquoi ? Eh bien, à cause des données avec lesquelles ces algorithmes étaient entraînés.
:tv: Les données manipulent les algorithmes. Science4All (2021).

[107] Les notions de « normal » et de « anormal » ont longtemps forgé le discours médical. Par exemple, l'homosexualité a longtemps été considéré anormale, ce qui a conduit l'Association Américaine de Psychologie à parler de « trouble mental ». En 1975, l'Association est revenu sur ce jugement, et ne considère plus qu'il s'agit d'un trouble mental.
:computer: Sexual Orientation & Homosexuality. The American Psychological Association (2021).

[108] Beaucoup des problèmes éthiques des algorithmes, comme l'amplification des appels à la haine ou de la mésinformation, semblent davantage à voir avec les effets secondaires imprévus (et donc non-intentionnels) des concepteurs de ces algorithmes. Dès lors, pour rendre ces algorithmes éthiques, il ne suffit pas de vouloir qu'ils soient « neutres », ni même de vouloir qu'ils soient « raisonnables » ; il est critique de chercher à activement anticiper leurs effets secondaires difficilement prévisibles, et d'investir massivement dans l'étude de ces effets secondaires.
:books: Le Fabuleux Chantier : Rendre l'Intelligence Artificielle Robustement Bénéfiques. Lê Nguyên Hoang & El Mahdi El Mhamdi. EDP Sciences (2019).

> nos investisseurs ! Et si on n'a pas plus d'investisseurs et un procès au cul, on risque de faire faillite. Et tout ça à quelques jours de SmartPoopCon... Il faut qu'on taise cette histoire.

Après un silence, Marc rajoute : « Je pense que c'est peine perdue. Les 163 victimes ont probablement presque toutes subi des souffrances dues à l'effet nocebo, dont on est responsables. »

> Comment ça « dont on est responsables » ? Tout ce qu'on a fait, c'est lancer des alertes. SmartPoop n'est qu'une aide à la décision.
>
> Le problème, c'est que nos utilisateurs font tellement confiance à SmartPoop, que nos aides à la décision sont devenues des décisions pour les utilisateurs. Tu l'as vu avec le cas d'Issa. Quand on lui a dit d'aller aux urgences, tu penses qu'il a pris ça pour une aide à la décision ? Non, il s'est dit, oh mince, il faut que j'aille aux urgences[109]. Et pire que cela, il a développé des symptômes !
>
> OK. Mais ce n'est pas comme si on avait torturé Issa !
>
> J'ai vu des expériences de nocebo terrifiantes. Dans l'une d'elle, une patiente est simplement assise. On ne lui fait rien. Mais on la met dans des conditions propices au nocebo. Sur une échelle de 0 à 10, la patiente rapportait une douleur de 9,5. Une douleur de 9,5 sur 10 !! Genre, c'était probablement pour elle comparable à la douleur de l'accouchement[110] !
>
> Attends, t'es en train de dire que, via SmartPoop, on a ciblé des noirs africains pour les torturer ? C'est n'importe quoi !
>
> Clairement, on ne l'a pas fait de manière intentionnelle. Mais oui, si on prend du recul, je pense qu'on peut dire que c'est ce qu'on a fait...
>
> Bon... Quoi qu'il en soit, il faut qu'on prenne les devants dans cette affaire avant que ça ne leake. Je vais réorganiser SmartPoopCon, pour parler du problème qu'on a, et de ce qu'on prévoit de faire pour éviter que nos algorithmes soient racistes et causent de telles souffrances.

Inclusion et diversité

Quelques jours plus tard, à SmartPoopCon 2024, la conférence annuelle majeure de l'entreprise SmartPoop, Katia prit la parole pour annoncer les nouvelles

[109] Lorsqu'un algorithme d'aide à la décision devient performant, il semble important d'y voir plus qu'une simple « aide » à la décision. À l'instar de Google Map, de tels algorithmes peuvent finir par être écoutés presque aveuglément par leurs utilisateurs, si bien que l'aide à la décision devient finalement essentiellement la décision elle-même.

[110] Marc fait ici référence à cette vidéo :
:tv: Touch - Mind Field (Ep 6). VSauce (2017).

mesures de diversité et inclusion de son entreprise.

> Dans notre industrie, nous avons tendance à vouloir constamment aller de l'avant, en présentant des nouveaux projets pour le futur. On attend de nous qu'on innove et qu'on surprenne. Mais ce n'est pas la mission de SmartPoop. La mission de SmartPoop, c'est de garantir à tous nos utilisateurs une prise en charge sanitaire sans égale. La mission de SmartPoop, c'est de veiller à la santé et au bien-être de *tous* les humains sur terre. Et aujourd'hui, j'aimerais affirmer publiquement toujours plus cette volonté de notre entreprise.
>
> Or, pour cela, pour vraiment arriver à prendre soin de tous les humains sur terre, plutôt que de toujours constamment regarder de l'avant, il est aussi critique de surveiller nos technologies et de vérifier qu'elles fonctionnent correctement. À SmartPoop, c'est ce que nous faisons systématiquement. Malheureusement, récemment, nous nous sommes rendus compte que nous avons un biais systémique dans la manière dont nous testons nos produits. Nos équipes d'ingénieurs étant principalement des hommes blancs, ils échouent souvent à penser aux préoccupations des minorités. Pour combler ce manque, nous sommes convaincus qu'il nous faut introduire beaucoup plus de mixité sociale dans nos équipes. C'est pourquoi nous nous engageons activement à combattre nos biais systémiques en promouvant le recrutement des populations actuellement sous-représentées parmi nos employés[111].
>
> Alors, j'aimerais pouvoir vous dire que nous avons pris cette décision parce que nous avons toujours priorisé l'éthique. Mais pour être honnête avec vous, je dois bien avouer que cette décision vient avant tout du constat, malheureusement trop tardif, que l'une de nos technologies n'était pas à la hauteur de nos exigences éthiques. SmartPoop, appliqué aux SmartToilet Deluxe, a subi ce qu'on appelle un distributional shift, c'est-à-dire une divergence entre ses données d'entraînement et les cas d'application pratique, qui l'a malheureusement conduit à des erreurs de diagnostics pour certaines populations[112]. À tous nos utilisateurs victimes de ces erreurs, mais aussi à toutes ces populations, nous présentons nos plus sincères

[111] Plusieurs études soulignent l'importance de la diversité pour la créativité d'un groupe. Il ne semble toutefois y avoir des résultats confus et contradictoires concernant l'impact de la diversité sur l'efficacité du groupe. Ce n'est sans doute pas étonnant, sachant que cet impact dépend très probablement de nombreux autres facteurs, comme la tâche assigneé au groupe, la sensibilité émotionnelle des membres du groupes et l'organisation de ce groupe.
:memo: Collective Intelligence and Group Performance. Anita Williams Woolley, Ishani Aggarwal & Thomas W. Malone. Current Directions in Psychological Science (2015).
:memo: Evidence for a Collective Intelligence Factor in the Performance of Human Groups. Anita Wooley, Christopher Chabris, Alex Pentland, Nada Hashmi & Thomas Malone. Science (2010).

[112] :memo: Preventing Failures Due to Dataset Shift: Learning Predictive Models That Transport. Adarsh Subbaswamy, Peter Schulam & Suchi Saria. AISTATS (2019).

excuses.

Malheureusement, en l'absence de suffisamment de données de certaines populations, il est en fait mathématiquement impossible de garantir la même qualité de service à ces populations que celle qu'on propose à d'autres populations. Or, à SmartPoop, nous pensons que les technologies doivent être inclusives. Vu l'impact de ces technologies sur la société moderne, il nous semble qu'il serait immoral de n'offrir ces technologies qu'à certaines populations, qui plus est à des populations déjà privilégiées. C'est pour cette raison que nous avons décidé de subventionner drastiquement les SmartToilets Deluxe aux populations dont nous manquons de données pour établir des alertes médicales fiables et pertinentes. Le budget de cette opération est estimé à 100 million de dollars. C'est une somme colossale pour nous ; mais c'est un petit prix à payer pour l'égalité sociale dans nos sociétés.

Le sujet de l'impact des technologies sur les inégalités sociales est délicat à aborder. Beaucoup d'entreprises préfèrent l'éviter. Cependant, contrairement à certains de nos concurrents[113], nous ne souhaitons pas cacher ces problèmes sous le tapis. Il est grand temps que l'industrie des technologies cherche activement à comprendre les conséquences néfastes de leurs technologies[114]. Nous nous engageons ainsi à soulever le moindre caillou, et à vous révéler de façon transparente les problèmes que nous rencontrerons[115]. Sachez, néanmoins, que ce faisant, nous révèlerons des problèmes souvent omniprésents dans toute l'industrie des technologies. Si nous en sommes victimes,

[113]Les *facebook files* révèlent que, contrairement à ce que Mark Zuckerberg a déclaré publiquement à plusieurs reprises, Facebook tient une *whitelist* secrète de personnalités exemptes de la politique de modération de contenus de Facebook, y compris des dirigeants de pays autoritaires. Voilà qui renforce le pouvoir de personnes qui en possèdent probablement déjà trop. Par exemple, le joueur de football Neymar étant une star planétaire, et sa présence sur Facebook attirant beaucoup de vues, Facebook autorise ce joueur à publier une vidéo de *revenge porn*, qui expose des images d'une femme nue sans son consentement. Une telle publication est interdite par les Facebook Community Standards, et auraient dû conduire au bannissement de Neymar de Facebook.
:headphones: The Facebook Files, Part 1: The Whitelist. The Journal (2021).

[114]Les *facebook files* révèlent aussi que la direction de Facebook a découragé à de multiples reprises l'investigation de potentiels problèmes sur la plateforme. De façon plus générale, en insistant sur le fait que toute entreprise peut être punie si elle facilite *sciemment* des actes illégaux, la loi encourage malencontreusement les entreprises à ne pas investiguer tout problème du genre, car une fois qu'il est possible de montrer que les entreprises étaient mises au courant de ces problèmes, elles ont alors le devoir légal d'investir massivement pour résoudre ces problèmes, ce qui est très coûteux pour elles.
:headphones: The Facebook Files, Part 3: 'This Shouldn't Happen on Facebook'. The Journal (2021).

[115]Très clairement, Google n'a pas du tout ce degré d'exigence. Pour rappel, les deux co-directrices de l'équipe d'éthique de Google ont été licenciées tour à tour, suite à la publication d'un article scientifique discutant des risques éthiques, environnementaux et sociaux, d'un déploiement des technologies avancées de traitement du langage.
:tv: Google démantèle son éthique (et tout le monde s'en fout...). Science4All (2021).

les autres entreprises le sont aussi. Quand vous ferez ainsi notre procès, pensez à faire celui de toute l'industrie des technologies, en vous disant que les entreprises qui révèleront le moins leurs problèmes sont en fait probablement celles qui en ont le plus[116].

Et là, j'aimerais que vous pensiez à vous poser la question la plus importante de l'industrie des technologies : ma confiance en tel ou tel produit de telle ou telle entreprise est-elle justifiée[117] ?

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.

[116]Malheureusement, aujourd'hui, les entreprises qui communiquent davantage sur les problèmes auxquelles elles sont confrontées sont souvent davantage critiquées que celles qui cachent leur problème. Typiquement, la politique plus transparente de Twitter semble avoir conduit à plus de critiques que la politique très opaque de Facebook. Ou pour prendre un autre exemple plus concret, en 2014, Facebook a partagé publiquement le résultat (fascinant) de son analyse de l'impact d'une réduction très légère de la publication, dans les fils d'actualités, de posts avec des émotions négatives, sur ce que les utilisateurs exposés à ces posts se mettaient à écrire à leur tour.
:headphones: Can Algorithms Choose our Emotions? Robustly Beneficial (2020).
:memo: Experimental evidence of massive-scale emotional contagion through social networks. Adam Kramer, Jamie Guillory & Jeffrey Hancock. PNAS (2014).
Cette étude a toutefois conduit à un lever de bouclier sur l'éthique de telles études.
:memo: Facebook's emotional contagion study and the ethical problem of co-opted identity in mediated environments where users lack control. Evan Selinger, Woodrow Hartzog. Research Ethics (2015).
Une conséquence malencontreuse de ce lever de bouclier, toutefois, c'est que Facebook a cessé de publier de telles études, quand bien même Facebook n'a très certainement pas cessé d'effectuer ces études, ainsi que des études beaucoup moins intéressantes sur l'éthique (comme l'A/B test des contenus les plus addictifs). Dès lors, de l'extérieur, il est très difficile de comprendre l'impact de Facebook. En fait, les *facebook files* montrent que ceci a surtout permis à Facebook de cacher l'ampleur de ses impacts néfastes, par exemple sur la santé mentale des adolescentes qui utilisent Instagram.
:headphones: The Facebook Files, Part 2: 'We Make Body Image Issues Worse'. The Journal (2021).
Plus généralement, on semble avoir là une tension entre *l'idéalisme* et le *conséquentialisme*. Alors que le conséquentialisme cherche à améliorer l'état du monde autant que possible, l'idéalisme se fixe un idéal et refuse toute action qui semble contradictoire avec cet idéal, même si cet idéal est inatteignable de manière pragmatique dans un futur proche.
:headphones: Radically Normal: How Gay Rights Activists Changed The Minds Of Their Opponents. Hidden Brain (2019).

[117]La question fondamentale que Katia soulève là est celle de la *calibration de la confiance*. Par exemple, si une personne est calibrée dans ses prédictions, sur l'ensemble des fois où elle affirme qu'un événement va arriver avec probabilité 80%, cette personne doit avoir vu juste 8 fois sur 10. À l'heure des nouvelles technologies, malheureusement, la confiance qu'on assigne à certains produits est injustifiablement élevée, tandis que la confiance qu'on assigna à d'autres produits est injustifiablement faible.
:tv: Le rôle de la psychologie pour rendre l'intelligence artificielle bénéfique - feat. Science4All. Université Grenoble Alpes (2021).
:memo: Practical Guidance for Evaluating Calibrated Trust. Patricia McDermott & Ronna ten Brink. Human Factors and Ergonomics (2019).

Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

Chapitre 4 — Les fuites sanitaires

C'est un séisme. Contre toute attente, le 4 septembre 2025, Jules Lartan, Président du Bokistan, décide de sortir son pays de la COP 31, la conférence internationale des parties sur les mesures à prendre pour lutter contre le changement climatique[118]. Pâle et visiblement épuisé, il tient le discours suivant devant la presse.

> C'est une décision difficile. Le changement climatique est une menace sérieuse au futur de l'humanité. Cependant, les exigences de nos partenaires internationaux sont injustement disproportionnées[119]. À cause de notre succès économique, beaucoup trop de ces partenaires nous identifient systématiquement comme étant la seule cause du dérèglement climatique actuel, ce qui les conduit à exiger du Bokistan des sacrifices déraisonnables. Les Bokistanais sont un peuple qui ne méritent pas un tel jugement, ni un tel mépris. J'ai ainsi préféré protéger les nôtres, plutôt que de succomber à la tyrannie pseudo-morale de nos partenaires internationaux. Nous persisterons à combattre le changement climatique. Cependant, nous le ferons avec calme et enthousiasme, et avec notre propre souveraineté ; pas selon le bon vouloir de nos partenaires. Vive le Bokistan libre.

Devant sa télévision, Célia Keita n'en croit pas ses oreilles. Cette journaliste d'investigation a suivi de près les débats en amont de la COP 31. Si le Président Lartan n'y a pas eu un rôle moteur, il semblait toutefois loin d'être réticent aux négociations pour lutter contre le changement climatique. Après tout, sa propre campagne présidentielle insistait beaucoup sur les valeurs familiales, et l'importance de préparer au mieux le futur des prochaines générations en luttant contre le changement climatique. D'ailleurs ses positions anti-avortement et

[118] La COP est une conférence annuelle internationale qui vise à s'attaquer aux questions environnementales.
:tv: C'est quoi, une COP ? Le tour de la question. La Croix (2017).

[119] :tv: Paris climate talks: Global problem, global deal. Nature video (2015).
:tv: Rich vs. Poor: Who Should Pay To Fix Climate Change? | Hot Mess (2018).

anti-euthanasie étaient justifiées par l'importance de rétablir à la Nature ses droits[120]. Comment a-t-il pu en venir à un tel volte-face ?

Même d'un point de vue stratégique, cette pirouette ne semble pas être justifiée. D'ailleurs, les jours qui suivent, c'est surtout l'incompréhension dans le parti du Président Lartan qui règne. La popularité du Président décroît, alors qu'on semblait se diriger vers un rebond de popularité en cas de succès de la COP 31. Quelque chose cloche[121].

La shitstorm

Célia décide de mener une investigation sur les raisons de ce volte-face. Elle interroge les autres parties prenantes de la COP 31. Certains décrivent alors le Président Lartan comme un homme difficile, refusant le compromis et essayant d'imposer ses propres idées. D'autres toutefois, qui le connaissent de longue date, affirment que celui-ci a eu un comportement de plus en plus difficile depuis son élection, et en particulier au cours des derniers mois. Petit à petit, Célia identifie une date de plus en plus précise, après laquelle le Président Lartan refusait le compromis. Quelque chose semble avoir eu lieu fin juillet 2025, qui a ensuite amené le Président Lartan à conduire la COP 31 droit dans une impasse.

Célia rassemble alors ses éléments, et publie finalement un article dans le plus grand journal Bokistanais, le journal La Terre[122]. L'article est intitulé « Le mystérieux volte-face du Président Lartan ». Le sous-titre est plus suggestif encore : « Comment cet écologiste convaincu a fini par saboter la plus grande initiative écologiste au monde ». L'article suggère vivement l'existence d'accords secrets qui ont changé les positions du Président Lartan.

Cet article est toutefois abondamment critiqué. « Sensationnaliste », « partisan », « spéculateur », voire même « conspirationniste ». On reproche à Célia de vouloir décrédibiliser tout un parti politique, et de vouloir semer le trouble dans la nation Bokistanaise. Célia reçoit de nombreuses insultes sur Twitter et Facebook, qui tournent ensuite au harcèlement quotidien. Chaque jour, des centaines de messages inondent Célia de qualificatifs détestables, allant de « troll », « putaclic » et « incompétente », à des attaques personnelles sur son physique et

[120] On peut remarquer que le clivage politique du Bokistan ne semble pas coïncider avec celui des États-Unis ou de la France, où les positions anti-avortement et anti-euthanasie sont corrélées avec des positions conservatrices, mais aussi capitalistes et peu enclines à combattre le changement climatique. Il est intéressant de se demander s'il y a une raison philosophique fondamentale qui justifie ces corrélations, ou s'il s'agit d'une coïncidence issue du bipolarisme politique auquel nos scrutins uninominaux nous poussent.
:tv: Nos démoracties divisent. Science4All (2017).

[121] Les modèles prédictifs de Bruce Bueno de Mesquita supposent que les politiciens sont avant tout stratégiques, et cherchent à être et à rester au pouvoir.
:tv: Le principe fondamental de la politique. Science4All (2017).
:tv: The Rules for Rulers. CPG Grey (2016).
:books: The Dictator's Handbook Why Bad Behavior is Almost Always Good Politics. Bruce Bueno de Mesquita & Alastair Smith. Public Affairs (2011).

[122] Le nom du journal est bien sûr une référence au journal français Le Monde.

ses origines[123].

Des milliers de montages pornographiques, avec le visage de Célia et des technologies de *deepfake*, sont même abondamment produits[124] et partagés sur des réseaux sociaux comme Reddit, accompagnés de menaces de viol et de mort. Terrifiée, Célia porte plainte contre ses harceleurs. Elle dépose aussi une plainte contre Twitter, Facebook et Reddit, qui relaient et promeuvent des messages haineux à son encontre[125]. Cependant, si les réseaux sociaux expriment publiquement des excuses et bloquent automatiquement les insultes envoyées à son égard, la police reste impuissante face à ces comptes anonymisés.

Sous le choc, Célia s'enferme chez elle. Elle est alors incapable de travailler. Elle finit par désinstaller tous les réseaux sociaux et rend ses comptes inaccessibles. Après un mois, elle décide de consulter un psychiatre, qui l'aide à retrouver une bonne humeur, une motivation pour travailler et une capacité de concentration[126].

Plusieurs mois plus tard, après avoir lu un article terrifiant sur les campagnes de désinformation organisées[127], Célia se demande enfin si elle en a été une victime. Et si, en plus de produire et de promouvoir de la désinformation via des faux comptes[128], ces campagnes cherchaient aussi à taire certaines informations,

[123]En 2021, la Déclaration de Bruxelles a été rédigée pour signaler le harcèlement que subissent de nombreux journalistes et pour appeler à la défense de ces journalistes.
:computer: La Déclaration de Bruxelles (2021).

[124]De nombreuses femmes ont été la cible de tels montages, comme la journaliste Rana Ayyub. Il est terrifiant de se rendre compte que l'une des grandes applications aujourd'hui du traitement automatisé d'images est le cyber-harcèlement de femmes sur le web, souvent dans le but de les réduire au silence, notamment lorsqu'elles veulent révéler des scandales.
:computer: I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me. Rana Ayyub. The Huffington Post (2018).
:computer: Opinion: The threat from deepfakes isn't hypothetical. Women feel it every day. The Washington Post (2021).

[125]En 2018, Facebook a modifié son algorithme de fils d'actualité. Ceci a conduit à une augmentation de la viralité des contenus haineux et polarisants, car ces contenus amenaient plus de gens à réagir à ces contenus, qui étaient alors davantage recommander. Les Facebook files montrent que ce problème a été soulevé par des ingénieurs, qui en ont parlé avec la direction de Facebook. Cependant, celle-ci a largement ignoré ces problèmes, le nouvel algorithme renforçant l'engagement des utilisateurs sur Facebook, et donc la profitabilité du modèle d'affaire de Facebook fondé sur la vente de publicités.
:headphones: The Facebook Files, Part 4: The Outrage Algorithm. The Journal (2021).

[126]Le podcast suivant raconte une histoire similaire, dans le cas d'une dépression suite à une utilisation obsessive d'Instagram.
:headphones: The Facebook Files, Part 2: 'We Make Body Image Issues Worse'. The Journal (2021).

[127]La désinformation organisée est un acte volontaire de manipulation de l'information qui sera reçue par le grand public. Sur les réseaux sociaux, elle peut consister à produire des fausses informations, ou à la relayer, mais aussi à trouver d'autres moyens d'actions pour taire les informations que la désinformation organisée ne veut pas voir partagée.
:tv: Who is Manipulating Facebook? - Smarter Every Day (2019).

[128]L'ampleur des faux comptes est colossale. À titre d'exemple, chaque année, Facebook retire aux alentours de 6 milliards de faux comptes de sa plateforme, en grande partie avec des algorithmes. Dès lors, à chaque instant, et malgré les efforts monumentaux de Facebook de garantir que chaque compte est possédé par un humain authentique, il est quasi-certain que la plupart des comptes sur Facebook sont des faux comptes.

en harcelant les journalistes qui s'y intéressent[129] ? Si tel a été leur objectif, ces campagnes ont réussi. Voilà maintenant 6 mois que Célia a suspendu son investigation !

C'est finalement l'idée d'avoir été abattue par des faux comptes qui pique la fierté de Célia et la motive pour reprendre son travail de journalisme d'investigation. Célia restaure son compte Twitter. Tout l'historique de ses conversations est alors accessible. Célia épluche alors les comptes qui l'ont harcelé. Elle se rend compte qu'une bonne proportion de ces comptes a été créée récemment, avec des images de profil typiques du site de création d'images d'individus fictifs thispersondoesnotexist.com[130]. Il ne fait plus aucun doute. Célia a été la cible d'une campagne de désinformation.

Le cas Paul Gremoux

Célia note alors des similarités entre certains des comptes qui l'ont harcelé, avec par exemple, des fautes d'orthographes aux mêmes endroits. Certaines expressions paraissent particulièrement étranges. Elle cherche ces expressions sur Google. Google référence alors ces tournures à un compte Tumblr. Étrangement, ce compte parle de mangas ; pas du tout de politique ou de journalisme. Curieuse, Célia navigue néanmoins sur le Tumblr, et reconnaît un style d'écriture similaire à celui d'un compte qui l'a harcelé. De façon plus intéressante encore, Célia découvre que le Tumblr contient le nom et le prénom de son auteur : un certain Paul Gremoux[131].

En cherchant désormais ce nom et ce prénom sur Google, Célia découvre qu'ils

:computer: Facebook has shut down 5.4 billion fake accounts this year. CNN Business (2019).
:computer: Meet the A.I. that helped Facebook remove billions of fake accounts. Fortune (2020).

[129] Selon un rapport de l'UNESCO de 2020, les femmes journalistes semblent particulièrement exposées à ces attaques ignobles.
:memo: Online violence Against Women Journalists: A Global Snapshot of Incidence and Impacts. Julie Posetti, Nermine Aboulez, Kalina Bontcheva, Jackie Harrison, and Silvio Waisbord. UNESCO (2021).

[130] Le site thispersondoesnotexist.com est un site qui, à chaque requête, affiche l'image d'une personne qui n'existe pas. L'image est ainsi fabriquée par un *réseau de neurones adverarial génératif* (GAN), qui, en gros, cherche à fabriquer des images « difficilement discernables » de son jeu de données de photographies d'humains. Les images de ce site web, reconnaissables à la position fixe des yeux, sont souvent utilisés par des faux comptes. Cependant, les technologies progressent, ce qui rend la détection des fausses images probablement impossible pour des faux comptes bien fabriqués.
Les réseaux adversariaux (GAN). Science4All (2018).

[131] Ce que fait Célia là est un travail de déanonymisation, avec une approche très simple et manuelle. Le travail de déanonymisation peut être beaucoup plus sophistiqué et puissant, en exploitant par exemple les méta-données, comme l'heure de publication à différents lieux. L'exemple le plus marquant de l'histoire est sans doute celui de la dé-anonymisation de Ross Ulbricht, l'homme derrière le marché noir *Silk Road* du dark web. :tv: Ross Ulbricht - Le baron de la drogue du dark web. La chaîne de P.A.U.L. (2017).
De façon générale, la théorie de la *confidentialité différentielle* (differential privacy) insiste sur le fait que toute publication pseudonymisée n'offre en fait aucune garantie ; pour garantir la protection des données sensibles, des méthodes beaucoup plus rigoureuses sont requises.
:tv: What is Privacy? Wandida (2017).

appartiennent à un auto-entrepreneur en menuiserie, dont le numéro de téléphone est accessible. Tel pourrait être le numéro de son harceleur. Célia respire alors un grand coup, et compose ce numéro sur son téléphone.

> Allô, bonjour, Paul Gremoux ?
>
> Oui c'est moi.
>
> Excellent, j'aimerais vous poser quelques questions.
>
> Qui est à l'appareil ?
>
> Avant que je ne vous dise qui je suis, j'aimerais vraiment insister sur le fait que je ne vous veux pas de mal. Je suis prête à protéger votre anonymat si vous répondez à mes questions et si vous le désirez.
>
> Euh... d'accord.
>
> Je suis Célia Keita, journaliste d'investigation pour le journal La Terre. Et encore une fois, j'insiste sur le fait que je souhaite vous protéger, à condition bien sûr que vous répondiez à mes questions.
>
> D'accord.
>
> Excellent. Je sais que vous tenez le compte Twitter @rwqg[132]. Ma question est la suivante : est-ce que vous travaillez pour une campagne de désinformation ? Et si oui, qu'est-ce qui vous a amené à accepter ce travail ?
>
> Oui madame Keita, je suis vraiment désolé pour cela. On m'a contacté pour vous harceler sur Twitter[133]. Tout ce que j'avais à faire, c'est vous écrire trois fois par jour avec des insultes et des menaces. Je suis vraiment désolé. C'est vraiment à contre-coeur que je l'ai fait. S'il vous plaît, ne me dénoncez pas. J'ai des gros risques d'AVC, diagnostiqués par SmartPoop. Et si mon assurance l'apprend, je vais avoir des factures que je ne pourrai pas payer. J'ai deux enfants à charge. S'il vous plaît, ne me dénoncez pas. Je vous en prie.
>
> Merci beaucoup Monsieur Gremoux. J'aimerais juste que vous me donniez plus d'informations sur la campagne de désinformation qui vous a contactée.
>
> Je ne les connais pas du tout. Ils m'ont juste envoyé un email un jour, affirmant qu'ils connaissaient mes données SmartPoop, et que j'étais donc à risque d'AVC. Ils ont menacé de le dire à mon assurance, et

[132] Ce compte, dont le nom a été choisi ici au hasard, est actuellement suspendu sur Twitter (en octobre 2021).

[133] Cette pratique semble très présente, notamment dans les pays autoritaires. Un rapport de l'Institut de Recherche Stratégique de l'École Militaire (IRSEM) en France fait état de 20 millions de chinois payés pour produire de la désinformation, dont 2 millions à temps plein. :memo: Les opérations d'influence chinoises. un moment machiavélien. ISEM (2021)

> m'ont proposé de vous harceler en échange. Je peux vous transférer l'email[134]. Mais je n'en sais pas plus. Je suis désolé. Depuis que j'ai perdu mon métier, je n'ai pas réussi à en retrouver un autre. J'ai essayé de lancer ma propre activité, mais le ROVID-19 a anéanti mon business. S'il vous plaît, j'ai deux enfants à charge.
>
> Merci beaucoup pour vos réponses. Je ne compte pas poursuivre mes plaintes contre vous. Je garderai votre histoire secrète. Je vous le promets. Je vous demanderais juste de me transférer cet email à mon adresse celia.keita@journal-laterre.com[135]. Merci.
>
> Merci infiniment à vous. Et toutes mes excuses à nouveau.

L'inflitration

Quelques minutes plus tard, Célia reçoit l'email transféré par Pau. L'email original, qui contient le chantage, a été envoyé depuis l'adresse foeqm@kinstova.com[136]. Le nom de domaine @kinstova.com correspond alors à un serveur dans un pays étranger, la Kormique. Ce pays est aussi bien connu pour son industrie pétrolière. Se pourrait-il que la campagne de désinformation ait été organisée par l'une de ces industries, pour freiner les réglementations contre ces industries ? Se pourrait-il que cette campagne se soit aussi attaquée au Président Lartan ? Et si le Président était l'objet d'un chantage similaire à celui de Paul Gremoux ?

Pour le savoir, Célia cherche à se rapprocher de Marie Routisse, la soeur de la femme du Président, qui se trouve aussi être l'amie d'un ami. En passant par son contact, Célia propose à Marie de prendre un café, pour parler d'un article sur la « place des femmes dans le monde professionnel ». Lors du premier contact, Célia ne cherche pas à obtenir directement de l'information de Marie. Elle cherche surtout à l'amener à baisser sa garde, en devenant amie avec elle[137].

Après trois ou quatre rencontres, Célia invite cette fois Marie dans un bar, et la pousse à la consommation d'alcool. En fin de soirée, elle en vient enfin au sujet qui l'intéresse.

> C'est vraiment difficile d'être une femme professionnelle tout en ayant une vie de famille. Je suis sûre que si ta sœur ne cherchait pas à être

[134] En 2021, le YouTubeur scientifique français Léo Grasset, de la chaîne Dirty Biology, a été contacté pour produire une désinformation au sujet des vaccins Pfizer-BioNTech. Il a dénoncé cette affaire, en partageant l'email avec des journalistes. Les journalistes ont ensuite pu remonter à des sources russes.
:tv: Comment une agence russe a essayé de m'utiliser. DirtyBiology (2021).

[135] N'utilisez pas cette adresse email en vrai ! Elle n'a été insérée ici que pour rendre le récit réaliste.

[136] Même remarque que pour l'adresse email de Célia.

[137] On parle parfois de *social engineering*. Il s'agit de techniques pour gagner la confiance d'une cible, soit pour récupérer de l'information d'elle, soit pour l'arnaquer. Le *social engineering* est souvent considéré être la principale vulnérabilité des systèmes informatiques.
:tv: Social Engineering - How to Scam Your Way into Anything. Brian Brushwood. TEDxSanAntonio (2011).

> une bonne mère de famille, c'est elle qui serait Présidente, non ?
>
> Ouais, je ne suis pas sûre que ma sœur soit une si bonne mère de famille...
>
> Comment ça ?
>
> Non, mais t'es une journaliste. Je ne devrais pas te le dire.
>
> Je te promets de ne jamais en parler publiquement.
>
> On m'a dit de ne jamais faire confiance aux journalistes.
>
> Marie, je t'invite à contacter toutes les personnes que j'ai citées dans mes articles. Tous ont eu un droit de regard sur mes articles, et si quiconque voulait que je retire une information de mes articles, je l'ai fait.
>
> Oui je sais. Mes avocats m'ont briefé sur toi. Mais ce que je vais te dire est très gros.
>
> Marie, je suspecte vivement le Président Lartan de trahir ses convictions politiques à cause de cela, et que cela vient d'un chantage. Je le suspecte d'autant plus car j'ai des sources qui ont été victimes de chantages similaires, venant probablement d'entreprises pétrolières Kormicaines. Ces personnes ont été amenées à me harceler quotidiennement contre leur gré. Et je crois que la source de leur chantage réside dans les données SmartPoop. Marie, c'est la gestion du changement climatique qui est en jeu, et donc le futur des prochaines générations. Le futur de tes enfants, et de leurs petits-enfants[138].

Après un long silence, Marie avoue : « ma soeur a avorté en juin 2025 ». Une tempête cérébrale agite alors les neurones de Célia, alors que toutes les pièces du puzzle se mettent en place, et que le brouillard s'éclaircit.

> C'est donc ça. Les industries Kormicaines ont certainement réussi à mettre la main sur des données SmartPoop de ta soeur. Et clairement, avec l'urine, il est possible de savoir si ta soeur est enceinte[139], et si elle cesse prématurément d'être enceinte ! Ils ont alors deviné qu'elle a effectué une interruption volontaire de grossesse. Or, si cela se sait, toute la communication anti-avortement du Président Lartan devient subitement décrédibilisée, et sa carrière est finie. Les industries Kormicaines ont certainement exigé que le Président Lartan sorte des accords de la COP 31, pour éviter un tel scandale. Wow !! C'est une énorme affaire !

[138] L'origine humaine du changement climatique et les risques majeurs qu'il fait encourir à toute l'humanité sont désormais des consensus scientifiques.
:tv: Comment le réchauffement climatique va bouleverser l'humanité (ft. Le Réveilleur). Le Monde (2021).
:tv: 97% of Climate Scientists Really Do Agree. It's Okay To Be Smart (2018).

[139] :tv: How do pregnancy tests work? - Tien Nguyen. TED-Ed (2015).

> Célia, je t'en supplie, ne le dit pas publiquement. Même si je suis choquée par ce qu'elle a fait, je ne veux pas que ma soeur soit humiliée publiquement et roulée dans la boue. Une telle humiliation personnelle aurait des répercussions énormes sur sa santé mentale déjà fragile, sans compter ses enfants, et toute sa famille y compris mes propres enfants. S'il te plaît, il y a de nombreuses vies en jeu[140].
>
> Marie, tu as ma parole. Mais je ne peux pas non plus ne pas révéler cette affaire. Le changement climatique menace lui aussi la vie de tes enfants. Je vais réfléchir à la meilleure façon d'avancer, avec ton avis. Mais on ne peut pas laisser des industries Kormicaines manipuler ainsi l'homme le plus puissant de notre pays ! Il faut que tu m'aides à arranger une rencontre avec le Président.

Le dilemme du Président

Quelques jours plus tard, Célia reçoit un appel du Président Lartan. Le Président l'invite à discuter dans son palais présidentiel, à l'abri de toute écoute de quiconque. Le soir même, Célia se rend au palais, et est reçue par le Président.

> Merci pour votre temps et votre attention. Je commence par vous demander : êtes-vous sûr que nous pouvons parler en sécurité ici, sans risquer d'être entendus ou enregistrés ?
>
> Oui, madame Keita, nous pouvons parler en toute sécurité[141].
>
> Marie vous l'a sans doute dit. Je sais que des entités externes sont en train de vous faire chanter, parce qu'elles savent que votre femme a eu une interruption volontaire de grossesse. Je sais qu'elles vous ont demandé en particulier de vous retirer de la COP 31. Je sais à quel point la situation doit être difficile pour vous et pour votre famille. Et j'ai promis à Marie de ne rien divulguer sans son accord. Néanmoins, je vous supplie de considérer l'ampleur de la vulnérabilité que ce chantage représente[142]. Vous êtes l'homme le plus puissant

[140]:memo: Depression high among youth victims of school cyber bullying, NIH researchers report. NIH (2010).

[141]Il est en fait désormais sans doute impossible, surtout pour un Président, de garantir qu'une discussion n'aura pas lieu sous écoute. Bien entendu, les téléphones écoutent en permanence, à l'affût d'un « OK Google » ou d'un « Hey Siri », et ils peuvent être hackés, comme on l'a vu dans l'affaire Pegasus. En plus de cela, un microphone peut aisément être caché dans une pièce. De façon plus surprenante encore, le son peut parfois être reconstruit à partir d'une simple vidéo, qui analyse les vibrations d'objets légers comme un paquet de chips.
:tv: Can You Recover Sound From Images? Veritasium (2019).

[142]Le cas de l'affaire Pegasus montre l'ampleur de telles attaques par espionnage des dirigeants politiques dans le monde moderne. Pegasus est un *spyware*, c'est-à-dire un algorithme, qui peut être utilisé pour infecter un téléphone cible, et surveiller tout ce que ce téléphone fait. Pegasus est développé par le groupe NSO en Israël, et l'on sait que de nombreux services de renseignement à travers le monde l'ont utilisé pour espionner de nombreux journalistes, militants et dirigeants politiques.
:tv: Comment PIRATER Jeff Bezos avec un même ? (histoire vraie). Micode (2021).

> du Bokistan[143].

En entendant ces mots, le Président Lartan se met à pleurer.

> C'est un cauchemar. J'aime ma femme, et j'ai promis de toujours la protéger. Je vous promets ne pas avoir été mis au courant de sa grossesse. Je n'ai découvert son interruption de grossesse qu'au moment où on a commencé à me faire chanter. Quand je lui ai demandé si elle l'avait fait, elle s'est mise à pleurer, et à me supplier de ne pas le divulguer publiquement. Je le lui ai promis. Et je n'ai jamais trahi aucune promesse que je lui ai faite. J'aime ma femme plus que tout, et l'idée de la faire souffrir me terrorise. Je ne sais plus quoi faire.
>
> Je comprends. Ça doit être une situation terriblement difficile à vivre.

Célia marque un silence, alors que le Président continue de pleurer.

> Monsieur le Président, dans le film Spiderman, quand on lui demande de choisir entre sa copine et un bus d'étudiants, Peter Parker choisit de sauver sa copine, et trouve le temps de venir aussi en aide au bus d'étudiants[144]. Mais... Mais nous ne sommes pas dans un film de superhéros. Vous ne pourrez pas sauver le changement climatique et le secret de votre femme. Vous devez faire un choix. Le futur des prochaines générations est en jeu, Monsieur le Président. Sachez que je vous fais confiance pour faire ce qui doit être fait.

C'est sur ces mots que Célia quitte le Président Lartan. Le lendemain, le Président prend la parole dans une conférence de presse, où il annonce sa démission. Il

---

Le Président français Emmanuel Macron semble avoir été ciblé par les services marocains. De façon étonnante, le gouvernement français n'a pas accablé les services marocains, ce qui peut laisser croire que les services marocains ont effectivement des informations compromettantes qui leur permettent de faire chanter ce gouvernement.
:computer: « Projet Pegasus » : l'exécutif se mure dans le silence malgré le ciblage d'Emmanuel Macron par le Maroc. Olivia Faye. Le Monde (2021).
Le journal *Le Point* révèle que 5 ministres français ont été infectés par ce logiciel. Ceci soulève sérieusement le risque que les démocraties modernes sont aujourd'hui largement hackées par des entités malveillantes capables de les faire chanter comme elles le souhaitent.
:computer: Les téléphones de 5 ministres français infectés par le logiciel espion Pegasus. Le Point (2021).

[143]Le documentaire suivant montre l'ampleur de la corruption d'un pays (ici l'Azerbaïdjan) sur les hommes politiques de démocraties étrangères (ici l'Union Européenne).
:tv: La caviar connection (2/2). La machine à corrompre. ARTE (2021).

[144]Cet exemple est utilisé par Monsieur Phi dans une vidéo où il insiste sur le fait que le dilemme auquel est confronté au Président Lartan n'est pas un dilemme entre deux fondements de la morale, qui opposerait par exemple l'utilitarisme à la déontologie. Il s'agit en fait davantage d'un dilemme entre une morale globale, qui favoriserait le bus d'étudiants, et une préférence individuelle, qui favoriserait les proches comme la copine de Peter Parker. Ou dit autrement, dans cette situation, faut-il faire ce qu'on préfère pour nous, ou ce que la plupart des gens préfèreraient qu'on fasse ?
:tv: Encore plus utilitariste ? Monsieur Phi (2017).

révèle également son chantage et la cause du chantage, et affirme son soutien inconditionnel à sa femme. Il présente enfin ses excuses les plus profondes, et supplie tous les journalistes et politiciens de lui en vouloir à lui, mais pas à sa femme.

Le lendemain, le vice-président reprendra la place de Lartan. Sa première mesure est de réintégrer le Bokistan dans la COP 31. Grâce à son travail d'investigation, Célia a relancé la plus grande initiative au monde pour le climat !

L'ultimatum de Célia

Cependant, Célia sait que ce qui est arrivé au Président Lartan peut désormais arriver à n'importe qui. Tant que les données SmartPoop seront vulnérables, la sécurité nationale des plus grands pays du monde sera en grand danger[145], surtout quand le pouvoir est aussi centralisé qu'au Bokistan, où le Président dispose de grandes largesses. C'est ainsi que le sujet d'investigation de Célia se déplace des affaires géopolitiques à la sécurité informatique de SmartPoop. En particulier, Célia en vient rapidement à appeler Katia Carpinski.

> Bonjour Docteure Carpinski, ici Célia Keita, journaliste d'investigation chez La Terre.
>
> Bonjour Madame Keita, que puis-je faire pour vous ?
>
> Je vous appelle car j'ai identifié des cas de chantage à partir de données SmartPoop. En particulier, je tiens de source fiable que le chantage du Président Lartan s'appuie sur la détection de la grossesse de sa femme à partir des données SmartPoop. J'aimerais en savoir plus sur la sécurité de vos données. Pensez-vous qu'une industrie étrangère ait pu hacker vos serveurs et récupérer des données ?
>
> C'est impossible, affirme Katia, tout en devenant pâle elle-même. Notre infrastructure repose sur les technologies de sécurité informatique les plus abouties qui soient.
>
> Est-ce que vous pouvez vérifier les données de la femme du Président ?
>
> Oui, dans le cadre d'un audit de sécurité, c'est possible. Cependant, la loi sur la protection des données personnelles m'interdit de partager

[145] De façon surprenante, en 2020, Vladimir Poutine a appelé à un accord pour limiter les cyber-attaques entre différents pays, alors même que la Russie a été reconnue comme l'auteur de beaucoup de ces attaques. Ceci suggère que même les auteurs de ces attaques sentent qu'elles finiront par tout déstabiliser, y compris eux-mêmes.
:computer: Putin says Russia and U.S. should agree not to meddle in each other's elections. Reuters (2020).
Pour la sécurité de tous, il est sans doute urgent d'établir une convention internationale qui, à l'instar des armes chimiques, biologiques et autonomes, interdirait aussi les cyber-attaques, tant celles-ci mettent en danger des millions de vies, voire des milliards de vies à travers le monde.
:tv: Pourquoi faut-il bannir les armes autonomes ? | The Flares (2019).

les résultats de l'audit avec vous.

Je comprends. Je me demandais par ailleurs. Effectuez-vous des audits externes de la sécurité de vos algorithmes ?

Nous avons des mesures d'inclusion et de diversité, qui visent à soulever les problèmes éthiques de nos produits.

Docteure Crapinski, vous ne répondez pas à ma question. Permettez-vous à des entités externes d'effectuer ce travail ?

Nous avons des collaborations avec de nombreuses entreprises d'audit externe.

Docteure Crapinski, ne me faites pas répéter ma question une troisième fois. Je m'apprête à écrire un article sur la sécurité de votre système d'information, et l'absence d'audit externe serait un point très néfaste. Je suis prête à adapter mon article si vous collaborez avec moi dans l'éclaircissement de cette affaire. Mais sachez que vos données représentent désormais un enjeu de sécurité nationale. Et si j'ai l'impression que vous ne faites pas des efforts à la hauteur de cet enjeu, dans l'intérêt de la société civile, je serai contrainte d'écrire un article extrêmement critique.

Pardonnez-moi, madame Keita. Non, la sécurité de notre système informatique n'est pas auditée par une entité externe.

Je vous inviterais à en faire.

Madame Keita, je vous assure que nous faisons de notre mieux pour protéger nos utilisateurs et leurs données.

Si vous faisiez de votre mieux, vous permettriez l'audit externe de vos algorithmes.

Oui, en effet. Je vous promets que nous collaborerons pleinement avec vous pour identifier et patcher toutes les vulnérabilités potentielles de notre système.

Je suis ravie de l'entendre. Je vous donne un mois pour me fournir un rapport sur la sécurité de votre système, si possible avec l'aval d'un audit externe. Et encore une fois, en cas de faux pas de votre part, je ne serai pas tendre.

Merci beaucoup pour ce délai. Je vous promets de faire de notre mieux.

SmartPoop est dangereux

Juste après avoir raccroché, Katia court dans le bureau de Marc. Elle est visiblement extrêmement préoccupée.

À quel point t'es confiant sur la sécurité de nos données ?

Je ne sais pas Katia, c'est toi l'experte en informatique parmi nous.

Je viens de parler avec une journaliste. Elle affirme que le scandale Lartan a commencé avec une fuite des données SmartPoop de la femme du Président. Elle prétend avoir des sources fiables, et ça m'a l'air sérieux. Si c'est vrai et si ça leake, on va passer des mois très difficiles.

Attends, t'es sérieuse ? On a fuité les données de la femme du Président Lartan ?

C'est ce qu'elle prétend. Ou plutôt, que quelqu'un nous les a volées.

C'est possible ?

Le plus probable, je pense, ça reste que la femme du Président a fait n'importe quoi, et qu'elle a envoyé un email à Y en pensant l'envoyer à X, ou qu'elle a été victime de je ne sais trop quelle autre arnaque par phishing ou par pièce jointe piégée. Ou peut-être encore que son téléphone s'est fait hacker. Les vulnérabilités imaginables sont très nombreuses.

Attends... Mais donc, la sortie de la COP 31, c'est lié à ça ?

Je n'en sais rien. Je n'y ai pas réfléchi. Mais c'est vrai que la journaliste a parlé de chantage.

Katia, tu te rends compte des conséquences géopolitiques du leak de nos données ?

Oui. Et si la journaliste en parle, ça risque de faire des gros dégâts pour SmartPoop. On risque de perdre la confiance des utilisateurs. Et si on perd la confiance des utilisateurs, on perdra celle de nos investisseurs. On risque de devoir fermer des branches, perdre des parts de marché et peut-être pire.

Katia, oublie SmartPoop quelques instants. On parle quand même du Président du Bokistan, le mec qui a fait capoter la COP 31 !

La journaliste a l'air toutefois très raisonnable. On peut essayer de lui montrer patte blanche, et faire en sorte que l'article ne soit pas trop violent pour nous.

Katia, tu ne te rends pas compte que la sécurité nationale du Bokistan est en jeu ? En fait, non, la sécurité de tous les états au monde est en jeu. C'est le futur de toute l'humanité qui est à risque !

Comment ça ?

> Tant que les données SmartPoop ne seront pas sécurisées, il faut s'attendre à ce que ce genre de manipulations de dirigeants devienne la norme. C'est terrifiant.

Katia reste songeuse. Visiblement, elle pense encore avant tout à sauver SmartPoop.

> Et Katia, ça ne se limite pas qu'à des affaires de chantages, poursuit Marc. Imagine qu'un pays parvienne à récupérer les données SmartPoop d'un autre pays. S'il arrive à anticiper une épidémie dans l'autre pays, il peut interrompre la chaîne de production des traitements contre l'épidémie, et ainsi provoquer une crise dans l'autre pays. La sécurité des données médicales, c'est vraiment très important. Sans compter les prix des assurances, la discrimination envers les malades et les outings familiaux imprévus. Il faut vraiment qu'on fasse gaffe à cela.
>
> Je vois. Mais comment on se protège ?
>
> Il faut un audit externe monumental de tout notre code informatique.
>
> Un audit complet ? Ça va nous coûter des milliards !
>
> C'est un petit prix à pays pour le futur de l'humanité.
>
> Non mais en fait, non. Auditer tout notre code, ça va coûter même beaucoup, beaucoup plus ! Je sais qu'il y a une banque australienne qui a payé 750 millions de dollars pour refaire le système informatique[146]. Et ça, c'est une petite banque, et leur service est des ordres de grandeur plus simple que SmartPoop. Je pense qu'on parle là en dizaines, voire en centaines de milliards de dollars.
>
> Katia, on n'a pas le choix. Ne serait-ce que pour sauver SmartPoop...

L'audit de SmartPoop

Un mois plus tard, SmartPoop publie un rapport de sécurité informatique. Ce rapport explique l'adoption d'une nouvelle mesure de protection des données personnelles, fondée sur le chiffrement homomorphe. Cette technologie cryptographique permet ainsi à des utilisateurs de participer à l'amélioration des algorithmes de SmartPoop, sans que SmartPoop n'ait jamais accès aux données en clair des utilisateurs. Plus précisément, quand l'utilisateur collecte des données via son smartphone ou ses toilettes, ces données sont automatiquement chiffrées, avec une clé dont seul l'utilisateur dispose. Néanmoins, ce chiffrement est effectué de telle manière que SmartPoop pourra apprendre de ces données, quand bien même SmartPoop ne pourra jamais déchiffrer ces données[147]. Ainsi,

[146]:computer: Banks scramble to fix old systems as IT 'cowboys' ride into sunset. Anna Irrera. Reuters (2017).

[147]Notez que les principales applications raisonnables aujourd'hui du chiffrement homomorphe aux algorithmes d'apprentissage concernent uniquement la phase d'inférence — pas

les données personnelles sont protégées contre SmartPoop même, ce qui empêche SmartPoop d'être une vulnérabilité !

Le rapport de 400 pages détaille aussi la liste des librairies de code utilisées par SmartPoop. Certaines librairies ont été développées par des initiatives OpenSource comme Django ou React, d'autres par des entreprises du numériques comme Apple ou Google, et d'autres ont été écrites par SmartPoop même. De façon cruciale, chaque librairie utilisée est une vulnérabilité potentielle.

Or le cumul de ces librairies de code représente désormais des centaines de millions de lignes de codes. Pire encore, ces librairies permettent de définir des algorithmes d'apprentissage qui, eux, possèdent près d'un million de milliards de paramètres[148]. Ces paramètres sont quotidiennement ajustés à partir des milliers de milliards de nouvelles données quotidiennes, issues des activités fécales de milliards d'utilisateurs SmartPoop.

> Nos algorithmes ont atteint la complexité du cerveau humain et ses millions de milliards de synapses connectant ses centaines de milliards de neurones, explique le rapport. Cette complexité dépasse très largement l'entendement humain. À titre de comparaison, le rapport que vous lisez représente environ un méga-octet, soit 100

d'apprentissage. Autrement dit, dans le cas de SmartPoop, ce chiffrement homomorphe permet de diagnostiquer un excrément d'un utilisateur de manière parfaitement chiffrée ; mais il ne permet d'exploiter les données de l'utilisateur pour améliorer l'algorithme de diagnostic.
:tv: Machine Learning with Encrypted Data | Homomorphic Encryption. SATSifaction (2020).
:memo: Contributions to data confidentiality in machine learning by means of homomorphic encryption. Martin Zuber. PhD Thesis at Université Paris-Saclay (2021).
Cependant, les travaux récents se tournent vers l'apprentissage homomorphe. Malheureusement, les techniques d'apprentissage homomorphe ne permettent pas aujourd'hui d'effectuer les formes d'apprentissages les plus performantes de manière efficace.
:memo: Privacy-Preserving Collective Learning With Homomorphic Encryption. Jestine Paul, Meenatchi Sundaram Muthu Selva Annamalai, William Ming, Ahmad Al Badawi, Bharadwaj Veeravalli & Khin Mi Mi Aung. IEEE Access (2021)
:memo: Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. Haokun Fang & Quan Qian. Future Internet (2021).

[148] Les algorithmes de traitement de langage modernes ne cessent de grossir en taille, à un rythme effreiné de x10 par an, depuis BERT, GPT-2, GPT-3, jusqu'aux Switch Transformers (avec mille milliards de paramètres), et bientôt Pathways.
:memo: Attention is All you Need. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser & Illia Polosukhin. NIPS (2017).
:memo: Better Language Models and Their Implications. OpenAI (2019).
:memo: Language Models are Few-Shot Learners. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei. NeurIPS (2020).
:memo: Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. William Fedus, Barret Zoph, Noam Shazeer. ArXiV (2021).
:computer: China's gigantic multi-modal AI is no one-trick pony. Engadget (2021).
:computer: Google is developing a new superintelligent AI but ethical questions remain. Quartz (2021).

> 000 lignes de codes, ou un million de paramètres. Pour décrire notre code, il faudrait mille livres ; et pour décrire notre algorithme obtenu par machine learning, il faudrait un milliard de livres[149]. Personne ne peut pleinement comprendre mille livres ; et personne ne peut ne serait-ce que survoler un milliard de livres. La sécurité de SmartPoop requiert un chantier d'ampleur astronomique.

Le rapport de SmartPoop finit par appeler à l'aide. Il demande des moyens monumentaux pour auditer, en interne et surtout en externe, les algorithmes qu'il utilise. Il reconnaît que ce travail exigera beaucoup plus de transparence de la part de l'entreprise, qui s'y engage publiquement[150].

> Des investissements massifs en termes de pédagogie sont indispensables, non seulement pour le grand public, mais aussi pour les professionnels, y compris les experts en sécurité informatique, ajoute le rapport. A minima, des centaines de milliards de dollars seront nécessaires pour obtenir un niveau de sécurité satisfaisant ; voire dix à cent fois plus encore. Tous les gouvernements et tous les partenaires économiques sont appelés à soutenir l'audit de SmartPoop. La santé, le bien-être et la sécurité de toute la population mondiale sont en jeu.

Une copie de ce rapport est envoyée à Célia, qui en profite pour remercier et féliciter Katia de cette initiative et de ces mots très forts. Célia envoie alors une version de l'article qu'elle compte publier à Katia. Cet article s'intitule « L'écologie dépend désormais de la sécurité informatique », avec comme sous-titre « Comment une vulnérabilité dans SmartPoop a failli causer un désastre environnemental[151] ». Si l'article est globalement très critique des failles de SmartPoop, il finit néanmoins avec des félicitations de SmartPoop pour son engagement pour plus de transparence, et les investissements massifs pour son audit interne et externe. « Qu'ils le veuillent ou non, SmartPoop est devenue une entreprise de sécurité informatique[152] », conclut l'article.

---

[149] En informatique, on parle là de *complexité de Solomonoff* (aussi connue sous le nom de *complexité algorithmique* ou de *complexité de Kolmogorov*) : il s'agit de la plus courte description d'un algorithme capable de résoudre une tâche donnée, comme fournir des diagnostics médicaux fiables aux milliards d'utilisateurs de SmartPoop (la difficulté étant d'être fiable pour tous).
:tv: Le rasoir d'Ockham :hot_pepper: Science4All (2020).
:tv: Le miracle de la complexité. Science4All (2019).

[150] Dans son interview dans *The Journal*, lorsque la journaliste lui demande quelle intervention serait la plus urgente à faire pour rendre Facebook plus bénéfique pour la société, la lanceuse d'alerte Frances Haugen insiste sur l'importance de la transparence, notamment sachant la difficulté énorme de modérer les discours de haine et la désinformation produits par des milliards de comptes.
:headphones: The Facebook Files, Part 6: The Whistleblower. The Journal (2021).

[151] :tv: La solution contre le changement climatique. Science4All (2018).

[152] Il s'agit là sans doute d'un aspect important dans la croissance de tout mouvement ou de toute entreprise. Plus une entité devient influente, plus il faut s'attendre à ce qu'elle attire des groupes malveillants, qui chercheront à exploiter ou à manipuler l'entité pour arriver à leurs fins. C'est pour cela qu'il est très maladroit de comparer des entités énormes comme Facebook à des petites initiatives méconnues et peu influentes. En particulier, quand une entité devient

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

énorme, pour demeurer bénéfique, il lui faut non seulement investir dans l'éthique, mais aussi se protéger des nombreux groupes malveillants qui voudront l'attaquer ou la manipuler. Tel est l'un des très grands défis de l'éthique des algorithmes et de l'information.
:books: Le fabuleux chantier. Rendre l'intelligence artificielle robustement bénéfique. Lê Nguyên Hoang & El Mahdi El Mhamdi. EDP Sciences (2019).

Chapitre 5 — Les FakePoops

> Les données sont encore très parcellaires, mais il y a de quoi être extrêmement inquiet, déclare le Premier Ministre Kormicain. Selon nos épidémiologistes, nous faisons probablement face à un retour du ROVID-19, possiblement un variant plus contagieux encore, plus difficile à détecter via SmartPoop. Avec désormais plusieurs centaines de cas identifiés, pour protéger toute la population, nous avons décidé d'imposer un confinement obligatoire. Seules les industries indispensables sont autorisées à rester ouvertes. Toutes les autres se doivent d'être en télétravail, ou de cesser leur activité. Dans le second cas, nous vous invitons à déclarer votre cessation d'activité sur le site web officiel du gouvernement, et à fournir des justificatifs, pour être compensés. Le confinement entrera en vigueur demain soir à 20h. Mais dès aujourd'hui, nous vous prions de prendre soin de vous, de vos proches et de vos voisins. Restons unis face au ROVID-19.

Cela fait une semaine que SmartPoop a détecté un premier cas de ROVID-19 en Kormique, en mai 2026. Malheureusement, le nombre de ces cas n'a cessé d'augmenter depuis. La crainte d'un retour aux confinements interminables de 2020 règne alors. Il n'a d'ailleurs pas fallu attendre le discours du Premier Ministre pour que les magasins se mettent à être dévalisés. Les autoroutes du pays sont toutes paralysées, alors que la population urbaine fuit les grandes villes[153]. Les services de nombreuses entreprises sont interrompus.

Le chaos

Malheureusement, plus les semaines passent, plus le chaos augmente. SmartPoop rapporte les décès de nombreuses personnes infectées, et le nombre de cas de ROVID-19 poursuit sa croissance exponentielle. Étrangement, toutefois, lors des semaines qui suivent, il n'y a pas d'hospitalisations de cas de ROVID-19, ce qui

[153] De tels exodes ont été observés en France pendant la pandémie COVID-19.
:memo: Urban exodus and the dynamics of COVID-19 pandemics. Gérard Weisbuch. Physica A: Statistical Mechanics and its Applications (2021).

soulève des hypothèses de morts violentes et soudaines, peu de temps après les premiers symptômes. Malheureusement, les médecins et épidémiologistes peinent à récupérer des données sur ces victimes du ROVID-19, si ce n'est via l'application SmartPoop. Cependant, notamment suite à l'implémentation du chiffrement homomorphe et des nouvelles mesures de protection des données personnelles, les statistiques individuelles des victimes ne peuvent pas être obtenues — y compris par SmartPoop[154].

Dans un contexte de confusion énorme, le gouvernement penche vers davantage de mesures de sécurité, en demandant à la police de patrouiller les lieux publics et d'arrêter tout individu se déplaçant sans les justificatifs adéquats. Cependant, les arrestations maladroites se multiplient. Des abus policiers sont rapportés par des témoins distants et des vidéos partagées sur les réseaux sociaux. Cependant, ces mêmes vidéos peinent à être authentifiées, et sont donc suspectées d'être des *deepfakes*. Un chaos informationnel émerge[155].

Sur les réseaux sociaux, c'est la cacophonie. Certains reprochent au gouvernement de cacher les cadavres. D'autres prétendent que le ROVID-19 n'existe pas, voire qu'il s'agit d'un canular du gouvernement pour contrôler la population. Quoi qu'il en soit, les cours boursiers sont en chute libre, et de nombreux étudiants qui n'ont pas pu effectuer d'exode urbain n'ont plus de quoi s'acheter à manger[156].

---

[154] Des applications comme WhatsApp, Signal ou Telegram permettent un chiffrement end-to-end, ce qui signifie que même ces applications n'ont pas accès aux messages échangés. Voilà qui garantit la protection des données personnelles contre ces organisations. Cependant, ceci signifie aussi qu'il est impossible pour ces applications d'effectuer de la modération de contenus, comme du cyber-harcèlement, des discours de haine, de la désinformation, des spams, de la pédophilie et des arnaques. Or, selon la loi de la plupart des démocraties, ces services ont le devoir de modérer ces contenus, et de rapporter à la police les utilisateurs qui commettent l'infraction de produire et partager ces contenus. D'ailleurs, l'entreprise ProtonMail a transmis à la police française les adresses IP de comptes suspectés d'avoir commis de telles infractions.
:computer: ProtonMail transmet des adresses IP à la police : 4 questions pour comprendre la polémique. Numerama (2021).
De même, Apple a mis en place des systèmes cryptographiques pour modérer la production d'images pédopornographiques de ses utilisateurs.
:computer: Here's Why Apple's New Child Safety Features Are So Controversial. The Verge (2021).
Selon RAINN, l'abus sexuel de mineurs serait très fréquent, puisqu'il affecterait une nouvelle victime toutes les 9 minutes.
:computer: Children and Teens: Statistics. RAINN.
Alors, techniquement, une organisation n'a ce devoir que si elle *sait* que ses utilisateurs commettent ses infractions ; cependant, il semble immoral pour ses organisations de faire des efforts pour ne pas le savoir. Globalement, il semble y avoir encore un flou juridique autour de cette tension entre la protection des données personnelles et la modération de contenus illégaux.

[155] Le mot « infodémic » a été utilisé par l'Organisation Mondiale de la Santé (OMS) pour décrire de tels chaos informationnels.
:computer: Infodemic. World Health Organization (2020).
:memo: Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Riccardo Gallotti, Francesco Valle, Nicola Castaldo, Pierluigi Sacco & Manlio De Domenico. Nature Human Behaviour (2020).

[156] La pandémie a causé beaucoup de précarité chez les étudiants, notamment ceux qui vivaient de « petits boulots » supprimés en période de confinement, comme les métiers de serveurs ou

Des associations de dons de nourriture se mettent en place, mais les mesures en vigueur rendent leur officialisation difficile.

Bientôt, la cacophonie s'empare de la rue. Des manifestations sont organisées sur les réseaux sociaux, et conduisent à des débordements[157]. De nombreuses altercations violentes avec les services de l'ordre ont alors lieu, et sont partagées massivement sur Twitter. Cependant, les revendications des manifestants sont aussi confuses que la situation sanitaire. Certains exigent plus de transparence du gouvernement, d'autres réclament des aides financières pour les démunis, d'autres encore crient au complot gouvernemental, couplé avec de l'incompétence des dirigeants[158]. Ces derniers appellent alors à cesser le confinement et les autres mesures liberticides.

L'économie est heurtée de plein fouet. Au fil des semaines, malgré les aides gouvernementales, de plus en plus de petits commerces sont contraints de déclarer faillite[159], tandis que les assureurs souffrent à leur tour et ne parviennent pas à payer leurs dettes[160]. De façon plus spectaculaire encore, alors que les transports sont immobilisés, les prix du pétrole chutent jusqu'à devenir négatifs[161], et les industries pétrolières Kormicaines demandent d'urgence des subventions publiques[162]. À l'étranger en particulier, de nombreux clients de ces industries annoncent de nouveaux investissements vers l'adoption de technologies compatibles avec les énergies renouvelables des concurrents Bokistanais. Voilà qui refroidit les investisseurs pétroliers, et qui pousse certaines industries pétrolières Kormicaines à la faillite. Des centaines de milliers d'employés se retrouvent soudainement au chômage.

Les irrégularités des ROVID-positifs

Cependant, pendant ce temps, les hôpitaux demeurent étonnamment vides. Même les morgues ne semblent pas se remplir de victimes du ROVID-19. De plus

---

de caissiers.
:tv: Covid-19 : comment la précarité frappe les étudiants. Le Monde (2021).

[157] À travers le monde, des manifestations contre les mesures sanitaires face au COVID-19 ont débordé.
:computer: Manifestation mouvementée contre le couvre-feu à Montréal. Radio Canada (2021).
:computer: Plus de 200 000 personnes manifestent contre le passe sanitaire. Le Monde (2021).
:computer: Protests over responses to the COVID-19 pandemic. Wikipedia (2021).

[158] On retrouve ces confusions dans de nombreuses manifestations populaires.
:tv: L'hooliganisme politique a gâché ma marche pour les sciences. Science4All (2017)
:computer: Pass sanitaire : les manifestations soutenues par des personnalités d'extrême-droite et d'extrême-gauche. Actualités Orange (2021).
:books: Twitter and tear gas: The Power and Fragility of Networked Protest. Zeynep Tufekci. Yale University Press (2017).

[159] :tv: Le plan pour sauver les entreprises - Heu?reka (2020).

[160] :tv: Les Subprimes 1ère Partie : La boulette ! - Heu?reka (2017).

[161] Ce phénomène est arrivé pendant la crise du COVID-19 aussi.
:tv: Prix négatifs du pétrole - Heu?reka (2020).

[162] Les subventions publiques des industries pétrolières demeurent très importantes.
:computer: America spends over $20bn per year on fossil fuel subsidies. Abolish them. The Guardian (2018).

en plus d'épidémiologistes soulèvent l'étrangeté de la situation, et suggèrent de plus en plus un dysfonctionnement de SmartPoop. Pourtant, les audits internes et externes de sécurité informatique de SmartPoop ne relèvent encore aucune faille de sécurité.

Toutefois, certains observateurs font remarquer que ces audits sont freinés par le chiffrement homomorphes. « Ce chiffrement est une bêtise. Il nous met dans le noir. Il rend impossible la détection d'anomalies dans les données, puisque ces données sont chiffrées et ne peuvent être utilisées que pour mettre à jour les algorithmes de SmartPoop, mais sans aucun contrôle de qualité possible[163] », affirme un expert en algorithmique, proche du gouvernement Kormicain.

Finalement, sous la pression du gouvernement Kormicain, Katia et Marc décident d'interrompre momentanément le chiffrement homomorphe, si l'utilisateur y consent. « Les données sont encore chiffrées entre l'utilisateur et SmartPoop. Mais avec le nouveau système, SmartPoop pourra déchiffrer les données, et les analyser, pour vérifier l'absence de dysfonctionnement dans notre système », précise Katia dans une interview télévisée.

Petit à petit, suite à des vagues de communications du gouvernement Kormicain, de plus en plus de Kormicains consentent à partager leurs données SmartPoop avec l'entreprise SmartPoop[164]. Étrangement, pendant longtemps, aucun cas de ROVID-19 n'est observé parmi ces données. Ce n'est qu'une semaine plus tard, alors que la moitié des Kormicains partagent désormais leurs données SmartPoop avec l'entreprise SmartPoop, que les premiers cas de ROVID-19 sont détectés

[163] Ce que souligne l'expert ici est le dilemme sécurité-privacy. Intuitivement, pour garantir autant que possible la sécurité, il est nécessaire de pouvoir tout vérifier dans un système, y compris (et surtout) les données d'entraînement des algorithmes. Voilà qui exige la *transparence* du système. Cependant, la privacy exige au contraire l'opacité de certains composants du systèmes. On retrouve ce dilemme par exemple dans le problème de l'adoption des technologies de notifications à l'exposition.
:tv: La notification à l'exposition sans surveillance (DP3T). Science4All (2020).
:tv: Applis Covid : dangereux pour la vie privée ? Philoxime (2020).
En particulier, le chiffrement homomorphe naïf ne semble pas pouvoir se combiner aisément avec les solutions d'apprentissage robustes aux fausses données, aussi connues sous le nom d'apprentissage Byzantin (notamment car le chiffrement homomorphe naïf ne permet pas des opérations autres que l'addition et la multiplication). Des techniques alternatives comme la confidentialité différentielle (*differential privacy*) se marrient également mal avec l'apprentissage Byzantin.
:memo: Differential Privacy and Byzantine Resilience in SGD: Do They Add Up? Rachid Guerraoui, Nirupam Gupta, Rafaël Pinot, Sébastien Rouault & John Stephan. PODC (2021).
Plus de recherches sont nécessaires pour mieux comprendre le tradeoff entre sécurité et privacy en machine learning. Mais il n'y aura sans doute pas de magie : il persistera une tension fondamentale entre ces deux propriétés désirables.

[164] Dans un genre similaire, en 2020, la *Mozilla Foundation* a lancé un programme participatif où les volontaires pouvaient partager leurs données YouTube pour permettre aux chercheurs de mieux comprendre l'algorithme de recommandation de YouTube. On a là encore un cas de dilemme sécurité-privacy. Tant que les données des utilisateurs et des algorithmes de YouTube resteront privés, il sera très difficile de comprendre les dynamiques de comportements des utilisateurs sur YouTube, et de prendre les mesures adéquates pour réduire le cyber-harcèlement, les appels à la haine et la désinformation.
:memo: YouTube Regrets. Mozilla Foundation (2021).

dana ces données de volontaires.

Katia mène elle-même l'analyse de ces cas, et Marc assiste à cette analyse. Marc demande alors :

> Est-ce que ces cas de ROVID-19 sont normaux ?
>
> À première vue, oui, répond Katia. Ils ne semblent pas sortir de la distribution statistique des ROVID-19 qu'on a observés en 2020.
>
> Pourquoi seraient-ils plus violents, à tel point que les victimes meurent chez elles ?
>
> Je ne sais pas.
>
> As-tu l'adresse des victimes, pour qu'on puisse les consulter, ou effectuer des analyses post-mortem ?
>
> Laisse-moi les imprimer.

Katia lance l'impression des analyses, et récupère les copies fraîchement sorties de la photocopieuse.

> Étrange, dit Katia. Aucun de ces comptes n'a renseigné son adresse, si ce n'est de façon très vague, avec juste l'adresse mais pas le numéro de l'appartement.
>
> On peut quand même envoyer des gens à l'adresse ?
>
> Bizarre, je ne trouve pas ces adresses sur Google Maps...
>
> Que se passe-t-il ?
>
> Laisse-moi voir leur date de création. Oui, c'est très bizarre ! Tous ces comptes ont été créés il y a moins de 7 mois. Ça pue les faux comptes[165] !
>
> Mais qui aurait pu créer ces faux comptes et pourquoi ?
>
> Je ne sais pas, répond Katia.
>
> À qui profite cette alerte de ROVID-19 ?
>
> Euh...
>
> Au Bokistan bien sûr, s'exclame Marc. Ou au moins aux industries d'énergie renouvelable bokistanaises[166].

[165] Ces faux comptes sont omniprésents sur Internet, et sont souvent contrôlés par des groupes stratégiques, avec leurs propres intentions. À titre d'exemple, en 2021, le New York Times a révélé une campagne de désinformation de Huawei s'appuyant sur des faux comptes Twitter. :computer: Inside a Pro-Huawei Influence Campaign. The New York Times (2021).

[166] Cette pratique semble très présente, notamment dans les pays autoritaires. Un rapport de l'Institut de Recherche Stratégique de l'École Militaire (IRSEM) en France fait état de 20 millions de chinois payés pour produire de la désinformation, dont 2 millions à temps plein. :memo: Les opérations d'influence chinoises. un moment machiavélien. ISEM (2021)

Mais donc, tu penses qu'ils ont créé des faux profils ?

Probablement.

Et qu'ils nous ont ensuite envoyé des fausses données d'excréments ? Des... FakePoops[167] ?

Oui, probablement générés à partir des données publiques sur le profil scatologique du ROVID-19, et peut-être même en particulier des Kormicains.

C'est brillant quand on y réfléchit. En renseignant des informations typiques des Kormicains, ils ont inséré des données qui ont été interprétées par notre système comme une nouvelle pandémie de ROVID-19 !

L'espionnage par FakePoops

Après quelques minutes de réflexion, Katia en vient à une seconde épiphanie.

En fait, ces FakePoops sont incroyablement dangereux. Je parie que c'est également ainsi que les industries Kormicaines ont découvert les secrets de la femme du Président Lartan !

Comment ça ? Explique.

Il suffit d'envoyer des FakePoops avec des profils très proches des déjections de la femme du Président. Comme SmartPoop s'appuie sur les données SmartPoop de tous les utilisateurs pour effectuer des diagnostics sur un échantillon d'excrément, si le profil le plus proche de l'échantillon d'excrément en question est celui de la femme du Président, alors SmartPoop pourrait fournir des diagnostics proches, voire identiques, aux diagnostics donnés à la femme du Président !

[167] On parle alors d'attaques par données empoisonnées (*data poisoning*). Ces attaques consistent à injecter de fausses données dans la base de données d'entraînement des algorithmes pour les amener à apprendre et à conclure des choses erronnées.
:tv: Hacker l'IA (ft. El Mahdi El Mhamdi). Science4All (2018).
L'attaque est ici simple : elle consiste simplement à lancer une alerte. Mais on peut imaginer des attaques plus sophistiquées, notamment pas « *backdoor* », qui conduirait l'algorithme à systématiquement se tromper pour certaines données.
:memo: Backdoor attacks against learning systems. Yujie Ji, Xinyang Zhang & Ting Wang. CNS (2017).
La défense contre le data poisoning semble surtout devoir s'appuyer sur la théorie de l'apprentissage Byzantine.
:memo: An equivalence between data poisoning and Byzantine gradient attacks. Anonymous authors. OpenReview (2021)
Cette théorie a montré des résultats positifs, avec des algorithmes disposant de garanties théoriques, mais aussi des résultats négatifs, avec des théorèmes d'impossibilité sur ce que les algorithmes peuvent garantir, notamment dans le cas où les données des utilisateurs sont hétérogènes, ce qui est clairement le cas quand il s'agit d'excréments, puisque les excréments de deux utilisateurs différents sont clairement très différents.
:memo: Collaborative learning in the jungle. El-Mahdi El-Mhamdi, Sadegh Farhadkhani, Rachid Guerraoui, Arsany Guirguis, Lê Nguyên Hoang & Sébastien Rouault. NeurIPS (2021).

Je ne suis pas sûr de comprendre.

Ils ont détecté quelque chose de similaire dans les algorithmes de langage, comme les algorithmes d'autocomplétion des claviers de téléphones. L'autocomplétion de ton téléphone s'appuie sur la manière dont tu tapes sur ton clavier, mais aussi sur la manière dont d'autres gens tapent sur leur clavier. Ainsi, si tu tapes exactement comme la femme du Président, tu vas avoir les mêmes autocomplétions que la femme du Président, qui correspondent à la manière dont elle tape sur son clavier[168]. Et si tout à coup, tu tapes "mon diagnostic SmartPoop de l'autre jour disait que", et si la femme du Président a précédemment tapé une telle phrase, alors tu pourrais avoir l'autocomplétion qui correspond mot pour mot à ce que la femme du Président avait tapé.

Attends, t'es en train de dire que SmartPoop ne parvient pas à distinguer les utilisateurs à qui il parle ?

C'est comme ça qu'il est entraîné. Souviens-toi, on apprend un modèle global pour tous les utilisateurs. Puis, ce modèle est ajusté aux données personnelles de chaque utilisateur. Mais si deux utilisateurs ont quasiment les mêmes données personnelles, alors ils vont avoir des diagnostics semblables.

Oui c'est le principe même de la médecine, non ? Voire de la science.

Oui, mais du coup, tu peux espionner les diagnostics SmartPoop d'un individu, en obtenant des données personnelles semblables[169].

Mais comment les espions pourraient-ils créer un profil similaire à celui de la femme du Président ?

Il suffit que la femme du Président ait été invitée à dîner chez l'espion, et qu'elle soit passée par ses toilettes. Ou mieux encore, que

[168]Ces algorithmes de traitement langage sont parfois qualifiés de « perroquets stochastiques », car ils apprennent à répéter le genre de phrases présentes dans leurs données d'entraînement.
:memo: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Emily Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. FAccT (2021).
On le voit particulièrement bien dans le cas des théories du complot, que ces algorithmes peuvent parfaitement répéter. Plus surprenant encore, la manière dont ils sont amenés à parler de ces théories dépend non seulement de la manière dont on leur demande d'en parler, voire, de façon plus subtile encore, de l'historique de la conversation, comme le montre l'article suivant.
:memo: The Radicalization Risks of GPT-3 and Advanced Neural Language Models. Kris McGuffie, Alex Newhouse (2020).

[169]On parle parfois d'attaques par extraction d'information. Les algorithmes de modernes mémorisant leurs données d'entraînement, ils sont en fait très vulnérables à ce genre d'attaque. En fait, il peut suffire de leur demander littéralement « quelle est l'adresse de Monsieur X » pour qu'ils révèlent une information sensible.
:memo: Extracting Training Data from Large Language Models. Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea & Colin Raffel. USENIX (2021)

les toilettes de l'hôte aient été modifiées par un espion, pour collecter des échantillons de ses excréments !

Ah oui ! Il faut vraiment faire gaffe à là où on chie ! Et donc, ça serait ça la vulnérabilité ? Ça serait une histoire de machine learning, pas de sécurité informatique classique ?

Dur à dire si c'est la seule. Mais c'est en tout cas une grosse vulnérabilité. Il faut qu'on voit comment on peut la patcher[170].

Il faut peut-être commencer par alerter les autorités Kormicaines ?

Ah oui, j'allais oublier. Par contre, on risque d'avoir un nouveau procès aux fesses.

Katia, des centaines de millions de vies sont chamboulées à cause de nous.

Oui, oui, je sais.

L'OMESA

Le lendemain, après avoir été prévenu par Katia et Marc, le Premier Ministre Kormicain annonce que la pandémie de ROVID-19 avait été fabriquée par une campagne de désinformation, probablement Bokistanaise, qui a exploité des vulnérabilités de SmartPoop. Il annonce surtout la fin du confinement, et le retour à la normale pour le pays.

Le Bokistan nie toutefois avoir été impliqué dans cette attaque, et insiste sur l'impossibilité de retrouver les coupables. « La démocratisation des technologies de cyber-attaques, la complexité des systèmes informatiques modernes, et leur place devenue centrale rendent notre monde moderne extrêmement vulnérables à des groupuscules malveillants », explique le nouveau Président Bokistanais. « J'appelle tous les pays du monde, y compris nos amis Kormicains, à bannir ces technologies et toutes les entreprises qui développent de telles technologies, ainsi qu'à investir massivement dans des technologies de cyber-défense, et non pas de cyber-attaque. La sécurité et le bien-être de l'ensemble de la population mondiale est en jeu[171] ».

[170] L'état de l'art du machine learning semble aujourd'hui très incapable de protéger les gros modèles, notamment les algorithmes de traitement du langage, de telles vulnérabilités.

[171] De façon surprenante, en 2020, Vladimir Poutine a appelé à un accord pour limiter les cyber-attaques entre différents pays, alors même que la Russie a été reconnue comme l'auteur de beaucoup de ces attaques. Ceci suggère que même les auteurs de ces attaques sentent qu'elles finiront par tout déstabiliser, y compris eux-mêmes.
:computer: Putin says Russia and U.S. should agree not to meddle in each other's elections. Reuters (2020).
Pour la sécurité de tous, il est sans doute urgent d'établir une convention internationale qui, à l'instar des armes chimiques, biologiques et autonomes, interdirait aussi les cyber-attaques, tant celles-ci mettent en danger des millions de vies, voire des milliards de vies à travers le monde.
:tv: Pourquoi faut-il bannir les armes autonomes ? | The Flares (2019).

Dans les mois à venir, une nouvelle organisation internationale est lancée, sous le nom d'Organisme Mondial de l'Éthique et de la Sécurité des Algorithmes (OMESA). L'une des missions principales de l'OMESA est alors d'organiser, chaque année, un Groupe International d'Étude sur la Sécurité et l'Éthique du Numérique (GIESEN). Inspiré du GIEC[172], ce groupe d'études vise à faire un point, chaque année, sur les grandes menaces de sécurité informatique et les grands dilemmes algorithmiques à résoudre, et à faire le point sur la recherche, le développement et le déploiement des solutions proposées. Plus que jamais, la paix mondiale semble dépendre de cet énorme effort de gouvernance.

Sans surprise, le premier rapport liste le problème des FakePoops comme étant l'une des plus grandes menaces pour le futur, et demande une audition publique de SmartPoop, et en particulier de Katia.

L'audition de Katia

Interrogée de cinq figures d'autorité sélectionnée par l'OMESA, Katia fait face aux caméras de télévision du monde entier. À l'écran, Katia est visiblement terrifiée.

> Pouvez-vous vous présenter ?
>
> Bonjour mesdames et messieurs de l'OMESA. Je m'appelle Katia Crapinski. Je suis Docteure en informatique, et désormais Présidente et Directrice Générale de l'entreprise SmartPoop.
>
> Bonjour Docteure Crapinski, ici Professeure Wang, chercheuse en machine learning. Pouvez-vous présenter le problème des FakePoops ?
>
> Professeure Wang, oui c'est une vulnérabilité qu'on n'a découvert que récemment, suite à des attaques par FakePoops. En gros, il s'agit d'entités malicieuses qui cherchent à empoisonner la base de données avec des fausses données. Mais alors, nos algorithmes, qui apprennent de cette base de données, ces algorithmes vont être à leur tour empoisonnés, ce qui peut les conduire à faire des prédictions imparfaites.
>
> Imparfaites ? Vous voulez dire dangereuses ? Ces prédictions sont même décidées par les entités malicieuses, si celles-ci fabriquent leurs fausses données adéquatement.
>
> Professeure Wang, oui, c'est une façon de voir les choses.
>
> Je vais vous dire comment il faut voir les choses. Vous avez conçu un produit et l'avez déployé à grande échelle, sans jamais vous soucier des abus et des réutilisations de ce produit à des mauvaises fins.

[172] Cette excellente vidéo décrit le fonctionnement du GIEC. La gouvernance des algorithmes gagnerait probabilité à s'en inspirer.
:tv: Comprendre le GIEC et ses rapports. Le Réveilleur (2021).

> Vous avez mis le monde entier en danger en poussant des milliards d'utilisateurs à utiliser vos produits, parce que ça faisait grossir vos chiffres d'affaires. Mais vous ne vous êtes jamais dit que vos produits pouvaient être une vulnérabilité pour la sécurité de tous les pays à travers le monde, et pouvaient être exploités par quiconque pour conduire à des millions de morts ! Vous avez été irresponsable[173]. Qu'avez-vous à dire pour vous justifier ?
>
> Professeure Wang, je... comment dire ? C'est très difficile...
>
> En 2019 seulement, Facebook a retiré 6 milliards de faux comptes de leurs plateformes[174]. En l'absence d'importants investissements dans la détection de faux comptes, l'écrasante majorité des comptes sur Facebook seraient des faux comptes. Chez SmartPoop, avez-vous des équipes de détection de faux comptes, et de leur impact sur les diagnostics SmartPoop ? Combien de faux comptes retirez-vous par an ?
>
> Professeure Wang... on... nous... C'est... je ne sais pas...
>
> Quantifiez-vous la vulnérabilité de SmartPoop ? Avez-vous une estimation du nombre de décès potentiellement causés par les FakePoops ?
>
> Professeure Wang... je... non...

À l'écran, Katia est toute pâle et en sueur. Elle est paralysée par ces accusations, et tremble en saisissant son verre d'eau[175]. Ces images terribles feront le tour du

[173]Ces critiques semblent largement s'appliquer aux grandes entreprises du numérique également. Comme le révèlent les *facebook files*, en 2018, un nouvel algorithme de fils d'actualité de Facebook a été déployé de manière précipitée (et avec des fausses justifications publiques), ce qui a conduit à une viralité accrue des discours de haine et de la mésinformation.
:headphones: The Facebook Files, Part 4: The Outrage Algorithm. The Journal (2021).
Plus récemment, Google semble vouloir déployer des algorithmes de traitement du langage très sophistiqués, mais dont la sécurité est extrêmement critiquée.
:tv: Google démantèle son éthique (et tout le monde s'en fout...). Science4All (2021).
Et pour cause, certains théorèmes d'impossibilité de sécurité semblent largement s'appliquer à ces algorithmes.
:memo: Collaborative learning in the jungle. El-Mahdi El-Mhamdi, Sadegh Farhadkhani, Rachid Guerraoui, Arsany Guirguis, Lê Nguyên Hoang & Sébastien Rouault. NeurIPS (2021).
Pire, Google semble vouloir concevoir un algorithme plus sophistiqué encore, appelé *Pathways*, qui répondrait à de nombreux services différents de Google, de Google Search aux recommandations YouTube, en passant par GMail, GDrive et GBoard.
:computer: Google is developing a new superintelligent AI but ethical questions remain. Quartz (2021).

[174]:computer: Facebook has shut down 5.4 billion fake accounts this year. CNN Business (2019).
:computer: Meet the A.I. that helped Facebook remove billions of fake accounts. Fortune (2020).

[175]Cette scène fait référence notamment à l'audition de Mark Zuckerberg, Jack Dorsey et Sundar Pichai, les PDG de Facebook, Twitter et Google, par le Congrès américain.
:tv: Republican Senator GRILLS Zuckerberg on Facebook, Google, and Twitter collaboration. CNET Highlights (2020).

monde, et deviendront des gifs abondamment partagés sur les réseaux sociaux.

La Proof of Personhood

L'audition se poursuit avec une longue discussion sur les conséquences sociétales des FakePoops. Plus d'une heure plus tard, et après une pause, les membres de l'OMESA en viennent enfin à parler de solutions potentielles. Katia, qui s'est mise de l'eau au visage pendant la pause, semble s'être un peu ressaisie.

> Bonjour Docteure Crapinski. Ici le Professeure Ferpo, chercheuse en sécurité informatique. Il me semble que vous faites face à un problème similaire à celui des crypto-monnaies, qui cherchent à se protéger de hackers malveillants. Je me trompe ?
>
> Professeure Ferpo, en effet, on a un problème similaire à celui des crypto-monnaies. En fait, le problème est très général. C'est ce que les économistes appelleraient la tragédie des communs[176]. Certaines activités ne sont possibles que si l'on collabore ensemble pour les réaliser. Mais tout projet collaboratif est a priori vulnérable à des attaques par des faux comptes, qui se font passer pour des contributeurs authentiques, mais dont le but est davantage de faire échouer le projet, ou de le biaiser dans une certaine direction[177]. On en a été victime.
>
> Est-ce qu'on ne peut pas s'inspirer d'elles, et par exemple, de la *Proof of Work*[178] ?
>
> Professeure Ferpo, oui on y a pensé. On a parlé de *Proof of Flush* dans le cas de SmartPoop. En gros, chaque utilisateur devait demander à sa machine de résoudre un problème difficile pour authentifier sa contribution à la base de données SmartPoop. Cependant, cette solution est extrêmement polluante, et elle n'est pas si sécurisée. Pour vraiment sécuriser la base de données SmartPoop, et ainsi sécuriser tous les algorithmes qui s'appuient sur cette base de données, il nous faut des solutions plus sécurisées.
>
> Comme quoi ?
>
> Professeur Ferpo, l'idéal serait une *Proof of Personhood*, qui consiste à associer à chaque individu sur terre une identité numérique unique,

[176]:tv: What is the tragedy of the commons? - Nicholas Amendolare. TED-Ed (2017).

[177]On parle parfois de *free riders*.
:tv: La morale des hooligans (LA NÔTRE !!). Science4All (2017).

[178]La *Proof of Work* est une tâche qu'une machine doit résoudre pour obtenir des droits, comme celui d'envoyer un email, ou celui d'écrire des transactions dans la Blockchain d'une cryptomonnaie.
:tv: Le Bitcoin et la Blockchain (avec Heu?Reka). Science Étonnante (2016).
:tv: Devenir riche grâce au minage des Bitcoins c'est possible ? - CRYPTO #06 - String Theory (2018).

vérifiable et sécurisée[179]. C'est d'ailleurs le sujet que j'aimerais inviter l'OMESA à investiguer. La sécurité de toutes les plateformes collaboratives et participatives, comme les réseaux sociaux, les cryptomonnaies et SmartPoop, me semble nécessiter inéluctablement une *Proof of Personhood*. En fait, en démocratie, la Proof of Personhood via les cartes électorales, c'est exactement ce qui permet de garantir le principe « un citoyen, une voix[180] ».

Bonjour Docteure Crapinski. Je suis Professeure Abdoul, chercheuse en géopolitique. Je ne suis pas sûre de comprendre. Est-ce que cette *Proof of Personhood* correspondrait à l'authentification de pièces d'identité ?

Professeure Abdoul, oui un peu. Mais il faudrait parvenir à faire mieux encore que les pièces d'identité classiques. Et même beaucoup mieux. Dans beaucoup de pays, certaines personnes n'ont pas de pièces d'identité. Dans d'autres, les citoyens sont identifiés par des proxys, comme des numéros de sécurité sociale, qui ne sont absolument pas sécurisés, faciles à deviner pour un attaquant, et constamment volés par des entités malveillantes[181]. Pire encore, il y a des pays où le gouvernement n'est pas digne de confiance pour gérer la distribution de pièces d'identité. Enfin, dans la plupart des cas, il est aujourd'hui possible pour des entités malveillantes de créer des fausses pièces d'identité ou des faux passeports, typiquement pour les espions qui cherchent à voyager incognito, ou pour les jeunes qui souhaitent consommer de l'alcool. Il nous faut un système beaucoup plus fiable, robuste et accessible d'authentification individuelle. Il doit devenir possible pour n'importe quel projet participatif de vérifier que chaque participant est un humain, et qu'aucun humain ne peut personnifier deux participants différents. C'est ce problème que le Proof of Personhood doit résoudre.

Bonjour Docteure Crapinski. Ici Professeur Smith, docteur en santé publique. Est-ce que les identifiants biologiques, comme l'image de l'iris, les empreintes digitales ou les compositions fécales, peuvent résoudre le *Proof of Personhood* ?

Professeur Smith, je pense que c'est un composant utile d'un système de *Proof of Personhood*. Mais il est loin d'être suffisant. L'un des problèmes est que, avec les *DeepFakes* et les *FakePoops*, il est facile de créer des nouveaux identifiants biologiques purement fabriqués.

[179]:memo: Proof-of-Personhood: Redemocratizing Permissionless Cryptocurrencies. Maria Borge, Eleftherios Kokoris-Kogias, Philipp Jovanovic, Linus Gasser, Nicolas Gailly & Bryan Ford. EuroS&P Workshops (2017).

[180]Ce principe est d'ailleurs aux fondements du projet Tournesol.
:tv: L'attaque des 51%. Science4All (2021).

[181]:tv: Social Security Cards Explained. CGP Grey (2017).

Par ailleurs, ces données peuvent être volées[182]. Or, une personne dont les identifiants biologiques ont été volés ne peut pas en créer de nouveau. Les identifiants biologiques, ce n'est pas comme les mots de passe. Bref. Il nous faut d'autres solutions aussi.

Bonjour Docteure Crapinski. Ici Professeur Raul, psychologue. Aujourd'hui, l'identité numérique repose en fait plus sur la mémoire, via les mots de passe. Est-ce fiable ?

Professeur Raul, vous le savez sans doute mieux que moi. Mais en fait, de nos jours, ce n'est plus tant la mémoire des utilisateurs qui leur sert d'identifiant ; c'est davantage la mémoire de leurs outils numériques, comme leurs téléphones[183]. Et ça, c'est aussi très dangereux. Il suffit d'un *spyware* pour voler de tels mots de passe[184].

Mais donc, comment on fait ?

Professeur Raul, il faut faire ce que les États ont cherché à faire pour donner un et exactement un droit de vote à chaque membre de leurs pays. Il faut faire des recensements précis, et réguliers pour résoudre la *Proof of Personhood*, via des organismes indépendants, auditables et audités. Et il faut les faire régulièrement, peut-être tous les ans, pour restaurer des identités numériques volées ou perdues[185].

Vous parlez d'un chantier monumental !

Professeur Raul, oui. SmartPoop a plus besoin de vous, que vous n'avez besoin d'auditer SmartPoop, je pense. Et ce, même s'il y a un besoin énorme à auditer SmartPoop. Le futur du monde numérique dépend de votre capacité à instaurer un *Proof of Personhood*[186]. Et

[182]:tv: The Most Horrific Case Of Identity Theft. The Infographics Show (2019).

[183]Le philosophe Michel Serres aimait insister sur l'impact des technologies de l'information, comme le papier, l'imprimerie ou les ordinateurs, sur *l'externalisation* de notre cognition. Il est ainsi remarquable de constater à quel point nos boîtes emails sont parvenues à externaliser une grosse partie de notre mémoire. En un sens, nos boîtes emails nous connaissent beaucoup mieux que nous nous connaissons nous-mêmes, non pas car elles sont « intelligentes », mais simplement car leur mémoire est beaucoup plus fiable que la mémoire humaine, et car rechercher dans ces boîtes emails est beaucoup plus efficace que rechercher dans ntore mémoire.
:tv: Michel Serres - Les nouvelles technologies : révolution culturelle et cognitive. I Moved to Diaspora (2012).

[184]Le cas de l'affaire Pegasus montre l'ampleur de telles attaques par espionnage des dirigeants politiques dans le monde moderne. Pegasus est un *spyware*, c'est-à-dire un algorithme, qui peut être utilisé pour infecter un téléphone cible, et surveiller tout ce que ce téléphone fait. Pegasus est développé par le groupe NSO en Israël, et l'on sait que de nombreux services de renseignement à travers le monde l'ont utilisé pour espionner de nombreux journalistes, militants et dirigeants politiques.
:tv: Comment PIRATER Jeff Bezos avec un même ? (histoire vraie). Micode (2021).

[185]:memo: Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood. Bryan Ford (2020).

[186]Une *Proof of Personhood* permettrait également d'instaurer facilement un revenu universel, notamment dans une cryptomonnaie. En effet, il suffirait de considérer que, à chaque instant, chaque compte doté d'une *Proof of Personhood* reçoit une certaine somme fixe d'argent, créée

le futur du monde entier dépend de ce monde numérique. Tant qu'on n'y sera pas, SmartPoop restera vulnérable aux *FakePoops*.

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

---

de nulle part.
:tv: Le revenu de base - Heu?reka (2016).
:tv: Et si l'argent tombait du ciel ? - Heu?reka (2020).

Chapitre 6 — Le marché marron

Frédéric Partoli met un pied à terre. Terriblement frustré, il sort du terrain, la main sur le cœur, lequel bat très fort. Il est accompagné directement à l'hôpital.

À l'hôpital, le médecin lui explique que Frédéric souffre d'une tachycardie[187], causée par une anémie[188]. Frédéric le sait. Depuis quelques jours, SmartPoop lui a annoncé son déficit en fer[189]. SmartPoop lui a conseillé d'éviter toute activité sportive. Mais le football était trop tentant pour Frédéric. Il a préféré jouer et s'y risquer.

Le médecin lui recommande alors de manger des lentilles vertes, des graines de sésame, du chocolat noir, des kiwis, des épinards, mais aussi du boudin, du foie, ou de la viande rouge. Il précise toutefois que la viande rouge doit être consommée avec modération. Enfin, il ajoute qu'il est important d'accompagner cela avec des apports en vitamine C, via des agrumes par exemple. Encore une fois, ces informations ne sont pas vraiment des nouvelles pour Frédéric. SmartPoop lui faisait les mêmes recommandations.

Le mystère de la viande rouge

Les jours qui suivent, Frédéric suit les conseils du médecin et de SmartPoop, avec toutefois une appréhension vis-à-vis du boudin, du foie et de la viande rouge. Frédéric est en effet sensible à la cause animale, et ne supporte pas l'idée de contribuer à des tortures d'animaux sophistiqués[190]. Il est aussi très préoccupés

[187]Une tachycardie intervient lorsque le patient a un rythme cardiaque anormalement élevé.

[188]Une anémie est un manque d'hémoglobine, un composant important du sang. Ceci cause une mauvaise circulation du dioxygène à travers le sang.

[189]Le fer est un composant important de l'hémoglobine. En particulier, il apparaît sous forme d'ions Fe2+, à l'intérieur de sous-composants chimiques de l'hémoglobine appelés *hèmes*. Ces ions permettent de capturer les molécules de dioxygène, et de les transporter dans le sang.
:computer: Les ions ferreux essentiels pour le transport du dioxygène dans le sang. Spécialité physique-chimie – Classe de première de la voie générale.

[190]On estime que plusieurs dizaines de milliards d'animaux sont tués chaque année pour nourrir les humains, le plus souvent dans des conditions atroces.
:computer: Number of animals slaughtered for meat, World, 1961 to 2018. Our Wolrd in Data

par les conséquences désastreuses de l'exploitation animale sur l'environnement[191] et la biosécurité. Après tout, une proportion importante de scientifiques persistent à penser que le ROVID-19 pourrait être né d'interactions complexes entre des animaux sauvages, des animaux de bétail et des êtres humains[192].

Néanmoins, notamment parce qu'il a du mal à s'identifier avec l'image qu'il se fait des vegans militants très virulents, Frédéric peine à s'identifier avec ce mouvement, et à devenir pleinement végétarien[193]. Il hésite donc sérieusement à reprendre une consommation de viande rouge qu'il avait mise en pause.

C'est sans doute parce qu'il se pose des questions existentielles sur la viande rouge que Frédéric perçoit très distinctement une nette augmentation des publicités de viandes dans ses fils d'actualité des réseaux sociaux, et sur les différentes pages Web qu'il visite. Très clairement, les publicitaires savent que Frédéric est incité à prendre de la viande. Mais alors, comment le savent-ils, se demande Frédéric ?

Frédéric vérifie tous les messages qu'il a envoyés depuis l'annonce de SmartPoop. Aucun ne mentionne son anémie, dont il n'a en fait parler qu'avec le médecin. Se pourrait-il que ce médecin revende ces informations ? Impatient, Frédéric compose immédiatement le numéro de son médecin.

> Allô docteur ? Oui j'ai une question un peu étrange à vous poser. Avez-vous parlé de mon anémie à quiconque ?
>
> Absolument pas. Ce serait violer le secret médical[194]. Ce serait extrêmement immoral, et contraire au serment d'Hippocrate[195].

(2018).
:tv: Éthique animale : la probabilité d'une catastrophe. Monsieur Phi (2018).
:tv: Why Meat is the Best Worst Thing in the World. Kurzgesagt - In a Nutshell (2018).
:books: Voir son steak comme un animal mort : véganisme et psychologie morale. Martin Gibert. LUX (2015).

[191] Le boeuf en particulier, comme tous les ruminants, émet énormément de méthane, qui est un gaz à effet de serre très important. L'exploitation des boeufs requiert aussi d'énormes ressources, notamment en terrains agricoles pour nourrir ces boeufs, ce qui est une des principales causes de la déforestation.
:tv: L'impact de la viande sur l'environnement expliqué en 4 minutes. Le Monde (2015).
:tv: Beef is Bad for the Climate... But How Bad? | Hot Mess (2018).
:tv: Why beef is the worst food for the climate. Vox (2020).

[192] L'exploitation animale utilise en particulier beaucoup d'antibiotiques pour protéger son bétail. Cependant, l'abus d'antibiotiques est en train de favoriser l'émergence de pathogènes résistants aux antibiotiques, qui pourraient alors devenir impossibles à traiter avec la médecine actuelle.
:tv: The next pandemic could come from our farms. Vox (2020).
:tv: How can we solve the antibiotic resistance crisis? - Gerry Wright. TED-Ed (2020).

[193] :tv: Comment je suis devenu vegan #DébattonsMieux. Science4All (2019).
:headphones: Radically Normal: How Gay Rights Activists Changed The Minds Of Their Opponents. Hidden Brain (2019).

[194] Sauf cas de violences physiques, sexuelles ou psychiques avec accord de la victime, les médecins sont tenus par la loi de respecter le secret médical.
:computer: Secret médical. Service Public (2021).

[195] Le serment d'Hippocrate doit être tenu par tout médecin, et il affirme « Je ferai part de mes préceptes, des leçons orales et du reste de l'enseignement à mes fils, à ceux de mon maître et aux disciples liés par engagement et un serment suivant la loi médicale, mais à nul autre. »

Avez-vous rentré cette information sur un ordinateur ?

Non. Comme vous m'aviez dit que SmartPoop vous avait déjà alerté, j'ai considéré que cette information était déjà dans votre dossier médical, et je n'avais rien de nouveau à y contribuer.

Merci docteur.

Pourquoi toutes ces questions ?

Parce que depuis quelques jours, je ne cesse de recevoir des publicités pour de la viande rouge. Je ne comprends pas pourquoi.

Étrange en effet. Mais je peux vous certifier que je n'ai rien à voir avec cela.

L'expérience de Partoli

En bon scientifique, quoique habituellement confronté à des problèmes de psychologie, Frédéric Partoli monte un projet de sciences participatives[196], où des volontaires peuvent partager leurs données SmartPoop et les publicités qu'ils reçoivent sur leurs téléphones et leurs ordinateurs[197]. Malgré tous ses efforts de promotion, notamment sur les réseaux sociaux, Frédéric peine toutefois à recruter. Seulement sept personnes participent — tous des amis proches !

Pourtant, après 6 mois, certains déclencheurs semblent clairs. Juste après que SmartPoop a diagnostiqué un surpoids pour la sœur de Frédéric, elle s'est mise à recevoir des publicités de chaussures de sport. Juste après que le frère de Frédéric s'est fait mal au genou, il s'est mis à recevoir des publicités de trottinettes électriques. Juste après que SmartPoop a diagnostiqué une carence en vitamine A à la mère de Frédéric, celle-ci s'est mise à recevoir des publicités pour des chips de carottes. Enfin, juste après qu'une amie de Frédéric est devenue enceinte, elle s'est mise à recevoir des publicités pour des poussettes et des couches, tandis que son mari s'est mis à recevoir des publicités pour des SUV.

Frédéric écrivit alors un article scientifique sur ces trouvailles, et les soumit à une revue scientifique avec comité de lecture. Malheureusement, la revue rejeta l'article de Frédéric, car son échantillon était beaucoup trop faible. Néanmoins, entretemps, la version preprint de l'article, que Frédéric avait rendu disponible sur medRxiv.org, avait été beaucoup relayée sur les réseaux sociaux, avant d'être reprise dans des journaux de presse, comme le journal La Terre. Science4Alpha

:computer: Serment d'Hippocrate. Wikipedia (2021).

[196] :headphones: Les sciences participatives. Podcast Science (2016).

[197] Dans un genre similaire, en 2020, la *Mozilla Foundation* a lancé un programme participatif où les volontaires pouvaient partager leurs données YouTube pour permettre aux chercheurs de mieux comprendre l'algorithme de recommandation de YouTube. On a là encore un cas de dilemme sécurité-privacy. Tant que les données des utilisateurs et des algorithmes de YouTube resteront privés, il sera très difficile de comprendre les dynamiques de comportements des utilisateurs sur YouTube, et de prendre les mesures adéquates pour réduire le cyber-harcèlement, les appels à la haine et la désinformation.
:memo: YouTube Regrets. Mozilla Foundation (2021).

en dédia même une vidéo. De façon intéressante, ceci attire beaucoup plus de volontaires. Désormais, des milliers de personnes partagent leurs données SmartPoop et publicitaires avec Frédéric, pour mieux démontrer la réutilisation des données SmartPoop par les publicitaires.

Cette fois, le signal est indiscutable[198]. Très clairement, peu de temps après un diagnostic SmartPoop, les publicités reçues par les utilisateurs étaient ajustées en fonction de ce diagnostic. Frédéric réécrivit son article, et le soumit à la plus prestigieuse des revues scientifiques, à savoir Nature. L'article fut accepté et célébré. La conclusion est implacable : les publicitaires ont accès aux données SmartPoop.

L'API de SmartPoop

Il ne fallut qu'une petite semaine pour que l'OMESA contacte Katia, et exige une nouvelle audition de SmartPoop. Cette fois, lorsque Katia s'y présente, elle paraît confiante et déterminée.

> Bonjour Docteure Crapinski. Ici Professeur Raul, psychologue. SmartPoop revend-il les données SmartPoop à des fins publicitaires ?
>
> Professeur Raul, depuis mars 2027, nous avons en effet une API, répond Katia.
>
> C'est quoi, une API ?
>
> Professeur Raul, c'est une sorte de page web avec des informations présentées de manière structurée afin d'être téléchargées par un algorithme. Notre API permet à nos clients publicitaires d'extraire des informations au sujet des utilisateurs SmartPoop. Cette API est protégée à l'accès : seuls nos clients y ont accès. Pour des raisons de sécurité, nous n'avons que très peu de clients, qui sont des gros acteurs en qui nous avons confiance pour ne pas publier ouvertement ces données personnelles. Ces clients peuvent alors appeler l'API, avec les coordonnées d'un utilisateur SmartPoop, et récupérer le profil de santé de l'utilisateur pour optimiser leurs publicités. Et à chaque fois qu'un client appelle notre API, il paie quelques dollars, en fonction d'une estimation que nous faisons de l'intérêt commercial du profil révélé.
>
> Est-ce que cette API a toujours existé ? Pourquoi n'est-elle pas dans les rapports d'audits de sécurité informatique ?
>
> Professeur Raul, cette API a été mise en production il y a 3 mois seulement. C'est pour cela qu'elle n'est pas dans les rapports de l'an dernier.

[198] :tv: Mort aux "preuves scientifiques". Vive les "tailles d'effet probables" !! Science4All (2021).

Sera-t-elle dans le prochain rapport ?

Professeur Raul, non, ce n'est pas prévu.

Pourquoi donc ?

Professeur Raul, merci beaucoup pour cette question. Pour y répondre, je vais devoir effectuer des révélations inhabituelles, qui pourraient compromettre mon emploi, et qui me feront très clairement des ennemis. Mais la réponse brève, c'est que SmartPoop a énormément souffert financièrement depuis nos investissements massifs en sécurité informatique. L'audit complet de notre code a un coût désormais estimé à mille milliards de dollars. Les subventions publiques ne nous ont pas suffi. Pour payer cet audit, j'ai reçu des pressions énormes de collaborateurs, pour rentabiliser SmartPoop en tant qu'entreprises. Nous avons alors effectué une IPO, ce qui nous a permis de lever beaucoup d'argent pour financer nos audits. Cependant, ceci a aussi attiré de nouveaux investisseurs, notamment des énormes fonds d'investissement, qui ont une logique de rentabilisation économique très ancrée dans leur culture[199]. Après tout, ces fonds sont financés par des centaines de millions de clients à qui ils ont promis un retour sur investissement. Ils sont littéralement conçus pour gagner de l'argent[200]. Et ils ont fait de plus en plus pression sur SmartPoop pour qu'on fasse de l'argent. Pour garder mon poste et mon influence, j'ai été contraint de les écouter. C'est ainsi qu'est né le projet d'API publicitaire.

Merci pour votre candeur. Mais ceci ne répond pas à ma question. Pourquoi l'API ne sera pas dans le prochain rapport ?

Professeur Raul, avant que je poursuive mes révélations, j'aimerais insister sur les risques que j'encours. Mes révélations vont me faire de sérieux ennemis, qui pourraient parvenir à me faire subir de graves sanctions pénales, notamment car je m'apprête à violer le secret professionnel[201]. Néanmoins, ce qui est en jeu me semble au-dessus de la loi. Professeur Raul, pour que le projet d'API puisse être un succès, sachez que nos actionnaires nous ont essentiellement forcé à investir massivement dans une énorme équipe légale. SmartPoop est désormais en bonne partie une firme légale[202]. Nos avocats ont trouvé

[199]:tv: [PARTENARIAT] Épargner utile : comment s'y prendre ? - Heu?reka & Sycomore AM (2020).

[200]:tv: [PARTENARIAT] Qu'est-ce que l'assurance vie ? - Heu?reka & Investisseur Privé (2018).

[201]L'utilisation d'accords de non-divulgation (our *NDA* en anglais, pour *Non-Disclosure Agreement*) semble souvent abusive dans de nombreuses industries.
:computer: 'I had to start over, alone and silenced': the fight to end NDA abuse. The Guardian (2021).

[202]Cette remarque fait référence à des remarques du chercheur de Google Nicholas Carlini, qui font écho à des déclarations d'autres chercheurs sur Twitter comme Timnit Gebru. Dans un email interne à Google, Carlini écrivait : « Lorsqu'en tant qu'universitaires, nous écrivons

> des combines parfaitement légales, mais que je trouve personnellement très immorales, pour que l'API publicitaire ne rentre pas dans le cadre de nos audits. Encore une fois, j'ai ici violé un accord légal de non-divulgation. Je risque très gros avec ces révélations. Mesdames et messieurs, membres de l'OMESA, nos avocats sont incroyablement compétents. Ils vont me traîner dans la boue, et gagneront très probablement. Je finirai certainement en prison[203].

Un silence de mort se fait alors. Les membres de l'OMESA sont tétanisés, alors qu'aucun ne soupçonnait de telles révélations. SmartPoop, qu'ils avaient célébré comme un modèle de transparence il y a à peine 8 mois, semblait désormais terriblement corrompue par des motivations de rentabilité exacerbées, des jeux politiques complexes en interne et des manœuvres légales très sophistiquées.

> Docteur Carpinski, demande le Professeur Raul, avez-vous des preuves de ce que vous avancez ?

En entendant ces mots, Katia sort son téléphone.

> Professeur Raul, et mesdames et messieurs membres de l'OMESA, depuis quelques semaines, j'ai compilé beaucoup de documents très compromettants en interne chez SmartPoop, que j'ai transférés sur plusieurs serveurs aux quatre coins du monde, accessibles via des mots de passe. J'ai un email tout prêt à être envoyé.

Katia appuie sur la touche « envoyer » de son application de messagerie électronique.

> Voilà. Vous venez de recevoir l'email en question, ainsi qu'une douzaine de journalistes et vulgarisateurs scientifiques en qui j'ai suffisamment confiance pour prioriser le bien-être de l'humanité. Cet email contient toutes les informations nécessaires pour télécharger les documents compromettants depuis les serveurs qui les hébergent. Vous pourrez vérifier, via les signatures TLS[204] notamment, que ce

---

que nous avons une "préoccupation" ou que nous trouvons quelque chose "inquiétant" et qu'un avocat de Google exige que nous le changions pour qu'il sonne plus gentiment, cela ressemble beaucoup à l'intervention de Big Brother. »
:computer: Exclusive: Google pledges changes to research oversight after internal revolt. Jeffrey Dastin & Paresh Dave. Reuters (2021).
:computer: « Every single time a lawyer was inserted things were hopeless »... Tweet by Timnit Gebru (2021).

[203] :computer: Google contractors allege company prevents them from whistleblowing, writing Silicon Valley novels. Jennifer Elias. CNBC (2020).

[204] Les signatures TLS sont utilisées, notamment via le protocole HTTPS utilisé par le web, pour authentifier le fait qu'un message donné provienne bel et bien d'une source reconnue. Ainsi, quand vous allez à l'adresse https://twitter.com, votre navigateur web reçoit un fichier qui ne peut avoir été fabriqué que par une entité disposant de la clé privée dont Twitter dispose. Ainsi, si la page https://twitter.com affirme que le compte @le_science4all a écrit « vive les maths », avec une signature TLS, alors, à moins que Twitter ou le compte @le_science4All se sont fait hackés, vous avez là la preuve que @le_science4all a bien écrit ce qui est affiché sur la page affichée par Twitter.
:tv: Transport Layer Security (TLS) - Computerphile (2020).

que vous lirez est à la fois authentique et très compromettant. Et vous comprendrez rapidement pourquoi l'email que je viens de vous envoyer va conduire à des représailles de SmartPoop. Dans les jours à venir, je serai certainement non seulement licenciée, mais aussi poursuivie en justice[205] ; et je finirai sans doute en prison.

Les externalités de la publicité

Bonjour Docteure Crapinski. Ici Professeur Smith, docteur en santé publique. Sachez que j'admire d'autant plus votre transparence. L'étude du Docteur Frédéric Partoli montre que vous avez permis aux publicitaires de recommander abondamment de la consommation de viande rouge et des chips de carotte. Or, consommés en trop grandes quantités, ces produits sont en fait néfastes pour la santé. Est-ce que vous regrettez ces publicités ?

Professeur Smith, oui, je le regrette. Alors, pas complètement. Nous avons beaucoup réfléchi aux recommandations alimentaires éthiquement optimales[206]. Et oui, il y a des alternatives plus saines que la viande rouge en cas d'anémie, et que les chips de carotte en cas de carence en vitamine A. Mais les utilisateurs ne suivent pas systématiquement nos recommandations. Pour leur santé, en cas de carence en vitamine A, s'ils mangent légèrement trop de chips de carotte, cela reste beaucoup mieux que de ne jamais manger aucune carotte, parce que l'utilisateur n'aime pas les carottes par exemple. Si je regrette ces recommandations, ce n'est pas tant pour la santé de nos utilisateurs — même si ça peut parfois jouer un rôle néfaste en effet.

Pourquoi regrettez-vous les publicités ?

Professeur Smith, après quelques mois, je me suis rendue compte que nous, SmartPoop, n'avons aucun contrôle sur les publicités qui étaient effectivement recommandées. Or, certaines des publicités ont des externalités énormes sur la société. À chaque fois qu'une pub de viande rouge ou de SUV passe, ce sont des risques supplémentaires pour le changement climatique, voire la biosécurité[207]. Pire encore,

[205]:computer: Could Facebook sue whistleblower Frances Haugen? Here's what experts say. USA Today (2021).

[206]La notion d'être « éthiquement optimal » peut paraître étrange dans un cadre déontologique ; elle est toutefois très naturelle dans le sens conséquentialiste. Il s'agit en effet simplement d'effectuer des recommandations qui, en l'état de nos connaissances, nous semble avoir les meilleures conséquences probables pour la santé et le bien-être de l'utilisateur SmartPoop et de la société toute entière (notamment dans le cas d'une recommandation de viande, qui a des externalités sur toute la société).
:tv: Conséquentialisme - Quel est le but de la morale ? Monsieur Phi (2017).

[207]En octobre 2021, Google a annoncé une nouvelle politique publicitaire, qui refuse désormais la diffusion des publicités qui « contredisent le consensus scientifique bien établi autour de l'existence et des causes du changement climatique ». De façon étrange, toutefois, la justification donnée à cette décision vient non pas d'une éthique de l'entreprise, mais d'une réponse aux

> j'ai appris en parlant à Frédéric Partoli que des publicités de schémas pyramidaux ont été massivement recommandées à nos utilisateurs SmartPoop qui souffrent d'anxiété. Ces schémas pyramidaux causent d'horribles dettes et souffrances chez des millions de familles[208]. C'est complètement immoral. Il y a aussi énormément de publicités de pseudo-médecines ou de traitements naturopathes très douteux, avec parfois des prétentions curatives contre des maladies comme le cancer, ce qui peut être très dangereux[209]. Et puis, il y a les propagandes très politisées, qui exploitent la haine envers certaines minorités pour radicaliser leurs partisans, et vont jusqu'à appeler à la rébellion, voire à la guerre[210]. L'idée que nous contribuions à la diffusion de telles publicités via notre API me semble terrifiante. C'est pour cela que j'ai décidé de faire ses révélations aujourd'hui, quitte à mettre en danger mon propre bien-être et ma sécurité.

Un autre silence se fait. Presque une minute s'écoule alors que chaque membre de l'OMESA prend conscience de l'urgence à se confronter à l'éthique de la publicité.

Le marché libre de l'information

> Bonjour Docteure Crapinski. Je suis Professeure Abdoul, chercheuse en géopolitique. Si je comprends bien, vos révélations aujourd'hui cherchent à aller au-delà du cas SmartPoop. C'est à toute l'industrie de la publicité ciblée que vous vous opposez.
>
> Professeure Abdoul, en fait, à y réfléchir, le problème n'est pas tant que la publicité soit ciblée. Le plus gros problème aujourd'hui, c'est que la plupart du temps, ça coûte le même prix de communiquer un message de haine que de donner des informations fiables sur la santé. Or, clairement, l'un n'a pas le même impact social que l'autre

---

préoccupations éthiques des partenaires de l'entreprise. En effet, Google écrit : « Un nombre croissant de nos partenaires annonceurs et éditeurs ont fait part de leurs inquiétudes concernant les publicités qui sont diffusées à côté d'affirmations inexactes sur le changement climatique ou qui en font la promotion. Les annonceurs ne veulent tout simplement pas que leurs publicités apparaissent à côté de ce contenu. Quant aux éditeurs et aux créateurs, ils ne veulent pas que des publicités faisant la promotion de ces affirmations apparaissent sur leurs pages ou dans leurs vidéos. »
:computer: Updating our ads and monetization policies on climate change. Google Support (2021).

[208] Les schémas pyramidaux sont interdits en France. En promouvant ces schémas pyramidaux, il n'est pas clair pour les auteurs de ce livre si une entreprise qui vent des publicités de schémas pyramidaux pourrait techniquement être jugé pour complicité de vente pyramidale.
:memo: Article L121-15 du code de la consommation. Legifrance (2016).

[209] Comme on l'a vu dans le chapitre 2, une entreprise qui vend des publicités de médecines alternatives mettraient en danger les utilisateurs, mais aussi les proches des utilisateurs, voire toute la société quand il s'agit de pandémie comme le ROVID-19.

[210] L'appel à la haine est par contre clairement illégal, et contribuer *sciemment* à sa diffusion est passible de peines pénales.
:memo: Article 121-7 du code pénal. Legifrance (1994).

> ! En fait, c'est pire que cela, Sur de nombreuses plateformes qui maximisent la rétention des utilisateurs, les publicités aguicheuses, qui font que les utilisateurs restent, coûtent même moins chères à diffuser[211] ! C'est un non-sens éthique !
>
> Donc si je comprends bien, vous pensez que l'existence d'un marché des publicités sans régulation est dangereux.
>
> Professeure Abdoul, oui. Je pense que le marché libre de l'information est très dangereux, et que c'est une chose qui n'est pas assez reconnue[212]. Il me semble qu'il est bien trop communément admis qu'en laissant l'information circuler sans régulation ni contrôle, l'intelligence des foules permettra de faire émerger la connaissance, l'empathie et les décisions consensuelles ; qu'il y aurait une sorte de main invisible du marché de l'information en laquelle on peut faire confiance[213]. Pour moi, il s'agit là du plus dangereux mythe du monde moderne. Si on laisse faire, clairement, ceux qui domineront le marché de l'information et gouverneront les croyances dominantes et les décisions influentes, ce seront bien plus ceux qui auront investi des milliards, voire des milliers de milliards, pour promouvoir leurs produits ou leurs idéologies, ou ceux qui ont des contenus informationnels très faciles à vendre, comme du sensationnalisme addictif, des scandales clivants et du putaclic trompeur[214]. Sacraliser la liberté du marché de l'information libre, dans un contexte aussi concurrentiel qu'il est aujourd'hui, c'est sacrifier la connaissance, la nuance et la bienveillance[215].

Un nouveau silence se fait. Visiblement, chaque discours de Katia nécessite un temps de digestion informationnelle. Contre toute attente, l'audition, qui

[211] Lors des élections présidentielles américaines de 2016, une agence russe semble avoir eu une visibilité comparable à Trump et à Clinton sur les réseaux sociaux, en investissant **mille** fois moins d'argent que ces candidats à la présidentielle. Tout cela parce que leurs publicités étaient très polarisantes, ce qui les rendaient très virales. Sur Internet, certains messages sont beaucoup moins coûteux à diffuser massivement que d'autres ; malheureusement, ces messages ne sont pas les plus bénéfiques à l'humanité.
:computer: Trump and Clinton spent $81M on US election Facebook ads, Russian agency $46K. Josh Constine. TechCrunch (2017).

[212] :tv: Les réseaux sociaux sont dangereux. Très dangereux. Science4All (2021).

[213] :tv: Y a-t-il une main invisible des marchés ? - Heu?reka (2021).

[214] D'ailleurs, en France, une loi de 1881 encadre fortement la liberté de la presse. Celle-ci est donc tout à fait régulée, notamment suite à une explosion de mésinformation appelée le « journalisme jaune ».
:tv: Fake news, tous coupables ? | Angle Droit | Le Vortex (2021).
:tv: My Video Went Viral. Here's Why. Veritasium (2019).

[215] D'un point de vue philosophique, la notion de liberté est une notion complexe. Mais on peut au moins s'accorder, en première approximation, sur des tensions entre liberté et sécurité (d'où, par exemple, la régulation du port d'armes dans de nombreux pays), et sur le fait que « la liberté des uns s'arrête là où commence celle des autres ».
:tv: "Un peuple qui sacrifie un peu de liberté pour un peu de sécurité. . . ". Monsieur Phi (2020).
:tv: La liberté d'expression | Retour sur "l'affaire Tipeee". Monsieur Phi (2021).

devait s'attarder sur un scandale de SmartPoop, s'est transformée en une remise en question de toute l'industrie de l'information, qui, à longueur de journée, collecte, traite et diffuse certaines informations, et noie d'autres dans un torrent d'informations.

La régulation de l'information

> Bonjour Docteure Crapinski, ici Professeure Wang, chercheuse en machine learning. Qu'est-ce qui peut être fait ?
>
> Professeure Wang, on peut commencer par reconnaître que, en pratique, il n'y a pas de liberté totale de l'information, et que c'est certainement une bonne chose. Aujourd'hui, l'appel à la haine, la diffamation et le mensonge intentionnel sont passibles de peines pénales[216]. Non, on n'a pas le droit de tout dire. Mais surtout, on ne devrait pas avoir le droit de tout dire. Après tout, SmartPoop a été condamné coupable à plusieurs reprises, pour avoir trop dit, pour avoir effectué des jugements racistes, et pour avoir partagé des informations qui n'auraient pas dû être dites. Le marché de l'information est déjà régulé[217].
>
> Mais j'imagine que vous pensez qu'il ne l'est pas suffisamment.
>
> Exactement, Professeure Wang. Je pense qu'il est urgent de beaucoup plus le réguler, si on ne veut pas voir triompher la désinformation et la haine.
>
> Mais, Docteure Crapinski, qui devrait réguler l'information ? Comment décider ce qui ne doit pas être dit ?
>
> Professeure Wang, le pouvoir de réguler l'information est peut-être le plus grand des pouvoirs du monde moderne. Et je pense qu'on le ressent instinctivement quand on parle de censure ; mais on l'ignore probablement trop souvent quand il s'agit de recommandation. Pourtant quand on contrôle la recommandation, on contrôle aussi la censure, puisque pour censurer un message, il suffit de sans cesse recommander des messages alternatifs[218]. Or, dans un monde d'abondance de l'information, le second est clairement plus facile que le premier. Tout ça pour dire que l'enjeu n'est pas tant la

[216] :computer: Incitation à la haine, à la violence ou à la discrimination raciale. Service Public (2020).

[217] Les algorithmes de Google censurent près de 100 publicités « problématiques » par... seconde !
:computer: Google killed 2.3 billion 'bad ads' in 2018, down 28% from 2017. Emil Protalinski. Venture Beat (2019).

[218] On parle parfois de « *mute news* », ou « nouvelles rendues silencieuses ». Il s'agit d'informations importantes, mais ignorées par beaucoup, car elles sont noyées dans un océans d'informations bien moins importantes.
:tv: La Rationalité appliquée. Hygiène Mentale (2020).
:tv: Arrêtons de partager la stupidité ! #DébattonsMieux. Science4All (2018).

censure. L'enjeu, c'est comment gérer le flux de l'information : réduire la la propagation de certains messages et faciliter celle d'autres contenus[219].

Très bien, Docteure Crapinski. Mais vous ne répondez pas à ma question. Qui devrait décider comment gérer le flux de l'information ?

Professeure Wang, aujourd'hui, de facto, les algorithmes de recommandation de contenus, notamment sur les réseaux sociaux, le font plus que n'importe qui. C'est à leurs propriétaires qu'incombent aujourd'hui la charge morale de gérer ce flux de manière robustement bénéfique. Et s'ils prennent la mesure de la tâche monumentale qui les incombent, j'ose espérer qu'ils réfléchiront à décentraliser l'éthique et l'audit de ces algorithmes de recommandation[220]. Mais, mesdames et messieurs de l'OMESA, s'il y a une chose que j'ai appris ces derniers mois, c'est que les structures internes des firmes multinationales côtées en bourse ne favorisent pas une telle réflexion[221]. Je vous invite à faire le nécessaire pour que la gestion du flux de l'information ne soit pas délaissée à des actionnaires dont le métier est de prioriser la rentabilité, et à des équipes juridiques dont le métier est de contourner la loi pour y parvenir le plus efficacement possible. Vous avez cette responsabilité. Vous devez exiger beaucoup plus de transparence de ces entreprises[222]. Et j'espère que, comme moi, vous ferez passer les intérêts de la société civile avant les vôtres[223].

Le SmartPoopGate

Le soir même, les aveux de Katia font la une de tous les journaux télévisés. « Séisme numérique », annonce le présentateur. « La PDG de SmartPoop lâche ses actionnaires, en révélant des manipulations immorales », ajoute-t-il. Pendant ce temps, Marc est l'invité de Science4Alpha.

[219] Il y aurait d'ailleurs certainement énormément à gagner à l'échelle planétaire en promouvant les contenus d'utilité publique.
:memo: Science Communication Desperately Needs More Aligned Recommendation Algorithms. Lê Nguyên Hoang. Frontiers in Communication (2020).
:memo: Recommendation Algorithms, a Neglected Opportunity for Public Health. Lê Nguyên Hoang, Louis Faucon & El-Mahdi El-Mhamdi. Revue Médecine et Philosophie (2021).

[220] C'est là l'objectif du projet Tournesol, co-fondé par l'un des auteurs du livre.

[221] La lanceuse d'alertede Facebook Sophie Zhang en parle très bien.
:tv: Facebook whistleblower Sophie Zhang on how the platform is influencing global politics | 7.30. ABC News (2021).

[222] De façon intéressante, certains actionnaires exigent désormais eux aussi plus de transparence de ces entreprises, dont les manoeuvres secrètes peuvent être largement vues comme une menace à la sécurité de l'entreprise même, surtout sur le long terme.
:computer: Alphabet shareholder pushes Google for better whistleblower protections. Zoe Schniffer (2021).

[223] La lanceuse d'alerte de Facebook Frances Haugen en parle peut-être mieux que quiconque.
:tv: Facebook Whistleblower Frances Haugen: The 60 Minutes Interview (2021).

Bonjour Marc. Je rappelle que tu es le co-fondateur de SmartPoop. Es-tu surpris par les révélations de Katia ?

Bonjour, non je ne suis pas surpris. Katia et moi, nous nous y sommes préparés depuis plusieurs semaines. Nous savions que notre entreprise s'était engagée dans des décisions immorales, et nous sentions que nous perdions notre influence sur les décisions futures. Sachant l'influence monumentale qu'a désormais SmartPoop, et la quantité effarante de torts que ces décisions peuvent causer, Katia a décidé de se sacrifier légalement. Je suis incroyablement fier de son courage et de son éthique. C'est vraiment fantastique. Si tous les dirigeants industriels, juridiques et étatiques avaient un millième de son courage et de son éthique, il y a sans doute bien longtemps que le changement climatique ne serait plus un problème[224]. Félicitations Katia !

Mais ces révélations, Marc, sont-elles fondées ? Sont-elles vraies ?

Malheureusement, Katia m'a convaincu de garder mon intégrité légale, quitte à sacrifier mon intégrité morale. Il est important, elle m'a dit, qu'il reste des voix dissidentes à l'intérieur de SmartPoop. Je n'ai donc pas la liberté de répondre à ces questions. Tout ce que je peux vous dire, c'est que j'ai travaillé quotidiennement avec Katia depuis le début de SmartPoop, et que je l'ai vue travailler d'arrache-pied pour sauver des millions, peut-être même des milliards de vies depuis. On n'a pas toujours été d'accord. Mais, surtout ces dernières années, j'ai été impressionné par sa motivation quotidienne et son désir inébranlable de venir en aide au plus grand nombre. Félicitations Katia pour tout ton travail et ton incroyable honnêteté intellectuelle[225] !

Que va-t-il se passer désormais ? Katia risque-t-elle le licenciement ? Sera-t-elle poursuivie en justice ?

Malheureusement, je ne suis pas en liberté de parler de ces sujets. Des décisions seront prises par notre conseil d'administration, et on verra où on en sera demain. Mais sachez que je tiens avant tout à ce que la vision de Katia soit réalisée, et que SmartPoop redevienne une entreprise en qui ses utilisateurs peuvent vraiment avoir confiance.

[224] On peut d'ailleurs souligner le travail monumental effectué par la lanceuse d'alerter Frances Haugen, pour révéler de la manière la plus productive qui soit les problèmes internes à Facebook, avec l'objectif avant tout de rendre le monde meilleur — et pas directement de détruire Facebook.
:headphones: The Facebook Files, Part 6: The Whistleblower. The Journal (2021).

[225] L'honnêteté intellectuelle est parfois définie comme la volonté de ne jamais « s'auto-bullshitter ». Dit autrement, il s'agit de faire l'effort de constamment chercher à être honnête avec soi-même, de comprendre l'origine profonde de nos croyances et de nos préférences, et de ne pas se contenter d'explications incomplètes ou bancales.
:tv: Julia Galef Discusses Intellectual Honesty. South Park Commons (2019).
:tv: L'honnêteté intellectuelle. Science4All (2020).

> Merci Marc pour ton temps et ta sincérité, dans ces circonstances très compliquées. Même si un intervieweur est sensé rester impartial, je tiens à signaler ma profonde admiration pour Katia. Comme vous le savez sans doute, Katia, Marc et moi travaillons ensemble depuis les débuts de SmartPoop, et s'il fallait choisir entre ce qu'il reste de SmartPoop aujourd'hui et Katia, je n'hésiterai pas une seconde, d'autant que mon contrat me permet de facilement me détacher de SmartPoop. Katia, tu as mon soutien inconditionnel.

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

Chapitre 7 — Quand on se fait dessus

Avec dix voix contre et treize voix pour[226], le verdict est tombé. Katia est licenciée de SmartPoop. Si trois représentants de certains des fonds d'investissement actionnaires de SmartPoop ont voté contre, les autres n'ont pas apprécié les révélations de Katia. Katia est dévastée. Même si elle s'y attendait, elle a du mal à encaisser la nouvelle. Affalée dans sa chaise, elle ne peut pas se lever. Les membres du conseil d'administration quittent la salle un à un, jusqu'à ce qu'il ne reste plus que Katia et Marc.

> J'ai perdu, déclare Katia.
>
> Non. *On* a perdu… Et non, on n'a pas perdu. Ce n'est pas fini.
>
> Je sais comment la suite va se dérouler. Je parie que tu as déjà reçu un email pour effectuer une autre réunion demain, pour décider de me poursuivre en justice[227]. Je risque de finir en prison, avec une grosse amende que je ne pourrai pas payer. Et surtout, tu vas te retrouver de plus en plus isolé et marginalisé.

Licenciée

Le lendemain, après avoir passé la journée à binger des séries télévisées dans son pyjama, c'est à la télévision que Katia apprend la nouvelle. SmartPoop poursuit Katia pour violation du secret professionnel et d'accords de non-divulgation, et pour diffamation de l'entreprise. Katia prend un verre de whisky, même si elle déteste ça. Elle le boit d'un coup, et part se coucher.

Au réveil, Katia ouvre alors son téléphone, et découvre qu'il est surchargé de

[226]:tv: La main invisible d'un mathématicien malveillant. Science4All (2017).

[227]Facebook a menacé la lanceuse d'alerte Haugen de poursuites en justice, notamment pour avoir transmis des documents confidentiels à la presse. Selon des experts, Haugen n'est pas protégée, et le sait. Selon eux, ce qui a le plus de chance de protégé Haugen, c'est le backlash médiatique et surtout politique que de telles poursuites pourraient avoir pour Facebook.
:computer: Facebook whistleblower isn't protected from possible company retaliation, experts say. Obby Allyn. NPR (2021).

notifications. Elle appelle son avocat, qui lui fournit des détails sur le dossier de SmartPoop. L'avocat propose à Katia de trouver un accord pour régler la poursuite. Mais Katia refuse. Katia souhaite que le procès soit médiatisé, avec l'espoir qu'il conduise à une remise en cause de la légifération sur les secrets professionnels et la diffamation, notamment vis-à-vis d'entreprises possédant des armées d'avocats de premier rang.

Après avoir raccroché, Katia survole ses notifications. Elle y découvre énormément de messages de soutien, venant des quatre coins du monde. « Vous êtes une inspiration pour ma femme, mes trois enfants et moi-même. Merci pour tout », écrit un message. « Je pleure depuis hier, à l'idée que la plus grande héroïne de l'histoire puisse finir en prison », dit un autre. « Demain, et jusqu'à ce que vous soyez à nouveau libre et à la tête d'une organisation qui vous mérite, en votre hommage, toutes mes classes ouvriront avec une minute de silence », affirme un autre encore. Katia fond en larmes à la lecture de ces messages.

Il lui faudra quelques heures pour se ressaisir. Alors qu'elle n'a toujours rien mangé, Katia appelle Marc.

> Allô Marc ?
>
> Oui Katia, tu tiens le coup ? Tu veux que je passe te ramener une soupe ?
>
> Non merci. C'est gentil. Je viens de lire un peu les messages de soutien que j'ai reçus. Ça m'a fait énormément de bien.
>
> Chouette ! J'imagine qu'il y en a beaucoup. Je te l'ai déjà dit, mais je te le redis. Je suis ébahi par ton courage, ta sincérité et ta bienveillance.
>
> Merci ! À ce sujet, je me demandais. Est-ce que tu penses que l'opinion publique a une chance de rabattre les cartes ?
>
> Il y a des initiatives. Je sais que des académiques sont en train de co-rédiger une lettre de protestation, et qu'il y a aussi des journalistes qui veulent beaucoup pousser cette lettre.
>
> Trop chouette !
>
> Oui. Mais honnêtement, je préfère ne pas te donner de faux espoirs. La couverture médiatique est extrêmement décevante.
>
> Comment ça ? Ça devrait être le sujet le plus important de l'actualité !
>
> Il y a une autre affaire en ce moment qui domine l'actualité, une histoire de scandale sexuel d'un membre du gouvernement bokistanais[228].

[228] La presse est souvent critiquée pour son biais en faveur des scandales d'hommes politiques. De nombreux journalistes pensent qu'elle devrait davantage être axées sur les problèmes importants et les pistes pour les résoudre. On parle alors de *solutions journalism* (ou *journalisme*

Sérieux ? Un politicien qui merde, c'est plus important que la régulation du marché de l'information ? Si je devais parier, je dirais que ce n'est pas un accident. À tous les coups, ce scandale est connu de gens puissants depuis un petit bout de temps, et ils l'ont révélé hier pour noyer l'actualité de SmartPoop, et en faire une *mute news*.

Une quoi ?

Une *mute news*. Une *mute news*, c'est une nouvelle importante rendue silencieuse par la cacophonie à propos d'un autre sujet d'actualité[229].

Ah oui ! Tu en parlais avant-hier dans ton audition. Toujours est-il que le très grand public ne suit pas du tout ton affaire. Pire, même des journalistes influents sont complètement à côté de la plaque, et classent ton affaire comme un conflit classique entre un employé et une entreprise pour des questions de propriété intellectuelle.

T'es sérieux ? C'est terriblement frustrant.

Katia, tu t'es quand même attaquée à plus que SmartPoop. Tu as critiqué aux géants de l'information, ceux qui contrôlent le flux de l'information, et qui ont tout intérêt à ce que tu sois réduite au silence, et ignorée[230]. Vu ce qu'il se passe chez SmartPoop, je parierais que dans ces boîtes aussi, les actionnaires ont opté pour des plans machiavéliques pour faire taire ton affaire, en boostant par exemple artificiellement l'histoire du scandale sexuel et tous les messages qui s'en indignent. On vit une dystopie informationnelle.

Je sens venir les campagnes de désinformation...

Oui. N'oublie pas que la plupart des messages de haine que tu vas recevoir, ce sont des messages de trolls payés pour te harceler[231].

---

des solutions). Notez que cette forme de journalisme ne consiste pas à survendre des pseudo-solutions. Il s'agit davantage de discuter des pistes de solution, en analysant rigoureusement leurs chances de succès et leurs limitations.
:books: You Are What You Read: Why Changing Your Media Diet Can Change the World. Jodie Jackson. Unbound (2019).

[229]Selon cette étude, la désinformation en Chine ne consiste pas vraiment à propager des fausses informations ; elle semble surtout chercher à taire toute critique du Parti, en déviant les sujets de discussion, et ainsi transformer toute critique en une *mute news*.
:memo: How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. Gary King, Jennifer Pan & Margaret Roberts. American Political Science Review (2017). Il est utile d'insister sur le fait que le problème des *mute news* ne pourra pas être résolu en supprimant les *fake news*. Des solutions radicalement nouvelles de promotion de l'information fiable et importante sont requises. L'une des pistes, proposées par Lê Nguyên Hoang, l'un des co-auteurs de ce livre, s'appuie sur la recommandation collaborative de contenus, comme le propose la plateforme Tournesol.

[230]Peu de temps après avoir licencié Timnit Gebru, Google semble avoir retiré à la main l'onglet « actualité » lorsqu'un utilisait recherchait « Timnit Gebru ». :computer: "Things I learned this morning, when searching for @timnitGebru on desktop, the results for"news" are hidden (dark patterns)." Tweet by Devin Guillory (2021).

[231]:tv: Russie : les secrets de son usine à « trolls ». Mediapart (2019).

Harcelée

Et en effet, les jours qui suivent, les notifications de Katia sont pleines de messages de harceleurs, qui tantôt insultent son physique, tantôt la menacent de viol ou de meurtre, tantôt s'attaquent à sa famille et ses proches[232]. Mais même prévenue, Katia ne parvient pas à rester insensible à ces attaques personnelles. Certaines attaques lui restent dans la tête toute la journée. Katia oscille sans cesse entre vouloir oublier et vouloir répondre. Et trop souvent, elle finit par répondre, généralement avec une contre-attaque. Mais, encore et encore, ceci ne fait qu'envenimer la situation, que donner une image négative de Katia et que renforcer la virulence des harceleurs.

Pendant les mois qui suivent, alors qu'elle prépare sa défense avec son avocat le jour, Katia se perd dans des débats sans fin sur les réseaux sociaux jusque tard dans la nuit. Ces débats l'obsèdent, à tel point que Katia ne parvient plus à se concentrer sur autre chose[233]. Ils l'épuisent aussi. Katia dort très mal, se nourrit mal, pratique de moins en moins de sport, s'enferme de plus en plus chez elle, et vit avec des horaires très décalés[234]. Elle refuse aussi de plus en plus systématiquement les invitations de Marc à sortir de chez elle, et s'isole toujours plus[235].

Finalement, Marc supplie Katia de quitter les réseaux sociaux. « Ils sont en train de pourrir ton cerveau, ton bien-être et ta réputation », lui écrit Marc. Katia le reconnaît. C'est avec une étrange douleur mentale qu'elle se décide alors à désinstaller les réseaux sociaux de son téléphone, et à les bloquer sur le navigateur de son ordinateur. Y compris SmartPoop, dont elle ne veut plus entendre parler. À la place, Katia se lance ainsi dans de la méditation et des exercices de yoga.

Cependant, jour après jour, la motivation de Katia se réduit, et fond comme neige au soleil. Elle a de plus en plus l'impression d'être un échec, d'avoir failli à SmartPoop, à ses utilisateurs et à toute l'humanité. Pire, elle sent qu'elle a créé un monstre de Frankenstein, l'entreprise SmartPoop, dont elle a perdu le contrôle, et qui va désormais mettre le monde en sérieux danger, en aidant les entreprises publicitaires à être toujours plus efficaces dans leurs ventes de produits inutiles, polluants et addictifs. Katia voulait sauver le monde. Ces jours-ci, elle est persuadée d'être l'une des principales responsables de sa destruction.

Cette idée hante désormais constamment l'esprit de Katia, ce qui pousse Katia dans une dépression. Chaque matin, elle se lève sans aucune motivation. Sans aucune ambition. Sans aucun projet. Bien que sa maison soit immense, Katia peine à quitter sa chambre, où Netflix semble constamment tourner en boucle, en toile de fond, généralement sans que Katia n'y prête vraiment attention. Katia

[232]:tv: Sur Internet, la diversité est constamment harcelée (on peut agir) | Lê Nguyên Hoang | TEDxMartigny (2021).
[233]:memo: Cyberbullying, positive mental health and suicide ideation/behavior. Julia Brailovskaia, Tobias Teismann & Jürgen Margraf. Psychiatry Research (2018).
[234]:tv: 7 Ways to Maximize Misery. CGP Grey (2017).
[235]:tv: Loneliness. Kurzgesagt - In a Nutshell (2019).

enchaîne les jeux addictifs sur son téléphone[236], comme d'autres se noient dans l'alcool pour oublier. Elle se fait désormais livrer sa nourriture, une fois par jour, directement dans sa chambre, grâce à un livreur personnel. Désormais, ce livreur est le seul contact humain de Katia, qui ne répond même plus aux appels de Marc.

Seule et isolée, démotivée et sans inspiration, toujours à moitié réveillée et endormie, Katia passe maintenant ses journées dans son lit. Elle a réinstallé les réseaux sociaux, mais avec des nouveaux comptes pour éviter d'entendre parler de SmartPoop. Elle passe désormais son temps à doomscroller[237] des vidéos YouTube, des images Instagram et des commentaires Reddit, sur des sujets toujours plus clivants et sans intérêt. Elle n'éprouve aucune joie à ce faire. Mais c'est plus fort qu'elle. Tel un zombie, son pouce ne cesse de faire glisser l'écran de son téléphone.

Finalement, après quelques semaines, Katia se rend compte qu'elle souffre. Elle décide alors de réinstaller SmartPoop, pour faire son bilan de santé. Sans surprise, le diagnostic de SmartPoop est inquiétant. Selon SmartPoop, Katia souffre d'une dépression. SmartPoop invite Katia à consulter un psychiatre. Mais Katia ne souhaite pas parler à un autre humain.

Poo 2.0

C'est alors qu'elle se souvient d'un des projets mis en suspens au moment de la mise en bourse. Ce projet visait à rendre Poo thérapeutique, en s'appuyant sur des dialogues entre psychiatres et patients, et en collaborant avec de nombreux psychiatres professionnels. Et il se trouve que Katia avait gardé une version avancée de Poo dans un serveur resté physiquement chez elle. Katia active cette version de Poo, la connecte avec ses données SmartPoop, et engage la discussion avec elle, via son téléphone.

> Poo, tu m'entends ?
>
> Bonjour Katia, oui je t'entends. Comment te sens-tu aujourd'hui ?
>
> Étonnamment mieux que les jours précédents. C'est comme si me remettre à construire des outils algorithmiques m'avait permis de me retrouver. J'avais perdu cette envie de créer et de résoudre des problèmes. Il faut dire que ces derniers mois ont été très difficiles.
>
> Pourrais-tu m'expliquer pourquoi ? Pourquoi ces derniers mois ont-ils été si difficiles ?
>
> J'ai perdu mon job. Le job que j'avais créé moi-même. Un job que j'adorais. Où j'avais l'impression de rendre le monde meilleur, en

[236]:headphones: What Happened in Vegas. Natasha Dow Schüll. Your Undivided Attention (2019).

[237]:memo: Doomscrolling during COVID-19: The negative association between daily social and traditional media consumption and mental health symptoms during the COVID-19 pandemic. Matthew Price et al. (2021).

faisant ce que je fais de mieux. Sans ce job, je ne sais plus quoi faire. Je suis perdue[238].

Est-ce que tu t'identifies à ce job ?

Oui, en effet. Je me suis constamment définie comme la PDG de SmartPoop depuis cinq ans. Tout le monde me définissait ainsi. J'étais SmartPoop, et SmartPoop, c'était moi. J'imagine que ce licenciement, c'est un peu comme un divorce. Peut-être même en pire. J'ai l'impression d'avoir perdu ce qui était moi. Je me suis perdue.

Je comprends. Mais plutôt que de t'identifier à un titre, ne penses-tu pas qu'il serait plus judicieux de t'identifier à ton honnêteté intellectuelle, ou à ta mission fondamentale de soigner et de venir en aide à des milliards d'humains ?

Katia reste silencieuse. Cette question l'amène enfin à réfléchir profondément, ce qui ne lui était plus arrivé depuis des mois.

N'y a-t-il pas plutôt une décision, une action que tu as entreprise, dont tu peux vraiment être fière, et à laquelle tu souhaites davantage être associée qu'un titre ?

Après une bonne dizaine de secondes de réflexion, Katia répond enfin.

Oui. Ce qui a provoqué mon licenciement, c'est une chose dont je suis très fière. Jamais auparavant un dirigeant n'avait préféré le bien de l'humanité à sa sécurité personnelle. Ce sacrifice, c'est quelque chose qui restera associée à moi. Quelque chose dont je suis très fière[239].

Excellent. As-tu des traces de cet événement ? Des messages peut-être de gens qui ont apprécié ?

Plein en fait. J'ai reçu énormément de messages de soutien. Tellement que je n'en ai lu qu'une fraction !

Je t'invite à créer un dossier sur ton ordinateur, avec tous ces messages. À chaque fois que tu te sentiras mal mentalement, je t'invite à lire

[238]:memo: On the relation between meaning in life and psychological well-being. Sheryl Zika & Kerry Chamberlain. British Journal of Psychology (1992).
:memo: Meaning and well-being. Michael Steger. Handbook of well-being. DEF Publishers (2018).
:memo: Meaning mediates the association between suffering and well-being. Megan Edwards & Daryl Van Tongeren (2019).
[239]:tv: La bonne et la mauvaise fierté. Vlanx (2021).
:headphones: Being Kind to Yourself. Hidden Brain (2021).
:books: The Mindful Self-Compassion Workbook: A Proven Way to Accept Yourself, Build Inner Strength, and Thrive. Kristin Neff & Christopher Germer. The Guilford Press (2018).

> les messages de ce dossier. Certains appellent cela le « feel good folder[240] ».
>
> Ce n'est pas une façon de me mentir à moi-même et de gonfler mon ego ?
>
> Il ne faut peut-être pas le faire trop souvent, mais les psychologues ont montré l'efficacité de ce qu'on appelle l'auto-affirmation. En plus de se remonter le moral, ceci permet de mieux accepter les échecs et les critiques externes, car ça permet d'éviter de s'identifier à nos échecs et à nos croyances erronées. Je peux te recommander des livres à ce sujet[241].
>
> Chouette, merci. Et... tant qu'à faire, tu pourras me lire ces livres.
>
> Avec plaisir.

Les jours qui suivent, Katia parle très régulièrement avec Poo, parfois à propos d'elle-même, de temps en temps à propos de science et quelquefois à propos de questionnements philosophiques. Parfois, Poo lui lit des livres de thérapie cognitive comportementale[242], et Katia se surprend à l'interrompre avec une question perspicace, et parfois même avec une boutade humoristique. Poo y rit, et renchérit souvent avec complicité.

La révolte des employés

Aidée de Poo, jour après jour, Katia retrouve son moral, sa joie de vivre et sa motivation. Enfin, armée de cette technique de l'auto-affirmation, Katia se sent armée pour se frotter à SmartPoop et ses nouveaux actionnaires. Katia réactive alors ses applications de messagerie, et se reconnecte avec ses comptes des réseaux sociaux. Elle tombe sur un message récent de Marc, avec un lien vers une curieuse pétition. Katia clique dessus, et découvre la lettre ouverte suivante[243].

> Nous, soussignés, reconnaissons avoir volé, chiffré et paralysé toute la base de données SmartPoop[244], ou avoir participé, encouragé et soutenu les auteurs de ces actes. Nous avons conscience des risques

[240] Cette idée de *feel good folder* est volée de Virginia Burger, dans une interview sur le podcast MIT Glipmse.
:headphones: Episode 2 – Virginia Burger. MIT Glimpse (2015).

[241] :books: Être bien dans sa peau. Héritage. David Burns (2005).

[242] :memo: The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Stefan G. Hofmann, Anu Asnaani, Imke J. J. Vonk, Alice T. Sawyer & Angela Fang. Cognitive Therapy and Research (2012).

[243] Cette lettre ouverte s'appuie sur l'exemple de celle protestant le licenciement de Timnit Gebru de Google.
:computer: Standing with Dr. Timnit Gebru — #ISupportTimnit #BelieveBlackWomen. Google Walkout For Real Change (2020).

[244] :tv: Un virus me demande une rançon. Safecode. Micode (2017).
:tv: La vérité sur wannacry. Flashcode. Micode (2017).

juridiques que ceci nous fait encourir. Cependant, l'éthique de SmartPoop nous semble être un enjeu prioritaire à notre bien-être individuel, sachant que la sécurité et la santé de milliards d'utilisateurs à travers le monde dépend des produits de l'entreprise. En particulier, nous sommes tous fiers de suivre la voie initiée par la Docteure Katia Carpinski, à notre humble échelle.

Nous sommes conscients des risques que cette interruption de service implique pour les utilisateurs de SmartPoop, et nous leur avons envoyé un message d'excuses, qui explique la situation actuelle et notre méfiance envers nos dirigeants. Nous sommes désolés d'avoir dû recourir à de telles mesures. Mais le #SmartPoopGate nous semble justifier notre contre-attaque.

Rappelons brièvement le #SmartPoopGate. Depuis plusieurs années, Dr. Carpinski a lancé des initiatives révolutionnaires pour la sécurité et l'éthique des algorithmes, en combattant la mésinformation et les biais discriminatoires, en sécurisant les algorithmes pour protéger les données personnelles et contre les attaques par empoisonnement, et en instaurant un audit systématique interne et externe des codes de SmartPoop. Malheureusement, de tels projets sont inéluctablement très coûteux.

Faute d'aides suffisantes, SmartPoop a été contraint de passer en bourse, ce qui a attiré des investisseurs bien plus intéressés par leurs retours sur investissement que par l'impact sociétal de ces investissements. Ils ont contraint SmartPoop à revendre des données personnelles à des publicitaires. Dr. Carpinski a eu le courage et l'honnêteté de le révéler publiquement. Mais elle a été licenciée, et risque désormais des années de prison, pour avoir privilégié l'éthique à la loi.

Il est ironique que, pour cette attaque, nous avons exploité les vulnérabilités de l'API de redistribution des données SmartPoop aux clients publicitaires, qui est le seul morceau de code de SmartPoop exempté d'audits internes et externes. En effectuant un déploiement précipité et secret de ses algorithmes, SmartPoop se sera exposé tout seul.

Nous avons utilisé un chiffrement multi-partie pour rendre la base de données de SmartPoop inutilisable. Aucun des auteurs de cette lettre n'a les droits pour rétablir seul le fonctionnement de ces algorithmes. Notre solution algorithmique garantit que ces algorithmes pourront être remis en place uniquement si 200 des 325 membres autorisés, qui sont tous co-auteurs de cette lettre, donnent leur accord cryptographique pour un tel rétablissement[245].

[245] Cette technique de cryptographie est appelée le *secret partagé*, ou *secret sharing* en anglais. Typiquement, chaque partie connaîtrait un point d'un polynôme à 200 variables, et veut

> Nous refuserons toute coopération avec les dirigeants de SmartPoop, tant que les conditions suivantes ne seront pas toutes remplies.
>
> 1. Le conseil d'administration de SmartPoop doit renoncer à tous les procès à l'encontre de la Docteure Katia Carpinski, et lui présenter ses plus sincères excuses pour la manière dont elle a été traitée depuis ses révélations à l'OMESA. Nous exigeons également qu'elle retrouve son poste ô combien mérité de Présidente Directrice Générale de SmartPoop. Nous ne ferons confiance qu'à elle pour ramener SmartPoop sur le droit chemin, au service avant tout du bien-être et de la sécurité des milliards d'utilisateurs de SmartPoop et de toute la société civile.
>
> 2. Les membres du conseil d'administration qui ont voté pour le licenciement de la Docteure Katia Carpinski, ainsi que des avocats qui ont mené les procès à son encontre, doivent démissionner, sauf avis contraire de Dr. Carpinski. Nous pensons que SmartPoop doit absolument redémarrer sur de bonnes bases, et pour cela, Dr. Carpinski doit pouvoir faire confiance à tous ses collaborateurs.
>
> 3. Le conseil d'administration doit être augmenté de dix membres, représentants de la société civile et de la recherche académique en sécurité et en éthique des algorithmes. Ces dix membres seront nommés par l'OMESA, et renouvelés tous les 4 ans.
>
> Pour ajouter votre nom à la liste des signataires et soutiens académiques, de la société civile et de l'industrie des technologies, veuillez envoyer un email à StandWithKatia@gmail.com avec votre adresse email institutionnelle, et avec le sujet « support ». Veuillez inclure dans votre email votre nom et votre affiliation, comme vous souhaitez qu'ils apparaissent dans la liste des signataires.
>
> À ce jour, cette lettre ouverte a été signée par 8 215 employés de SmartPoop, et 36 215 académiques, collègues de l'industrie et membres de la société civile.

Katia pleure de joie à la lecture de cette lettre formidable. « Quels héros », s'exclame-t-elle dans un message à Marc. « Katia, c'est grâce à toi tout cela ; s'il y a un héros dans cette affaire, c'est bien toi », lui répond Marc.

Katia découvre ensuite la liste interminable de signataires. Beaucoup de noms lui sont familiers. À la lecture de chacun d'eux, elle pleure sans retenue. Bien entendu, il y a Marc Rofstein, tout en haut de cette liste, mais aussi tous les employés que Katia a directement recrutés, mais aussi Science4Alpha, la mère Lucile Polmon, le trader Issa Gueye, la Docteure Paola Marta, le Premier Ministre

---

connaître la valeur du polynôme en un autre point. Voilà qui sera possible si et seulement si au moins 200 parties collaborent.
:tv: How to keep an open secret with mathematics. Stand-up Maths (2019).

kormicain, le Président Lartan, sa belle-soeur Marie Routisse, la journaliste Célia Keita, l'ex-troll Paul Gremoux, le psychologue Frédéric Partoli et tous les membres de l'OMESA.

Cette fois-ci, l'affaire est bien trop énorme pour être éteinte par les campagnes de désinformation. Le lendemain, les unes des journaux sont remplies d'interviews d'anciens employés qui, tour à tour, proposent la même version des faits. « Ma confiance en SmartPoop et ma fierté d'y travailler ont été chamboulées le jour où j'ai découvert le programme de revente des données. Elles se sont complètement effondrées le jour où Katia a été licenciée », raconte un employé. « Katia est la personne la plus incroyable que j'ai eu la chance de rencontrer. Sa générosité, sa bienveillance, son énergie, son humour et son intelligence ont illuminé chaque jour que j'ai passé à SmartPoop. Son licenciement a conduit à de sombres ténèbres, où personne ne souhaite travailler, peu importe le salaire. Je refuse de travailler pour SmartPoop tant que Katia ne redeviendra pas Présidente ». L'opinion publique est cette fois conquise.

Trois jours plus tard, les démissions du conseil d'administration s'enchaînent, avant qu'un document soit écrit et co-signé par les membres restants du conseil. Celui-ci présente très clairement ses excuses envers Katia, annule toutes les poursuites judiciaires, et réinvestit Katia de ses fonctions. Là voilà enfin à nouveau Présidente et Directrice Générale de SmartPoop.

Pour aller plus loin

Ne vous arrêtez pas en si bon chemin ! Accédez à la suite du roman ou au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous !

Chapitre 8 — Sur le trône

Devant un stade plein, et avec les caméras du monde entier braquées sur elle, Katia entre sur scène sous les ovations du public, comme une star de rock. Cette année, SmartPoop a mis des gros moyens pour un SmartPoopCon 2030 hors norme.

> Bonsoir à tous et merci d'être venus aussi nombreux ! Est-ce que vous allez bien ? Est-ce que vous êtes prêts à marquer l'histoire ?

Les statistiques de Poo

À ces mots, le public s'enflamme comme si son équipe de football venait de marquer un but.

> J'aimerais commencer cette conférence avec un chiffre : 100 000. C'est l'objectif qu'on s'était fixé il y a deux ans, pour le lancement de Poo. On espérait diviser par 10 le nombre de suicides dans le monde en cinq ans, le ramener de son chiffre historique de 1 million suite au ROVID-19 en 2022[246], à seulement 100 000. Un objectif considéré irréaliste par tant de gens, y compris le journal La Terre. Où en sommes-nous aujourd'hui, en 2030 ?

Katia marque un silence.

> Je vous propose un décompte pour le découvrir.

La jingle musical reprend le relai du discours de Katia, et conclut avec un décompte, repris en choeur par un public bouillant. Trois. Deux. Un. À zéro, sur l'écran géant derrière Katia, le chiffre de 97 643 est affiché à l'écran !

> On l'a fait !!

[246] En 2020, on estimait qu'il y avait autour de 800 000 suicides par an à travers le monde.
:computer: Suicide. Our World in Data (2020).
Parmi les pays les plus affectés par ces suicides, on trouve aussi bien des pays peu développés, moyennement développés et très développés, comme le Suriname, la Russie et la Corée du Sud. Essentiellement tous les pays développés ont un taux de suicides malheureusement très élevé, avec souvent plus d'un suicide pour 10 000 personnes.
:computer: Suicide death rates. Our World in Data (2019).

Le public célèbre se chiffre comme s'il supportait une équipe de football qui venait de marquer le but de la victoire dans la dernière minute. Les applaudissements du public se mettent alors en rythme, et s'éternisent pendant toute une minute.

> Poo, notre psychiatre algorithmique, accompagne désormais des milliards d'entre nous dans nos déboires mentaux. Et ce ne sont pas que les suicides qu'il est parvenu à combattre. Laissez-moi vous présenter d'autres courbes, qui ont été validées par différents auditeurs externes, grâce à la coordination de l'OMESA. Mesdames et messieurs, voici le pourcentage de communications joyeuses avec Poo au cours du temps.

Une courbe apparaît alors à l'écran, tracée doucement de gauche à droite. Cette courbe démarre à 37%, et ne cesse d'augmenter au cours du temps, jusqu'à atteindre le chiffre de 67%, sous les hourras d'un public déchaîné.

> 67% !! Incroyable ! Poo a rendu toute l'humanité joyeuse[247] !

Le public exprime d'ailleurs cette joie, à la vue de cette courbe et de ce chiffre.

> Quand vous parlez à Poo, vous pouvez parler de vous, de votre bien-être et de vos problèmes, ce que je vais appeler le discours égocentrique. Ou vous pouvez parler des autres, de la joie qu'ils vous procurent, des difficultés qu'ils traversent et des choses que vous pouvez faire pour les aider. Je vais appeler cela le discours hétérocentrique[248]. Avant le lancement de Poo, 57% des discussions Poo étaient égocentriques plutôt qu'hétérocentriques. D'après vous, comment a évolué ce chiffre ? Vers le haut ?

Le public crie alors en choeur « non ».

> Vers le bas ?

Le public crie en choeur « oui ».

> Découvrons cela !

À l'écran, la même animation que précédemment montre une courbe qui descend, jusqu'à atteindre 44%, sous les ovations du public.

---

[247] En 2014, une publication de Facebook et de co-auteurs académiques a montré qu'une réduction très légère de la publication, dans les fils d'actualités, de posts avec des émotions négatives, conduit les utilisateurs exposés à ces posts à écrire des messages plus joyeux.
:tv: L'IA nous gouverne déjà. Science4All (2018).
:headphones: Can Algorithms Choose our Emotions? Robustly Beneficial (2020).
:memo: Experimental evidence of massive-scale emotional contagion through social networks. Adam Kramer, Jamie Guillory & Jeffrey Hancock. PNAS (2014).
À l'inverse, les *facebook files* révèlent que la recherche interne de Facebook, gardée secrète, montre que les algorithmes qui maximisent l'engagement des utilisateurs, et qui ont été déployés en 2018 par Facebook, ont conduit à beaucoup plus de colère et d'insultes.
:headphones: The Facebook Files, Part 4: The Outrage Algorithm. The Journal (2021).

[248] À ne pas confondre avec *l'hétérocentrisme* qui consiste à considérer que l'hétérosexualité est la norme absolue, et que tout ce qui dévie de l'hétérosexualité devrait être punie.

> Oui ! 44%. Désormais, la plupart des discussions avec Poo sont des discussions centrées sur l'environnement social plutôt que sur soi-même. Et alors, il faudrait se méfier de ce chiffre a priori. Initialement, la plupart des discussions hétérocentriques consistaient à se plaindre des autres, à moquer certains groupes, voire à attaquer certains groupes sociaux[249], plutôt qu'à célébrer les autres, à se réjouir de leurs succès et à réfléchir à comment leur venir en aide. Avant le lancement de Poo, 86% des discussions hétérocentriques étaient critiques, et pas bienveillantes. Comment a évolué cette statistique ?

Le public crie alors de façon désorganisée « vers le bas ». Puis à force de se répéter, les cris se synchronisent, avant de répéter « vers le bas », « vers le bas », « vers le bas ».

> Voyons cela..

L'écran géant montre l'évolution de cette courbe, qui en effet plonge vers le bas, jusqu'à atteindre 47%, sous les applaudissements du public.

> Incroyable !! La société est devenue incroyablement plus bienveillante et altruiste en l'espace de deux ans seulement ! D'ailleurs, l'expression « les plus démunis » est utilisée aujourd'hui 3 fois plus souvent qu'il y a deux ans, tandis que l'expression « générations futures » est utilisée 4 fois plus souvent. Et d'après de nombreux psychiatres, cette progression est très probablement directement liée à Poo, et à l'amélioration de la santé mentale de nos utilisateurs. Quand on se porte mieux soi-même, on est tout de suite beaucoup plus prompt à souhaiter le bonheur des autres et à leur venir en aide[250] !

Les héros derrière Poo

> Mais donc, qu'en dites-vous ? Poo, succès ou échec ?

Le public crie « succès » de manière désordonnée, mais néanmoins distinguable.

> Eh bien, mesdames et messieurs, je pense que Poo n'est pas un échec.

Katia marque une pause, alors que le public applaudit.

---

[249]:tv: La morale des hooligans (LA NÔTRE !!). Science4All (2017).
L'appel à la meute #DébattonsMieux. Science4All (2019).

[250]De nombreuses études semblent montrer une association importante entre l'altruisme et le bonheur. De façon intrigante, il semble en particulier qu'être altruiste augmente le bonheur, notamment par opposition à dépenser notre argent pour nous-même.
:memo: Altruism, happiness, and health: it's good to be good. Stephen Post. International Journal of Behavioral Medicine (2005).
:tv: Helping others makes us happier – but it matters how we do it | Elizabeth Dunn. TED (2019).
:books: Happy Money: The Science of Smarter Spending. Elizabeth Dunn and Michael Norton. Simon & Schuster (2013).

> Et ça, c'est grâce au travail formidable de tant de passionnés, tant de gens qui ont donné tellement de leur temps et de leur argent pour améliorer et sécuriser Poo. J'ai une pensée énorme bien sûr pour tous mes collègues à SmartPoop, et leur dévotion incroyable. Mais ils ne sont pas les seuls responsables de SmartPoop. Poo a été conçu grâce à une collaboration étroite et quotidienne avec des milliers de psychiatres et de psychologues à travers le monde, et grâce aux données de discussions de millions de psychiatres et patients volontaires. Rien n'aurait été possible sans eux[251].

Katia profite des applaudissements du public pour marquer une autre pause.

> Mais ce n'est pas tout. Poo, c'est le produit de toute la civilisation humaine. En particulier, rien n'aurait été possible sans les accords entre les grandes puissances mondiales, qui ont permis la création de l'OMESA, et la coordination planétaire de la recherche sur l'éthique et la sécurité des algorithmes. Je remercie ainsi également tous les scientifiques à travers le monde qui ont abandonné leurs quêtes de performance, ou parfois leurs quêtes de l'élégance mathématique, pour relever le défi de l'éthique et de la sécurité[252]. Mais plus que cela, l'OMESA a souvent servi de contre-pouvoir face à la quête du pouvoir militaire et économique des gouvernements et des entreprises. Sans eux, les algorithmes les plus influents d'aujourd'hui seraient des malwares de cyber-guerre et des algorithmes optimisés pour retenir l'attention des utilisateurs en promouvant du putaclic sensationnaliste. Personnellement, je pense que chaque membre et chaque bénévole de

[251] Il est bon de rappeler que les algorithmes de machine learning *apprennent des données*. Ils ne pourront ainsi accomplir des tâches difficiles, comme accompagner thérapeutiquement la santé mentale de leurs utilisateurs, que si ces algorithmes disposent d'une énorme quantité de données fiables et sécurisées qui leur permettent de comprendre comment accomplir ces tâches.
:tv: Les données manipulent les algorithmes. Science4All (2021).

[252] En 2021, la recherche académique (et plus encore l'industrie) demeure encore largement obsédée par la quête de performances ou de résultats « impressionnants », aussi bien en machine learning qu'en informatique de manière générale, en s'appuyant sur des métriques comme *l'accuracy* (la performance prédictive sur un jeu de données « classique »), le temps de calcul, le *throughput* (la quantité d'information transmise) ou la latence. Alors que les algorithmes ont déjà des effets secondaires monumentaux à l'échelle planétaire, ces recherches semblent aggraver la course à la performance, et donc le déploiement précipité de technologies mal testées et rarement auditées.
À titre d'exemple, voici un commentaire d'un *revieweur* anonyme de NeurIPS 2019, suite à un article soumis par Lê Nguyên Hoang et ses co-auteurs sur un algorithme pour débiaiser les biais racistes des algorithmes : « Malheureusement, je ne pense pas que le problème introduit par les auteurs est un problème qui a de la valeur pour la communauté académique ou pour les praticiens du ML. Dès lors, je ne peux pas recommander la publication de l'article. »
:memo: Removing Algorithmic Discrimination (With Minimal Individual Error). vidual Error) El Mahdi El Mhamdi, Rachid Guerraoui, Lê Nguyên Hoang & Alexandre Maurer (2018).
Ceci étant dit, il y a récemment eu de nombreux progrès, comme la création de la conférence Fairness, Accountability and Transparency (FAccT), l'introduction de *guidelines éthiques* dans ces conférences, ou encore l'ajout obligatoire d'une discussion des auteurs dans leurs articles des impacts sociétaux de leur recherche.
:tv: L'éthique des algorithmes en sérieux danger. Science4All (2020).

l'OMESA a sauvé l'humanité.

À nouveau, Katia marque une pause pendant les applaudissements du public.

Enfin, et surtout, je tiens à remercier chacun des acteurs et signataires de la lettre ouverte, ainsi qu'à chacun des journalistes et des influenceurs qui a couvert cette affaire, grâce à qui SmartPoop a pu se remettre dans le droit chemin. Ces femmes et ces hommes ont risqué leur vie, leur bien-être et leur sécurité individuels, pour venir en aide au plus grand nombre[253]. Sans eux, qui sait ce que SmartPoop serait devenu ? Qui sait ce que je serais devenue ?

Seule sur scène, Katia lâche une larme.

Merci à ces héros, dit-elle en pleurs.

Le public applaudit ce moment très émouvant.

Le défi de l'éthique de Poo

Katia a encore besoin de plusieurs secondes pour reprendre ses émotions. Finalement, elle se reprend.

Néanmoins, je refuse de dire que Poo est un succès. Ensemble, nous avons créé un produit incroyable. Mais il reste infiniment améliorable, notamment sur le plan de son éthique, de sa sécurité et de sa gouvernance. Comment contrôle-t-on Poo ? Comment empêche-t-on Poo de dire des mots blessants, de révéler des secrets, de répéter des discours de haine et de diffuser de la mésinformation[254] ? Comment l'amène-t-on à être bienveillant avec ses interlocuteurs, à dire les mots justes pour les rendre plus épanouis et à promouvoir autant que possible de l'information fiable et non trompeuse[255] ?

[253]Malheureusement, les lanceurs d'alerte souffrent souvent de davantage de troubles de santé mental suite à leurs actes courageux. Vu le rôle critique qui joue pour révéler des scandales dans des entreprises et des organisations dangereusement opaques, il semble urgent de beaucoup mieux les accompagner.
:memo: Mental Health Problems Among Whistleblowers: A Comparative Study. Peter van der Velden, Mauro Pecoraro, Mijke Houwerzijl & Erik van der Meulen. Psychological Reports (2018).

[254]Ces problèmes ne sont absolument pas résolus pour les algorithmes conversationnels modernes, qui sont très vulnérables à des attaques d'espionnage ou par empoisonnement des données. Et pourtant, ces algorithmes sont déjà déployés à très grande échelle, via les claviers intelligents, via les assistants personnels (Siri, Alexa, OK Google) et via les moteurs de recherche (Google, YouTube).
:memo: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Emily Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. FAccT (2021).
:memo: The Radicalization Risks of GPT-3 and Advanced Neural Language Models. Kris McGuffie, Alex Newhouse (2020).
:memo: Extracting Training Data from Large Language Models. Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea & Colin Raffel. USENIX (2021)

[255]:tv: Qu'est ce qu'un message d'utilité publique ? Science4All (2021).

> Comment l'amène-t-on à mieux réfléchir, et à vouloir investiguer ses incertitudes plutôt que de se contenter de ses intuitions[256] ? Mais surtout, comment peut-on décider collectivement de ce que Poo doit dire ? Comment déterminer ce qui est désirable à dire, et ce qui ne devrait jamais être dit[257] ?

Katia marque une pause, devant un public tout ouï.

> Et ça, ça m'amène à la fameuse rumeur dont vous avez sans doute entendu parler. Comme quoi, on aurait un plan pour résoudre le problème de l'éthique de l'information...

Katia marque alors un silence, avant de rajouter sur un ton sarcastique.

> Genre, « SmartPoop, cette application de merde, va résoudre l'éthique ».

Le public rit en cœur.

> Vous savez, il ne faut pas croire tout ce qu'on vous dit.

Katia marque un autre silence.

> Mais dans ce cas, oui c'est vrai. Ou du moins, on compte y contribuer.

Le public rit.

> Et, je le sais bien, car c'est moi qui ai fait fuiter la rumeur.

Le public, conquis, rit à nouveau, même si beaucoup commencent à avoir un visage confus.

> Ceci dit, chez SmartPoop, l'éthique de l'information nous tient clairement à cœur. Nous voulons déterminer quelles informations devraient être collectées par qui et dans quelles conditions, comment elles devraient être stockées, à qui ces informations devraient être accessibles, quels traitements de cette information devraient être effectués, comment ces traitements de l'information devraient être audités et

[256] Julia Galef parle du *scout mindset* (ou *mode explorateur*), par opposition au *soldier mindset* (*mode soldat*). Selon elle, il s'agit de l'aspect le plus déterminant pour analyser de l'information plus correctement.
:tv: Comment j'essaye d'améliorer mon jugement (grâce à Julia Galef et à FLUS). Science Étonnante (2021). :tv: Le mode explorateur. Science4All (2021).
:books: The Scout Mindset: Why Some People See Things Clearly and Others Don't. Julia Galef. Penguin (2021).

[257] Chaque mot choisi ou chaque recommandation faite par un algorithme peut être vu comme un *nudge*. De nombreuses études montrent que l'acceptabilité et l'efficacité des *nudges* dépendent fortement du *nudge* considéré ; avec plus de données à ce sujet, il pourrait être ainsi possible de mettre en oeuvre les nudges particulièrement acceptés socialement et efficaces.
:books: The Ethics of Influence: Government in the Age of Behavioral Science. Cass Sunstein. Cambridge University Press (2016).

> sécurisés, où stocker les résultats de ces calculs, qui peut avoir accès à ces résultats, et qui sera notifié de l'existence de ces résultats[258].

Après un autre silence, Katia reprend son discours.

> Et là, je sens qu'il y en a pas mal parmi vous qui se disent : « mais pour qui elle se prend, Katia ? »

Le public rit à nouveau.

> Bien entendu, beaucoup de ces opérations devraient être largement configurables par les utilisateurs. Ceci étant dit, la plupart des utilisateurs ne voudront pas gérer toutes les configurations de tous leurs systèmes d'information, et vérifier, par exemple, que ces configurations respectent le règlement pour la protection des données personnelles, ou empêchent la diffusion massive de discours de haine. Non, la plupart des utilisateurs sont comme moi : ce sont des fainéants.

Le rire du public permet alors à Katia de reprendre son souffle.

> Mais surtout, sur Internet, beaucoup d'utilisateurs veulent nuire à d'autres utilisateurs ; ou au moins les influencer d'une certaine manière. Pensez à toutes les campagnes de désinformation qui sévissent sur les réseaux sociaux[259]. Or, même l'un des plus grands défenseurs du libéralisme, le philosophe John Stuart Mill, pense que la liberté des uns doit s'arrêter là où elle nuit aux autres. Tel est le principe de l'absence de tort.

Katia marque à nouveau une pause, et regarde son public, en déplaçant son regard de gauche à droite.

> L'un des grands défis de l'éthique de l'information, c'est l'implémentation du principe de l'absence de tort. Car pour autant que je sache, ce principe est quasi-consensuel en philosophie morale[260] — et c'est un sacré exploit d'être consensuel en philosophie morale.

Katia marque un nouveau silence.

> Bon ça, c'est en principe. En pratique, c'est difficile de se mettre d'accord sur ce qui constitue un tort. Est-ce qu'un commentaire agressif cause un tort ? Est-ce qu'un petit mensonge cause un tort ? Est-ce qu'une blague au dépens d'une communauté cause un tort ? Malheureusement, en pratique, on ne sera pas d'accord. Nous avons des préférences éthiques difficilement réconciliables, voire parfois

[258]:tv: Michel Serres - Les nouvelles technologies : révolution culturelle et cognitive. I Moved to Diaspora (2012).
[259]:tv: Les réseaux sociaux sont dangereux. Très dangereux. Science4All (2021).
[260]:tv: Ce principe sur lequel tout le monde s'entend. Monsieur Phi (2021).

> clairement irréconciliables. Que faire alors[261] ?

Katia semble vraiment poser cette question à son public, comme si elle attendait une réponse. Le public a l'air pensif, et attend impatiemment une réponse de Katia.

Qui décidera de l'éthique Poo ?

Katia relance avec une autre question.

> Est-ce que SmartPoop devrait trancher ?

Le public reste silencieux. Clairement, il s'agit d'un groupe de fans de SmartPoop. Mais même eux ne semblent pas emballés à cette idée.

> Si vous voulez mon avis, la réponse est clairement non. On l'a vu il y a deux ans. Une structure comme SmartPoop peut perdre son éthique et sa voie. Et même si nous avons fait énormément de progrès dans notre gouvernance pour éviter que ceci ne se reproduise, je ne pense pas que SmartPoop soit suffisamment robustement bénéfique pour une telle tâche.

Katia marque une pause, comme si elle invitait sérieusement le public à y réfléchir.

> Mais donc, qui ? Qui devrait déterminer l'éthique de l'information et du traitement de l'information ?

Encore une fois, cette question est posée comme si Katia n'en avait aucune réponse, et comme si elle attendait du public une réponse. Après de longues secondes, Katia offre sa réponse.

> Eh bien, je propose que ce soit vous. Je propose que ce soit nous tous. Je propose que, ensemble, toute l'humanité décide collectivement de l'éthique de l'information.

Le public applaudit.

> Mais... Comment peut-on amener des milliards d'humains à décider collectivement de quelque chose d'aussi complexe que l'éthique de l'information ?

Le public est maintenant circonspect.

> Réfléchissez-y. Comment fait-on aujourd'hui pour prendre des décisions collectives ?

Katia marque à nouveau une pause, jusqu'à ce qu'elle entende quelqu'un dans le public crier « le vote ».

[261]:tv: [Conférence SML] Les mathématiques de la démocratie - Lê Nguyên Hoang. Maison des mathématiques et de l'informatique (2018).

> Le vote, oui ! Dans beaucoup de pays à travers le monde, quand une décision collective doit être prise, on essaie souvent de se ramener à une question avec une réponse oui ou non, et on demande au peuple de voter pour oui, ou pour non ? C'est ainsi que l'on tranche des désaccords irréconciliables.

Katia s'arrête quelques secondes.

> Mais le problème du vote, en tout cas tel qu'il est pratiqué aujourd'hui, c'est qu'il ne permet à chaque citoyen d'envoyer que quelques bits d'information par vote seulement. La COP 30, pour ou contre ? La vaccination obligatoire, pour ou contre ? La régulation des algorithmes, pour ou contre ? Lequel des 15 candidats suivants devrait être élu ? Il n'y a pas 36 000 réponses possibles à ces questions. Or, pour résoudre l'éthique des algorithmes, il va falloir fournir des réponses complexes. Il y a même des milliards de milliards de discours qu'un utilisateur de SmartPoop peut produire. Parmi les milliards de réponses imaginables, laquelle Poo devra-t-elle adopter ?

Katia reprend son souffle, avant la suite de son discours.

> Et si on concevait désormais des votes où la voix de chacun n'était pas réduite à une réponse binaire[262], donnée uniquement une fois par an ? Et si on permettait à chacun de partager toute la complexité de son jugement éthique ? Et si on parvenait à tenir compte de toute cette complexité pour décider collaborativement de l'éthique de l'information[263] ?

Katia marque une nouvelle pause.

Le fabuleux chantier[264]

Katia change alors de ton, en prenant une voix plus grave et posée.

> Mesdames et messieurs, aujourd'hui est un jour historique, car je vais vous présenter le résultat de deux ans de travail, en collaboration intime avec l'OMESA. Au moment où je parle, une nouvelle plate-

[262] L'une des solutions pour voter en grande dimension est de s'appuyer sur le principe « un électeur, une force unitaire », ce qui peut typiquement conduire à utiliser la *médiane géométrique*.
:memo: On the Strategyproofness of the Geometric Median. El-Mahdi El-Mhamdi, Sadegh Farhadkhani, Rachid Guerraoui & Lê-Nguyên Hoang (2021).

[263] Ceci nécessitera certainement de combiner des systèmes de scrutins avec des méthodes d'apprentissage. C'est ce que propose Licchavi.
:memo: Strategyproof Learning: Building Trustworthy User-Generated Datasets. Sadegh Farhadkhani, Rachid Guerraoui & Lê-Nguyên Hoang (2021)

[264] Le fabuleux chantier est le nom d'un livre précédent de Lê Nguyên Hoang, l'un des auteurs de ce livre.
:books: Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique. Lê Nguyên Hoang & El Mahdi El Mhamdi. EDP Sciences (2019).

> forme vient d'être mise en ligne, appelée girasol.app[265]. Girasol est un site web entièrement Open Source, sous licence libre[266], qui va coordonner la conception de l'éthique de l'information, en permettant aux utilisateurs de fournir des jugements éthiques, et en utilisant des algorithmes de vote pour construire collaborativement l'éthique de l'information à partir des jugements éthiques des utilisateurs ! Le premier étage d'une éthique démocratique de l'information a été posé !

Le public applaudit cette annonce de Katia.

> Je précise que la gouvernance de ce projet est entièrement sous le contrôle de l'OMESA aujourd'hui, et SmartPoop ne sert, et ne servira, que de contributeur bénévole à la base de codes et à la promotion du projet. Tout le code est audité par de nombreuses entités, si bien qu'il est quasiment impossible que le projet soit détourné par une entité maléfique — et ça inclut de potentiels investisseurs de SmartPoop !

Katia marque une nouvelle pause.

> Alors, il y aurait énormément d'autres détails importants à préciser sur ce projet complexe. Mais sachez que, en collaboration avec l'OMESA, nous avons fait de notre mieux pour que chacun de ces détails devienne un sujet de recherche sur lequel travaillent différentes équipes pluridisciplinaires à travers le monde. Ces détails incluent des problématiques comme l'authentification des comptes, pour éviter les faux comptes, en exploitant des mécanismes de Proof of Personhood, et comme garantir le fait que chaque compte authentifié a le même droit de vote que tout autre compte authentifié. Ils incluent aussi l'identification de l'expertise et des excès de confiance des utilisateurs, pour éviter que des théories anti-scientifiques polluent l'éthique de l'information. On peut mentionner aussi l'optimisation de la plate-forme par individu, pour que cet individu exploite la manière la plus appropriée pour lui d'exprimer ses jugements éthiques. Ou encore les algorithmes de rectification du biais de participation, pour bien tenir compte des préférences de ceux qui n'ont pas pu participer à Girasol, faute de temps ou d'accès à Internet[267].

[265] Girasol n'existe pas, mais il s'agit là en fait clairement d'une référence au projet tournesol.app lancé par Lê Nguyên Hoang, l'un des auteurs de ce livre. Le reste du livre décrit finalement la vision global de Tournesol. Vous trouverez beaucoup plus d'informations sur le wiki de Tournesol. Il est utile de noter que, surtout pour l'instant, l'objectif de Tournesol est davantage de servir de « microscope des jugements humains », c'est-à-dire d'outils de collecte de données sur ce que les humains jugent éthiquement préférables. En particulier, et entre autres, Tournesol espère ainsi détecter des consensus moraux aujourd'hui difficilement observables, faute de données.

[266] Le code de Tournesol est sous licence AGPL, tandis que la base de données publique est sous license ODbL (à confirmer).

[267] Le projet Tournesol soulève énormément de défis, allant de la recherche au développement,

Katia s'arrête quelques instants.

> Bref. Il y a plein de défis de recherches remarquables que Girasol devra résoudre, pour ensuite permettre une conception collaborative adéquate de l'éthique de l'information. Girasol est clairement un énorme chantier très difficile à mener. Et c'est aussi un chantier urgent à résoudre.

Katia reprend son souffle pour conclure son discours.

> Mais avant tout, Girasol est un *fabuleux* chantier. Si vous me demandez, il s'agit pour moi du plus fabuleux de tous les chantiers menés par l'humanité, plus grandiose encore que construire des pyramides, plus ambitieux que d'éradiquer des pandémies comme la variole[268], et plus spectaculaire que d'envoyer des humains sur la Lune. Girasol, c'est unir toute l'humanité derrière le plus important de tous les aspects de la civilisation humaine : maîtriser collaborativement le flux de l'information[269], et garantir que celui-ci coule comme nous, l'humanité, souhaiterions vraiment qu'il coule, lorsqu'on y réfléchit à tête reposée, avec bienveillance et rigueur. Mesdames et messieurs, ensemble, résolvons l'éthique de l'information[270] !

À ces mots, le public explose de joie et d'enthousiasme, alors que le vacarme laisse petit à petit place au nom de Katia, lequel est scandé en rythme par tout le stade. Seule sur scène, Katia profite de l'instant, avec un sourire radieux, et salue le public de la main. À ce moment, elle pense à tout ce que SmartPoop a accompli jusque là. Mais aussi et surtout, Katia est incroyablement enthousiasmée par la vision d'une civilisation humaine qui, grâce à Girasol, va enfin prendre le destin de sa civilisation entre ses mains.

Ensemble.

Pour aller plus loin

Voilà, le roman est fini ! Vous pouvez revenir au sommaire.
Si vous avez apprécié, pensez à partager et à promouvoir ce roman de science-fiction auprès de vous. Nous vous en serions très reconnaissants !

---

en passant par la promotion, par le financement et par les partenariats, entre autres, que les membres de Tournesol ne pourront absolument pas résoudre seuls. *Vous* pouvez aider. Pour en savoir plus, notamment sur l'aspect recherche et développement, nous vous encourageons à lire le *white paper* technique du projet.
:memo: Tournesol: A quest for a large, secure and trustworthy database of reliable human judgments. Lê-Nguyên Hoang, Louis Faucon, Aidan Jungo, Sergei Volodin, Dalia Papuc, Orfeas Liossatos, Ben Crulis, Mariame Tighanimine, Isabela Constantin, Anastasiia Kucherenko, Alexandre Maurer, Felix Grimberg, Vlad Nitu, Chris Vossen, Sébastien Rouault & El-Mahdi El-Mhamdi (2021).

[268]:tv: Le plus grand triomphe de l'humanité. Science4All (2020).

[269]Ceci pourrait d'ailleurs être le sujet d'un prochain livre de Lê Nguyên Hoang… #teaser

[270]:tv: Résolvons l'éthique ensemble !! Science4All (2021).